# THIRTY YEARS OF TONE ORTHOGRAPHY TESTING IN WEST AFRICAN LANGUAGES (1977-2007)[1]

**David Roberts**

LLACAN[2] and SIL-Togo[3]

There is an ongoing debate about how tone should be represented in the emerging orthographies of African languages. One of the most significant strands in the debate is a small but growing body of literature describing formal experiments which test the different options. In this article, I present an overview of the existing repertoire which covers ten experiments and three decades. I adopt a comparative approach, examining all the experiments in parallel. I focus in turn on aims, design, sample profile, sample size, experience, training, test materials, tasks, scoring, results and interpretation. In conclusion, I offer some practical advice for future experimenters. I also attempt to identify whether any consensus is emerging about the profile of an optimal tone orthography.

Le débat sur la représentation du ton dans les orthographes émergeantes des langues africaines a déjà fait couler beaucoup d'encre. L'un des plus importants volets dans ce débat est une littérature, modeste mais croissante, décrivant des expériences formelles entreprises pour tester les différentes options. Nous présentons un survol de ce répertoire qui couvre dix expériences et trois décennies. Dans cet article, nous adoptons une approche comparative, en examinant l'ensemble des expériences en parallèle. Nous nous focalisons sur les objectifs, la conception, le profil de l'échantillon, la taille de l'échantillon, le degré d'expérience, le niveau de formation, les matériaux expérimentaux, les tâches, le scoring, les résultats et l'interprétation. En conclusion, nous offrons quelques conseils pratiques pour de futurs expérimentateurs. Nous tentons également de dégager un consensus en ce qui concerne le profil d'une graphie tonale optimale.

## 0. INTRODUCTION

It was over thirty years ago that an appeal was made for more empirical testing on the world's emerging orthographies (Gudschinsky, 1970: 24):

> "We urgently need a large number of perceptive and sensitive tests to find out to what extent the orthographies currently in use are in fact adequate […] Until such testing is done, we will continue to pay an outrageously high price in readability for relatively small gains in simplicity and to remain quite unaware of it."

Since then, numerous researchers working on the tone orthography of African languages have echoed this plea:

---

[1] 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. I would also like to thank Russell Bernard, Bernard Caron, Philip Davison, Joseph Mfonyam, Jacques Nicole, Steve Walter and an anonymous reviewer for reading and commenting on a draft version of this article or sections of it. Their comments have contributed significantly to its clarity and accuracy. An earlier French version (Roberts, 2008: 63-111) can be downloaded from www.orthographyclearinghouse.org/phdma.html.

[2] Langage, Langues et Cultures d'Afrique Noire, 7 rue Guy Môquet, 94801 Ville-juif, France (UMR 8135 du CNRS).

[3] SIL, B.P. 57, Kara, Togo.

> "An immediate area of further research is the area of tone orthography evaluation. There is a need to develop an objective and reliable method of evaluating the efficiency of tone orthographies […]" (Mfonyam, 1989: 535)

> "[…O]bjective experimental work on the writing of tone languages is rarely undertaken; this domain of investigation is uncharted territory […] Rigorous testing of a variety of tone marking options should be a core part of tone orthography design […] More experimentation is required." (Bird, 1999b: 86, 107-8)

> "[…O]nly a worldwide comparative program of experiments can resolve the question whether to mark tones in any given language at all, and if so, how much. One of our goals in this paper has been to reinforce the importance of continued experimentation […]" (Bernard et al., 2002: 346)

However, in spite of repeated rallying cries of this kind, anyone familiar with the subject is only too aware that this entire area of research has been sadly neglected. To my knowledge, there are only ten previous published experiments on the tone orthography of African languages (table 1). This list covers the full range of research: quantitative to qualitative; a few brief paragraphs to a work of several hundred pages; the most superficial to the most penetrating; the most informal to the most scientifically rigorous.

Table 1: Formal, published testing of the representation of tone in West African languages

| Language | Reference |
|---|---|
| Efik | (Essien, 1977) |
| Yoruba | (Klem, 1982) |
| Western Krahn | (Duitsman, 1986) |
| Bura | (Badejo, 1989) |
| Yoruba | (Fagborun, 1989) |
| Bafut | (Mfonyam, 1989: 309-348) |
| Limbum | (Mfonyam, 1989: 459-473) |
| Kom | (Bernard et al., 1995) |
| Dschang (Yemba) | (Bird, 1999b) |
| Kom | (Bernard et al., 2002) |

It will be helpful to identify certain links between the different experiments. The Western Krahn experiment was preceded by pre-test 1982, apparently never published (Duitsman, 1986: 3). Yoruba is the only language whose tone orthography has been the subject of two experiments by two different researchers. Klem's (1982) analysis being limited to three paragraphs in a longer article, it is very sketchy and leaves numerous questions concerning his methodology. Fagborun (1989) fills in at least some of the gaps a few years later.

The Bafut and Limbum experiments are twins, because the same author applied the same methodology to two closely related languages under one title (Mfonyam, 1989: 309-348, 459-473). Both experiments can be better appreciated if the reader is familiar with the rest of the doctoral thesis in which they are located, as well as other research that the author has contributed on the same theme (Mfonyam, 1982, 1986, 1990a, 1990b, 1996).

Kom, too, has benefited from two experiments designed jointly by the same authors, the first (Bernard et al., 1995) being the prototype for the second (Bernard et al., 2002). The title of the second article has had a somewhat chequered history. The version published in 2002, to which I will refer, is entitled "Does marking tone make tone languages easier to read?" A pre-publication version, available on the internet, is entitled "Language survival, popular literacy and tone marking". Bird (1999b: 91-94) refers to the 1997 manuscript version, entitled "Does tone need to be marked?"

Finally, Bird's (1999b) research describing an experiment in the Dschang language is one of a trilogy of articles that deal with different aspects of the same theme. The other two articles provide the background to the first. One outlines the different strategies available for representing tone in African languages (Bird, 1999a). The other describes the sociolinguistic context in Cameroon where the author did his field research (Bird, 2001). It is helpful to read each of these articles in the light of the others. But because of space limitations, I will only be referring to the first of the three articles, the one which describes the formal experiment itself.

Why do I consider that a critical overview of the literature is necessary thirty years after Essien's (1977) pioneering research, when three other authors have already included such overviews in their own research (Bernard et al., 1995: 28-30, 2002: 339-340; Bird, 1999b: 86-94; Wiesemann, 1981: 38-41)? Well, firstly the survey which follows takes a comparative approach. Rather than describe each experiment one after each other as previous researchers have done, I will examine each aspect of the testing methodology in turn, comparing the different approaches of all the authors in parallel. Secondly, this summary includes experiments in four languages to which none of the other overviews makes reference, namely Bura, Western Krahn, Yoruba and Limbum (Badejo, 1989; Duitsman, 1986; Fagborun, 1989; Klem, 1982; Mfonyam, 1989).

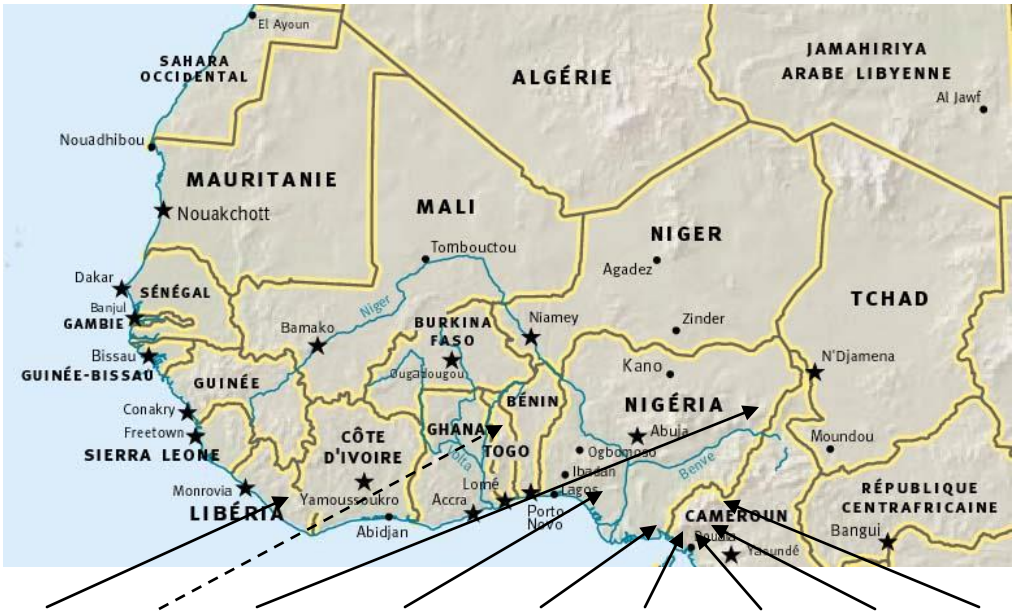
I indicate orthographic script between vertical lines | a | and phonetic script between square brackets [ a ].

## 1. BACKGROUND INFORMATION

### 1.1 GEOGRAPHY

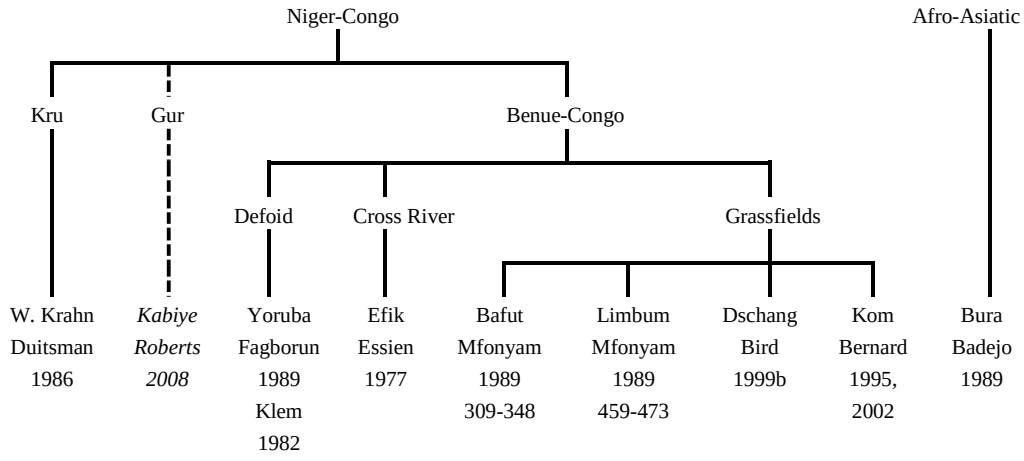
Not only is the literature extremely limited in size, but also in its geographical scope. The title of this article refers to West Africa, yet six of the experiments focus on an even more limited area: the borderland between Nigeria and Cameroon (figure 1 page 202). My own research on Kabiye (Roberts, 2008) brings witness from Togo, a hitherto unrepresented country. I will not be referring to this work, because it falls outside of the three decades encompassed by this article. In any case, it is for others to assess its merits.

Figure 1: Geographical area represented in the literature




| W. Krahn | *Kabiye* | Bura | Yoruba | | Efik | Bafut | Limbum | Kom | Dschang |
|---|---|---|---|---|---|---|---|---|---|
| Duitsman | *Roberts* | Badejo | Fagborun | Klem | Essien | Mfonyam | Mfonyam | Bernard | Bird |
| 1986 | *2008* | 1989 | 1989 | 1982 | 1977 | 1989<br>309-348 | 1989<br>459-473 | 1995,<br>2002 | 1999b |

Figure 2: Linguistic families represented in the literature



Furthermore, West Africa appears to be the only region of the continent which has seen any published tone orthography experiments at all. I invite correction on this point, but as far as I am aware, the whole of Central, East and Southern Africa is entirely unrepresented in the literature. And to my knowledge, neither Asia nor the Americas nor the Pacific fare any better. In the current state of things, West Africa remains the world's crucible for tone orthography experimentation.

### 1.2 GENEALOGY

The geographical limitation gives rise, not surprisingly, to a genealogical restriction. All but one of the experiments target languages of the great Niger-Congo phylum (see figure 2, page 202). The exception is Bura (Badejo, 1989), which is an Afro-Asiatic language. Bird (1999b: 4) relegates this experiment to a footnote, claiming that it is not sufficiently detailed, neither in its methodological description nor in its analysis, to merit inclusion in his overview. But I include it because it is the only language in the repertoire which falls outside of the dominant phylum.

Furthermore, the five most recent studies, which represent half of all that exists and all that has been published in the last 20 years, have focused solely on the Grassfields subfamily of Bantu languages. My own research on Kabiye (Roberts, 2008), though outside the chronological scope of this article, brings a new linguistic family to the repertoire, the Gur family.

## 2. TONE SYSTEMS AND TONE ORTHOGRAPHY

In this section I will review what the different authors provide by way of descriptions of the various tone systems and the tone orthographies which represent them.

Phonetic data appears between square brackets [ a ] and orthographic data between vertical lines | a |. In most cases, one of the orthographies tested is the standard orthography, or at least was at the time when the experiment was performed. In these cases, I add the letter (S) to identify it. Whenever I cite an experimental orthography, I add the letter (E) to identify it.

The reader should bear in mind that some of the languages have not (or had not at the time they were described) benefited from an autosegmental analysis, and for this reason the description of the tone system is sometimes lacking. For example, the question of the number of phonemic tones is not clearly elucidated in Bura and Efik (Badejo, 1989: 45-46; Essien, 1977: 155). Neither author distinguishes between a phonemic M tone and a downstepped H tone. As it happens, in neither Bura nor Efik does the standard orthography mark tone. Still, some of the data which Badejo (1989: 46) cites is helpful in proving tonal contrasts:

Bura lexical tone and tone orthography

| | | Pronunciation | Orthography Accentless (S) | Accentual (E) | |
|---|---|---|---|---|---|
| 1 | H | [ **fá** ] | \| **fa** \| | \| **fá** \| | long life |
| 2 | L | [ **fà** ] | \| **fa** \| | \| **fà** \| | to remove / collect |
| 3 | H | [ **tsá** ] | \| **tsa** \| | \| **tsá** \| | to beat |
| 4 | L | [ **tsà** ] | \| **tsa** \| | \| **tsà** \| | he / she / it |
| 5 | HL | [ **wúlà** ] | \| **wula** \| | \| **wúlà** \| | see |
| 6 | L | [ **wùlà** ] | \| **wula** \| | \| **wùlà** \| | awake |

And in Efik, Essien (1977: 156) offers the following sentences as examples of the interplay between lexical and grammatical tone:

Efik lexical and grammatical tone and tone orthography

| | | | |
|---|---|---|---|
| 7 | Accentless orthography (S): | \| **ekpat ubɔk anwan mi okpon** \| | |
| 8 | Pronunciation: | [ **ékpát úbɔ̀k ànwàn mì ókpōn** ] | My wife's arm is big |
| 9 | Pronunciation: | [ **ékpát úbɔ̀k ànwàn mì ókpón** ] | It is my wif'e's arm which is big |
| 10 | Pronunciation: | [ **èkpàt úbɔ̀k ànwàn mì ókpōn** ] | My wife's handbag is big |
| 11 | Pronunciation: | [ **èkpàt úbɔ̀k ànwàn mì ókpón** ] | It is my wife's handbag which is big |

The Yoruba orthography has a long pedigree, the use of diacritics having been recommended in the first ever Yoruba orthography standardisation meeting in 1875 (Fagborun, 1989: 82). In principle, the three phonemic tones are represented orthographically, H tone by an acute accent, L tone with a grave accent, and M tone by absence of an accent (ibid. 1989: 76):

Yoruba lexical tone and tone orthography

| | Melody | Pronunciation | Accentual orthography (S) | |
|---|---|---|---|---|
| 12 | MH | [ **ígbá** ] | \| **igbá** \| | calabash |
| 13 | L | [ **ìgbà** ] | \| **ìgbà** \| | time |
| 14 | M | [ **ígbā** ] | \| **igba** \| | two hundred |
| 15 | LH | [ **ìgbá** ] | \| **ìgbá** \| | garden egg |
| 16 | ML | [ **ígbà** ] | \| **igbà** \| | sort of climbing rope |

Klem (1982: 20), citing Abraham (1958: 511), offers a similar set of examples. It is not clear why Klem omits some of the acute and grave accents in his examples. I have added them, so that this data set concurs with Fagborun's (1989: 76) explanation of the Yoruba standard orthography conventions. Alongside this, I also cite one of Klem's experimental orthographies, the zero representation. (It is not possible to deduce from

Klem's description what his other experimental orthography, a partial representation, would look like).

Yoruba lexical tone and tone orthography

| | Melody | Pronunciation | Orthography | | |
|---|---|---|---|---|---|
| | | | Accentual (S) | Accentless (E) | |
| 17 | ML | [ **ōkò** ] | \| **okò** \| | \| **oko** \| | boat |
| 18 | MH | [ **ōkó** ] | \| **okó** \| | \| **oko** \| | hoe |
| 19 | M | [ **ōkō** ] | \| **oko** \| | \| **oko** \| | husband |
| 20 | L | [ **òkò** ] | \| **òkò** \| | \| **oko** \| | spear |

However, Yoruba tone orthography conventions are variable in practice, and often completely abandoned. This is a major source of difficulty when the reader, even if experienced, encounters a previously unknown text (Fagborun, 1989: 77-78; Klem, 1982: 22-23).

The Western Krahn orthography differs on either side of the Liberia ~ Ivory Coast border (Gordon, 2005). Duitsman investigates the Liberian variety. Established in 1973, it employs four accents to represent three discrete tones (H, M, L) and a rising contour tone (MH) (Duitsman, 1986: 2; Jim Laesch, personal communication). Duitsman's concern is to test this system against the widely practiced Ivorian convention of using punctuation symbols in word initial and final position to signal tone.[4] For example (based on Kutsch Lojenga, 1993: 13-14):

Table 2 : Marking tone with punctuation in Ivorian languages

| | |
|---|---|
| "CV | very high tone |
| 'CV | high tone |
| CV | mid tone (unmarked) |
| -CV | low tone |
| =CV | very low tone |
| -CV' | low to high contour tone |
| CV- | mid to low contour tone |

The languages in question are typically monosyllabic. They are highly isolating in their root structures, so if polysyllabic orthographic words occur at all, they tend to be compounds. They usually have more than three phonemic tones and numerous tonal contours on single syllables. Kutsch Lojenga provides a brief assessment of the advantages and disadvantages of punctuation marks to mark tone.

The convention was first adopted in the Blowo and Gweetao varieties of Yacouba (or Dan; Bolli, 1978; 1989; 1991; Hartell, 1993a, 1993b; Kutsch Lojenga, 1993; SIL, 1981, 1982). Since then, it has been widely replicated in numerous Kru, Mande and Kwa languages (ILA, 1979: 18-21). Each language chooses what is necessary from the basic set

[4] I would like to express my thanks to Eddie Arthur, Margaret Bolli, Jonathan Burmeister, Philip Davison, Constance Kutsch Lojenga, James Laesch and Philip Saunders for providing the information in this section.

of symbols, making up combinations appropriate to the tone system in question. The enthusiasm with which Ivorian field workers often speak of this system does not go unnoticed, even though it has never been adopted outside of this limited geographic area.

Duitsman himself does not cite any data examples in his experimental orthography. However, he does cite a few examples in the accentual standard orthography. So I have inferred how these words should be written in the experimental orthography from Kutsch Lojenga's explanation above and in discussion with an experienced field worker (James Laesch, personal communication):

Liberian Western Krahn lexical tone and tone orthography

| | | Pronunciation | Orthography | | |
|---|---|---|---|---|---|
| | | | Accentual (S) | Punctuation (E) | |
| 21 | H | [ **kó** ] | \| **kó** \| | \| **'ko** \| | frog |
| 22 | MH | [ **kǒ** ] | \| **kǒ** \| | \| **ko'** \| | chalk |
| 23 | M | [ **kōō** ] | \| **koo** \| | \| **koo** \| | mongoose |
| 24 | L | [ **kòò** ] | \| **kòò** \| | \| **-koo** \| | bath house |

Western Krahn has never implemented the punctuation system, however it is in use by the Tchien (Eastern) Krahn in Liberia (James Laesch, personal communication).

It only remains now to introduce the four Grassfields languages, which of all the languages in the repertoire have the deepest tone systems. In Bafut, eleven surface tone melodies occur on tone bearing units: three discrete level tones (H, M, L); four simple contour tones (HL, ML, LM, LML), non-automatic downstep either before a discrete level tone ($^{\downarrow}$H) or in combination with a contour tone ($^{\downarrow}$HL, H$^{\downarrow}$H) and upstep ($^{\uparrow}$L) (Mfonyam, 1989: 49-90). The following examples demonstrate the role of grammatical tone in Bafut. Since no tone orthography had been developed at that time of the experiment, I will cite only the surface representation, which Mfonyam himself devised for the purposes of the experiment (1989: 333):

Bafut grammatical tone

| | Surface orthography (E) | |
|---|---|---|
| 25 | \| **a ghɛ́ɛ̄ mfá nɨbɔ'ɔ nyà** \| | he has gone to give the pumpkin |
| 26 | \| **a ghɛ̄ɛ̄ mfá nɨbɔ'ɔ nyà** \| | he has gone and given the pumpkin |
| 27 | \| **á ghɛɛ̄ mfá nɨbɔ'ɔ nyà** \| | he is (in the process of) going to give the pumpkin |

The tonal system of Limbum has three discrete tones (H, M and L) five contour tones (L$^{\downarrow}$L, HL, HM, LM$_{=LH}$ and ML) and non-automatic downstep ($^{\downarrow}$H) (ibid: 435, 442). Like Bafut, at the time of doing his research, no tone orthographies had been adopted for Limbum. Mfonyam undertook his two experiments with a view to establishing what would be the optimal solutions. In Limbum, at least, his recommendations were adopted afterwards (ibid: 529). Mfonyam cites the following examples to demonstrate lexical tone in Limbum. I will cite only the surface experimental orthography:

Limbum lexical tone and tone orthography

| | Surface Orthography (E) | |
|---|---|---|
| 28 | \| **báa** \| | madness |
| 29 | \| **bâa** \| | two |
| 30 | \| **baa** \| | corn fufu |
| 31 | \| **bǎa** \| | father |
| 32 | \| **bàa** \| | hate! |
| 33 | \| **bàa** \| | bag |

Kom generates eight surface tone melodies, including three discrete level tones (H, M and L) and five contour tones (HL, ML, LH, LM and MH). Only two of these tones are represented in the orthography: the grave accent |`| signals the level L tone, and the circumflex | ̂ signals the HL contour (Bernard et al., 2002: 341). Bernard et al. provide no linguistic data to demonstrate this.

Dschang has two tones, which are sometimes faithful to the underlying form (H, L) and sometimes preceded by non-automatic downstep (ꜜH, ꜜL, ꜜꜜH). In addition, floating tones are attested (L°) which can themselves be preceded by non-automatic downstep (ꜜL°).

Bird (1999b: 85), citing the accentless orthography he later uses for testing purposes, provides evidence of the lexical and grammatical functions of tone:

Dschang lexical tone

| | Accentless orthography (E) | Musical pitch | |
|---|---|---|---|
| 34 | \| **lətɔŋ** \| | 2 B | feather |
| 35 | \| **lətɔŋ** \| | 2 3 | reading |
| 36 | \| **lətɔŋ** \| | 2 2̲ | navel |
| 37 | \| **lətɔŋ** \| | 2 | finishing |

Dschang grammatical tone

| | Accentless orthography (E) | Musical pitch | |
|---|---|---|---|
| 38 | \| **əfɔ tɔŋɔ mɔ** \| | 2B23 | The chief called the child (near past) |
| 39 | \| **əfɔ tɔŋɔ mɔ** \| | 22223 | The chief calls the child (present) |
| 40 | \| **əfɔ tɔŋɔ mɔ** \| | 22112 | The chief will call the child (near future) |

The Dschang tone system, like those of the other Grassfields languages, is extremely deep. However, the tone orthography adopted in the 1980's is shallow and graphically dense. It includes three diacritics: the acute accent | ́ |, the macron | ̄ | and the apostrophe | ' | (Bird, 1999b: 94, 111):

Dschang tone orthography (S)

41 | **Kaŋ pɔ́ mbhū́ é lelá' ŋ́gɔ̄ mésō, ḿbú ńzīŋ́ɛ́ ta' enɔ.** |
Once upon a time, a squirrel and a dog were friends and always went about together.

## 3. FIELD METHODOLOGY

### 3.1 AIMS

Experimental design is wholly driven by the goals of the experiment, so a clear statement of goals is necessary at the outset. Any design critique must be made in the context of the experimenter's stated aims, and the limitations in interpretation with respect to any other aims pointed out. I will explore this principle in the framework of a tone orthography typology which I have described in detail elsewhere (Roberts, 2008: 40-62):

Table 3: A tone orthography typology

| | Parameter | Example |
|---|---|---|
| i | Domain | sound-based (phonographic)<br>meaning-based (semiographic) |
| ii | Target | if sound-based: H, M, L, contour tones, downstep etc.<br>if meaning-based: person, number, gender, tense, aspect, mode, etc. |
| iii | Symbol | accents, punctuation, letters, numbers etc. |
| iv | Position | superscript, subscript, to right or to left of letter, morpheme, word, sentence etc. |
| v | Density | on a scale from zero to exhaustive |
| vi | Depth | on a scale from surface, through shallow, to deep |

To start with the simplest approach, if the researcher's stated goal is simply to compare orthography A with orthography B, then it is the researcher's right to design the experiment accordingly. For example, Duitsman (1986) tests an accentual orthography, the standard, against one which uses the Ivorian punctuation mark convention (see figure 2, page 205). Duitsman never claims to understand individual parameter effects which the experiment has not been designed to test. His experiment lies solely within the 3rd parameter, the choice of symbol:

Table 4: 3rd parameter, the choice of symbol

| Language | Reference | 1st orthography | 2nd orthography |
|---|---|---|---|
| Western Krahn | (Duitsman, 1986) | accents (S) | punctuation |

Similarly, numerous experiments focus on the 5th parameter, the choice of graphic density:

Table 5: 4th parameter, the choice of graphic density

| Language | Reference | zero | partial | exhaustive |
|---|---|---|---|---|
| Yoruba | (Klem, 1982) | ✓ | ✓ | ✓(S) |
| Bura | (Badejo, 1989) | ✓(S) | | ✓ |
| Kom | (Bernard et al., 1995) | ✓ | ✓(S) | |
| Kom | (Bernard et al., 2002) | ✓ | ✓(S) | |
| Dschang | (Bird, 1999b) | ✓ | | ✓(S) |

Of course, experiments which only explore one parameter of interest are intrinsically unable to provide holistic information. In many experiments, multiple parameters are expected to affect the outcome of interest. Particularly where the setting of

one parameter is likely to affect the impact of another, a well designed experiment will almost certainly involve the introduction of multiple parameters. The skill is to do this in such a way that at the end of the experiment it is possible to disentangle the effects of the individual parameters and understand their interactions. To explore this principle further, I will take two of the experiments as case studies, Efik and Limbum.

In Efik, Essien himself labels both orthographies as partial and exhaustive (1977: p.160, table 1; p.161), which implies that the primary purpose of the experiment is to test the 5th parameter, that of graphic density. However after a careful re-reading, one might conclude that Essien's purpose was to test a lexical representation against a grammatical one (ibid. p. 158, 163), elements which sit within the 1st parameter, the choice of language domain represented. Furthermore, certain references in the article suggest that the experiment is designed to test a deep representation, the isolated word, against a shallow one (ibid. p. 155-156), which is specific to the 6th parameter, the level of opacity. So the Efik experiment, for lack of a clearly stated aim, does not always succeed in disentangling the interaction between the different parameters.

Table 6: Targeted parameters in the Efik experiment

| Language | Reference | Parameters | | | | | |
|---|---|---|---|---|---|---|---|
| | | i | ii | iii | iv | v | vi |
| | | Domain | Target | Symbol | Position | Density | Opacity |
| Efik | (Essien, 1977) | ✓ | | | | ✓ | ✓ |

In the Limbum experiment, closely following the methodology of his prototypical Bafut experiment, Mfonyam prepared four experimental tonal orthographies. The first, which he calls “stable”, signals L tone and the contour tones containing a L tone (i.e. L, HL, $LM_{(=LH)}$ et ML). The second does the opposite, signalling H tone and the contour tones which contain a H tone (i.e. H, HL, HM[5] and $LH_{=LM}$). The third is a minimal representation, which signals discrete level L tones only. The fourth marks tone as exhaustively as possible[6]. I summarise these four orthographies in table 7. Recall that in the case of the minimal representation, there is no one to one grapheme ~ phoneme correspondence, because the two accents are used only sporadically (Mfonyam, 1989 p. 460-461):

[5] The melody HM is represented as H. It only occurs only on grammatical morphemes and constructions and is in free variation with H (Joseph Mfonyam, personal communication).

[6] $^{\downarrow}$L is represented as L. L and $^{\downarrow}$L are allotones, with $^{\downarrow}$L occuring only in utterance final position, and L elsewhere. Similarly, $^{\downarrow}$H is represented as H (Joseph Mfonyam, personal communication).

Table 7: The four Limbum experimental tone orthographies

| | H | M | L | HL | HM | $LM_{=LH}$ | ML |
|---|---|---|---|---|---|---|---|
| stable | | | ̀ | ̂ | | ̌ | ̂ |
| unstable | ́ | | | ̂ | ́ | ̌ | |
| minimal | | | ̀ | | | | |
| surface | ́ | | ̀ | ̂ | ̼ | ̌ | ̂ |

So what are the aims of this experiment in terms of the six parameters presented in table 3 above? First, Mfonyam compares a stable representation with an unstable one, in order to determine whether it would be better to signal L tone with a grave accent or H tone with an acute accent. These are the concerns of the 2nd and 3rd parameters, the choice of target and symbol. In addition, the four experimental orthographies, even though they are all partial representations, are all of differing degrees of density, which is the concern of the 5th parameter. Then the experience also examines the effect of orthographic depth, which is the 6th parameter. And in addition, the minimal representation introduces the 1st parameter, the language domain represented, because it represents the lexicon and grammar rather than phonology. In fact the 4th parameter of diacritic position is the only parameter which is not treated the experiment. All the experimental orthographies place accents in the classic superscript position.

Table 8: Targeted parameters in the Limbum experiment

| | i<br>Domain | ii<br>Target | iii<br>Symbol | iv<br>Position | v<br>Density | vi<br>Opacity |
|---|---|---|---|---|---|---|
| stable | | ✓ | ✓ | | ✓ | ✓ |
| unstable | | ✓ | ✓ | | ✓ | ✓ |
| minimal | ✓ | | | | ✓ | |
| surface | | | | | ✓ | ✓ |

Mfonyam's research is particularly ambitious. No other researcher to date has done so much preparatory background research before undertaking experimentation. The author deserves praise for having successfully tackled a thorny issue with clarity and in such detail. In an experiment the stated aims of which require the introduction of multiple parameters, the results might have been more robust and convincing if the author had found a way of disentangling the impact that each of those parameters had on the others. I do not underestimate the difficulties inherent in this requirement.

Before proceeding to the next section, I will a brief word about Fagborun's research, the only one the literature to which the six parameter typology cannot be applied. The value of his approach in his comparison of the de jure orthographic conventions with the de facto practice in Yoruba (Fagborun, 1989: 77-78).

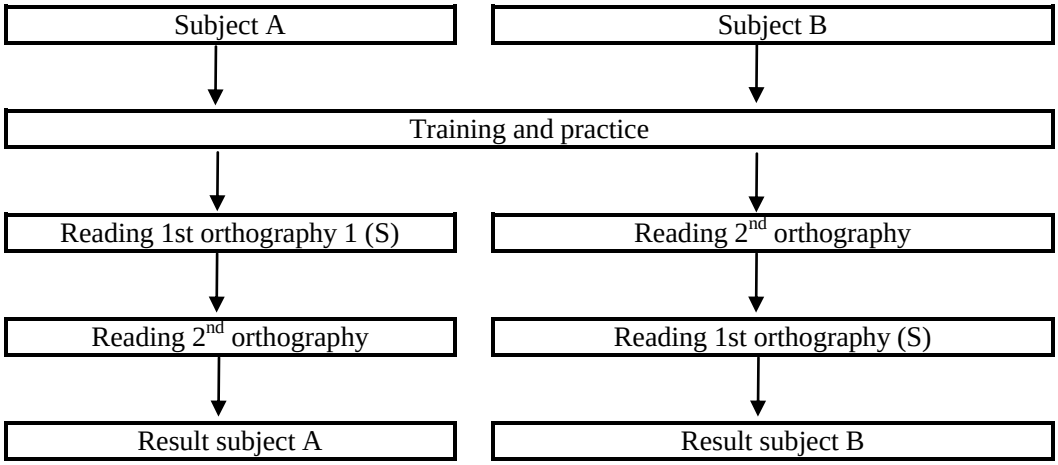
## 3.2 DESIGN

It is possible to divide the different experiments into two large types on the basis of their methodological design, which can be either a within-subject or a between-groups design.

In a within-subject design, the performance of each subject is tested with two orthographies. This is the case with the Bura experiment, where Badejo's (1989) aim was to test a zero representation, the standard, against an experimental accentual representation. Bird (1999b) simplifies his design still further in Dschang. Subjects had prior knowledge of an accentual orthography, the Standard, and the aim was to observe the effect on performance of these same subjects when the accents were removed. So Bird was able to remove the training phase entirely, because the subjects had nothing new to learn. The texts were presented one at a time, and the subjects had no opportunity to prepare them in advance. The second experiment in Kom also applies this model with slight modifications (Bernard et al., 2002).

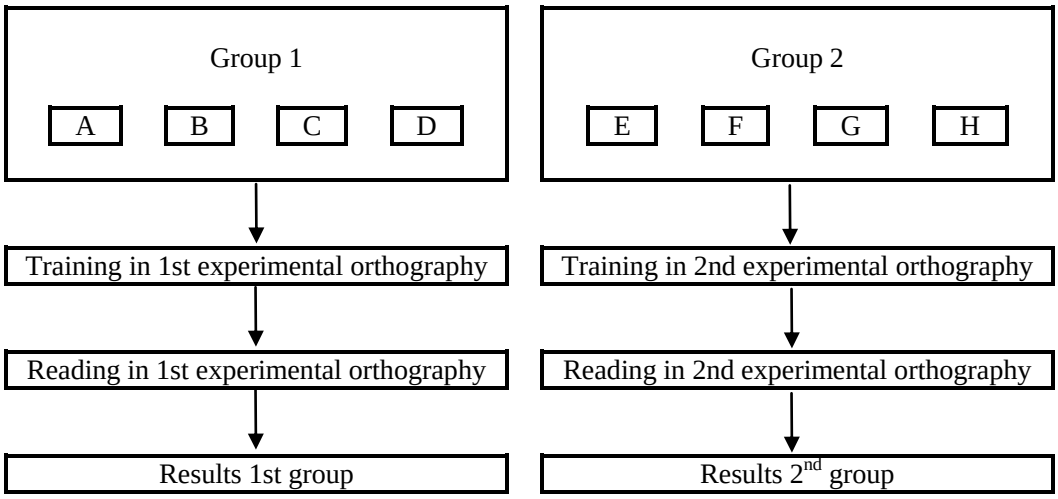
The design of the Kom and Dschang experiments are elegant and simple, all the more so because these two researchers realised the importance of taking precautions to avoid an ordering effect, as shown in figure 3. Bird varied the order in which the texts were presented (Bird, 1999b: 96), whilst Bernard et al. changed the presentation order in the second Kom experiment. Instead of showing the 50 sentences without accents on the first day and the same 50 phrases with accents the next day, they mixed both orthographies at random, numbering them from 1 to 100 (Bernard et al., 2002: 341-342). The absence of this precaution in Bura (Badejo, 1989) is a weakness in the experiment design.

Figure 3: Within-subject design




The second approach is a between-groups design. Here the sample is divided into several parallel groups (1, 2, 3...). For between-groups studies, an appropriate randomisation procedure of subjects to groups is crucial (A, B, C…). We will return to the question of what constitutes balanced groups later on (page 212). One group learns one experimental orthography at the same time as the other group learns another experimental orthography. Then each group is tested on what they have learned.

Figure 4: Between-groups design




Four experiments in the repertoire adopted the between-groups design, each of them adapting it to their own requirements. In Yoruba, Klem (1982) divided the sample into three balanced groups, and each was tested on one of the orthographies. Mfonyam (1989) follows more or less the same approach. For both the Bafut and Limbum experiments, he divides the sample into four groups, and each group learns one experimental tone orthography. In Western Krahn, Duitsman (1986) divided his sample into two balanced groups. One group was taught to mark tone with accents, and the other to mark tone with punctuation marks.

The between-groups design has the particular merit of ensuring the total separation of the different experimental orthographies. One cannot influence the other. But a disadvantage should also be noted, i.e. that in spite of all efforts to effectively randomise the procedure, some of the observed differences in performance may be due to unanticipated imbalance between the groups.

The only experiment which matches neither the within-subject nor the between-groups designs is Fagborun's (1989). It is a study of the variation between subjects within a single sample. Each subject's de facto performance in the dictation task is compared not with their own performance on a different task, nor with the performance of another group, but with the de jure conventions, the Yoruba standard orthography. The structure of this experiment is the simplest of all. Quite simply, Fagborun distributed a text to Yoruba teachers in several universities, then collected the responses and analysed them.

## 3.3 SAMPLE PROFILE

In the next two sections I will examine the question of sample selection strategy. The results of the experiment will only be valid and generalisable if both the profile and the size of the sample are right. Let us explore what constitutes a valid profile first.

On the one hand, some samples are clearly socially heterogeneous with regard to age, gender, origin, profession and education level and so on. It is the express choice of the researchers in Bura, Yoruba, Bafut and Limbum (Badejo, 1989; Klem, 1982;

Mfonyam, 1989). Others are more homogeneous, especially when the experiment takes place in an educational establishment. The Efik experiment targets a sample of expatriate university students (Essien, 1977: 158). The Western Krahn experiment targets school children (Duitsman, 1986: 5). Fagborun targets university students who have chosen to study Yoruba language and literature (Fagborun, 1989: 75).

But the reality is that, even within a more socially homogeneous sample, several variables come into play. Perfect homogeneity is not feasible, neither is it necessarily desirable. In Western Krahn, Duitsman works in three different schools, and in four year groups in each location. In his self-assessment of the experiment, he concedes that some students had prior knowledge of the accentual tone orthography, while others did not (Duitsman, 1986: 8). The students studying Yoruba language and literature were enrolled in several different Nigerian universities.

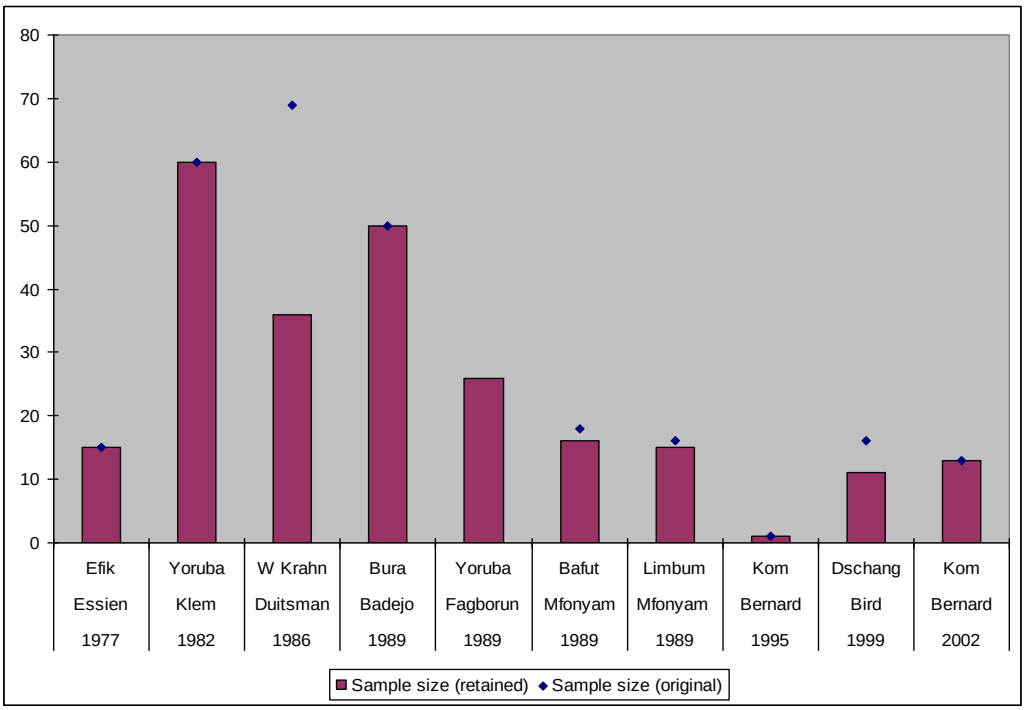
Authors sometimes incompletely describe the social profile of the sample. In Efik, it would have been instructive to know whether the sample is composed of men and women, and how much contact each of them has with their Efik homeland (Essien, 1977). And in Yoruba, neither author refers to the origin of their subjects. This is important information, since it is a language with several dialects (Fagborun, 1989; Klem, 1982).

Finally, and this is the most important point, it is not enough to list these variables for information only. The investigator also needs to control for them, to establish whether each independent variable has had an effect on the results. It is only the three most recent experiments, Kom (Bernard et al., 1995, 2002) and Dschang (Bird, 1999b) which include this analytical step of paramount importance. We will return later to present the results of their analyses in more detail.

### 3.4 SAMPLE SIZE

Once the sample profile has been identified, the next question concerns the appropriate sample size. The size of the samples varies widely, from 1 to 60 subjects. Such a large range immediately raises the question: what is the optimum sample size and how does one decide?

Figure 5: Sample size




Clearly, an initial large number acts as a kind of insurance against certain unpredictable technical and practical problems, such as missed recordings or unexpected absences. However, high numbers do not necessarily ensure successful results and are not a prerequisite for it. Three experiments with some of the most compelling results, the Kom and Dschang experiments (Bernard et al., 1995, 2002; Bird, 1999b), are also those based on the smallest samples. Bird had originally hoped to engage 40 Dschang subjects, but was prevented from doing so by purely practical considerations. Then of the 16 people who finally participated in the experiment, he had to eliminate five who were not up to the task (Bird, 1999b-96). So the final analysis is based on the results of just eleven subjects. But what this experiment lacks in sample size, it more than makes up for in its statistically rigorous approach.

It is worth noting that eliminating subjects can lead to biased results (though I am not suggesting that this happened in Bird's experiment). For example, people with less education might cope with a less dense orthography but not with a dense one, whilst people with more education might cope with either. In such a case, to eliminate subjects "not up to the task" could give results that are not illustrative of the real problem caused by the dense orthography. The reasons why data is missing can in itself be revealing.

If a large number of participants insures against unforeseen technical problems, the other side of the coin is that quantitative experiments are extremely data productive. One should guard against generating excessive amounts of data, because there is a risk of being swamped when it comes to the analysis. The most extreme case of this Fagborun (1989:

79), who originally contacted no less than 605 Yoruba students (a figure too high to represent on graph 1), of whom he finally retained 26, less than 5% of the total.

Let us turn finally to the only single subject experiment in the repertoire, the first Kom experiment (Bernard et al., 1995: 30). Numerous literacy researchers sing the praises of single subject research (for example Yetta Goodman, 1972b; Neuman and McCormick, 2000) but let us remember that it is usually reserved for qualitative experiments, particularly miscue analysis, not quantitative ones.

Bernard et al. themselves acknowledge the limitations of their approach (1995: 38-39). Clearly, if every individual in the relevant population could be guaranteed to react in the same way to the tests presented, it would only ever be necessary to test one person. However, for a variety of reasons people react in a variety of ways. The more variation there is between individuals, the larger sample size is needed. It ensures that the results obtained from the sample reflect the results which would be obtained if everyone to whom the results might ultimately be generalised were to be tested.

Nevertheless, the single-subject approach is by no means ill-conceived as a starting point. Its great value is that the first experiment became the prototype for the second which targets thirteen subjects, and the methodology of the second was modified in the light of the first. The first precautionary experiment was not only a pilot field test, but also a pilot analysis.

### 3.5 EXPERIENCE AND TRAINING

Experiments may target beginners or experienced readers. Given the different pedagogical implications, it is entirely normal and even desirable that these two broad categories of test be kept separate in experiment design, and this has been the case in all the experiments so far:

Table 9: Experience and training

| Language | Reference | Beginners | Experienced | Length of training and practice |
|---|---|---|---|---|
| Efik | (Essien, 1977) | ✓ | | 10 minutes |
| Yoruba | (Klem, 1982) | | ✓ | None |
| W. Krahn | (Duitsman, 1986) | ✓ | | 1½ hours |
| Bura | (Badejo, 1989) | | ✓ | Unknown |
| Yoruba | (Fagborun, 1989) | | ✓ | None |
| Bafut | (Mfonyam, 1989: 309-348) | ✓ | | 15 days |
| Limbum | (Mfonyam, 1989: 459 – 473) | ✓ | | 10 days |
| Kom | (Bernard et al., 1995) | ✓ | | 15 minutes |
| Dschang | (Bird, 1999b) | ✓ | ✓ | None |
| Kom | (Bernard et al., 2002) | | ✓ | “little” |

The issue of experience is closely linked to that of training. Certain experiments do not require a training phase. Fagborun (1989) does not need it because he wants to examine the existing competence of university students studying Yoruba language and literature, who in principle are already experienced. In Dschang and the previous Yoruba experiment, there was no need to devote time to teaching because the aim was to test the

effect when accents are taken away from an already known Standard Orthography (Bird, 1999b; Klem, 1982). But for all others, a minimum of instruction, training and practice was necessary. However, among all aspects of experiment design, this is the one which shows that the biggest range: from 10 minutes to 15 days.

In principle, one would expect to see an inverse correlation between the level of experience and the time allocated to training and practice. In other words, the less experienced the subjects the more they need training in advance. Unfortunately this has not always been the case. Ten minutes is not enough time to teach inexperienced Efik speakers most of whom have never encountered a tone orthography, as the experimenter himself concedes and others have noted (Bird, 1999b: 88; Essien, 1977: 162; Wiesemann, 1981: 41).

Amongst all the experimenters, it is Mfonyam that has the most impressive pedagogical track record. He build in no less than 15 days for the Bafut experiment and 10 days for the Limbum one (1989: 328, 461). Apart from him, no other researcher, even those targeting beginners (Duitsman, 1986: 5), has devoted more than a few hours to teaching and practice.

Mfonyam himself assumed the role of teacher in the Bafut and Limbum experiments, albeit with the help of a native speaker classroom assistant in the second case (1989: 329, 463). The drawback of this strategy is that the "observer paradox" (Labov, 1970: 32) is likely to exert an influence on performance. Some might accuse the researcher, rightly or wrongly, of having unwittingly shared his preference for one of the experimental tone orthographies as he was teaching. The potential for this could have been avoided by entrusting the task of teaching to trained assistants. But the other side of the coin is that the researcher's presence in the classroom permitted him to live through the experiment side by side with the subjects and closely follow their evolution. Mfonyam's two experiments are rich in qualitative observations which would not otherwise have been possible.

The observer paradox is even more pertinent in the case of an expatriate researcher. Bird gave over the task of recording to a Dschang local assistant, in order to remain invisible himself (1999b: 96). As for Duitsman (1986: 5), he engaged two Western Krahn teachers in the experiment. One had previous experience with the accentual orthography, while the other had to learn it for the purposes of the experiment. But before beginning, the two teachers were entirely comfortable with the two systems. If either one of them had any personal bias in favour of accents or punctuation marks, these were tempered by a strict control of pedagogical materials in terms of content, presentation and lesson duration.

### 3.6 MATERIALS

Written materials are of three types: educational, informational and experimental.

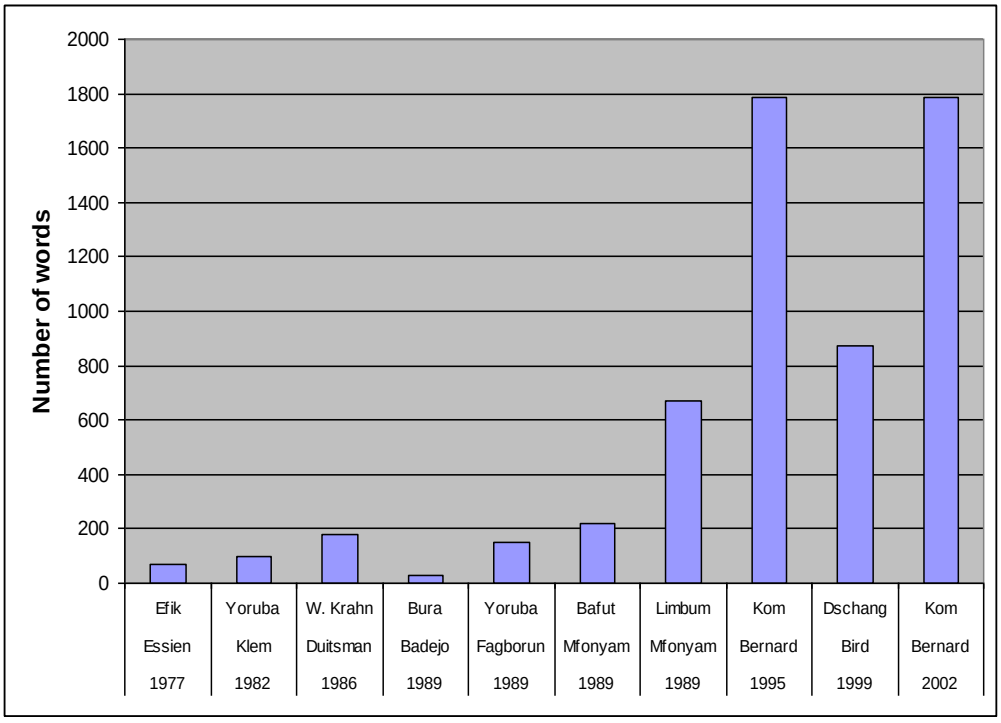
Pedagogical materials: If the experiment is to contain any training at all, it goes without saying that prior preparation of pedagogical materials cannot be bypassed. In Western Krahn Duitsman developed two courses identical in content apart from the tone orthography in question. He reproduces a sample lesson in his appendix (1986: 5-6, 9). As for Mfonyam, the fact that he devoted so much time to the teaching phase necessitated the development of an 80 page course for Bafut and a 60 page one for Limbum. He uses the

keyword approach to introduce the different tone melodies. He reproduces the full texts of these courses in his appendices (1989: 537-670).

Informational Materials: Two experimenters included guidelines to help the subjects understand what was being asked of them. These were written in English for the Efik experiment (Essien, 1977: 159) and in French for the Dschang experiment (Bird, 1999b: 96). Bird also included a questionnaire to glean sociolinguistic information, also in French.

Experimental Materials: All the experiments require the preparation of materials used in the execution of the tasks themselves. Figure 6 shows the size of the whole corpus used for all the tasks in any one experiment:

Figure 6: Size of corpus




The most striking feature of this graph is the large gap between the number of words used from the Limbum experiment onwards compared to all those that preceded it. The most extreme is the Kom corpus which contains 1,788 words (Bernard, personal communication. Neither published article mentions the exact corpus size.). The Dschang corpus does not reach these proportions, but it is still six times larger than the average of the five earliest works (Bird, 1999b, 111-114). Most researchers favour lists of words and sentences rather than texts:

Table 10: Experimental materials

| Language | Reference | Lists of sentences | Texts |
|---|---|---|---|
| Efik | (Essien, 1977) | ✓ | |
| Yoruba | (Klem, 1982) | | ✓ |
| W. Krahn | (Duitsman, 1986) | | ✓ |
| Bura | (Badejo, 1989) | ✓ | |
| Yoruba | (Fagborun, 1989) | ✓ | |
| Bafut | (Mfonyam, 1989: 309-348) | ✓ | ✓ |
| Limbum | (Mfonyam, 1989: 459 – 473) | ✓ | ✓ |
| Kom | (Bernard et al., 1995) | ✓ | |
| Dschang | (Bird, 1999b) | | ✓ |
| Kom | (Bernard et al., 2002) | ✓ | |

Essien (1977: 164-166) prepared a list of fourteen minimal sentences in Efik, each of which have multiple meanings. He wrote each of them (i) with zero tone representation (ii) with partial representation (lexical tones marked on nouns and verbs) and (iii) with exhaustive representation, thus highlighting the grammatical distinctions. For example:

The different tone orthographies used in the Efik experiment: minimal sentences

| | | | |
|---|---|---|---|
| 42 | Zero: | \| **mmɔ etɔŋɔ ndɪdiɔŋ usuŋ** \| | They have begun to repair |
| 43 | Partial: | \| **mmɔ etɔ́ŋɔ́ ndidìɔŋ úsúŋ** \| | the road |
| 44 | Exhaustive: | \| **ḿmɔ̀ étɔ̄ŋɔ̄ ńdídìɔŋ úsúŋ** \| | |
| 45 | Zero: | \| **mmɔ etɔŋɔ ndɪdiɔŋ usuŋ** \| | They are beginning to repair |
| 46 | Partial: | \| **mmɔ etɔ́ŋɔ́ ndidìɔŋ úsúŋ** \| | the road |
| 47 | Exhaustive: | \| **ḿmɔ̀ étɔ̀ŋɔ̀ ńdídíɔŋ úsúŋ** \| | |
| 48 | Zero: | \| **mmɔ etɔŋɔ ndɪdiɔŋ usuŋ** \| | They have begun to to |
| 49 | Partial: | \| **mmɔ etɔ́ŋɔ́ ndidìɔŋ ùsùŋ** \| | prepare fufu |
| 50 | Exhaustive: | \| **ḿmɔ̀ étɔ̄ŋɔ̄ ńdídìɔŋ ùsùŋ** \| | |
| 51 | Zero: | \| **mmɔ etɔŋɔ ndɪdiɔŋ usuŋ** \| | They are beginning to |
| 52 | Partial: | \| **mmɔ etɔ́ŋɔ́ ndidìɔŋ ùsùŋ** \| | prepare fufu |
| 53 | Exhaustive: | \| **ḿmɔ̀ étɔ̀ŋɔ̀ ńdídìɔŋ ùsùŋ** \| | |

Mfonyam uses a similar set in the Bafut experiment (1989: 333). In this task, subjects were asked to add accents on unmarked sentences:

The different tone orthographies used in the Bafut experiment: minimal sentences

| | Orthography | | |
|---|---|---|---|
| 54 | stable (E) | \| **à ghèɛ mfa nɨbɔ̀'ɔ̀ nyâ** \| | He has gone to give the pumpkin |
| 55 | basic (E) | \| **a ghɛē mfá nɨbɔ'ɔ nyā́** \| | |
| 56 | minimal (E) | \| **à ghèɛ mfa nɨbɔ'ɔ nya** \| | |
| 57 | surface (E) | \| **a ghɛ́ɛ̄ mfá nɨbɔ'ɔ nyà** \| | |
| 58 | stable (E) | \| **à ghɛɛ mfa nɨbɔ̀'ɔ̀ nyâ** \| | He has gone and given the pumpkin |
| 59 | basic (E) | \| **a ghɛ̄ɛ̄ mfá nɨbɔ'ɔ nyā́** \| | |
| 60 | minimal (E) | \| **à ghɛ̄ɛ̄ mfa nɨbɔ'ɔ nya** \| | |
| 61 | surface (E) | \| **a ghɛ̄ɛ̄ mfá nɨ́bɔ'ɔ nyà** \| | |
| 62 | stable (E) | \| **a ghèɛ mfa nɨbɔ̀'ɔ̀ nyâ** \| | He is (in the process of) going to |
| 63 | basic (E) | \| **á ghɛē mfá nɨbɔ'ɔ nyá** \| | give the pumpkin |
| 64 | minimal (E) | \| **a ghèɛ̄ mfa nɨbɔ'ɔ nya** \| | |
| 65 | surface (E) | \| **á ghɛē mfá nɨbɔ'ɔ nyà**\| | |

Numerous researchers have criticized this approach for two reasons. On the one hand, it is not advisable to use artificially generated minimal word or sentence pairs, because they are most unlikely to coincide in natural reading contexts. Moreover, reading a series of words or sentences out of context, whether artificial or otherwise, is not a natural literacy activity (Bernard et al., 1995: 30, 2002: 340; Bird, 1999b: 88, 93; Wiesemann, 1981: 41). That is why, elsewhere in the same experiment, Mfonyam uses natural texts for the oral reading task[7]:

The different tone orthographies used in the Bafut experiment : natural texts

| | Orthography | |
|---|---|---|
| 66 | stable (E) | \| **À fɛ̀'è mâ, m̀bəŋ kân loo, a ghɛè ŋkwɛrə akoŋə a mbo mumaà yì ntsɨrə ta karə lɔ̀ɔ njoo jìi mə à kɨ tswe nî nyùuà.** \| |
| 67 | basic (E) | \| **A fɛ'ɛ̄ mâ, mbəŋ kán lóó, a ghɛɛ̄ ŋkwɛ̄rə̄ akōŋə̄ á mbō múmáa yi ntsɨ́rə̄ tā kárə́ lɔ̄ɔ̄ njōō jiī má a kɨ tswé nɨ nyúúa.** \| |
| 68 | minimal (E) | \| **A fɛ'ɛ mə, mbəŋ kan loo, a ghɛɛ ŋkwɛ̄rə̄ akoŋə a mbo mumaa yi ntsɨrə tā karə lɔɔ njoo jìi mə a kɨ tswe nɨ nyuua.** \| |
| 69 | surface (E) | \| **A fɛ'ɛ mə̀, mbəŋ kà n lōō, á ghɛ́ɛ ŋkwɛ̄rə̄ akóŋə̄ á mbó múmáa yi ntsɨ́rə̄ tā kárə́ lɔ̄ɔ̄ njōō jiī mə a kɨ tswē nî nyuūa.** \| |
| | | When she went it started raining and she went and took an umbrella from her sibling and covered herself and went about looking for the things she had to buy. |

[7] I am grateful to Joseph Mfonyam for providing the English glosses for examples 66 - 73.

Mfonyam also chooses natural texts for the oral reading task in the Limbum experiment (Mfonyam, 1989: 464-465):

| | Orthography | |
|---|---|---|
| 70 | stable (E) | \| **Tàta à m yuu mbàŋ, a koo ŋgunyăm yì tâ tɛ e bo dù ntaa àwo. E m lòr kwâa a lɛ njep bàa.** \| |
| 71 | basic (E) | \| **Tata à m yúu mbaŋ, á kóo ŋgunyăm yi tâ tɛ́ é bó du ntaa awo. É m lor kwâa á lɛ́ njép baa.** \| |
| 72 | minimal (E) | \| **Tàta à m yuu mbàŋ, a koo ŋgunyam yì ta tɛ e bo dù ntaa àwo. E m lòr kwaa a lɛ njep bàa.** \| |
| 73 | surface (E) | \| **Tàta à m yúu mbàŋ, á kóo ŋgunyăm yì tâ tɛ́ é bó dù ntaa àwo. É m lòr kwâa á lɛ́ njép bàa.** \| |

Tata has bought kernels. He caught the pig of his father. He will go to the market with it. He took corn fufu and put it in the bag.

Fagborun, writing in the same year, follows suite in his experiment (1989: 88), though the Yoruba speaker who provided this English translation considers his sentences to be unnatural (Seun Gloria Adewara, personal communication). For example:

Text in Yoruba standard orthography

74 \| **Idí rèé tí kì í fi í sọ nǹkan kan bí ọwọ́ bá ti tẹ̀ ẹ́. Se ni yó dà bí igi gẹ̀dú tí ẹni kan kò náání lẹ́sẹ̀ kan náà tí yó sì tẹ̀ lé àwọn géńdé márùn-ún tó dúró bí asọ́nà àwọn mẹ́wẹ̀ẹ̀wá tó ń gbé ilé olójúlé mẹ́wàá tí bàbá rẹ̀ kọ́ fún un.** \|

This is the reason why he ususally doesn't talk when he is caught. He then follows the five out of the ten guards that watch the ten houses his father built for him.

Bird also recognises the importance of using natural contexts. He chooses four texts, for example:

Standard and experimental orthographies in Dschang

| | | |
|---|---|---|
| 75 | Accentual (S) | \| **Kaŋ pɔ́ mbhɯ̄ é lelá' ŋ́gɔ̄ mésō, ḿbú ńzīŋɛ́ ta' enɔ. Pɔ́ lelá' ńnāŋ tɛ eshɯ̄' amɔ̄' álī'í, ḿbɛ́ á ápa, ńdɔk ŋgɯ̄ ɔ́ á ŋ́kā' ŋiŋ njɯ́ɯ́ a apumā.** \| |
| 76 | Accentless (E) | \| **Kaŋ pɔ mbhɯ e lela ŋgɔ meso, mbu nziŋɛ ta enɔ. Pɔ lela nnaŋ tɛ eshɯ amɔ alii, mbɛ a apa, ndɔk ŋgɯɔ a ŋka ŋiŋ njɯɯ a apuma.** \| |

Once upon a time, a squirrel and a dog were friends and always went about together. One day, they decided to get a sack and take themselves off to a grove to steal oranges.

Another dimension to be addressed in the selection of experimental materials is whether the text is already known or not. In Western Krahn and Limbum, the oral reading tasks were performed twice, first on a text of which the subjects had prior knowledge, and the other on an unknown text. But the authors do not distinguish between the two in the presentation of results (Duitsman, 1986: 6; Mfonyam, 1989: 464). In Kom, on the other hand, a variable was incorporated to indicate whether or not a phrase is a proverb (Bernard

et al., 1995: 32, 2002: 342-343). In Dschang, Bird incorporated the same binary variable for complete texts (Bird, 1999b: 97).

### 3.7 TASKS

In Table 11, the clear preference for the oral reading task is very obvious. In fact, all the experimenters except Fagborun (1989) included it:

Table 11: Tasks

| Language | Reference | Oral reading | Adding accents | Writing | Dictation | Self-evaluation |
|---|---|---|---|---|---|---|
| Efik | (Essien, 1977) | ✓ | | | | |
| Yoruba | (Klem, 1982) | ✓ | | | | |
| W. Krahn | (Duitsman, 1986) | ✓ | | | | |
| Bura | (Badejo, 1989) | ✓ | | | | |
| Yoruba | (Fagborun, 1989) | | | | ✓ | |
| Bafut | (Mfonyam, 1989: 309-348) | ✓ | ✓ | ✓ | | |
| Limbum | (Mfonyam, 1989: 459 – 473) | ✓ | ✓ | | | |
| Kom | (Bernard et al., 1995) | ✓ | | | | |
| Dschang | (Bird, 1999b) | ✓ | ✓ | | | ✓ |
| Kom | (Bernard et al., 2002) | ✓ | | | | |

As far as writing is concerned, adding accents on a prepared text is favoured in Dschang, Bafut and Limbum (Bird, 1999b: 96; Mfonyam, 1989: 331-333, 464, 467). This task has the great disadvantage of not being a natural literacy activity. Field workers often remark anecdotally that writers form all the consonants and vowels of the entire sentence first, then go back to add accents. This is an undesirable practice in the sense that it reflects how the writer sees the accents as not being critically important, as Fagborun observed in Yoruba (1989: 85).

Two writing tasks have been hitherto neglected. First, Fagborun's (1989) Yoruba experiment remains the sole representative of a dictation task. The advantage of this task is that the pace of the exercise is governed by the test administrator, so the subject is compelled to add accents while writing rather than filling them in later. Secondly, there is an creative writing task in the Bafut experiment (Mfonyam, 1989: 333). This is the writing activity par excellence, the most natural of all. The Bafut and Limbum pedagogical materials also include translation exercises from English and the completion of incomplete sentences (Mfonyam, 1989: 576-669).

The Dschang experiment is the only one requiring subjects to fill in a self-assessment evaluation (Bird, 1999b: 96). The questionnaire focused on the area of personal correspondence because it was felt that this literacy activity provides a good measure of ability when the subject is in an informal and unstructured context. Such a questionnaire is very valuable, because it is a rich source of sociolinguistic information, for example attitudes, skills and preferences, which can then be grafted into the analysis as independent variables.

## 3.8 SCORING

Overall, the different experiments take into account three dimensions in scoring, namely, accuracy, speed (measured in seconds, and sometimes divided into two phases, perception and oralisation) and comprehension:

Table 12: Scoring

| Language | Reference | Accuracy | Perception speed | Oralisation speed | Compre-hension |
|---|---|---|---|---|---|
| Efik | (Essien, 1977) | ✓ | ✓ | ✓ | |
| Yoruba | (Klem, 1982) | ✓ | | ✓ | |
| W. Krahn | (Duitsman, 1986) | ✓ | | | |
| Bura | (Badejo, 1989) | ✓ | | ✓ | |
| Yoruba | (Fagborun, 1989) | ✓ | | | |
| Bafut | (Mfonyam, 1989: 309-348) | ✓ | | | |
| Limbum | (Mfonyam, 1989: 459 – 473) | ✓ | | | |
| Kom | (Bernard et al., 1995) | ✓ | ✓ | ✓ | |
| Dschang | (Bird, 1999b) | ✓ | | ✓ | ✓ |
| Kom | (Bernard et al., 2002) | ✓ | ✓ | ✓ | |

All the experiments incorporate, one way or another, a measure of accuracy, but with different approaches. In Yoruba and Bafut, the number of miscues is counted (Klem, 1982: 24; Mfonyam, 1989: 334). The Dschang experiment distinguishes between lexical and grammatical tonal miscues (Bird, 1999b: 99). In Kom and Efik, the researchers use a binary measure, whether or not the pronunciation of the entire sentence was correct (Bernard et al., 1995: 31, 2002: 342; Essien, 1977: 160-161). Essien adds that it is not possible to measure the accuracy of the oral reading task in a Standard Orthography which does not mark tone, because we know in advance that it is ambiguous.

Most oral reading tasks measure the speed of oralisation, but the experimenters in Kom and Efik go further by measuring the speed of perception, in other words the time devoted to the study the text in advance (Bernard et al., 1995: 31, 2002: 342; Essien, 1977: 159). In these cases, the subject was encouraged to study each sentence as long as necessary to find the right meaning before starting to read.

A reliable comprehension measure is notable by its absence throughout the literature. This is rather surprising given that comprehension is the sine qua non of the reading process. Bird is the only one to have attempted it, in the Dschang experiment (1999b: 103), but he himself admits that a more accurate comprehension measure is desirable for future research.

Duitsman, in Western Krahn, lists the classic miscue analysis taxonomy as defined by Kenneth Goodman (1965; 1969; 1972) namely: repetition, self correction, hesitation, wrong pronunciation, omission, insertion and substitution. However, Duitsman's meticulous data collection is largely superfluous because the presentation of the results gathers all the different kinds of miscue together into one, with the exception of repetition.

## 4. RESULTS AND INTERPRETATION

### 4.1 THE EARLY RESEARCH (1977-1989)

The majority of the researchers up to the 1990s arrived at their conclusions simply by calculating averages and percentages. This is the case in Bura, Efik and Yoruba (Badejo, 1989: 47; Essien, 1977: 159-160; Klem, 1982: 24):

Table 13: Average speed per sentence[8] in seconds and accuracy percentages, oral reading in Bura

| | oralisation speed | accuracy of targeted lexemes |
|---|---|---|
| zero (S) | 8 | 60% |
| exhaustive | 13 | 100% |

Table 14: Average speed per sentence in seconds, oral reading in Efik

| | perception speed | oralisation speed | Accuracy of sentences |
|---|---|---|---|
| zero (S) : | 12.25 | 7.25 | - |
| partial : | 9.5 | 6.5 | 28.8% |
| exhaustive : | 14.5 | 8.5 | 26.6% |

Table 15: Average speeds in seconds and average number of miscues; oral reading of a 100 word Yoruba text

| | oralisation speed | miscues |
|---|---|---|
| zero | 58.5 | 3.8 |
| partial | 52.6 | 1.6 |
| exhaustive (S) | 64.0 | 2.3 |

Essien excluded the zero representation accuracy score from these results (table 14), because the orthography is known to be ambiguous, and therefore unmarkable.

The consensus of these three authors is that a partial tone representation would be the optimum choice. Badejo noted that removing the accents in Bura considerably reduced the accuracy rate, but also that writing accent slows the reading speed. Essien noted that the perception and oralisation results in Efik are symmetrical. In both cases, speed is slower with exhaustive tone marking, while the shortest speed is the partial tone marking. Klem observes that the zero representation in Yoruba generated the most miscues, while the exhaustive marking, even though it reduced the number of miscues substantially, slows the oral reading speed. In contrast, performance is faster and more accurate with a partial representation.

To strengthen his argument in Efik, Essien adds another dimension to his analysis, presenting the minimum and maximum speeds and the range between them (1977: 160) :

[8] This is my interpretation of Badejo's description which is not entirely clear on this point.

Table 16: Maximum and minimum speed per sentence in seconds: oral reading in Efik

| | perception speed | | | oralisation speed | | |
|---|---|---|---|---|---|---|
| | maximum | minimum | range | maximum | minimum | range |
| zero (S): | 130 | 5 | 125 | 70 | 15 | 55 |
| partial : | 75 | 10 | 65 | 40 | 13 | 27 |
| exhaustive : | 103 | 25 | 78 | 75 | 15 | 60 |

Essien notes the wide range between the minimum and maximum perception speeds with a zero representation. He infers that if certain subjects take a long time for the perception, it is because they are thinking about the range of possible meanings before making their choice. But at the other end of the scale are subjects, less hampered by the lack of accents, who are probably not even aware of the ambiguity, and pronounced the first meaning that comes into their heads. As for the oralisation speed, the widest range is in exhaustive representation, and the narrowest range in the partial representation.

Essien deserves some credit for paying attention to the score range. It is only possible to really understand whether the differences between means for different orthographies carry any significance if some estimate of variability is available. A plot showing the distribution of results would have been useful, to make it clear whether there were just a few outlying values, or whether there was a wide and general spread.

Duitsman's approach shadows the other experimenters of his time by simply calculating average scores. He concludes that the results of his experiment offer no convincing reasons to change from the standard accentual orthography to a punctuation system:

Table 17: Average number of miscues; oral reading of a 100 word text in Western Krahn

| year grade | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| accentual system (S) | 27 | 24 | 20 | 17 |
| punctuation system | 27 | 27 | 28 | 25 |

Let us look now at the results of the Bafut and Limbum experiments. From the outset, Mfonyam had excluded the possibility of either the zero or exhaustive extremes in his representation of the 5th parameter, that of graphic density (Mfonyam, 1989: 326). Thus, although he developed no less than eight experimental orthographies for the two languages, all of them represent tone partially. In his analysis, he follows the same approach as the other early experimenters by calculating averages and percentage success rates. Mfonyam did not evaluate the group 3 for the essay writing task, because the subjects did not write tone at all (1989: 334):

Table 18: Results of the Bafut experiment (ibid: 337)

| | Oral reading | | Essay writing | | Adding accents | | Total |
|---|---|---|---|---|---|---|---|
| | average | % | average | % | average | % | % |
| 1. Stable | 146.00 | 89.80 | 19.00 | 63.30 | 35.34 | 72.31 | 73.16 |
| 2. Basic | 101.00 | 61.50 | 19.00 | 63.30 | 25.56 | 51.81 | 53.88 |
| 3. Minimal | 97.00 | 59.50 | - | - | 18.87 | 55.90 | 56.64 |
| 4. Surface | 93.00 | 57.10 | 10.50 | 35.00 | 21.14 | 41.70 | 42.57 |

The results for the stable experimental tone orthography are the highest, proving, according to Mfonyam, that this system would be optimal for Bafut. This is the orthography, remember, which marks L tone, the most stable tone. So Mfonyam rejects one of its own initial hypotheses, that an optimal tone orthography would mark H and M tones (ibid: 338, 342-344, 346).

Mfonyam concedes that the basic tone orthography, which retains fixed word images, might be an effective solution in the case of a language having simple and easily defined tonal processes. But in Bafut, this is far from being the case. It is not at all easy to identify the basic tone, let alone to teach them to novices, as they are not equivalent to the citation form (ibid: 338-340).

Mfonyam admits that the participants themselves expressed a strong preference for the minimal representation, which is not surprising because it is the system requiring the least effort (at least from the writer's point of view) and the one which most resembles French, the language of formal education. Mfonyam considers this orthography to be impractical because it requires an encyclopaedic knowledge of the language. In addition, readers do not necessarily reflect on all possible lexical contrasts before pronouncing a word (ibid: 338, 341).

As for the surface representation, although it reduces ambiguity to the bare minimum, the score is the lowest of all the results (ibid: 338, 339). Between all the different options, Mfonyam reports, this is the one which provoked the strongest negative reaction among the participants. This echoes the Efik experiment (Essien, 1977: 160).

Mfonyam follows the same analytical approach in Limbum as in Bafut, but with an added dimension. The second experiment sought to confirm that it would be preferable to mark the more stable tone, that is the L tone, and not the least stable, that is the H and M tones (Mfonyam, 1989: 460):

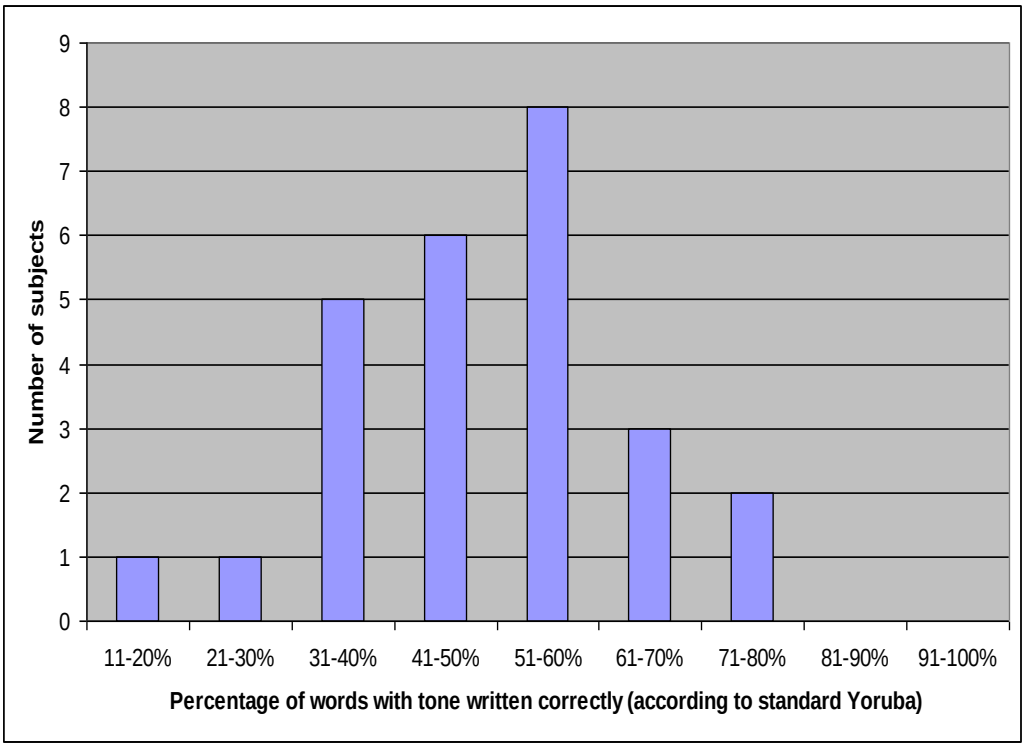
Table 19: Results of the Limbum experiment

| | | Oral reading | | Adding tones | | Global score | |
|---|---|---|---|---|---|---|---|
| | | average | % | average | % | average | % |
| 1. | stable | 92.21 | 80.89 | 326.33 | 92.95 | 418.54 | 86.92 |
| 2. | unstable | 71.31 | 62.55 | 311.38 | 88.81 | 373.93 | 75.68 |
| 3. | minimal | 83.96 | 73.65 | 304.88 | 87.09 | 388.84 | 80.37 |
| 4. | surface | 57.91 | 50.80 | 292.67 | 83.59 | 350.58 | 67.20 |

Based on these results, Mfonyam concludes, as he had for the prototypical test in Bafut, that it is the stable tone orthography which would be the optimal choice for Limbum.

Fagborun's experiment in Yoruba is the only one of the early experiments to distance itself from the approach dominated by calculating averages scores and percentages success rates. He groups results of accuracy in dictation into 10% bands. He scored by counting the number of words with correctly placed accents, when matched against Standard Yoruba. Words written without accents and words with misplaced accents were counted as incorrect (This is my interpretation of Fagborun's (1989: 84) ambiguous statement “All the wrongly tone-marked words were not counted”). The results are summarized in figure 7, based on Fagborun (1989: 86):

Figure 7: Tone marking in a dictation task by 26 Yoruba students




Fagborun draws attention to the enormous disparity between practice and orthographic convention. Thirteen students, representing half the sample, score under 50% and no student obtained more than 73%. Fagborun concludes that tone marks are commonly perceived as being incidental to spelling and that probably the younger Yoruba generation isn't even aware of the established conventions.

In this section, I have taken stock of the different analytical approaches represented in the early research up to 1989. In all these experiments, any reference to statistical significance and the influence of different variables is noticeable by its absence. Any one contribution to scientific research represents only a small step in the accumulation of knowledge, and we have to wait until the 1990s before seeing any implementation of a more rigorous statistical methodology. Indisputably, it is Bernard et al. (1995; 2002) who set a new standard in the field, and Bird (1999b) closely follows their lead. That is why I have chosen to examine the results of these three experiments separately and in greater detail in the section which follows. These experimenters use classic statistical techniques as described with relation to their usefulness in language studies by Woods, Fletcher and Hughes (1986). A few brief definitions of basic terminology will serve as a starting point for the summaries of the three experiments which follow.

## 4.2 STATISTICS: SOME BASIC NOTIONS

A widely used statistical technique is that of regression analysis. Simple linear regression examines the strength of the relationship between two scalar variables. When

there is evidence that a causal relationship exists between the two variables, it is normal to refer to the "causing" variable as the independent variable. It is called independent because it does not depend on the experiment. The information contained in it can be collected beforehand, and would exist even if the experiment did not happen. The "affected" variable, on the other hand, is known as the dependent variable. It is called dependent because it emerges from the experiment itself.

Additional regression methods have been developed for examining the relationship between more complex sets of data. For example, multiple regression explores the relationship between two or more independent variables and a dependent variable. For situations where the response variable is binary (for example passed ~ failed; died ~ survived), a family of methods known as logistic regression is used to describe the relationship the response and the independent variable(s).

The experimenters use several types of measure to analyse the results.

- The constant is the value of the dependent variable when all the independent variables are at their "zero" setting.
- The regression co-efficient is the key relationship produced by a regression analysis. It is a measure of the extent to which a particular variable contributed to performance.
- A regression equation is generated by multiplying the regression coefficient and the value of the independent variable (or some transformation thereof) then adding this to the constant.
- The P value associated with a specific regression coefficient or variable is considered statistically significant when it drops below a threshold which must be specified at the outset of the study. A threshold of 0.05 is often chosen.
- The confidence interval is composed of two values between which one can be 95% sure that the regression coefficient in question is situated.
- The odds ratio is a measure of effect size and is used in logistic regression. It is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. An odds ratio greater than 1 indicates that the condition or event is more likely in the first group. An odds ratio less than 1 indicates that the condition or event is less likely in the first group.
- A confusion matrix is a data table which compares the predicted set of values (columns) against the reported set (rows).

### 4.3 KOM (BERNARD *ET AL.*, 1995)

Bernard et al. put forward a provocative hypothesis from the outset, namely that adding accents to represent tone only serves to complicate the reading process, whereas it should be simple (1995: 28). Both experiments were set up to confirm or reject this hypothesis.

In the prototypical experiment, drawing on Essien (1977: 159-161), the experimenters identified three dependent variables (Bernard et al., 1995: 29, 31).

Table 20: Dependent variables in the first Kom experiment

| | |
|---|---|
| Perceive | perception speed measured in seconds |
| Vocal | oralisation speed measured in seconds |
| Correct | whether or not the sentence was read correctly. |

In addition to this, they identified four independent variables:

Table 21: Independent variables in the first Kom experiment

| | |
|---|---|
| Words | the number of words in the sentence |
| Sequence | sentence presentation order (1 - 50) |
| Proverb | whether or not the sentence is a proverb |
| Tone | whether or not the orthography includes accents |

Bernard et al. state that their aim is to discover whether any of the independent variables, particularly TONE, would have any influence on the performance of the single subject tested, a highly educated male Kom speaker. To answer this question, the performed multiple regression of the two dependent variables PERCEIVE and VOCAL (ibid: 33).

Table 22: Multiple regression of the dependent variables PERCEIVE and VOCAL; reading task, first Kom experiment

| Variable | PERCEIVE | | VOCAL | |
|---|---|---|---|---|
| | Coefficient | p. value | Coefficient | p. value |
| CONSTANT | 7.201 | 0.026 | -2.416 | 0.069 |
| WORDS | 1.033 | 0.000 | 0.627 | 0.000 |
| TONE | 13.440 | 0.000 | 0.473 | 0.629 |
| PROVERB | -5.317 | 0.045 | -0.838 | 0.436 |
| SEQUENCE | -0.173 | 0.016 | -0.034 | 0.259 |
| PERCEIVE | - | - | 0.187 | 0.000 |

According to the authors' interpretation of these results, the variable that had by far the most predominant effect on perception speed is the addition of tone marks. It adds an average of more than 13 seconds to perception speed. The estimated probability of observing a regression coefficient this extreme, were tone to have no effect on perception time, is less than 0.0005. (ibid: 32-34).

Bernard et al. interpret the fourth and fifth columns of table 22 as showing that once the reader has taken the time to decide the correct pronunciation of the sentence, the presence or absence of tone marks does not have any great effect on oralisation speed.

Bernard et al. also claim that longer sentences offer a more developed context, and this reduces the need to decipher the tone marks. When context increases, the presence of accents serves only to confuse the reader and the time devoted to oralisation increases proportionally (ibid: 33-34).

After this, logistic regression analysis was performed on CORRECT, the variable which indicates accuracy (ibid: 37):

Table 23: Logistic regression of the dependent variable CORRECT; oral reading, 1st Kom experiment

| Variable | Coefficient | p. value |
|---|---|---|
| CONSTANT | -0.676 | 0.499 |
| WORDS | 3.045 | 0.002 |
| PERCEIVE | -1.898 | 0.058 |
| VOCAL | -2.666 | 0.008 |
| TONE | 0.877 | 0.380 |
| PROVERB | 1.245 | 0.213 |
| SEQUENCE | 1.910 | 0.056 |

According to the experimenters' interpretation of these results, the longer the sentence the more likely it is to be read accurately. This effect takes place irrespective of whether the tone marks are present or absent (ibid: 36-38).

Since the authors' interpretation of the results of the first Kom experiment support the initial hypothesis, namely that adding accents to represent tone only serves to complicate the reading process, they wanted to confirm that the results did not merely arise from the fact that it was a single subject experiment (Bernard, personal communication). So they proceeded to apply the same methodology to a large sample in the second experiment.

### 4.4 KOM (BERNARD *ET AL.*, 2002)

In the second experiment, three more independent variables were added (Bernard et al., 2002: 342):

Table 24: Independent variables added for the second Kom experiment

| | |
|---|---|
| AGE | Age |
| CERTIF | Number of years of schooling |
| LIT | Degree of functional literacy |

The last of these variables, LIT, is itself an aggregation of eight variables that the authors preferred to treat as a single variable (ibid: 342):

Table 25: Values included in the independent variable LIT for the 2nd Kom experiment

| | |
|---|---|
| INFORM | the subject's identity |
| READFREQ | how often the subject reads Kom |
| LTONE | whether or not the subject had ever learned to read with accents |
| OUT | whether or not the subject had ever lived outside of the Kom homeland |
| SEX | the sex of the subject |
| TEACH | whether or not the subject teaches the Kom orthography to others |
| TRAINING | whether or not the subject has ever been trained to read and write Kom |
| READ | whether or not the subject can read Kom |

The following table summarises the outcome of the multiple regression of perception and oralisation speed (ibid: 343). The variables LP, LV and LW in the second

experiment correspond to the variables PERCEIVE, VOCAL and WORDS in the first experiment (see tables 20 and 21, page 228):

Table 26: multiple regression of LP and LV; oral reading in the second Kom experiment

| | LP | | LV | |
|---|---|---|---|---|
| Variable | Coefficient | p. value | Coefficient | p. value |
| CONSTANT | 2.245 | .000 | -1.291 | .000 |
| LW | 0.474 | .000 | 1.212 | .000 |
| LIT | 0.455 | .000 | - | - |
| CERTIF | -0.137 | .000 | - | - |
| PROVERB | -1.699 | .000 | -1.095 | .000 |
| TONE | 1.081 | .000 | 0.480 | .000 |

Neither AGE nor SEQUENCE appear in this table, because they did not exhibit any measurable effect (Bernard et al., 2002: 343). Then putting aside the two variables, CERTIF and LIT, which had a slight effect only on perception speed, Bernard et al. identify three variables that appear to have a greater impact on speed:

- As sentence length rises, so does perception time and vocalisation time;
- Subjects perceive and read proverbs faster than other types of sentences;
- Subjects perceive and read sentences written with tone marks more slowly than sentences written without tone marks. For example, for the shortest sentence (six words), subjects spend more than double the time to study it in advance if it has accents. This gap is reduced proportionately as the length of the sentence increases. But even for the longest sentence, which was 67 words long, the presence of accents adds more than a second to perception speed.

Table 27 summarises the results of a logistic regression analysis performed on the likelihood of a sentence being read accurately (ibid: 344):

Table 27: Logistical regression of the variable CORRECT; oral reading in the 2nd Kom experiment

| Variable | Odds ratio | 95% confidence interval | |
|---|---|---|---|
| | | minimum | maximum |
| LP | 0.572 | 0.459 | 0.713 |
| LV | 0.082 | 0.052 | 0.128 |
| LW | 16.599 | 8.890 | 30.995 |
| PROVERB | 45.452 | 2.178 | 948.374 |

On the basis of these results Bernard et al. draw the following conclusions:

- Accuracy rate increases with sentence length, because the reader benefits from the context;
- Subjects read proverbs much more accurate than other types of sentence;
- The more time a subject takes to perceive and read the sentence, the less likely it is that he will read it accurately;

- The presence of accents does not contribute to improved accuracy. Rather, it indirectly exerts a negative effect because it increases the length of perception and oralisation time which, in turn, reduce the probability that the sentence will be read correctly.

An anonymous reviewer has pointed out that the analysis and interpretation of both Kom experiments are contestable because Bernard et al. included two dependent variables PERCEIVE and VOCAL in the model for "CORRECT" as if they were independent variables, which they are not. So the reader is denied a simple interpretation. This concern notwithstanding, the second Kom experiment with thirteen subjects reinforces the results of the single subject first experiment.

### 4.5 DSCHANG (BIRD, 1999B)

The analytical phase of the Dschang experiment closely follows the work of Bernard et al. in Kom. First, Bird noted four dependent variables for the reading task (1999b: 97):

Table 28: Dependent reading variables in the Dschang experiment

| | |
|---|---|
| TIME | Duration: The average time taken to read 100 words |
| DISFL | Disfluencies: The number of repetitions and hesitations per 100 words |
| COMP | Comprehension: the number of tonal miscues resulting in a wrong meaning |
| PERF | Performance: the number of tonal miscues resulting in nonsense |

In addition to this, a profile was generated for each subject, according to the following six independent variables (ibid: 97):

Table 29: Independent subject profile variables in the Dschang experiment

| | |
|---|---|
| AGE | age |
| EDUC | years of schooling |
| TRAINING | degree of exposure to the primer |
| CONF | self evaluation by the subject of his reading competence |
| CATEGORY | degree of fluency in reading |
| SEX | sex |

Finally, a profile was generated for each recording concerning the identify of the text itself, according to five binary independent variables (ibid: 97-98):

Table 30: Independent text variables identified in the Dschang experiment

| | |
|---|---|
| TONE | whether or not the text recorded is written in tone orthography |
| FAMILIAR | whether or not the subject had prior knowledge of the text |
| ANKA | whether or not the text is "anka" |
| ATAK | whether or not the text is "atak" |
| LEKAN | whether or not the text is "lekan" |

For the reading task, multiple regression of the first two criteria of competence, that is speed and fluency, was performed. After removing the variables that proved to be insignificant, five variable remained that had a significant effect on the results (ibid: 98):

Table 31: Multiple regression of the dependent variable Time; oral reading in Dschang

| Variable | Coefficient | p. value | 95% confidence interval | |
|---|---|---|---|---|
| | | | minimum | maximum |
| Constant | 87.179 | .0001 | 75.062 | 99.296 |
| TONE | 7.534 | .0332 | 0.632 | 14.435 |
| CATEGORY | -4.962 | .0044 | -8.277 | -1.648 |
| ATAK | 14.848 | .0000 | 6.878 | 22.818 |
| TRAINING | -13.623 | .0446 | -26.898 | -0.348 |
| CONF | -4.292 | .0557 | -8.693 | 0.110 |

Table 32: Multiple regression of the dependent variable DISFL; oral reading in Dschang

| Variable | Coefficient | p. value | 95% confidence interval | |
|---|---|---|---|---|
| | | | minimum | maximum |
| Constant | 18.776 | .0001 | 14.323 | 23.229 |
| TONE | 2.662 | .0537 | -0.0448 | 5.370 |
| CATEGORY | -3.733 | .0001 | -4.922 | -2.545 |
| ATAK | 4.973 | .0026 | 1.847 | 8.100 |
| CONF | 1.952 | .0189 | 0.339 | 3.565 |

As Bird explains, the regression coefficients show that, when reading 100 words, the presence of accents in the text adds an average of 7.5 seconds to the speed of reading 100 words, and an average of 2.7 hesitations or repetitions. P values indicate that the first result is statistically significant, and that the second is not far from being statistically significant. Bird concludes that adding accents has a negative effect on reading speed and fluency.

Bird goes on to analyse the individual tone comprehension errors made for both zero and full marking, and this was also revealing. A native speaker classified the errors into lexical and grammatical errors, by listening to the reader's pronunciation and deciding whether a comprehension error was due to a lexical or grammatical misunderstanding (ibid: 99):

Table 33: Analysis of raw tone comprehension errors in Dschang

| | lexical errors | grammatical errors |
|---|---|---|
| zero | 0 | 18 |
| exhaustive (S) | 3 | 11 |

Bird points out how remarkable it is that there were no lexical errors while reading the zero representation. Both orthographies produce numerous grammatical errors, but they are most numerous in the accentless orthography.

As for the writing task, subjects were given 20 minutes to add accents to prepared texts, a pilot test having demonstrated that an unlimited duration was impractical. Bird computerised the written texts, and then used software to compare them with the original and calculate the number of disparities.

On the basis of the results obtained, Bird judged it best to divide the sample into two clusters, "experienced" and "inexperienced", and it is like this that the results were entered into confusion matrices. Scores were calculated first on the basis of raw percentages, then

were adjusted to show how much the raw score is an improvement over random tone marking (ibid: 100):

Table 34: Adding accents to a text by experienced Dschang readers

| Intended | Observed | | | Success rate (%) | |
|---|---|---|---|---|---|
| | H | M | 0 | Raw | Adjusted |
| H | 1,434 | 36 | 395 | 76.9 | 58.3 |
| M | 38 | 350 | 138 | 66.5 | 61.7 |
| 0 | 63 | 22 | 1,709 | 95.3 | 91.7 |
| | | | Mean: | 83.5 | 73.1 |

Table 35: Adding accents to a text by inexperienced Dschang readers

| Intended | Observed | | | Success rate (%) | |
|---|---|---|---|---|---|
| | H | M | 0 | Raw | Adjusted |
| H | 494 | 59 | 1,068 | 30.5 | -24.9 |
| M | 86 | 88 | 262 | 20.2 | 9.4 |
| 0 | 159 | 83 | 1,356 | 84.9 | 73.1 |
| | | | Mean: | 53.0 | 22.0 |

The most important finding, according to Bird, is the poor performance of the experienced subjects, whose average adjusted score is just 73.1%. Bird also considers that the performance of inexperienced subjects in marking the acute accent is worse than chance. If the results in both the "absence of accents" columns are quite high, it is probably because this choice operates as the default. The subject reasons that "if in doubt, it's better to write nothing."

Bird goes on to introduce the concept of the tonal stability rate, which he defines as the probability that a syllable carrying the tone in question when the word is pronounced in isolation will also carry that tone in context. He calculates that it is the M tone, which has the lowest stability rate in Dschang, and infers that this is the reason why the macron marking is the lowest of all the scores. Driven by this hypothesis, the author examines prefixes more closely, which are the most tonally variable of all morphemes, being heavily influenced by the tone of the preceding word (ibid: 101).

In conclusion Bird raises two questions. Firstly, why is the performance of the subjects so strong when accents are absent? Mere counting of tonal minimal pairs leads to the conclusion that tone has an extremely high functional load in Dschang. Yet the fact that comprehension errors (table 33, page 232) are not that high suggests otherwise. The experiment demonstrates that tone plays a key role in a limited number of grammatical constructions and not at all in the lexicon.

Bird's second question is, why is the performance of subjects so poor when accents are present? In response to this question, Bird contends that it is because it is a shallow orthography representing a deep tone system (i.e. one with many morphotonological processes). This is not conducive to preserving the fixed word images which aid fluent reading, and in particular for comprehension (ibid: 102).

The influence of Bird's research has been considerable. It has already extended not only to related languages as he himself predicted (Bird, 2001: 150), but from Pakistan

(Gojri, Indo-Aryan; Losey, 2002) to Alaska (Athabaskan; Holton, 2003) via Tanzania (Eastern Bantu; Stegen, 2005), not to mention the more general debate about orthography (Seifart, 2006). Moreover, it was reading Bird's trilogy of articles that inspired this author to undertake experimentation on Kabiye tone orthography as a subject for his own PhD research (Roberts, 2008).

## 5. SUMMARY

In Tables 36 - 39 pages (234 - 235), I present summaries of the ten experiments according to language, aims, sample, corpus, and procedure:

Table 36: Summary of language

| Language | Reference | Country | Family | Tones | Standard orthography |
|---|---|---|---|---|---|
| Efik | (Essien, 1977) | Nigeria | Cross river | 2 (?) | Zero |
| Yoruba | (Klem, 1982) | Nigeria | Defoid | 3 | Exhaustive |
| W. Krahn | (Duitsman, 1986) | Liberia | Kru | 3 | Accents |
| Bura | (Badejo, 1989) | Nigeria | Afro-Asiatic | 2 (?) | Zero |
| Yoruba | (Fagborun, 1989) | Nigeria | Defoid | 3 | Exhaustive |
| Bafut | (Mfonyam, 1989: 309-348) | Cameroon | Grassfields | 3 | No standard existed prior to the experiment |
| Limbum | (Mfonyam, 1989: 459 – 473) | Cameroon | Grassfields | 3 | No standard existed prior to the experiment |
| Partial CHECK | (Bernard et al., 1995) | Cameroon | Grassfields | 3 | Partial |
| Dschang (Yemba) | (Bird, 1999b) | Cameroon | Grassfields | 3 | Exhaustive |
| Kom | (Bernard et al., 2002) | Cameroon | Grassfields | 3 | Partial |

Table 37: Summary of aims

| Language | Reference | Contest | Parameters |
|---|---|---|---|
| Efik | (Essien, 1977) | Zero v. partial v. exhaustive | Density |
| Yoruba | (Klem, 1982) | Zero v. partial v. exhaustive | Density |
| W. Krahn | (Duitsman, 1986) | Accents v. punctuation | Symbol |
| Bura | (Badejo, 1989) | Zero v. partial v. exhaustive | Density |
| Yoruba | (Fagborun, 1989) | De jure v. de facto | Aims cannot be stated in terms of six parameters |
| Bafut | (Mfonyam, 1989: 309-348) | Stable v. unstable v. minimal v. surface | Domain, target, symbol, density, depth |
| Limbum | (Mfonyam, 1989: 459 – 473) | Stable v. unstable v. minimal v. surface | Domain, target, symbol, density, depth |
| Kom | (Bernard et al., 1995) | Partial v. exhaustive | Density |
| Dschang | (Bird, 1999b) | Zero v. exhaustive | Density, depth |
| Kom | (Bernard et al., 2002) | Partial v. exhaustive | Density |

Table 38: Summary of sample and corpus

| Language | Reference | Sample Profile | Size | Experience | Corpus Profile | Size (words) |
|---|---|---|---|---|---|---|
| Efik | (Essien, 1977) | Expatriate university students | 15 | Inexperienced | sentences | 70 |
| Yoruba | (Klem, 1982) | Mixed | 60 | Experienced | texts | 100 |
| W. Krahn | (Duitsman, 1986) | School children | 36 | Inexperienced | texts | 177 |
| Bura | (Badejo, 1989) | Mixed | 50 | Experienced | sentences | 29 |
| Yoruba | (Fagborun, 1989) | University students | 26 | Experienced | sentences | 152 |
| Bafut | (Mfonyam, 1989: 309-348) | Mixed | 16 | Inexperienced | sentences | 222 |
| Limbum | (Mfonyam, 1989: 459 – 473) | Mixed | 15 | Inexperienced | sentences | 669 |
| Kom | (Bernard et al., 1995) | Highly educated male | 1 | Inexperienced | sentences | 1,788 |
| Dschang | (Bird, 1999b) | Mixed | 11 | Mixed | texts | 873 |
| Kom | (Bernard et al., 2002) | Mixed | 13 | Experienced | sentences | 1,788 |

Table 39: Summary of test procedure

| Language | Reference | Design | Training | Tasks | Scoring |
|---|---|---|---|---|---|
| Efik | (Essien, 1977) | Within subject | 10 min. | Reading | Accuracy, speed |
| Yoruba | (Klem, 1982) | Between groups | None | Reading | Accuracy, speed |
| W. Krahn | (Duitsman, 1986) | Between groups | 1½ hours | Reading | Accuracy |
| Bura | (Badejo, 1989) | Within subject | Unknown | Reading | Accuracy, speed |
| Yoruba | (Fagborun, 1989) | Between subjects | None | Dictation | Accuracy |
| Bafut | (Mfonyam, 1989: 309-348) | Between groups | 15 days | Reading, writing | Accuracy |
| Limbum | (Mfonyam, 1989: 459 – 473) | Between groups | 10 days | Reading, writing | Accuracy |
| Kom | (Bernard et al., 1995) | Single subject | 15 min. | Reading | Accuracy, speed |
| Dschang | (Bird, 1999b) | Within subject | None | Reading, writing, self-evaluation | Accuracy, speed, comprehension |
| Kom | (Bernard et al., 2002) | Within subject | Little | Reading | Accuracy, speed |

## 6. CONCLUSIONS AND FUTURE PROSPECTS

### 6.1 A MATURING METHODOLOGY

It is in the nature of social science experiments to be methodologically deficient in one way or another. No researcher reaches the ideal, which in any case is ephemeral, to reproduce sterile laboratory conditions in the classroom, as Duitsman (1986: 8) reminds us. This is especially the case with a specialisation still in its infancy. No model yet exists which has been approved by the small community of researchers who work on it. No reference book has yet been written showing in a handy format the essentials of experiment design. Nevertheless, we are in a better position now to achieve robust methodologies than we were thirty years ago, thanks to the researchers whose efforts I have just summarised.

The pioneering work of Essien in Efik is important in that the author had the insight to pinpoint a problem for the first time, and to test it with the methodological tools available to him at the time. It is Essien who opened the door to a new avenue of research. But he himself stresses several times that his conclusion is preliminary and tentative and needs to be confirmed or rejected by later experimentation (1977: 163). This has been the case, and the ensuing three decades have showed a marked improvement in scientific rigour as the literature has accumulated. I take the four Grassfields languages to illustrate this point.

The twin experiments in Bafut and Limbum (Mfonyam, 1989) introduce two new elements in experiment design. Firstly, Mfonyam's thorough knowledge of autosegmental phonology enables him to distinguish between the depths of different tone systems. Secondly, he is the only researcher to have insisted on a generous period of time devoted to training and practice, and he does so with an appropriate pedagogical approach.

The three most recent works, in Kom and Dschang (Bernard et al., 1995, 2002; Bird, 1999b), take into account the influence of different variables on the results as well as the notion of statistical significance. It is to be hoped that future experimenters will take these works as a reference point, and that no one will think it appropriate to turn back the clock on this issue.

Still in Dschang, Bird's research is strongly rooted in the theory of the cognitive psychology of reading (he cites Fowler and Liberman, 1995; Frost and Katz, 1992; Henderson, 1984; Katz and Frost, 1992; Liberman et al., 1980). This breathes some much needed fresh air into a debate hitherto heavily dominated by phonology throughout the course of the 20th century.

Given this long process of methodological refinement, out of all different aspects found in the previous research, what should be assimilated into future experiments? I will make six recommendations.

Firstly, in many (if not most) experiments, the stated aims will require the introduction of more than one of the six typological parameters introduced at the outset (table 3, page 208), and these parameters are expected to affect the outcome of interest. In such cases, a well-designed experiment must find ways, drawing on current statistical methodology, of disentangling the interaction that each of those parameters may have on the others.

Secondly, the hunt is still on for a good comprehension measure. The Dschang experiment is the only one so far to have taken this variable into account, but Bird (1999b:

99, 103) concedes that the results are far from conclusive and that it would have been better to include additional exercises that target this single aspect of performance.

Thirdly, one can only regret the almost total absence of references to pilot tests in the repertoire, other than the first Kom experiment (Bernard et al., 1995) and a fleeting reference by Bird in the Dschang experiment (1999b: 96). It is essential to conduct multiple pilot tests to ensure the smooth running of the final experiment. And since this step is an integral part of the test methodology, why would the researcher not also describe it in the published write-up? Detailed descriptions of test pilots, rich in qualitative observations, would help optimise the design of future experiments.

Fourthly, future experimenters should always furnish linguistic examples, for example by explicitly placing phonetic data and the various orthographic options side by side, or by providing 100 word text samples in each of the experimental orthographies. Oddly enough, several of the articles in the literature do not present the testing materials (Bernard et al., 1995, 2002; Duitsman, 1986; Klem, 1982), and those that do often leave us with an incomplete picture (Badejo, 1989: 47, 50-51; Essien, 1977: 164-166). Neither of the two Kom experiments provide any linguistic data at all (Bernard et al., 1995, 2002). Mfonyam (1989), painting on the ample canvas of a doctoral thesis, is by far the most thorough experimenter in this regard. But Bird (1999b) shows that this can also be achieved within the limited constraints of a journal article.

Fifthly, I advocate a more generous use of graphs. Data plotted in graphic form tends to be easier to interpret than statistical tables, especially for the uninitiated reader. Most statistics software packages generate them automatically. Yet graphs are few and far between in the existing repertoire. Fagborun (1989: 86) and Bernard et al. (1995: 39; 2002: 345) provide one graph each in their articles; the others present none at all.

Sixthly, I will wave the flag for collaborative research. The pressing need to employ a statistically rigorous methodological approach might be off-putting to the field linguist who has no particular desire to get to grips with an entirely new specialist field, when already overworked in his or her own domain. But it should not be a hindrance. It is more than just possible, it is highly desirable for the linguist to enter into collaboration with a statistician, both of them bringing the fruits of their respective specialities to the analysis. Until now, Bernard's team – which marshalls skills and experience in anthropology, statistics and linguistics – remains the only example of such collaboration in the repertoire (1995; 2002). The three members of this team combined their forces and the quality of the result shows. That a linguist should enter into collaboration with a native speaker to produce a good phonological analysis has never been called into question. Why would (s)he hesitate to enter into collaboration with other specialists, such as statisticians, to produce an quantitative analysis of an orthography experiment?

### 6.2 A WIDENING RANGE

However it is viewed, the literature remains extremely limited at the threshold of the 21st century. Despite the efforts of various researchers, the whole domain is still characterized by vast gaps in our knowledge. Apart from the glaringly obvious need to expand the geographical area and typological range represented, this is a plea for a richer variety of experiments regarding the objectives, tasks and methodological approaches.

**Objectives:** The current literature is dominated by experiments which test the fifth and sixth parameters, those of orthographic density and depth. Deep tone systems are

comparatively well represented, namely those of Efik, Kom, Dschang, Bafut and Limbum (Bernard et al., 1995, 2002; Bird, 1999b; Essien, 1977; Mfonyam, 1989). As for the other languages in the repertoire, the depth the tone system is not clearly identified. It would be useful to add to the literature experiments on languages whose tone systems are known to be shallow. As far as the third parameter, the choice of graphic symbol, apart from the very brief experiment in Western Krahn (Duitsman, 1986), which examines the use of punctuation marks, we have no evidence of the effectiveness of using characters other than accents. Moreover, it would be valuable to know whether there is any difference in performance between uneducated and educated subjects. None of the existing research addresses this crucial question. All researchers who report on this variable state that subjects were already literate in English or French.

**Tasks:** The fact that the oral reading task dominates the literature is no bad thing in itself, but it should be remembered that it puts the subject in the place of a "passive actor". Writing tasks require the subject to be creative and (s)he becomes an "active actor". Mfonyam notes that the performance of all four groups in Limbum is better in reading than in writing (Mfonyam, 1989: 469). This suggests that the core of the problem might be found by examining those activities which require the most initiative on the part of the subject. For these reasons, it would be desirable for future experiments to include more writing tasks, such as dictation and creative writing.

**Approaches:** Along with the quantitative approach favoured in most experiments until today, it would be good to see more qualitative research. The advantage of this approach is that it allows the researcher to make detailed field observations, producing data which is measurable but not countable such as attitudes and preferences, and to examine in detail the source of miscues. It is the experiments in Western Krahn and Yoruba (Duitsman, 1986; Fagborun, 1989), which show us the way forward on this point.

One would hesitate to call the Western Krahn experiment a miscue analysis in the strict sense of the term (Yetta Goodman, 1972), because it is limited to a gross counting of miscues. It does not generate the qualitative data which lies at the heart of a true miscue analysis, going in search of the source and making use of the classic miscue notation (Schreiner, 1979) to present the most interesting results. However, since Duitsman's is the only experiment to even come close to a miscue analysis in the literature, we should acknowledge the foresight of the author for having applied the methodology for the first time in the field. I sincerely hope that future researchers will become aware of the enormous potential of miscue analysis.

As for Yoruba, Fagborun compares the de jure orthographic conventions with the de facto practice. It is clear that Fagborun has a profound knowledge of the socio-political realities of his field. At ease in this context, he takes time to focus on specific examples in a very detailed manner. This is an approach which has been almost completely overlooked by others. It is very promising for future research because it is draws out the instincts of native speakers.

## 6.3 AN EMERGING CONSENSUS

Experiments in the social sciences are rarely unanimous in their conclusions, and this is certainly the case with the repertoire of tone orthography experiments. Mfonyam concludes that the "stable" orthography, which is a kind of surface representation, would

be the optimal choice in Bafut and Limbum (1989: 338, 346, 470, 534), whereas Bird considers surface representation to be worse than zero representation in Dschang, a closely related language (Bird, 1999b, 107). In Bura and Yoruba the experimenters claim that the tone marks contribute to the accuracy in reading (Badejo, 1989: 48; Klem, 1982: 24), while experiments Kom and Dschang contend that the presence of accents does not contribute, and even contributes negatively, to accuracy (Bernard et al., 2002: 355; Bird, 1999b: 98).

In addition, some researchers have reached conclusions which, it should be said, are difficult to justify. Badejo contends that the functional load of tone in Bura is high (1989: 48), but it is difficult to find anything in his results which support this claim. The concept of functional load is relative, and it is only possible to assess its importance in a given language by comparing it with at least one other language. Mfonyam seeks to extrapolate from the Bafut and Limbum results principles which he proposes to generalize to all Bantu languages (Mfonyam, 1989: 515, 517, 535). This seems untenable, given the enormously wide range of tone systems within this immense linguistic family, with its more than 600 languages, representing approximately 10% of the world's languages.

For the moment, then, we are still desperately far from any kind of general consensus. Nevertheless I will enumerate three points, hopefully without risk of contradiction. In what follows, the reader should necessarily give more weight to those experiments whose research methodology is reliable because it is statistically rigorous.

**1. Multiple tone marks hinder reading speed.** This is the unanimous conclusion of all the experimenters who measured it, namely in Bura, Kom, Dschang, Efik and Yoruba (Badejo, 1989: 48; Bernard et al., 2002: 355; Bird, 1999b: 98; Essien, 1977: 159; Klem, 1982: 24). Moreover, the two experiments that measured perception speed, namely Kom and Efik, found that it too is prolonged by the presence of multiple tone marks (Bernard et al., 1995: 34, 2002: 355; Essien, 1977: 159).

**2. Exhaustive tone marking is not optimal.** All the researchers who have examined this issue are in agreement. Most of them argue for a partial representation, namely in Bura, Efik, Yoruba, Bafut and Limbum (Badejo, 1989: 49; Essien, 1977: 159; Klem, 1982: 24; Mfonyam, 1989). In Kom and Dschang the experimenters conclude that zero representation is preferable to exhaustive representation (Bernard et al., 1995: 38, 2002: 345; Bird, 1999b, 107).

**3. The percentage of sentences read and written correctly with tone marks is pitifully low.** This is the case when reading in Efik (Essien, 1977: 159-160) and when writing in Dschang and Yoruba (Bird, 1999b: 88, 100-101; Fagborun, 1989: 84, 86). The testimony of the latter is particularly alarming because the sample consists of university students studying Yoruba language and literature who are, in principle at least, highly motivated and experienced. Either they do not know the rules, or they consider them too difficult to master, or they perceive them to be unnecessary. It seems like this is also the case in Dschang and Bafut, where the subjects expressed their preference for a minimal representation, or even a zero one (Bird, 1999b: 101; Mfonyam, 1989: 341, 345).

In the foregoing discussion I have tried to trace the accumulation of knowledge on tone orthography experimentation from its origins thirty years ago up to the present day. I have noted the strengths and weaknesses of the various methodologies employed, and have tried to detect any elements of consensus from the various contributions. What is

certain is that the most convincing evidence will gradually emerge from an assessment of the overall literature, from which we can aggregate general principles. If such a review is undertaken periodically, it will illuminate the path that lies ahead of us.

## REFERENCES


Abraham, Roy Clive (1958): Dictionary of Modern Yoruba. London, University of London Press.

Badejo, B. Rotimi Maiduguri (1989): An experimental study of tone marking in Bura . In Frankfurter Afrikanistische Blätter, 1, pages 44-51.

Bernard, Russell H., George N. Mbeh and W. Penn Handwerker (1995): The Tone Problem . In The Complete Linguist: papers in memory of Patrick J. Dickens, eds. A. Traill, R. Vossen and M. Biesele. Cologne, Rüdiger Köppe Verlag, pages 27-44.

Bernard, Russell H., George N. Mbeh and W. Penn Handwerker (2002): Does marking tone make tone languages easier to read? In Human organisation, a journal of the Society for Applied Anthropology, 61 (4), pages 339-349.

Bird, Steven (1999a): Strategies for Representing Tone in African Writing Systems . In Written Language and Literacy, 2 (1), pages 1-44.

Bird, Steven (1999b): When Marking Tone Reduces Fluency: An Orthography Experiment in Cameroon . In Language and Speech, 42, pages 83-115.

Bird, Steven (2001): Orthography and Identity in Cameroon . In Written Language and Literacy, 4 (2), pages 131-162 and reprinted in SIL Notes on Literacy 26(1-2): 3-44.

Bolli, Margaret (1978): Writing tone with punctuation marks . In SIL Notes on Literacy, 23, pages 16-18.

Bolli, Margaret (1989): The teaching of tone in the Dan language. Manuscript, Abidjan, SIL.

Bolli, Margaret (1991): Orthography difficulties to be overcome by Dan people literate in French . In SIL Notes on Literacy, 65, pages 25-34.

Duitsman, John (1986): Testing two systems for marking tone in Western Krahn . In SIL Notes on Literacy, 49, pages 2-10.

Essien, Udo E. (1977): To end ambiguity in a tone language . In Languages and Linguistics problems in Africa. Proceedings of the 7th Conference on African linguistics, eds. Paul F. A. Kotey and Haig Der-Houssikian, pages 155-167, Columbia, South Carolina, Hornbeam Press.

Fagborun, J. Gbenga (1989): Disparities in tonal and vowel representation: some practical problems in Yoruba orthography . In Journal of West African languages, 19 (2), pages 74-92.

Fowler, A.E. and I. Y. Liberman (1995): The role of phonology and orthography in morphological awareness . In Morphological aspects of language processing, ed. L.B. Feldman. Hillsdale, New Jersey, Lawrence Erlbaum Associates, pages 157–188.

Frost, Ram and Leonard Katz eds. (1992): Orthography, Phonology, Morphology and Meaning. vol. 94. Advances in Psychology. Amsterdam, North-Holland.

Goodman, Kenneth S. (1965): A linguistic study of cues and miscues in reading . In Elementary English, 42 (6), pages 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 . In Reading Research Quarterly, 5, pages 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, Kenneth S. (1972): Orthography in a theory of reading instruction . In Elementary English, 49 (8), pages 1254-1261.

Goodman, Yetta M. (1972): Reading diagnosis: qualitative or quantitative? In The Reading Teacher, 26, pages 27-32, and reprinted in 1997, volume 50(7), pages 534-8.

Gordon, Raymond J. ed. (2005): Ethnologue: languages of the world. Dallas, SIL International.

Gudschinsky, Sarah C. (1970): More on formulating efficient orthographies . In The Bible translator, a journal of the United Bible Societies, 21 (2), pages 21-25.

Hartell, Rhonda L. ed. (1993a): Alphabets des langues africaines. Dakar, UNESCO et SIL.

Hartell, Rhonda L. ed. (1993b): Alphabets of Africa. Dakar, UNESCO and SIL.

Henderson, L. ed. (1984): Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics. Hillsdale, New Jersey, Lawrence Erlbaum Associates.
Holton, Gary (2003): Shallow tone marking and literacy in Alaska Athabascan languages . In Proceedings of the 2003 Athabascan Languages Conference, ANLC Working Paper 3, Arcata, California., ed. S. Tuttle, pages 1-14, Fairbanks, Alaska, Alaska Native Language Center. (Original title of conference paper: On the representation of tone in Alaska Athabaskan practical orthographies ).
ILA (1979): Une orthographe pratique des langues ivoriennes. Abidjan, Institut de linguistique appliquée, Université d'Abidjan.
Katz, Leonard and Ram Frost (1992): The reading process is different for different orthographies: The orthographic depth hypothesis. In Haskins Laboratories Status Report on Speech Research, 111-112, pages 147–160.
Klem, Herbert V. (1982): Oral communication of the Scriptures: insights from African oral art. Pasadena, CA, William Carey Library.
Kutsch Lojenga, Constance (1993): The writing and reading of tone in Bantu languages . In SIL Notes on Literacy, 19, pages 19.
Labov, William (1970): The study of language in its social context . In Studium Generale, 23, pages 30-87.
Liberman, I., A. M. Liberman, I. Mattingly and D. Shankweiler (1980): Orthography and the beginning reader . In Orthography, reading and dyslexia, eds. J. Kavanagh and R. Venezky. Baltimore, University Park Press, pages 137–153.
Losey, Wayne E. (2002): Writing Gojri: linguistic and sociolinguistic constraints on a standardised orthography for the Gujars of South Asia. University of North Dakota, M.A. thesis.
Mfonyam, Joseph Ngwa (1982): The Tone in the orthography of Bafut. Université de Yaoundé, Cameroun, Thèse de doctorat du 3e cycle.
Mfonyam, Joseph Ngwa (1986): Tone in the orthography of two Grassfields Bantu languages. Paper presented at the 17th congress of the West African Linguistics Society, Ibadan, Nigeria, March 17-21, 1986.
Mfonyam, Joseph Ngwa (1989): Tone in orthography: the case of Bafut and related languages. Université de Yaoundé, Cameroun, Thèse d'état.
Mfonyam, Joseph Ngwa (1990a): Levels of Tone Representation. Paper presented at the 19th congress of the West African Linguistics Society Congre, Legon, Ghana, April 2-6, 1990.
Mfonyam, Joseph Ngwa (1990b): Tone analysis and tone orthography . In Journal of West African Languages, 20 (2), pages 19-30.
Mfonyam, Joseph Ngwa (1996): Reading and writing tone in African languages . In Guide to readability in African languages Linguistics edition, ed. Emmanuel Nges Chia. München, Lincom Europa, pages 11.
Neuman, Susan B. and Sandra McCormick (2000): A case for single subject experiments in literacy research . In Handbook of reading research, volume 3, eds. Michael L. Kamil, Peter B. Mosenthal, P. David Pearson and Rebecca Barr. Mahwah, New Jersey, Lawrence Erlbaum Associates, pages 181-194.
Roberts, David (2008): L'orthographe du ton en kabiyè au banc d'essai. Paris, Institut national de langues et de civilisations orientales (INALCO), Thèse de doctorat. http://www.orthographyclearinghouse.org/phdma.html.
Schreiner, Robert (1979): Reading Tests and Teachers: a practical guide. Newark, Delware, International Reading Association.
Seifart, Frank (2006): Orthography development . In Essentials of language documentation, eds. Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel. Berlin, New York, Mouton de Gruyter, pages 275-301.
SIL (1981): Guide d'orthographe pour la langue dan: dialecte blo -wo. Abidjan, SIL.
SIL (1982): Guide d'orthographe pour la langue dan: dialecte gwɛtaawo. Abidjan, SIL.
Stegen, Oliver (2005): Tone in Eastern Bantu Orthographies. Paper presented at the SIL Bantu orthography meeting, Dallas, 7-12 November, 2005.

Wiesemann, Ursula (1981): Native speaker reaction and the development of writing systems for unwritten languages . In Cahiers du département des langues africaines et linguistique, Université de Yaoundé, Cameroun, 1, pages 29-44.
Woods, Anthony, Paul Fletcher and Arthur Hughes (1986): Statistics in language studies. Cambridge Textbooks in Linguistics. Cambridge, Cambridge University Press.