

Exploring written ambiguities can help assess where to mark tone

David Roberts

Prepublication version. Published in *Writing Systems Research*, 2(1), 25-40.

Abstract: In certain tone languages, orthography stakeholders are open to the possibility of some kind of partial representation of tone. But how can the researcher know for sure which parts of the language need to be targeted for disambiguation? This paper proposes that analysis of written ambiguity can help to answer this question. It is a method involving three stages: the development of a homograph corpus, a frequency and distribution analysis of homographs in natural texts and a miscue analysis of oral reading performance. The method is applied to Kabiye (Gur, Togo), the standard orthography of which does not currently mark tone. The conditional clause is traced through each of these three stages, ending with a proposal for its modification. This method demonstrates the extent to which the Linguistics of Writing can enrich a debate that has long been dominated by utterance-based phonological analysis.

1. Introduction

1.1 Preamble

In many languages tone plays an important role, but its functional load is probably not high enough to justify an exhaustive diacritic representation in the orthography. In many such cases, decision makers have chosen to target only certain localised ambiguities, leaving the rest of the orthography untouched. This process often leads them beyond any presupposition that tone must be represented by diacritics (Roberts, 2013: 5-9).

But quite apart from the question of appropriate symbolization, how can one know with certainty which ambiguities should be targeted in the first place? This article proposes that analysis of written ambiguity can help answer this question. Using Kabiye (Gur, Togo) as an example, it proceeds in three stages. The first treats *ambiguity of isolated words* and involves generating an exhaustive list of homographs. This process highlights general tendencies, establishes sets of similar ambiguities, and serves as a basic reference list for the rest of the research. The second stage scrutinizes *ambiguity in context*. On the basis of a large computerized corpus of published literature, the frequency and distribution of homographs are analysed in a rich variety of contexts. Non-essential problems are weeded out by analyzing the contribution of context. These first two theoretical steps are followed by a final practical one, which studies *ambiguity in practice*. It involves a classroom experiment involving miscue analysis of oral reading to highlight which parts of the orthography are perturbing readers.

Since it is the concept of ambiguity that binds the three stages of the method together, a definition is in order. Ambiguity is defined as the effect produced by any orthographic element the interpretation of which is uncertain because it has two or more distinct meanings. This eliminates polysemy, which concerns only words with closely related meanings. The study focuses on homographs, and distinguishes between two kinds: heterophonic homographs (tonal minimal pairs) and homophonic homographs. Only the first can perturb oral reading, but both can perturb comprehension. Only the first can be disambiguated in a phonographic (sound-based) orthography, but both can be disambiguated in a semiographic (meaning-based) orthography.

As for the term 'homonym', it is best avoided for the purposes of this article, because according to some authorities it does not specify the medium (cf. Collins English Dictionary: "one of a group of words *pronounced or spelt* in the same way but having different meanings"; but Oxford English Dictionary: "Each of two or more words having *the same written form* but of different meaning and origin, a homograph." Italics mine).

A high degree of ambiguity in a natural text slows down oral reading speed, generates miscues and can lead to incomprehension. Catach (1988: 11) describes ambiguity as "Writing's number one enemy". At the same time, fluent readers have a certain tolerance for ambiguity, which is why this method carries no assumption that each word should have a unique meaning in order to achieve maximally efficient communication. The aim is not to rid the orthography of all ambiguity, but to target only those ambiguities that are likely to be disruptive.

With this in mind, it will be helpful to distinguish between two levels of ambiguity. In this study, "absolute ambiguity" means that there is no possible semantic or syntactic clue in the surrounding sentence that could help the reader make the correct choice. This is in contrast to "relative ambiguity", where although the word in isolation is ambiguous, the correct choice may be understood by a careful study of the surrounding context. Relative ambiguities may still perturb the reading process, especially if several converge on one sentence.

The term 'functional load' is also an important concept in what follows. It describes the extent to which any given phonological contrast contributes to keeping utterances apart (King, 1967). However, no viable method of quantifying functional load exists.

Researchers tend to use the term subjectively, claiming that a given phonological contrast in a language has a 'high' or a 'low' functional load, but without offering proof. It would certainly be wrong to equate functional load with mere frequency. However, we do know that frequency, which is a measure of how much use a language makes of a linguistic unit, will play an important part in any assessment of functional load. It is in this perspective that the central role of frequency in the method that follows should be understood.

1.2 *The linguistics of writing*

The method presented in this paper draws inspiration from the Linguistics of Writing, a domain of linguistics that has waited a long time to become a respected field of research. Throughout the 20th century, ever since the influential work of Saussure (1916), there has been a strong emphasis on utterance based linguistic analysis.

The debate about how tone should be represented orthographically, from the earliest researchers (eg. Christaller, 1875) to the most recent (Schroeder, 2008) has found its inspiration in this utterance-based approach. Many linguists study tone languages motivated by the pressing need to develop practical orthographies for previously unwritten languages. But ironically, even these researchers have typically come at the question from a decisively phonological standpoint (Gudschinsky, 1959; Kutsch Lojenga, 1993; Mfonyam, 1989; Pike, 1947, 1948; Wiesemann et al., 1988; Williamson, 1984). Any references to the Linguistics of Writing in the literature on tone orthography are few and far between. A good orthography has often been viewed as an extension of a good phonological analysis, and a rather unscientific one at that. Thus, in language after language, the transition from the oral to the written has been made without any theory of writing ever being invoked.

However, by the turn of the millennium the Linguistics of Writing found itself at a turning point in its evolution (Jaffré, 1997: 17), with numerous researchers helping to elevate it to its rightful place (Brissaud et al., 2008; Catach, 1988; Coulmas, 2003; Daniels, 2001; Jaffré, 2003; Rogers, 2006). This investigation of the orthographical representation of Kabiye tone draws inspiration from this new current of research.

1.3 *Kabiye spelling conventions*

It will be helpful to distinguish several types of Kabiye spelling. Firstly, the *standard orthography* is the written form of the language as it was officially adopted in the early 1980s by the Comité de Langue Nationale Kabyè (henceforth CLNK), and organ of the Togolese Ministry of Education. This is the primary entry in the Kabiye - French dictionary (CLNK & SIL-Togo, 1999). The standard orthography does not mark tone. Some words permitted by the CLNK to reflect dialect variants. In the dictionary they are listed as secondary entries and cross-referenced to the primary entry.

As for unauthorized spellings, the term *popular orthography* refers to any spelling which represents a conscious decision on the part of the writer to deviate from the standard for a specific reason. Such spellings are not formally adopted by the CLNK, but they appear fairly frequently in practice, even among writers who sit on the committee. The term *incorrect orthography* is reserved for spontaneous spelling errors, even those that result in a different correct word. And finally, a peculiarity of the Kabiye sociolinguistic context is the existence of an *alternative orthography* that is beyond the scope of this article, but is the focus of another one (Roberts, 2008b).

1.4 A response to felt needs

The Kabiye tone system is characterized by two discrete level tones, H and L, automatic and non-automatic downstep. Numerous lexical and post-lexical morphotonological processes occur once words are placed in context (Delord, 1976; Kassan, 2000; Lébiakaza, 2003; Lébiakaza, 1994; Lébiakaza, 1999; Roberts, 2002, 2003a, 2003b). Although tone plays an important role in both the lexicon and the grammar, it is in the latter that it bears the greater burden. By most estimates the functional load of tone is neither as high nor as low as it is in some other African languages, though this claim awaits further objective investigation.

The CLNK is well aware of the problems caused by the lack of tone representation in the standard orthography (CLNK, 1995: 4-5, 16-17 my translation of the French):

"The Commission for Linguistic Research, seeking to address the question of Kabiye tone, felt that, in view of the enormous difficulties encountered when reading Kabiye, we should return once again to this old chestnut... "

The three stages of the methodology find their inspiration in certain stated aspirations of the CLNK. First of all, the Commission proposes "[...] *that we continue to collect tonal minimal pairs [...]*". The first stage of the methodology, the homograph corpus, is a response to this. The Commission continues: "[...] *placing them in a corpus of sentences [...]*". The second stage of the methodology, the frequency and distribution analysis, answers this point. The Commission concludes by pleading for "[...] *the results of this research to be tested in schools, so that we can draw conclusions.*" The third stage of the methodology, the miscue analysis, contributes to meeting this need.

I highlight these citations to show that this research is firmly rooted in sociolinguistic realities. It is an implementation of a methodology originally proposed by the CLNK, who are the sole decision makers in Kabiye orthography matters. It was carried out under their auspices, and involved a process of networking with stakeholders in the Kabiye community over a period of five years (Roberts, 2008a: 147-167). In this respect, the linguistic analysis that follows should be understood within a wider framework of ‘orthography as social practice’ (Sebba, 2007: 26-57). I will now describe each of the three methodological stages in turn.

2. Ambiguity in isolation

2.1 *The homograph corpus*

The first stage of the analysis required the creation of a computerized corpus of homographs in the standard orthography (Roberts, 2008a: A5-169)¹. It is based on the Kabiye - French dictionary which contains 6,541 entries. To achieve a near exhaustive list (Catach, 1998: 92), the first step was to simply search for adjacent homographic entries in the dictionary. They can either be heterophonic homographs (1) or homophonic homographs (2):

- | | | | |
|---|----------|-----------|-------------------------------------|
| 1 | <kpaan> | [kpáá-ŋ] | <i>porch-N3</i> |
| | | [kpa-a-ŋ] | <i>madness-N3</i> |
| 2 | <habiyε> | [hábí-yé] | <i>road-N7</i> |
| | | [hábí-yé] | <i>type of traditional dance-N7</i> |

It was also necessary to take into account the multiplicity of homographs that do not have a separate dictionary entry. So the corpus also contains homographic roots (3), prefixes (4) and suffixes (5):

- | | | | |
|---|--------|------------------------|----------------|
| 3 | <kat>- | [kat-] (tone class H) | <i>dare-VR</i> |
| | | [kat-] (tone class HL) | <i>meet-VR</i> |

¹ I am grateful to David Rowe and Neal Breakey for their help with generating the two corpuses.

4	<τι>-	[tí-]	habitually (HAB)
		[tí-]	in spite of (ADV)
		[tí-]-	meanwhile (EXP)
5	-<aa>	[-áa]	nominal suffix, N2
		[-aa] ²	verbal suffix, PER

Each root and affix generates a multitude of other homographs once they undergo inflection, for example:

6	<εkataa>	[ε-gat-áa]	SP3/1-dare-PER	He ~ she dared
7	<εkataa>	[ε-gát-aa]	SP3/1-met-PER	He ~ she met
8	<εtukataa>	[ε-dii-gat-áa]	SP3/1-ADV-dare-PER	He ~ she dared even so
9	<εtukataa>	[ε-dii-gát-aa]	SP3/1-ADV-met-PER	He ~ she met even so

None of these inflected forms appears in the homograph corpus, though an exception was made for cases of morphemic mismatch, for example:

10	<kataa>	[kat-áa]	Dare-PER	(he ~ she) dared
11	<kataa>	[ká-tá-a]	SP3/5-anoint-AOR	(and) he ~ she anointed

But in general, inflected forms are so numerous that to include them in the corpus would be impractical. This means that the total number of possible homographs in the written language is far greater than the 1,205 listed. This does not imply that the corpus is deficient, because all possible homographs in the written language are identifiable using the corpus as a starting point. These inflected forms are crucial to the analysis, because it is them that readers will identify as orthographic words. On this point, let us also note that a methodology that is essentially based on written data does not preclude the need for oral elicitation, since it is important to orally elicit all the derived forms that do not have a dictionary entry.

One of the most useful aspects of creating a homograph corpus is that it permits the researcher to generate tone paradigms and to trace the *patterns* of ambiguity that occur

² The melody of this suffix depends on the tone class of the verb to which it is attached (Roberts, 2002).

in them. I will take the conditional clause as an example, and trace it through all three stages of the analysis, beginning with its representation in the homograph corpus.

2.2 The conditional clause: homographs in isolation

Using the homograph corpus as a starting point, it is possible to assemble a tone paradigm of concordant markers in the noun class system. This reveals that all subject pronouns are ambiguous with relation to the declarative and conditional clauses:

	DEC		CND	
12	<ma> ⁻³	[ma]- I	[má] if I	SP1s
13	<ŋ>	[ŋ]- you	[ŋ] if you	SP2s
14	<ɔ̄> ⁻⁴	[ɔ̄]- we	[tɔ̄]- if we	SP1p
15	<ɪ> ⁻⁵	[é]- you	[é]- if you	SP2p
16	<ε>	[ε]- he ⁶	[é]- if he	SP3/1
17	<pa> ⁻⁷	[pa]- they	[pá] if they	SP3/2
18	<kɪ>	[kɪ]- he	[kí]- if he	SP3/3
19	<ɪ>	[ɪ]- they	[í]- if they	SP3/4
20	<ka> ⁻⁸	[ka]- he	[ká] if he	SP3/5
21	<sɪ>	[sɪ]- they	[sí]- if they	SP3/6
22	<ɔ̄>	[ɔ̄]- he	[tɔ̄]- if he	SP3/7
23	<a>	[a] they	[á] if they	SP3/8
24	<tɪ>	[tɪ]- he	[tí]- if he	SP3/9
25	<pɪ>	[pɪ]- he	[pí]- if he	SP3/10

3 Or <ma mɛ me mə mo man mɛn men mɔ̄n mon>- depending on vowel harmony and morphophonemic rules.

⁴ Or <ɔ̄>- depending on vowel harmony rules. Likewise in examples 18, 19, 21, 22, 24 and 25.

5 The standard orthography of the second person plural is a morphonographic representation (Roberts, 2008a: 25-28, 224, 309-311).

6 Or 'she' or 'it' in each case.

7 Or <pa pɛ pe pɔ̄ po>- depending on vowel harmony rules.

8 Or <ka kɛ ke kɔ̄ ko>- depending on vowel harmony rules.

26	<ña>-	[ñá]-	he himself	[ɲá]-	if he himself ⁹	SP3FOC
27	<a>-	[á]-	who?	[á]-	if who?	SP3INT

Most of the pairs in this paradigm are heterophonic homographs (i.e. tonal minimal pairs). But there are also three sets of homophonic homographs (15, 26, 27). Now an utterance based phonemic analysis would automatically exclude these from a list of tonal minimal pairs and the resulting paradigm would be incomplete. But an analysis based on written data, because it takes account of both heterophonic and homophonic homographs, is able to treat the series as a unified whole, which the H tone regularity of the conditional column of the paradigm begs for.

In addition to the H tone on the subject pronouns, certain conditional clauses are also characterized by a clause final marker. Some of these are homographic, so they appear in the homograph corpus too:

28	<y>	[^l yó]	CND	[yó]	SUB
29	<le>	[^l lé]	CND ~ TMP	[lé]	then
30	<elε>	[élé]	CND ~ but ¹⁰	[élé]	ABS3/1-4 ¹¹

But a mere list of isolated homographs gives us no indication of their functional load. A much more important consideration is the frequency and distribution of some of them once they appear in natural contexts.

9 This pronoun is neutral with respect to class and number, so there are multiple translations.

10 In some cases, the oppositional conjunction <le> *but* is better translated with a consecutive meaning *and then...*

11 <elε> also forms part of a series of very infrequent pronouns that Lébikaza calls « démonstratifs de déixis temporelle proximale » (1999: 464-466). He cites [énó ~ ónó] for class 1. But my two assistants have always insisted on <elε>, which is the form cited in the dictionary.

3. Ambiguity in context

3.1 *The literature corpus*

Reading lists of isolated words is a world away from reading natural texts. It is context that invests words with meaning, as many researchers have already pointed out concerning tonal minimal pairs (Crofts, 1976: 127; Grimes et al., 1963: 114-119; Hollenbach, 1978: 56; Powlison, 1968: 74-91; Seifart, 2006: 280).

The literature corpus provides a way of weighing the homograph corpus in the balance. Frequency and distribution analysis contributes to an assessment of functional load. It distinguishes between real ambiguities and those whose meaning is discernable with the help of context. Among the multitude of words in the homograph corpus that probably do not pose any difficulty, it identifies the real culprits: a small minority which should be targeted for modification.

The literature corpus contained a total of 142,483 words, and 18,961 distinct word forms. This was sufficient to conduct a detailed analysis of all grammatical homographs, which is where the real functional load of tone lies. The literature corpus contained only of natural and authentic texts (Catach, 1989: 230). It included a rich variety of genres (narrative 59%, pedagogical 34%, correspondence 3%, proverbs 3%, poetry 1%) and themes (religious 39%, culture 20%, miscellaneous 16%, agriculture and health 14%, politics 7%, mathematics 4%). The corpus only included texts written in standard orthography. However, it did not exclude texts containing spelling mistakes, out of a concern to reflect the social reality that Kabiye texts, as in any literate culture, tend to be imperfectly written.

The corpus was stored in Shoebox, an integrated data management and analysis program.¹² The analysis exploited two functions. Firstly, the word list function generates

¹² Shoebox has been superseded since this research was undertaken. Toolbox (<http://www.sil.org/computing/toolbox/>) is an enhanced version of Shoebox. Language Explorer is the lexical and text tools component of the FieldWorks suite of programs (<http://www.sil.org/computing/fieldworks/>).

a list of all the words in the corpus, sorting them in descending order of frequency. This reveals that the three conditional clause final markers (28, 29, 30) appear among the top ten most frequent words in the language (31, 32,33):

			frequency	rank
31	<yɔ>	CND ~ SUB	4,844	3
32	<lɛ>	CND ~ TMP ~ then	3,204	5
33	<ɛlɛ>	CND ~ but ~ ABS3/1-4	932	10

Secondly, the concordance function in Shoebox can search all the occurrences of any orthographic sequence in the literature corpus. It displays not only the target but its context to the left and right. This operation targeted all grammatical homographs (including not only whole words but also affixes) in the homograph corpus.

Although the concordance operation is a powerful tool, there are limits to its contribution. It can target any homograph, but it is not capable of identifying the correct meaning of that homograph in a given context. Only a mother-tongue Kabiye reader can do this. This was the work of three research assistants, who determined the meaning of each targeted homograph by studying the immediate context. The fact that they could do this, by the way, does not suggest that the forms were unambiguous in the first place. The process required hours of meticulous scrutiny and group discussion - a world away from the natural reading process.

The results permitted a calculation in percentage terms of the distribution of the various meanings of each homograph. Let us return to the case of the conditional clause, and see how it fared in the frequency and distribution analysis.

3.2 The conditional clause: frequency and distribution

Among the 940 conditional clauses in the literature corpus, there are no less than twelve different linguistic strategies for expressing the conditional. These are generated by segmental elision processes that force the burden of differentiating meaning up onto the tonal tier. The strategies can be grouped into two sets, which we will explore in turn. Examples 34 - 39 show the first set. The figures in the third column show the frequency of the form as a percentage of all conditional clauses in the literature corpus.

Two explanations are necessary to enable correct interpretation of the data. First, the percentages in examples 34-39 and 42-47 do not add up to 100, because all cases of clause final marker elision from both sets were conflated into two results, i.e. <ye_ Ø> (36, 44) and <Ø _ Ø> (39, 47). Second, the prolific non-automatic downstep [[↓]] is due to a rule stipulating that any underlying /HLH/ melody always surfaces as [H[↓]HH] or [HH[↓]H] depending on the skeletal structure of the segments to which it associates.

34	<Yee ητεκε abalu κῶϰ, ηηκυυ τῶϰυυ.>	[yéé η-τε-κε abal-ó [↓] κῶϰ, η-η-κῶ-ó [↓] τῶϰ-óó]	2.13
35	<Yee ητεκε abalu γῶ, ηηκυυ τῶϰυυ.>	[yéé η-τε-κε abal-ó [↓] γῶ, η-η-κῶ-ó [↓] τῶϰ-óó]	36.17
36	<Yee ητεκε abalu, ηηκυυ τῶϰυυ.>	[yéé η-τε-κε abal-ó η-η-κῶ-ó [↓] τῶϰ-óó]	15.85
37	<ητεκε abalu κῶϰ, ηηκυυ τῶϰυυ.>	[η-τε-κε abal-ó [↓] κῶϰ η-η-κῶ-ó [↓] τῶϰ-óó]	3.19
38	<ητεκε abalu γῶ, ηηκυυ τῶϰυυ.>	[η-τε-κε abal-ó [↓] γῶ η-η-κῶ-ó [↓] τῶϰ-óó]	32.66
39	<ητεκε abalu, ηηκυυ τῶϰυυ.>	[η-τε-κε abal-ó η-η-κῶ-ó [↓] τῶϰ-óó]	7.34
	<i>If you are not a man, you will not kill the lion.</i>	(CND) you-NEG-be man-N1 (CND) you-NEG-kill-IMP lion-N3	

The full conditional clause in the first set is framed by two markers <ye> [yéé]¹³ and <κῶϰ> [κῶϰ]¹⁴ (34). But this form is rare. Much more commonly, the first syllable of the final marker elides leaving <γῶ> [[↓]γῶ] preceded by a floating low tone which triggers non-automatic downstep (35). In addition, the final marker is optional and can elide altogether (36).

It is also possible for the initial marker <ye> [yéé] to elide. In speech, the elided segments leaves its H tone on the subject pronoun, but in writing the only remaining indication of the conditional meaning is the final marker. In these cases the final marker rarely appears in its complete form (37); it is the elided form that is preferred (38). In writing, this double elision generates ambiguity, because the final particle <γῶ> [γῶ] (not

13 The standard orthography is <ye>, but the majority of writers prefer the tolerated orthography <ye> that corresponds to speech [yéé]. Both were incorporated into the results.

14 Some writers adopt a popular orthography, <keyῶ> that represents a dialect variant [keyῶ]. These were incorporated into the results.

preceded by non-automatic downstep in speech) is the classic subordinate clause marker, producing a counter-factual meaning in this context (40):

- 40 <ἦτεκε abalu γῶ, ἠγκυυ τῶγυυ.>
 [ἦ-τε-κε abal-ύ γό ἦ-ἦ-κό-ύ †τόόγυ-ύύ]
 you-NEG-be man-N1 CF you-NEG-kill-IMP lion-N3
Since you are not a man, you will not kill the lion.

Lastly, it is possible for both the initial and the final markers to elide. In this case, in speech, the H tone on the subject pronoun is the only indicator of the conditional meaning (39). This generates ambiguities, since the same written sentence may also have a declarative meaning (41):

- 41 <ἦτεκε abalu, ἠγκυυ τῶγυ-υυ.>
 [ἦ-τε-κε abal-ύ ἦ-ἦ-κό-ύ †τόόγυ-ύύ]
 you-NEG-be man-N1 you-NEG-kill-IMP lion-N3
You are not a man; you will not kill the lion.

Now let us turn to the second set of conditional clauses (42-47):

		%
42	<Υεε νῶγῶ εσαγ-ἦ τῶμ ελε, ἠκοο νε ἠγα-μ !>	0.00
	[yéé †νόό†γῶ ἐ-κάα-ἦ τῶμ ελέ ἦ-κόο νε ἦ-γάμ]	
43	<Υεε νῶγῶ εσαγ-ἦ τῶμ λε, ἠκοο νε ἠγα-μ !>	2.66
	[yéé †νόό†γῶ ἐ-κάα-ἦ τῶμ †λέ ἦ-κόο νε ἦ-γάμ]	
44	<Υεε νῶγῶ εσαγ-ἦ τῶμ, ἠκοο νε ἠγα-μ !>	15.85
	[yéé †νόό†γῶ ἐ-κάα-ἦ τῶμ ἦ- κόο νε ἦ-γάμ]	
45	<Νῶγῶ εσαγ-ἦ τῶμ ελε, ἠκοο νε ἠγα-μ !>	0.00
	[νωό†γῶ ἐ-κάα-ἦ τῶμ ελέ ἦ-κόο νε ἦ-γάμ]	
46	<Νῶγῶ εσαγ-ἦ τῶμ λε, ἠκοο νε ἠγα-μ !>	0.00
	[νωό†γῶ ἐ-κάα-ἦ τῶμ †λέ ἦ-κόο νε ἦ-γάμ]	
47	<Νῶγῶ εσαγ-ἦ τῶμ, ἠκοο νε ἠγα-μ !>	7.34
	[νωό†γῶ ἐ-κάα-ἦ τῶμ ἦ-κόο νε ἦ-γάμ]	
	<i>If anyone provokes you, call me!</i>	
	(CND) certain he-wants-IMP-you word (CND) you-shout-AOR and you-call-AOR-me	

These occur much less frequently than the first set. The full conditional clause in the second set is framed by the two markers <υεε> [yéé] and <ελε> [ελέ]. (The precise

semantic difference between the two markers <yə> and <ɛɛ > would make an interesting subject for further study). There are no examples of the complete form in the literature corpus, but example 42 was elicited for the sake of completeness.

The final marker <ɛɛ> sometimes undergoes elision to become <lɛ> [ˈlɛ] (43), and can also elide altogether (44). It is also possible for the initial marker <yee> to elide. Again, there are no occurrences in the literature corpus, but examples 45 and 46 were elicited for the sake of completeness. And finally, it is possible for both markers to elide (47). In these cases, the H tone on the subject pronoun is the only signal of the conditional meaning, and this is not represented orthographically.

To summarise, the frequency and distribution analysis reveals that the conditional clause generates two types of ambiguity, both linked to the elision of the segmental clause initial marker <yee>.

Firstly, there are conditional clauses that are identifiable only by the presence of the partially elided final marker <yə>. These are ambiguous with relation to the counterfactual clause. Secondly, those conditional clauses that contain neither the initial marker <yee> nor one of the final markers <kəyə, yə, ɛɛ, lɛ> are ambiguous with relation to the declarative clause. These two types of ambiguities represent 40% of the conditional clauses in the literature corpus. They are often relative rather than absolute ambiguities once the wider context, beyond sentence level, is taken into account. But in a literacy context where there are many beginner readers, and where most readers do not practice their skills on a daily basis, such a complex range of linguistic strategies in speech cries out for the introduction of a single, unifying strategy in writing.

What of the three homographic clause final markers <yə, ɛɛ, lɛ>? The first step of the methodology showed that they are all ambiguous in isolation. The second step revealed that they are among the most frequent words in the written language. But further

frequency and distribution analysis reveals that none of them is ambiguous in natural texts. The syntactic context will always decide the meaning, because they always occupy clause final position.

By way of contrast, the subject pronouns are deeply ambiguous in context once the clause initial marker <yee> elides. In silent reading, a process that permits greater tolerance of imprecision, this may not be too perturbing. But proficient oral reading requires a very specific choice: to pronounce either a H or a L tone on the subject pronoun.

On this point, it will be helpful to view the data from quite a different angle: that of the study of eye movements in the cognitive psychology of reading. When the reader fixates on the clause initial subject pronoun, there is a moment of choice. If the clause initial particle <yee> is absent, as it is in 43.19% of cases, the only other clue is the clause final marker, if it is present at all. In the cases where this is a homograph, the syntactic context supplies enough information to decide whether or not the meaning is conditional. But crucially, there will be many sentences in which the clause final marker is beyond the parafoveal vision of the reader, counted as 16 letters of normal type size to the right of the fixation (Taylor & Taylor, 1983: 121-139 ; Underwood & Batt, 1996: 144-188). This psychological consideration must be taken into account in assessing how best to modify the conditional clause.

The first two stages of the methodology have been theoretical. The aim has been to predict, on the basis of homograph frequency and distribution in natural texts, where readers will have difficulties in performance and comprehension. But as yet there is no empirical proof. To find that, we turn to the third, practical stage: a miscue analysis of oral reading performance.

4. Ambiguity in practice

4.1 *Miscue analysis*

Miscue Analysis is a technique for qualitatively evaluating a reader's oral reading performance in natural contexts (Brown et al., 1996; Goodman, 1965, 1969; 1997; Goodman & Burke, 1972). It is widely used in the world of education, often as a way of assessing individual children with learning difficulties. But it has rarely been used to assess orthographies themselves, in spite of recommendation (Bird, 1999: 28). The aim of this experiment was to adapt the classic model for this purpose.

Miscue Analysis often focuses on a single subject. But since the aim of this experiment was to identify weaknesses in the orthography itself, it targeted multiple readers. The first sample was a group of twenty female adult volunteer literacy monitors working with AFASA (Association des Femmes pour l'Alphabétisation, la Santé et les Activités génératrices de revenus). Most of them were highly motivated, relatively experienced readers (henceforth "the monitors"). The second sample was a group of 19, grade 10 secondary school pupils who had only recently chosen written Kabiye as an optional subject (henceforth "the pupils").

The experiment used ten varied texts of 100 words each, all extracted from the literature corpus. Previously known texts were excluded. When texts contained spelling mistakes, these were left unmodified out of a concern for authenticity. Two mother-tongue research assistants prepared interlinearized versions of each text in standard orthography, a word for word translation, and a free translation. Accents symbolising tone were added to these pre-test materials, but only to facilitate the research team's work. They pre-analyzed each text to identify homographs, tolerated and incorrect

spellings. This preparatory phase provided solid predictions about where readers were likely to miscue. The main objective was not to answer the question "to what extent does the presence of homographs cause miscues?" but rather the much more general question "What causes miscues?"

The experiment took place over several days. Subjects were recorded individually reading aloud each text once in standard orthography with no tone marks added. The experiment administrator did not interrupt or give help.

The post-experiment phase involved annotating interlinearized versions of each text in two stages. One set recorded miscues. These included repetitions, substitutions, hesitations, omissions, insertions, methathesis and ignoring punctuation. Figure 1 lists only those symbols that are pertinent to the examples cited in this article (cf. Schreiner, 1979: 59):

Figure 1: Text annotation

Type	Code	Standard orthography
Repetition	←	← kowolo Εσο cολο>
Meaningful substitution	×	[tɛtaa] × <tɛtɔ>
Meaningless substitution	××	[tɛtɛta] ×× <tɛtɔ>
Self-correction of an incorrect spelling	✓	[tɛtɔ] ✓ <tɛtɔ>
Hesitation	//	[ɛtasɯɯ]-kɛ se // kɛfɛyi]

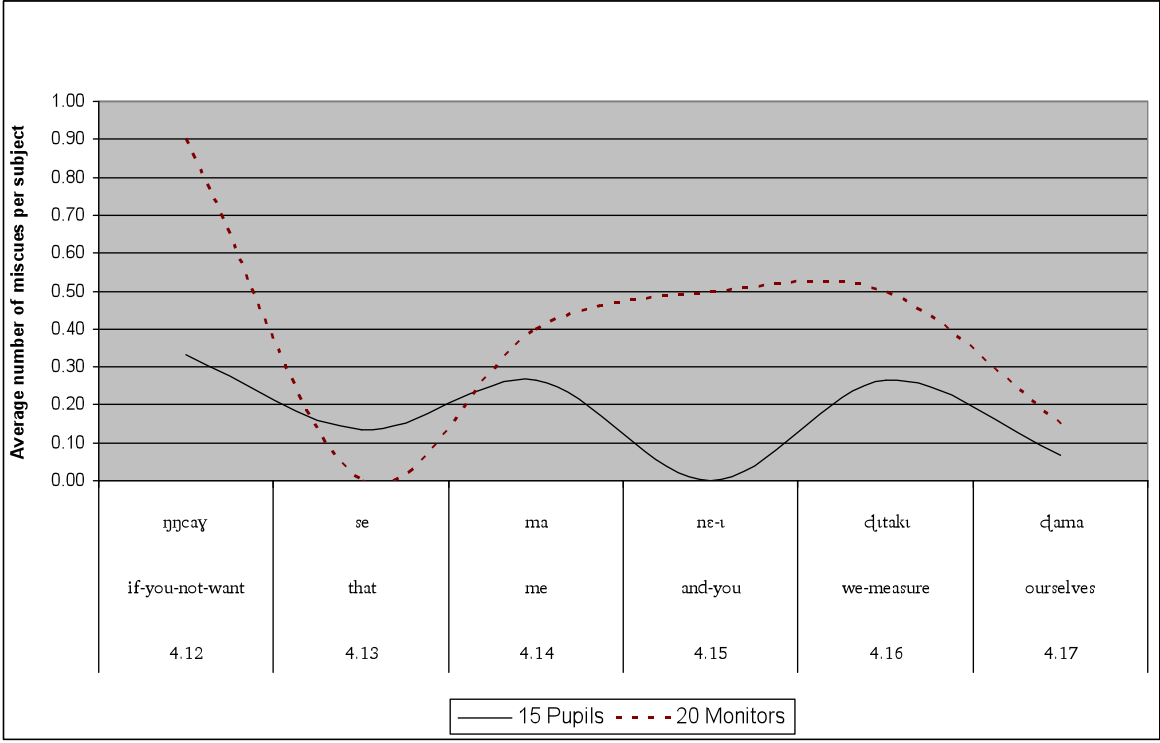
The other set of texts recorded the raw number of miscues and the average number of miscues per subject on each word, a process that guides the researcher to the most interesting areas of study. The recordings contained many miscues in all ten texts, and it is clear that not all of these are attributable to the lack of tone marking in the

orthography. But the miscue analysis shows that at least some of them were. Let us examine what it revealed about the conditional clause.

4.2 The conditional clause: miscues in oral reading

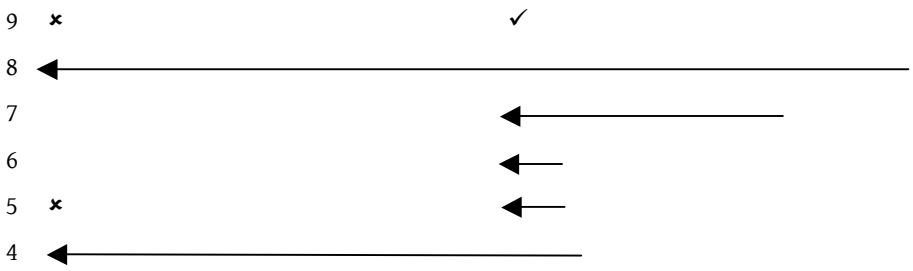
Given the optional nature of the segmental markers, it is not surprising that conditional clause contributes to oral reading miscues. Figure 2 shows a text excerpt in which the monitors' miscue rate on the first word, a conditional, is relatively high (an average of one miscue per monitor).

Figure 2: "If you don't want us to get into a fight ..." (extract from Azoti, 2008)



Let us examine the performance of one subject more closely (figure 3):

Figure 3: Miscues in the environment of a conditional clause



3 [nε-ή] ✓ ðítaa // kí] **

2 ←

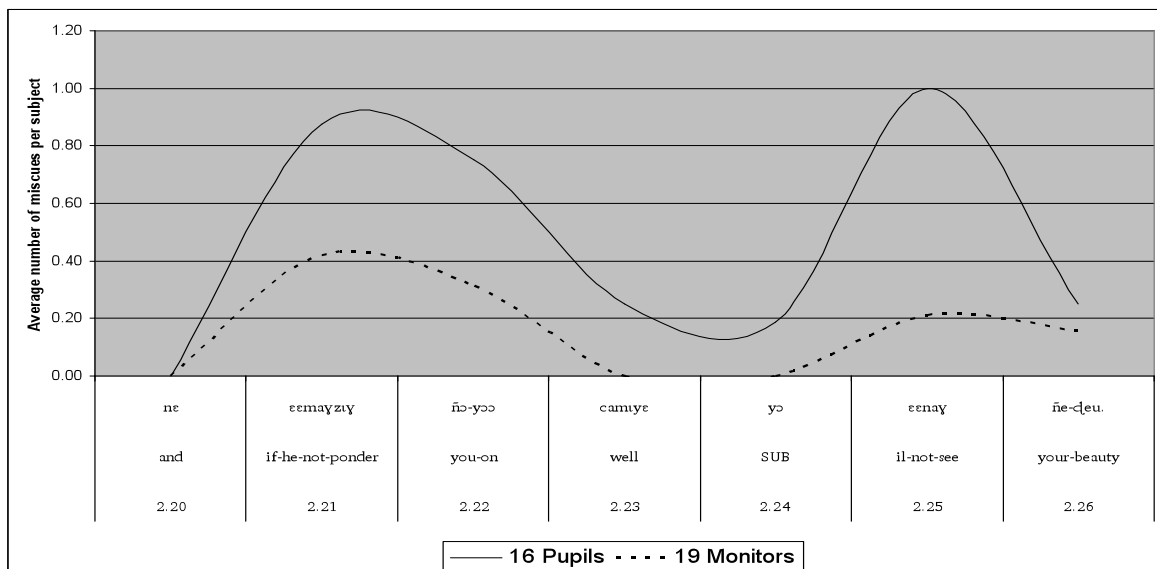
1 [ήçáα * na] **

48 <ήçα> se ma nε-ι ðítaki ðama ...>
 if-you-not-want that me and you we-touch ourselves
If you don't want us to get into a fight...

The subject in figure 3 interprets the negative conditional clause as a simple affirmative clause <ήçα...> [ήçáα] you want (line 1). This miscue triggers a chain reaction in what follows. First, it takes two attempts before she can correctly pronounce the spelling mistake *<nε-ι> and you (lines 2 and 3). Then she hesitates at length on the verb <ðítaki> we touch (line 3), seeking in vain for a signal to the left (line 4). She repeats the word again twice without success (line 5 and 6). She seeks a signal in the context to the right, again in vain (line 7). Finally, by starting the sentence all over again sentence (line 8) she succeeds in correctly pronouncing the second verb (line 9). But she never manages to identify the conditional nature of the first verb, which is one source of her difficulties.

A similar case crops up in another text. The sentence in question consists of two verbs that attract many miscues especially among the pupils:

Figure 4: "... and if he does not ponder you deeply, he will not see your beauty ..." (Alou, 1990)



Quite by chance, all the conditional clauses represented in the miscue analysis texts are in negative sentences. Kabiye expresses the negative imperfective by the lengthening of the subject pronoun, and this introduces a second level of ambiguity into the texts. Not

only are the conditional and the declarative ambiguous, but also the negative and the subordinate future. So each occurrence has three possible interpretations: negative declarative (49, 52), negative conditional (50, 53) and subordinate future (51, 54):

49	<εεμαζιγ>	[ε-ε-μααζ-υιύ]	SP3/1-NEG-think-IMP	he does not think
50	<εεμαζιγ>	[έ-έ-μααζ-υιύ]	SP3/1-NEG-CND-think-IMP	if he does not think
51	<εεμαζιγ>	[ε-έ-μααζ-υιύ]	SP3/1-FUT-think-IMP	when he will think
52	<εεναγ>	[ε-ε-νά-α]	SP3/1-NEG-see-IMP	he does not see
53	<εεναγ>	[έ-έ- [↓] νά-α]	SP3/1-NEG-CND-see-IMP	if he does not see
54	<εεναγ>	[ε-έ-νά-α]	SP3/1-FUT-see-IMP	when he will see

The most common miscue on the first verb phrase in the text extract in figure 4 is to substitute the affirmative conditional (55). But some pronounced the indicative mood, either in the affirmative (56) or in the negative (57), despite the presence of the clause initial conditional marker <yee> earlier in the text:

55	[έ-μάάζ-υιυ]	SP3/1-CND-think-IMP	if he thinks (10 subjects)
56	[ε-μάάζ-υιυ]	SP3/1-think-IMP	he thinks (4 subjects)
57	[ε-ε-μααζ-υιύ]	SP3/1-NEG-think-IMP	he does not think (2 subjects)

This confusion on the first verb phrase has an echo effect on the second, even though it is a very common word that should not in principle pose any difficulties. Notice that both words begin with the same lengthened subject pronoun <εε>-, which contributes to the confusion. The most common miscue is to pronounce a simple affirmative (58). But others substituted a completely different word (59, 60):

58	[ε-νά-α]	SP3/1-see-IMP	he sees (19 subjects)
59	[ε-νόσ]	SP3/1-understand-IMP	he understands (2 subjects)
60	[ε-ε-νσ-ό]	SP3/1-NEG-understand-IMP	he does not understand (1 subject)

In conclusion, it is important to return again to the distinction between oral and silent reading. This experiment has investigated ambiguity through the lens of miscue analysis. By definition this entails oral reading, a literacy task in which the reader is actively active. The reader is faced with a tone choice. The choice may be correct or incorrect. If it is incorrect, it may be meaningfully or meaninglessly so. In silent reading on the other hand, the reader is only passively active. There is no imperative to commit to an articulated choice, and this leaves more room for tolerating subtle ambiguities. The

miscue analysis cannot speak to the question of silent reading, but it does suggest that the standard orthography of the conditional clause is perturbing in oral reading. True, to prove the point beyond any doubt would require a comparison between the conditional clause and other types of subordinate clause, or a comparison between the standard orthography and a modified orthography. Miscue Analysis is never conclusive, and these would be interesting avenues for future research. In the meantime the experiment has furnished some strong initial proof confirming the findings of the first two theoretical stages.

5. Disambiguating the conditional clause

This article is essentially concerned with describing analysis of written ambiguity as a research methodology for evaluating tone orthography. But it would be incomplete without presenting the consequences of having followed this methodology. So what orthographic modification might be appropriate for the conditional clause?

A maximal phonographic representation would be undesirable because almost all the accents would be redundant. Multiple diacritics would only succeed in obscuring the apostrophe which marks non-automatic downstep before the clause final marker <yɔ> in 61, and its absence in 62:

Hypothetical Kabiye orthography: maximal phonographic representation

61 <ɛteke abaló 'yó, ɛɛkúú tóóyúú.>
 [ɛ-te-ke abal-ó 'yó ɛ-ɛ-kú-ú 'tóóy-úú]
 CND/SP2s-NEG-be man-N1 CND SP2s-NEG-kill-AOR lion-N3
If you are not a man, you will not kill the lion.

62 <ɛteke abaló yó, ɛɛkúú tóóyúú.>
 [ɛ-te-ke abal-ó yó ɛ-ɛ-kú-ú 'tóóy-úú]
 SP2s-NEG-be man-N1 CF SP2s-NEG-kill-AOR lion-N3
Since you are not a man, you will not kill the lion.

So would it be preferable to opt for a minimal representation that is faithful to the minimal difference? This would entail marking only the subject pronoun with an acute accent, and non-automatic downstep with an apostrophe:

Hypothetical Kabiye orthography: partial phonographic representation

63 <ŋteke abalu 'yɔ, ŋŋkɔ tɔɔyɔ.>
[ŋ-te-ke abal-ó 'yó ŋ-ŋ-kó-ó 'tɔ́y-óó]
CND/SP2s-NEG-be man-N1 CND SP2s-NEG-kill-AOR lion-N3
If you are not a man, you will not kill the lion.

64 <ŋteke abalu yɔ, ŋŋkɔ tɔɔyɔ.>
[ŋ-te-ke abal-ó yó ŋ-ŋ-kó-ó 'tɔ́y-óó]
SP2s-NEG-be man-N1 CF SP2s-NEG-kill-AOR lion-N3
Since you are not a man, you will not kill the lion.

The use of the apostrophe to mark non-automatic downstep after elision has the distinct advantage that it echoes its use in the orthography of French, the official language (e.g. 'il n'aime pas'). But three psycholinguistic concerns militate against introducing this convention in Kabiye. Firstly, minimal graphic strokes provide only minimal visual impact in response to an ambiguity that has been shown to be both frequent and disturbing. Secondly, this strategy would often require placing an acute accent on a sentence initial capital letter which gets messy (see examples 61 and 63). Thirdly, and most importantly, the apostrophe on the clause final marker would frequently be beyond the range of the reader's parafoveal vision and of no use practical use.

The research team hosted three regional orthography stakeholder consultations to discuss how best to represent tone in the orthography (Pidassa & Roberts, 2005, 2006, 2008). During these, one participant proposed that the orthography should always include the elided forms and readers left to choose whether or not to pronounce them. But this suggestion assumes that written language should be of a more formal register than spoken, whereas an optimal orthography should offer writers the flexibility to distinguish between different registers.

A stronger option might be a semiographic solution. For example the logogram <±> could be systematically added to the beginning of all conditional clauses, whether or not they are ambiguous:

Hypothetical Kabiye orthography: semiographic representation

65 <± Yee ŋteke abalu kɔyɔ, ŋŋkɔ tɔɔyɔ.>

- 66 <± Yee ητεκε abalu γα, ηηκυσ τάγυυ.>
 67 <± Yee ητεκε abalu, ηηκυσ τάγυυ.>
 68 <± ητεκε abalu κάγ, ηηκυσ τάγυυ.>
 69 <± ητεκε abalu γα, ηηκυσ τάγυυ.>
 70 <± ητεκε abalu, ηηκυσ τάγυυ.>
If you are not a man, you will not kill the lion.
- 71 <± Yee νάγυ εσαγ-η τάμ ελε, ηκοο νε ηγα-μ !>
 72 <± Yee νάγυ εσαγ-η τάμ λε, ηκοο νε ηγα-μ !>
 73 <± Yee νάγυ εσαγ-η τάμ, ηκοο νε ηγα-μ !>
 74 <± Νάγυ εσαγ-η τάμ ελε, ηκοο νε ηγα-μ !>
 75 <± Νάγυ εσαγ-η τάμ λε, ηκοο νε ηγα-μ !>
 76 <± Νάγυ εσαγ-η τάμ, ηκοο νε ηγα-μ !>
If anyone provokes you, call me!

The advantages of this solution are numerous. Firstly, the position of the symbol in clause initial position is within the foveal vision of the reader at the moment of choice (the tone on the subject pronoun). Secondly, the notion of pre-signalisation has a long pedigree in Spanish, an international language with a vast literature. Thirdly, the logogram <±> is composed of two mathematical symbols, addition <+> and subtraction <-> that will already be familiar to readers. Fourthly, their affirmative and negative values take on an iconic significance, since in most conditional clauses it is not known whether the action will or will not take place.

In one of the regional orthography consultations, another stakeholder suggested only including the logogram <±> in cases where the initial marker <yee> elides. This misses the point. The purpose of introducing the logogram throughout the paradigm is to bequeath greater visual regularity to all kinds of conditional clause, no matter what elision occurs. The logogram would signal ‘conditional clause’, and would be taught as a semiographic, meaning-based representation of the grammar, not as a tonal morpheme.

This proposal was incorporated into a quantitative experiment which pitched two experimental orthographies against each other – one representing grammar, the other representing tones. The conditional clause results were some of the most striking of all. In a dictation task, almost 80% of the grammar sub-group succeeded in writing the

logogram, but no one in the tone sub-group succeeded in writing the verb with the correct accents (Roberts, 2008a: 511)

But whatever the merits or otherwise of this proposal, they are not the focus of this article. Rather, the aim has been to reveal precisely what parts of the orthography should be targeted for modification. Exactly how they are modified is an important issue, but of secondary concern here.

As for the clause final markers <yǝ, lɛ, ɛlɛ>, there is no evidence that they need modifying, even though they are amongst the most frequent words in the written language and ambiguous in isolation. Even in the standard orthography, the conditional clause provides enough syntactic clues for the reader to make the right choice. The introduction of the logogram <±> would reinforce the existing contextual support.

By way of contrast, the triple meaning of the series of lengthened subject pronouns urgently needs untangling in the orthography. The logogram <±> adequately separates the negative conditional (examples 50 and 53) from the other two. I have proposed a solution for separating the subordinate future (51, 54) from the negative declarative (49, 52) elsewhere (Roberts, 2008a: 400-402). This is a good example of the extent to which the researcher is often grappling with several *dimensions of ambiguity*. Each layer should be isolated and targeted for separate treatment. But at the same time, whatever modification is proposed for one layer may have unforeseen consequences for another layer. That is why it would be unwise to introduce modifications on the basis of a half-completed frequency and distribution analysis.

6. Conclusion

This article has presented a three stage method for identifying the parts of an orthography that need to be disambiguated, based on analysis of written ambiguities. Some might question whether it is worth making a major investment of time for a quantitative study of this scale and rigour which is then only interpreted qualitatively. Certainly, the limitations of the quantitative approach should be clearly understood. Unseth and Unseth (1991: 46) warn against any attempt to develop a quantitative algorithm for precisely measuring orthographic ambiguity. The method presented here is not in conflict with this warning. It uses calculation, but only as a means to an end. Its

purpose is to guide the researcher to the most important areas of study, and it does this effectively. But the overall method remains essentially qualitative in nature. It is a process of observation and evaluation that depends on the researcher engaging wholeheartedly with the language itself in both its written and oral form. It also demands a continuous process of dialogue and negotiation with mother-tongue research assistants and orthography stakeholders.

Analysis of written ambiguity is particularly appropriate for evaluating orthographies in languages the functional load of tone of which is known to be largely grammatical. It may also be possible to develop the technique for dealing with languages where lexical tone has a high functional load. But this would be much a more ambitious study. Even if this were possible to generate the vast corpus necessary, it is not clear that it would be viable since most items would remain low frequency due to the exponential distribution of words (Zipf, 1935, 1949).

This paper has applied analysis of written ambiguity to an orthography that does not mark tone. But it could also be applied to a language that already marks tone, either by temporarily stripping out the diacritics or by leaving them in, depending on the research aims. It might be adapted still further for a context with two or more orthographies under consideration, by assessing two data sets in parallel. Furthermore there is no reason why it should not be applied to other segmental or supra-segmental features. In any of these cases, it is essentially second generational research.

It is not necessary to perform the three stages of the methodology in the particular order outlined in this article. For example, let us imagine a language community that already has a functioning literacy programme, but no dictionary on which to base a homograph corpus (step 1) and no text collection on which to base a frequency analysis (step 2). In such a context, a classroom experiment including miscue analysis (step 3) would be a rich source of qualitative observations of readers and teachers about the strengths and weaknesses of the orthography. This in turn could lead to a decision to collect examples (step 2) followed by identification of problematic words (step 1). Indeed, in the application of the method as presented in this paper, the homograph corpus (step 1) continued to evolve simultaneously alongside the frequency and distribution analysis (step 2) and the miscue analysis (step 3), as results from one stage fed into another.

Nor is it necessary to perform all three stages of the methodology. The homograph corpus could stand by itself, though it is of limited value because we so rarely read words in isolation. The frequency and distribution analysis goes further and might stand alone so long as we remember that it only delivers predictions and hypotheses. As for the miscue analysis, it would conceivably be possible to perform this without the preparatory work of the first two stages, especially if the experimenter was a mother-tongue speaker.

But analysis of written ambiguity is best performed if all three stages are undertaken one after each other, because the each stage trains the researcher to undertake the next stage. If all three stages are followed systematically, the researcher will be better placed to make recommendations concerning which parts of the orthography deserve to be targeted for modification.¹⁵

Author's contact details

website: www.nyinyiyu.com

e-mail: dave_roberts@sil.org

address: B.P. 57, Kara, Togo

Abbreviations

ABS	absolute pronoun
ADV	adversative
AOR	aorist

15 I would like to thank Bernard Caron, Jacques Nicole and Gray Plunkett for commenting on a draft of this article. I am grateful to Steven Bird, Jean-Pierre Jaffré, Ettien Koffi, Constance Kutsch Lojenga, Thomas McCormick, JeDene Reeder, James Roberts and Keith Snider for stimulating discussions that helped shape the end result. I acknowledge my indebtedness to the members of the Kabiye National Language Committee for the interest they have shown in this research, in particular Alou Kpatcha (president), Batchati Bawubadi (vice-president), Simtaro Dadja (secretary) and Aritiba Adji (member). This article also benefitted from discussions following its presentation at the 26th West African Languages Conference, Winneba, Ghana July 28 - August 03, 2008. It is based on an earlier French version (Roberts, 2008a: 442-486). Lastly, I am deeply indebted to my three research assistants, Pidassa Emmanuel, Pakoubètè Noël and Pidassa Jonas, without whose efforts this research would never have been completed.

CF	counter-factual
CND	conditional
DEC	declarative
EXP	expectative
FOC	focaliser
H	high tone
HAB	habitual
IMP	imperfective
INF	infinitive
INT	interrogative
L	low tone
NEG	negative
N1	Noun of class 1 (and likewise for the other classes)
PER	Perfective
p	plural
s	singular
SP	subject pronoun
SUB	Subordinate clause marker (except conditional and temporal)
TMP	Temporal clause marker
VR	verb root
1	First person
2	Second person
3/1	third person, class 1 (and likewise for the other classes)
↓	non-automatic downstep
< >	orthographic script
[]	phonetic script
-	Morpheme boundary

References

- Alou, Kpatcha (1990), Man-Kabiye Kɔnɔɔ (My Kabiye Language), In *Ebe Laba ? (What's happening?) the bi-annual journal of the Comité de Langue Nationale Kabiye* 8.2-3 and reproduced in CLNK, (1997), 22: 33.
- Azoti, S.B. (2008), *Paamaala : suye maɔv səsə*. Kara: AFASA (Association des Femmes pour Alphabétisation, la Santé et les Activités génératrices de revenus).
- Bird, Steven (1999), Strategies for representing tone in African writing systems, *Written Language and Literacy* 2:1.1-44.
- Brissaud, Catherine, Jean-Pierre Jaffré & Jean-Christophe Pellat (eds) (2008), *Nouvelles recherches en orthographe. Actes de la journée d'étude « L'orthographe aujourd'hui comme hier ? Acquisition et didactique de l'orthographe »*. Université Marc Bloch, Strasbourg 2. Limoges: Editions Lambert-Lucas.
- Brown, J., Kenneth S. Goodman & A. Marek (1996), *Studies in Miscue Analysis: an Annotated Bibliography*. Newark, Delaware: International Reading Association.
- Roberts, D. (2010). Exploring written ambiguities can help assess where to mark tone. *Writing Systems Research*, 2(1), 25-40.

- Catach, Nina (ed.) (1988), *Pour une théorie de la langue écrite : actes de la table ronde internationale CNRS-HESO, 23-24 octobre 1986*. Paris: Editions du CNRS.
- Catach, Nina (1989), *Les délires de l'orthographe*. Paris: PLON.
- Catach, Nina (1998), *L'orthographe*. Paris: PUF.
- Christaller, J.G. (1875), *A Grammar of the Asante and Fante Language called Tshi (Chwee, Twi) based on the Akuapem Dialect, with Reference to the Other (Akan and Fante) Dialects*. Basel: Basel Evangelical Missionary Society, XXIV-203. Reprinted 1964, Gregg Press: Ridgewood, New Jersey.
- 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 kabiyè-fransu (Dictionnaire kabiyè-français)*. Lomé: Ministère de l'Éducation Nationale et de la Recherche, Comité de Langue Nationale Kabiyè et SIL-Togo.
- Coulmas, Florian (2003), *Writing Systems - an Introduction to their Linguistic Analysis*. Cambridge: CUP.
- Crofts, Marjorie (1976), Must tone always be written in a tone language?, *The Bible translator* **27/1**.127-134.
- Daniels, Peter T. (2001), Writing systems in *The Handbook of Linguistics*, ed. M. Aronoff & J. Rees-Miller, 43-80. London: Blackwell.
- Delord, Jacques (1976), *Le kabiyè*. Lomé: Institut national de la recherche scientifique.
- Goodman, Kenneth S. (1965), A linguistic study of cues and miscues in reading, *Elementary English* **42**:6.639-643 and reprinted in *Theoretical Models and Processes of Reading*, 3rd edition, eds. Harry Singer and Robert B. Ruddell. Newark, Delaware, International Reading Association, pages 129-134.
- Goodman, Kenneth S. (1969), Analysis of oral reading cues : applied psycholinguistics, *Reading research quarterly* **5**.9-30 and reprinted in *Language and Literacy: the Selected Writings of Kenneth Goodman*, ed. F. Gollasch. 1982, Boston, Routledge & Kegan Paul, pages 123-134. .
- Goodman, Yetta M. (1997), Reading diagnosis : qualitative or quantitative ?, *The Reading Teacher* **50**:7.534-38. Originally published in volume 26, 1972, pages 27-32.
- Goodman, Yetta M. & C.L. Burke (1972), *Reading miscue inventory*. London: Collier-Macmillan.
- Grimes, Joseph E., Augustus B. Marwief & Amy Bauernschmidt (1963), Several Kru orthographies in *Orthography Studies: Articles on New Writing Systems*, ed. W.A. Smalley, 114-119. London: United Bible Societies.
- Gudschinsky, Sarah C. (1959), Recent trends in primer construction, *Fundamentals and Adult Education* **II**.67-96.
- Hollenbach, Barbara E. (1978), Choosing a tone orthography for Copala Trique, *SIL Notes on Literacy* **24**.52-61.
- Jaffré, Jean-Pierre (1997), From writing to orthography: the functions and limits of the notion of system in *Learning to Spell: Research, Theory, and Practice across Languages* ed. C.A. Perfetti, L. Rieben & M. Fayol, 3-20. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Jaffré, Jean-Pierre (ed.) (2003), *Dynamiques de l'écriture : approches pluridisciplinaires* (Faits des langues : revue de linguistique 22). Paris: Ophrys.
- Kassan, Balaïbaou Badameli (2000), De l'influence du ton consécutif dans les formes de l'aoriste en kabiyè, *Cahiers voltaïques / Gur Papers* **5**.13-22.
- Roberts, D. (2010). Exploring written ambiguities can help assess where to mark tone. *Writing Systems Research*, 2(1), 25-40.

- King, Robert D. (1967), Functional Load and Sound change, *Language* **43**.831-852.
- Kutsch Lojenga, Constance (1993), The writing and reading of tone in Bantu languages, *SIL Notes on Literacy* **19**.19.
- Lébibkaza, Kézié (2003), Tones, syllables and consonants in Kabiye: interaction of suprasegmental features on consonants. Paper presented at the 33rd International Colloquium on African Languages and Linguistics, Leiden, Pays-bas.
- Lébibkaza, Kézié Koyenzi (1994), Les tons des verbes kabiye dans les formes de l'inaccompli in *Sprachen und sprachzeugnisse in Afrika: Eine Sammlung philologischer Beiträge, Wilhelm J.G. Möhlig zum 60. Geburtstag zugeeignet*, ed. T. Geider & R. Kastenholz, 263-279. Cologne: Rüdiger Köppe.
- Lébibkaza, Kézié Koyenzi (1999), *Grammaire kabiye: une analyse systématique - phonologie, tonologie et morphosyntaxe*. Cologne: Rüdiger Köppe.
- Mfonyam, Joseph Ngwa (1989), *Tone in Orthography: the Case of Bafut and Related Languages*: Université de Yaoundé, Cameroun. Thèse d'état.
- Pidassa, Emmanuel & David Roberts (2005), *Compte rendu d'une consultation régionale sur l'orthographe kabiye au Centre Régional des Affaires Sociales de Kara*. Kara: SIL.
- Pidassa, Emmanuel & David Roberts (2006), *Compte rendu d'une consultation régionale sur l'orthographe kabiye au Centre Régional des Affaires Sociales de Kara*. Kara: SIL.
- Pidassa, Emmanuel & David Roberts (2008), *Compte rendu d'une consultation régionale sur l'orthographe kabiye au Collège Protestant de Lomé*. Kara: SIL.
- Pike, Kenneth (1947), *Phonemics: a Technique for Reducing Languages to Writing*. Ann Arbor, Michigan: University of Michigan Press.
- Pike, Kenneth (1948), *Tone Languages: a Technique for Determining the Number and the Type of Pitch Contrasts in a Language*. Ann Arbor: University of Michigan press.
- Powlison, Paul S. (1968), Bases for formulating an efficient orthography, *The Bible translator* **19**:2.74-91.
- Roberts, David (2002), *Les classes tonales du verbe en kabiye*: Université de la Sorbonne nouvelle, Paris III. Mémoire de maîtrise.
- Roberts, David (2003a), *La tonologie des préfixes de modalité en kabiye*: Université de la Sorbonne nouvelle, Paris III. Mémoire de DEA.
- Roberts, David (2003b), Tone spreading in the Kabiye associative noun phrase, *Cahiers voltaïques / Gur Papers* **6**.95-100.
- Roberts, David (2008a), *L'orthographe du ton en kabiye au banc d'essai. Thèse de doctorat (INALCO, Paris)*. Villeneuve d'Ascq: Atelier national de reproduction des thèses.
- Roberts, David (2008b), The Two Kabiye Orthographies: A Sociolinguistic and Linguistic Comparison, *Written Language & Literacy* **11**:1.49-72.
- Roberts, David (2013), A tone orthography typology in *Typology of Writing Systems*, ed. S.R. Borgwaldt & T. Joyce, 85-111. Amsterdam: John Benjamins.
- Rogers, Henry (2006), *Writing Systems: a Linguistic Approach*. Oxford: Blackwell.
- Saussure de, Ferdinand (1916), *Cours de linguistique générale*. Paris: Payot.
- Roberts, D. (2010). Exploring written ambiguities can help assess where to mark tone. *Writing Systems Research*, 2(1), 25-40.

- Schreiner, Robert (1979), *Reading Tests and Teachers : a Practical Guide*. Newark, Delaware: International Reading Association.
- Schroeder, Leila (2008), *Bantu Orthography Manual*. Dallas: SIL International.
- Sebba, Mark (2007), *Spelling and Society: the Culture and Politics of Orthography*. Cambridge: Cambridge University Press.
- Seifart, Frank (2006), Orthography development in *Essentials of Language Documentation*, ed. J. Gippert, N.P. Himmelmann & U. Mosel, 275-301. Berlin, New York: Mouton de Gruyter.
- Taylor, Insup & M. Martin Taylor (1983), *The Psychology of Reading*. London: Academic Press.
- Underwood, Geoffrey & Vivienne Batt (1996), The role of eye movements in reading in *Reading and Understanding*, 144-188. Oxford: Blackwell.
- Unseth, Peter & Carole Unseth (1991), Analyzing ambiguity in orthographies, *SIL Notes on Literacy* **65**.35-52.
- Wiesemann, Ursula, Etienne Sadembouo & Maurice Tadadjeu (1988), *Guide pour le développement des systèmes d'écriture des langues africaines*. Yaoundé: SIL et l'Université de Yaoundé.
- Williamson, Kay (1984), *Practical orthography in Nigeria*. Ibadan: Heinemann Educational Books.
- Zipf, George K. (1935), *The Psychobiology of Language*. Boston, Massachusetts: Houghton-Mifflin.
- Zipf, George K. (1949), *Human Behavior and the Principle of Least-Effort*. Cambridge, Massachusetts: Addison-Wesley.