Does marking tone make tone languages easier to read?
H Russell Bernard; George N Mbeh; W Penn Handwerker
Human Organization; Winter 2002; 61, 4; ABI/INFORM Global
pg. 339

Human Organization, Vol. 61, No. 4, 2002
Copyright © 2002 by the Society for Applied Anthropology
0018-7259/02/040339-11$1.60/1

Does Marking Tone Make Tone Languages Easier to Read?

H. Russell Bernard, George N. Mbeh, and W. Penn Handwerker

Among designers of orthographies for previously nonwritten languages, the prevailing wisdom is that native speakers of tone languages need to see at least some tones marked to achieve full literacy in those languages. A counter argument is that marking tone confuses native speakers of tone languages. We report here on an experiment on tone marking in Kom, a language of the Grassfields region of Cameroon. The results show that, for Kom at least, tone marking makes written sentences harder to figure out initially, harder to say, and harder to say correctly. Comparative research using refinements of the methodology described here will help us answer important questions, like whether or not these effects vary with the number of tones in a language or with the linguistic function of tones. The knowledge gained from comparative experiments will help speakers of the world's nonliterary languages achieve popular literacy in those languages. Popular literacy, we argue, is one of a small number of things that promotes the survival of language diversity.

Key words: tone languages, orthographies, indigenous literacy, field experiments, Kom

We know from much evidence that native speakers of all written languages use context to disambiguate homophone pairs that have a single written form. In English, for example, we do not mark the vowel in the word "wind" to indicate whether it sounds like the vowel sound in "shy" (as in "wind up the watch") or like the vowel sound in "sit" (as in "an ill wind blows"). We rely instead on the reader's understanding of context. In an extreme case of context dependence, adult, native speakers of modern Hebrew read and write their language without marking any vowels at all.

In many languages of the world, tone distinguishes between minimal pairs of words (lexical tone) and it can also mark grammatical features, like tense (grammatical tone). This raises the question: Does marking tone in writing previously nonwritten languages help or hinder native speakers who learn to read those languages?

H. Russell Bernard (Ph.D. 1968, University of Illinois) is professor of anthropology at the University of Florida. George Ngong Mbeh received his Ph.D. in anthropology in 2000 from the University of Florida. W. Penn Handwerker (Ph.D. 1971, University of Oregon) is professor of anthropology and head of the Department of Anthropology at the University of Connecticut. *We gratefully acknowledge the Wenner-Gren Foundation for Anthropological Research for support of this effort and Dr. Paul Nchoji Nkwi (Department of Anthropology, University of Yaoundé) for his collaboration in the project. We also gratefully acknowledge the collegial advice and criticism of Stephen Bird (Linguistic Data Consortium, University of Pennsylvania) in developing the arguments in this paper. We are indebted to the participants in Kom for generously offering their time to take part in this experiment; and we thank Gery Ryan (RAND Corp.) who offered helpful suggestions during the design of the experiment.*

At least since Pike's (1948, 1964) seminal work on the phonemics of tone languages, designers of orthographies have grappled with this problem. Pike (1964) argued in favor of marking at least some tones in these languages, as have some other linguists who study tone languages (see, for example, Awedoba 1990; Gudschinsky 1970; Lojenga 1993; Longacre 1964; Wiessemann 1989, 1995). The counterargument is that marking tone adds a layer of effort to writing and to reading and thus hinders both the production and the consumption of texts (Bernard 1974, 1980; Bird 1999b; Crofts 1976; Fagborun 1989; Voorhoeve 1962).

Whether marking tone helps or hinders native speakers in reading their languages is subject to experimental test. We present here the results of our effort to test this for Kom, a nonliterary, Niger-Congo language of Cameroon. We use the term *nonliterary* language rather than *nonwritten* language deliberately. Orthographies exist for nearly all languages in the world. Most languages, then, are written, but most also lack a substantial literary tradition. To anticipate the findings: 1) marking tone slows adult Kom readers down; and 2) readers make more mistakes in reading tone-marked text than they make in reading the same text without the tones marked. Later, in the Discussion section, we consider the implications of our findings for literacy training in tone languages across the world.

The Background—Essien's Experiment

Our method is based on the pioneering effort of Udo Essien (1977) in his study of tone in Efik, a member of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Niger-Congo family of languages. In 1998, Efik was spoken by about 2.4 million people in Nigeria, including 400,000 for whom it was the primary language (Grimes 2000). Essien created 14 short sentences (from three to seven words each) that could be ambiguous in Efik, depending on the tones. For example, depending on the tone, *nnyin imama ubom esie* could be read to mean "we like its size" or "we like his canoe."

Essien presented three sets of 4 of the 14 sentences to 15 college-educated, native speakers of Efik. All the participants in Essien's experiment were unaccustomed to reading or writing Efik, but were highly competent in reading and writing English. Thus, participants were skilled in literacy but had not applied their skills as adults to their first language, Efik.

Essien administered his experiment to each participant individually. He tutored the participants for an average of 10 minutes each on how to read tone marks in Efik. Then, he presented one set of four sentences without any tones marked. Each of the four sentences was on a separate 3"x5" card. Next, Essien presented participants with the same set of four sentences, but this time, all possible tone-marked variations of each sentence were included and every syllable of each variation was marked for tone. In other words, if a sentence could have two possible readings, depending on tones, then both readings were presented with all syllables marked for tone. If it could have three readings, then all three readings were presented, with all tones marked. Finally, Essien presented all possible variations of each sentence with tones marked on just the nouns and verb roots.

Essien told participants to study the four sentences on the cards and to recite the sentences out loud when they had determined the meaning of all the sentences in a set. He timed the participants on how long they took to study each set of sentences before starting to read out loud and how long they actually took to say each set of the sentences out loud. Essien reports averages of these times, and from here on we will refer to these quantities as *average perception time* and *average vocalization time*.

Table 1 shows six of Essien's key results: 1) It took the 15 Efik speakers an average of 14.50 seconds to begin vocalizing sentences that had *tone marked on every syllable*. 2) It took them 12.25 seconds, on average, before vocalizing sentences with *no tones marked*. 3) Sentences that had *tone marked only on nouns and verb roots* required 9.50 seconds, on average, of perception time. 4) *Fully marked* sentences took 8.50 seconds in vocalization time. (5) *Unmarked sentences* took 7.25 seconds. 6) *Sentences with tones only on nouns and verb roots* took 6.50 seconds.

These results argue for marking tones only on nouns and verb roots—a middle ground between what Bird (1999a), in his comprehensive paper on principles of African tone orthography, calls the "zero option" (no tone marking) and complete marking of tones on every syllable.

But there was one more important finding. Essien graded the participants on whether they read each tone-marked sentence correctly. (Since the sentences were selected especially for their ambiguity, he did not grade vocalization of unmarked sentences.) Of the sentences that were marked, less than 30 percent were read correctly, irrespective of whether the sentences had all syllables marked for tone or just the nouns and verb roots. This finding argues for the zero option because, while marking tones on nouns and verb roots may have decreased perception time, it appears to have produced errors in comprehension.

Table 1. Summary of Essien's Results (from Essien 1977)

	Marking Tones		
	Every Syllable	Only Nouns and Verbs	None
Average perception time (in seconds)	14.5	9.5	12.25
Average vocalization time (in seconds)	8.5	6.5	7.25

Overall, then, Essien discovered that too much information (full marking) and not enough information (no marking) about tone slowed native speakers down, while partial marking (nouns and verb roots) helped them perceive and vocalize written sentences. On the other hand, both full and partial marking of tone made it more difficult for educated native speakers of Efik to read sentences aloud *correctly*. This intriguing finding—that selected marking of tone on nouns and verb roots helped readers of Efik perceive and vocalize written sentences while at the same time causing them to make mistakes in reading those sentences out loud—stimulated us to replicate and extend Essien's work.

The 14 sentences that Essien chose for his experiment were short (an average of 5.1 words) and had built-in ambiguities based on differences in tone. Out of context, and juxtaposed against one another, such sentences are bound to cause readers a lot of difficulty. This was, of course, the object of the exercise, but we felt that participants would find the experiment more realistic if we presented them with less artificial sentences as stimuli.

What we needed, then, were real examples of spontaneously written text in a tone language that, like Efik, has no written literary tradition of its own.

Kom

The language map of Cameroon is very complex. Some 279 African languages (mostly of the Bantu family) are spoken by about 15 million Cameroonians (Grimes 2000:28). At independence in 1961, the Republic of Cameroon was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

formed from a French colony and an English colony. About three-fourths of all adult Cameroonians are bilingual, either in English or in French.

In 1978, Maurice Tadadjeu, a Cameroonian linguist, conceived a bold project—Projet de Recherche Operationnelle Pour l'Enseignement des Langues au Cameroun, or PROPELCA—to promote the teaching of literacy in every language recognized in the country at the time (Tadadjeu 1990; Tadajeu and Chia 1982; Tadadjeu, Gfeller, and Mba 1991). Tones were marked in the didactic texts developed by PROPELCA, and native speakers were taught to read and write them. In 1986, PROPELCA operated in 11 schools comprising four languages, including Ewondo, Lamnso', Douala, and Fe'efe'e (Mbah 1990).

Kom is a matrilineal kingdom in the Boyo Division of North West Province of Cameroon. Like Efik, Kom is also a Niger-Congo language, but it is part of the Western Grassfields group in Cameroon. There were 127,000 speakers of Kom at the last census in 1982 (Grimes 2000:41).

In the early 1980s, the Yaounde Branch of the Kom Bum Development Union (an association of Kom people who were living in the capital city of Yaounde) decided to promote the writing and teaching of Kom. They selected a few people and had them trained at seminars, known as Decouvre Ta Langue, given by the Summer Institute of Linguistics and the University of Yaounde. Several competing orthographies for Kom were extant, but none had become a standard. The Kom Bum Development Union charged a language committee to develop a Kom orthography and, in the late 1980s, the teaching of Kom was incorporated into the PROPELCA experiment.

Kom can have as many as eight phonetic tones including contours, or combinations of tones. These are high, mid, low, falling high-to-low, falling mid-to-low, rising low-to-high, rising low-to-mid, and rising mid-to-high. The Kom Language Development Committee recognized that native speakers do not need to see all the tones marked and kept only two of the eight tones, low [`] and falling from high to low [^], in the most recent orthography adopted for teaching Kom literacy in local schools (Chia 1984; Chia and Kimbi 1992).

Replicating Essien's Experiment

In 1989, Bernard taught five bilingual Kom-English speakers to use a word processor to write their native language. Mbeh was one of the five. The others were a Catholic priest, a lawyer recently graduated from the national university in Cameroon, and two primary school teachers. One of the school teachers produced a text of 102 Kom proverbs. Others in the project produced free-flowing texts on themes in Kom life. Mbeh, who was studying medical anthropology, wrote about Kom medical practices; the lawyer wrote about Kom marriage contracts; the priest wrote about Kom religious practices; and the second teacher wrote about scenes from her native village. In addition to the collection of proverbs, the group produced about 25,000 words of Kom text during the two-week session.

From this corpus of text, we selected 50 sentences randomly, including 12 of the 102 proverbs, that capture differences in writing styles and sentence structures in normal Kom discourse. Proverbs are scripted sentences that are repeated frequently in any society. We wanted to test whether proverbs, with and without tones marked, would be read faster than would ordinary sentences.

Mbeh typed each sentence on two cards, once with the tones marked and once without. No modifications were made to the sentences, except that the orthography was standardized. Some of the authors of the texts had used the [+], while others had used the [ü] for one of the high-central vowels, for example. (This had no effect on subsequent analyses.)

We then had two sets of cards. One set had 50 tone-marked, spontaneously written sentences of greatly varying length (from 6 to 67 words each) representing greatly varying styles and containing greatly varying content; a second set had exactly the same sentences but without the tones marked. There was one exception. One of our randomly selected sentences was a proverb that normally would not take any tone marks in Kom orthography. We should have selected another sentence but instead we decided not to mark the sentence artificially for tone (we now see this as an error). This proverb thus appears without tone marked in the set of 50 tone-marked sentences as well as in the set of 50 tone-unmarked sentences. In replications of the experiment for other languages, of course, we will not repeat the error. However, as will become evident, the results of the experiment are so strong that our mistake could not have made any difference.

We ran a prototype experiment using these changes in Essien's design with one native speaker of Kom (Bernard, Mbeh, and Handwerker 1995). The results were quite strong: Tones slowed down the reader's perception time and this added significantly to the probability that he would make a mistake in saying a sentence out loud. On the other hand, the longer the sentence, the more likely it was to be read correctly, irrespective of whether the tones were marked. In other words, tone got in the way, but context (more words) fixed the problem.

The Current Experiment

Of course, with just one native speaker (a man in his mid-40s who holds a doctorate in history from a U.S. university), we could not generalize even to speakers of Kom, much less to speakers of other tone languages. We repeated the experiment with 13 additional speakers of Kom. In the prototype experiment, the consultant read and responded to the pack of tone-unmarked sentences on one day and then the pack of tone-marked sentences the next day. This reflected Essien's design. In our second experiment, we mixed the two packs of sentence cards (tone unmarked and tone marked) into a single pack of 100 randomly ordered sentences, and numbered the sentences from 1 to 100. This eliminated order

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

effects that might have been produced by showing participants a tone-marked or tone-unmarked sentence first. However, all 13 participants saw the 100 sentences (tone-marked and tone-unmarked) in the same random order.

The test was administered by Mbeh in Njinikom and Belo, the heart of Kom country, during the spring and summer of 1994. All 13 participants could read and write English. Ten of the 13 had been through formal training on reading and writing Kom by PROPELCA, and six had even taught this system to others. Mbeh taught in PROPELCA from 1985 to 1989. In fact, 12 of the 13 could read Kom. Two of the three who had no formal training had picked it up from others.

Participants in our experiment were people who were willing and ready to participate. They varied widely in age, education, and professional backgrounds. There were seven women and six men, ranging in age from 10 to 62, and in education from 4 to 18 years (that is, from primary school to university level). Their occupations included one primary school pupil, two secondary school students, two farmers, three teachers, one Catholic catechist, one retired police officer, and three unemployed persons.

This is not a random sample of Kom speakers; however, all 13 participants were competent native speakers of Kom. Thus, we have 13 repetitions of an experiment from persons who are, by definition, equally capable of understanding and generating Kom. The participants were, of course, not equally competent in reading. However, our findings below make clear that this was not an issue for interpreting the results of the experiment regarding the influence of marking tone on perceiving and vocalizing correctly any given sentence.

At the beginning of each interview, we explained that we were studying the role of tones in reading Kom. Since all but one of our participants could read Kom, they required little instruction. We reminded them about how tones function in Kom and showed them how to read the two tones (low and falling) in tone-marked sentences. We also briefed them on the special characters (+, ŋ, ´) and the digraphs (gh, ny, oe, ae, ue) used in writing Kom (the ´ represents the glottal stop). The one participant who had no prior training in reading and writing Kom required more instruction, of course.

After the introduction, we gave participants eight sentences for practice: four with tones marked and four without tones marked. Finally, participants were given time to ask questions before the test began. The introductory exercise averaged about 15 minutes.

Participants were asked to pick up each card, look at the sentence written on it, and read it, silently, to themselves. They were told to take whatever time they needed and to read each sentence aloud only after they had figured out how to read it correctly.

The experiment took from 45 to 120 minutes per participant. There was always a break after the first 50 sentences, but in three cases the experiment had to be administered in two parts across two consecutive days. Some participants complained that the experiment took too long, but all said that they enjoyed the experience. One participant said the exercise had stimulated her to consider learning to read and write Kom more seriously.

Analysis: The Variable

Perception time was measured in the field with a digital stopwatch as the time between an informant's initial look at the sentence and initial vocalization. Vocalization time was measured with a digital stopwatch from tape recordings as the time between an informant's initial and final words. Both of these variables (perception time and vocalization time) and one of the independent variables (the number of words in each sentence) are measured in their natural logarithms because these yield homoscedastic residuals.[1] These variables are represented as **LP** (that is, the natural **L**og of the **P**erception time), **LV** (the natural **L**og of the **V**ocalization time), and **LW** (the natural **L**og of the number of **W**ords), respectively.

The third dependent variable, CORRECT, is binary and measures whether the participant correctly read each sentence aloud (1 = yes, 0 = no).[2] Mbeh listened to each sentence and decided whether the vocalization was free from error (scored as 1) or contained an error (scored as 0). If a sentence that was unmarked for tone could have more than one rendering, then any sensible rendering was scored as free from error.

The independent variable TONE refers to whether or not tone is marked in each sentence (1 = yes, 0 = no); INFORM is a character variable indicating the identity of each participant; PROVERB measures whether a sentence is a proverb (1 = yes, 0 = no); and SEQ is the sequence number, from 1 to 100, of the sentences read by participants. We use this last variable to test for learning effects in the experiment.

We coded 10 other variables: AGE is the informant's age in years; CERTIF is the number of years of schooling for each informant; LTONE is whether an informant had learned to read with tones (1 = yes and 0 = no); OUT is whether the person had spent time outside of Kom country (1 = yes and 0 = no); SEX is the sex of each informant; WRITE is whether the informant knows how to write Kom (1 = yes and 0 = no); TEACH is whether the informant teaches others to write Kom (1 = yes and 0 = no); TRAINING is whether the informant had received any training in reading and writing Kom (1 = yes and 0 = no); READ is whether the informant reads Kom (1 = yes, 0 = no); and READFREQ is how often the person claimed to read materials in Kom (an ordinal scale coded 0, 1, 2, 3).

AGE exhibits no effect in any subsequent analysis; CERTIF exhibits limited effects in subsequent analyses; and a factor analysis shows that the other eight variables (LTONE, OUT, SEX, WRITE, TEACH, TRAINING, READ, and READFREQ) reflect a single, underlying variable, which we interpret as a scale of functional literacy (Cronbach's alpha = .80). We labeled that scale LIT, and in the analyses that follow, we use the factor scores on that scale as an independent variable to predict LP, LV, and CORRECT.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Analysis: The Models

Tables 2 and 3 show ordinary least squares (OLS) regression solutions for perception time (LP) and vocalization time (LV) models, respectively.[3]

Analysis of both sets of residuals shows they are approximately normal and homoscedastic with no significant autocorrelation. Incorporating the informant variable (INFORM) effectively corrected for the autocorrelation built into the data by the experimental design (presenting 13 informants with 100 sentences in identical order). Condition indexes < 30.0 revealed the absence of multicollinearity disturbances (Belsley, Kuh, and Welsch 1980).

Note that in Tables 2 and 3 extraneous variables have been eliminated. Including extraneous independent variables in these models biases the predictions on the dependent variable (see Gujarati 1995). For example, neither the learning variable SEQ (p = .062) nor AGE (p = .081) showed any measurable direct effects on perception time, so these variables do not appear in Table 2. And there were no measurable direct effects on vocalization time from SEQ (p = .062), AGE (p = .875), CERTIF (p = .955), or LIT (p = .182), so these variables do not appear in Table 3.

In Table 2, we see that schooling experience (CERTIF) *shortens* perception time slightly. Since the values for the regression coefficients are in natural logarithms, we need to take their exponents to convert them to real measures. The coefficient –0.137 for CERTIF in Table 2 means that perception time goes down by 1.1 seconds (the natural logarithm of 1.1 is 0.137) for each year of schooling. However, functional literacy (LIT), *lengthens* perception time slightly. For each increment in functional literacy (as measured in factor scores), perception time grows by 1.6 seconds (the exponent of 0.455 is 1.6).[4]

Sentence length exerts a major effect on perception time. Sentences ranged from 6 to 67 words. Independently of the effects of other variables, six-word sentences took, on average,

exp((Ln(6)*0.474))

or 2.34 seconds to perceive and the 67-word sentence took, on average,

exp((Ln(67)*0.474))

or 7.34 seconds to perceive.

Proverbs were, as we would expect, easier to perceive, all things being equal. However, as with all sentences, perception time for proverbs was contingent on sentence length, and the advantage for proverbs turned into a penalty as sentences grew longer. For example, the six-word sentence that took 2.34 seconds to perceive if not a proverb took

exp((Ln(6)*0.474)+(-1.699)+(0.722*(Ln(6))))

or 1.56 seconds to perceive if it were a proverb. The 67-word

Table 2. Multiple Regression with Perception Time as the Dependent Variable

	Coefficient	P
CONSTANT	2.245	.000
LW	0.474	.000
INFORM		.000
INFORM 1	3.183	
INFORM 2	-0.184	
INFORM 3	-0.640	
INFORM 4	0.589	
INFORM 5	-2.202	
INFORM 6	0.368	
INFORM 7	-0.659	
INFORM 8	0.565	
INFORM 9	0.190	
INFORM 10	-0.256	
INFORM 11	0.455	
INFORM 12	-0.137	
LIT	0.455	.000
CERTIF	-0.137	.000
PROVERB	-1.699	.000
PROVERB*LIT	-0.131	.004
PROVERB*LW	0.722	.000
TONE	1.081	.000
TONE*SEQ	-0.009	.000
TONE*LW	-0.212	.003

Table 3. Multiple Regression with Vocalization Time as the Dependent Variable

	Coefficient	P
CONSTANT	-1.291	.000
LW	1.212	.000
LP	0.095	.000
INFORM		.000
INFORM 1	-0.092	
INFORM 2	0.247	
INFORM 3	-0.450	
INFORM 4	-0.280	
INFORM 5	-0.192	
INFORM 6	0.060	
INFORM 7	-0.211	
INFORM 8	-0.108	
INFORM 9	0.036	
INFORM 10	-0.126	
INFORM 11	-0.002	
INFORM 12	0.492	
PROVERB	-1.095	.000
PROVERB*LIT	-0.056	.028
PROVERB*LW	0.361	.000
TONE	0.480	.000
TONE*LW	-0.111	.005
TONE*SEQ	-0.003	.000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sentence that took 7.34 seconds to perceive if not a proverb took

exp((Ln(67)*0.474)+(-1.699)+(0.722*(Ln(67))))

or 27.93 seconds to perceive if it were a proverb.

Tone, independently of other effects, increases perception time. Like proverbs, the effect of tone is contingent on sentence length. But the penalty for tone did not disappear as sentences grew longer—the increase in perception time per word just grew smaller. For example, the six-word sentence that averaged 2.34 seconds to perceive took

exp((Ln(6)*0.474)+(1.081)+(-0.212*(Ln(6))))

or 4.71 seconds to perceive if it were marked by tone. The 67-word sentence that averaged 7.34 seconds to perceive took

exp((Ln(67)*0.474)+(1.081)+(-0.212*(Ln(67))))

or 8.87 seconds to perceive if it were marked by tone.

Variables measured qualitatively, like CORRECT, elicit questions like: What is the likelihood that a person says a sentence correctly (the dependent variable) under different sets of conditions (independent variables)? Logistic regression answers questions like these directly with odds ratios (OR). Odds ratios tell you the probability that a person says a sentence correctly, relative to the probability that the person says the sentence incorrectly. Two events that are equally likely to occur (a 50/50 split) produce an odds ratio of 1.00. If one event is more likely to occur than another (e.g., a 75/25 split), the odds ratio will be over 1.00. Similarly, if one event is less likely to occur than another (e.g., a 25/75 split), the odds ratio will be under 1.00. Table 4 shows the results.[5]

Summary test statistics (Hosmer-Lemeshow, Pearson, and Deviance) test the null hypothesis and confirm that observed values (1/0) correspond well with predicted values (in the range 1,0) (Hosmer and Lemeshow 1989). Examination of diagnostic statistics revealed no cases with high leverage or poor fit that distort model coefficients. A log-likelihood posthoc test for: 1) direct effects of AGE, CERTIF, LIT, SEQ, and TONE; and 2) interaction effects of TONE*LW and TONE*SEQ, yields a probability of 0.199. We infer that adding these variables to the model contributes nothing of importance.

Table 4. Logistic Regression: Likelihood of Saying Sentence Correctly

		95.0% Bounds	
Parameter	**Odds Ratio**	**Upper**	**Lower**
2 LP	0.572	0.713	0.459
3 LV	0.082	0.128	0.052
4 LW	16.599	30.995	8.890
5 INFORM	2.026	3.377	1.215
6 INFORM	0.341	0.648	0.179
7 INFORM	0.388	0.719	0.209
8 INFORM	0.643	1.119	0.369
9 INFORM	0.947	1.582	0.567
10 INFORM	1.206	2.011	0.724
11 INFORM	0.546	1.281	0.233
12 INFORM	1.389	2.336	0.825
13 INFORM	3.577	6.967	1.836
14 INFORM	5.303	11.031	2.549
15 INFORM	0.609	1.010	0.367
16 INFORM	1.324	2.468	0.710
17 PROVERB	45.452	948.374	2.178
18 PROVERB*LIT	1.664	2.429	1.140
19 PROVERB*LW	0.159	0.612	0.041

Log Likelihood of constants only model = LL(0) = -762.525

2*[LL(N)-LL(0)] = 540.385 with 18 df Chi-Square p-value = 0.000

McFadden's Rho-Squared = 0.354

	Statistic	**p-value**	**df**
Hosmer-Lemeshow	7.075	0.529	8.000
Pearson	1181.625	0.971	1276.000
Deviance	984.665	1.000	1276.000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Tables 2, 3, and 4 show that, as expected, individuals vary enormously in their capacity to perceive sentences quickly, say them quickly, and say them right. Still, *over and above variability attributable solely to individuals*, the following relationships hold:

1. As sentence length rises, so does perception time and vocalization time. As perception time rises, so does vocalization time.

2. Informants perceive and say proverbs more quickly than they do other sentences. They do so even more quickly to the extent that they report functional literacy skills (measured as each informant's factor score on LIT). However, the perception and vocalization speed advantage for proverbs falls as sentences grow longer.

3. Informants perceive and say sentences marked with tones less quickly than they do the same sentences unmarked for tone, but this penalty falls as sentences grow longer and falls slightly with repetition.

4. The longer it takes to perceive and say a sentence, the less likely it is that informants say it right.

5. Informants correctly say proverbs far more often than they correctly say other sentences. This effect is strengthened if informants report functional literacy skills (the LIT variable), but the advantage built into proverbs falls as sentences grow longer.

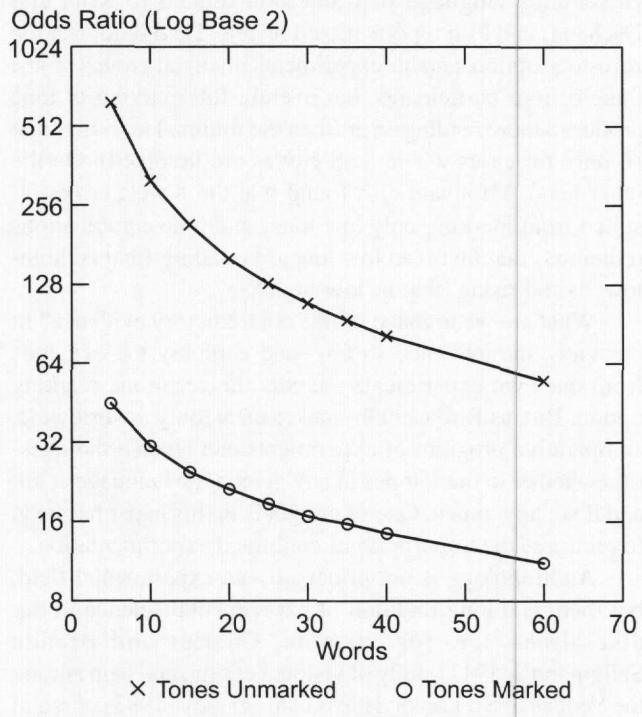
6. The longer the sentence, the more likely the informant will say it right, but informants take longer to perceive and vocalize long sentences. The balance of direct effects (longer sentences increase the probability of saying it right) and indirect effects (longer sentences increase perception and vocalization times, which decrease the probability of saying it right) goes to the latter: the observed (zero-order) probability of saying a sentence correctly declines with sentence length.

7. Marking tone contributes nothing directly to a reader's ability to say a sentence correctly. However, marking tone contributes indirectly to a reader's inability to say a sentence correctly by increasing perception and vocalization times. Tone marking penalizes readers of even the shortest sentences and the penalty grows as the length of sentences grow.

Figure 1 presents the odds ratios for correctly saying sentences of different lengths, unmarked and marked by tones, and thus highlights the practical difference that tone marking makes.

As sentence length grows, readers take longer to perceive and say the sentence whether tones are marked or not, and the odds of saying the sentence correctly fall. However, readers take less time to perceive and say sentences without tone marks. The odds of saying a sentence correctly are far higher for a sentence unmarked for tones than for the same sentence marked by tones. Indeed, the odds of correctly saying the longest sentences unmarked for tone (OR = 54.995) are higher than the odds of saying correctly the shortest sentences marked with tones (OR = 45.034). In short, marking

Figure 1. Odds Ratios for Saying Sentence Correctly, for Sentences of Various Lengths, Marked with Tones or Not



tone penalizes the reader. Fortunately for readers of tone-marked text, as sentence length increases—as context widens—they are able to recover from this penalty.

Discussion

Our findings are consistent with some of Essien's (1977) results, but inconsistent with others. Essien found that selected marking of tone on nouns and verb roots helped readers of Efik perceive and vocalize written sentences, but he also found that it caused readers to make mistakes in reading those sentences out loud.

Experiments by Bird (1999b) also tend to support our findings. He tested the ability of 11 speakers of Dschang (another Grassfields language of Cameroon related to Kom) to read texts that were either tone marked or tone unmarked. Tone marking penalized readers, adding an average of about 7.5 seconds per 100 words to the time it took native speakers to read the texts. It also added about 2.7 disfluencies (false starts, repetition of syllables, and errors), on average, per 100 words. On the other hand, Bird found that experienced readers made many fewer errors than inexperienced readers. Participants in his experiment marked two tone-unmarked, 200-word passages for tone. Six of the people who participated in Bird's experiment were experienced writers, including three full-time

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

language development workers. Among these six, the overall accuracy for marking tone was only 73 percent and just 22 percent for those without experience in writing Dschang.

Mfonyam (1989, reported in Bird 1999b) conducted an experiment on tone marking for Bafut (another Niger-Congo, Grassfields language of Cameroon related to Kom and Dschang) and also reports mixed results. He did not test the zero-tone option, and his experiments involved groups of just three or four participants, but overall, full marking of tone produced more reading errors than did minimal marking (that is, only for cases where ambiguities are involved). On the other hand, Mfonyam also found that the fewest errors resulted from marking only low tones and tone combinations (contours) that involved low tone, like falling (that is, high-low ^) and rising (that is, low-high ˅).

What are we to make of this contradictory evidence? In our view, the evidence so far—and certainly the evidence from our own experiments—favors the zero-tone marking option. But, as Bird (1999b) makes clear, only a worldwide, comparative program of experiments can resolve the question whether to mark tones in any given tone language at all, and if so, how much. One of our goals in this paper has been to reinforce the importance of continued experimentation.

Anthropology is not principally an experimental field, but there is a long tradition of experimental science in our discipline—see, for example, Charles and Brenda Seligmann's (1911) study of vision, hearing, and pain among the Veddas of Sri Lanka; Harris et al.'s study (1993) of racial categories in Brazil; recent work by Henrich (2000) testing differences in economic decision making among the Machiguenga of Peru and people in Los Angeles and by Gil-White (2002) on cognition about ethnicity in Mongolia. For the tone-marking problem, the experimental method is surely the method of choice.

There is more, however. While we wait for a program of experiments to resolve the scientific questions, language diversity, and with it, cultural diversity, diminishes almost daily. Language death is by now very well documented (Crystal 2000; Dorian 1981; Fishman 2001). Less well known, perhaps, are the success stories in language preservation.

On the one hand, language preservation means staving off language death by getting people of all ages to learn and to use a language as part of everyday life. Language nest programs (schools that involve total immersion in an ancestral language) have produced thousands of new speakers of Hawaiian (Donaghy 1998) and Maori (Fishman 2001). In California, master-apprentice programs, in which elders teach younger people the tribal ancestral language (Hinton 1998), are producing new speakers of American Indian languages, as are innovative grade-school programs in Arizona (Watahomogie and Yamamoto 1988).

On the other hand, language preservation means producing native-language texts that store and protect information about a language and the culture it reflects. We are enriched in our knowledge of many American Indian cultures by the indigenous-language texts that early anthropologists produced in collaboration with native speakers of those languages (see Bernard and Ryan 1998 for a review). Those texts are a precious resource today, not just for scholars of those languages and cultures, but for the people of those cultural groups who may be trying to revitalize their ancestral languages.

In Oaxaca, Mexico, the Centro Editorial de Literatura Indígena, A.C. (CELIAC) has taught 160 bilingual educators from six countries in Latin America to read and write their indigenous languages. At least a dozen of the Mexicans who have been through the program have produced books in Totonac, Ñähñu, Zapotec, Chinantec, and Mazatec (see Bernard 1985, 1997; González Ventura 1997; Salinas Pedraza 1997). The corpora of texts produced at CELIAC are available for study by linguists who need large amounts of text to test theories about grammars. Furthermore, since the books are written on word processors, the texts are machine readable and can be used for the construction of simple word lists and key-word-in-context dictionaries.

Languages that develop a literary tradition are less likely to succumb. One key to publishing in indigenous languages, we argue, is the use of orthographies that people find easy to read and write. For tone languages, this means not insisting on marking tone and not waiting for experiments to determine the optimum tone-marking solution for each language.[6]

Stimulating the local production of native-language texts often involves negotiating a thicket of local political battles over, for example, which of several dialects should represent a language in designing an orthography or which of several competing orthographies (sometimes the products of competing missionaries from different religions) should be chosen to write a language (see King 1996; Kroskrity 2000). The orthographies of the world's major languages, however, became standardized in the context of publishing books, using any orthography that people would read.

Books published in the United States 200 years ago still used the *f* grapheme for the sound /s/. Not very long ago, publishers in the United States used the character œ in the word "œdipus." Now the word is spelled "oedipus" and the digraph is gone from American English orthography. The circumflex "o" in "rôle" and the ç in "façade" are gone, too, but still appear in some British publications. The æ in "archæology" has become "ae," but early in the 20th century the digraph was dropped in U.S. government publications and many publishers now spell the word "archeology." Eventually, one spelling will emerge the winner. The word "millennium" appeared on the World Wide Web as "millennium" 41,814 times in June 1997 and as "millenium" 31,829 times (Newman 1997). Out of such orthographic *Sturm und Drang* are orthographic standards made.

This does not deny the importance of work by linguists on problems of orthography. Watson (1991:273), for example, has focused on the seven Moru-Ma'di languages, which stretch across parts of Sudan, Zaire, and Uganda. His objective is to promote, wherever possible, standardization of orthographies across Nilo-Saharan. It is clear that the linguistic insights of such meticulous work are of value in their

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

own right and will be duplicated neither by the free-for-all environment of the Internet nor by the competitive environment of publishing.

In fact, those insights will sometimes be decisive in the standardization of orthographies for indigenous languages, as is the case for the PROPELCA experiment in Cameroon (Tadajeu 1982). In other cases, linguistic insights will be ignored for political reasons. So, our argument is simple: If the objective is to help indigenous people keep their languages vital through the vehicle of literacy, then linguists would do well to help indigenous people start local publishing industries that will experiment in their own ways with orthographies that make sense.

With Bird (1999b), we advocate a program of sustained, comparative research to address the utility of tone marking for indigenous readers and writers. In the meantime, languages—and the cultures they carry—are lost every year.

Notes

[1]When the variance of the dependent variable in a regression model remains the same for all values of an independent variable, the model is homoscedastic ("evenly spread," in Greek) for that independent variable (Gujarati 1988:316). Homoscedasticity may be understood, then, as a constant level of prediction error for all values of the dependent variable.

[2]Of 1,300 cases, 100 for each of 13 participants, there are five cases of missing data: three cases for the variable LV and two cases for the variable CORRECT. These cases of missing data resulted from gaps in the tape after pauses in the experiment. Cases with missing data are excluded from analyses reported here.

[3]We used SYSTAT software to generate the output reported here (Steinberg and Colla 1991; Wilkinson 1992).

[4]Note that scores for only 12 of the 13 informants are listed in Tables 1 and 2. The procedure of converting a *k*-category variable into binary variables for regression analysis produces *k*-1 new variables. The regression coefficient for each binary variable expresses the difference between any one of the *k*-1 variables and the reference group or person. SYSTAT automates the process and, by default, assigns the *k*th category as the reference group/person.

[5]Logistic regression also avoids the heteroscedasticity (see Note 1) that arises when OLS regression is applied to binary variables (see Gujarati 1988 for a discussion of this issue). The logistic regression solution presented in Table 4 was solved with White's (1982) quasi-maximum likelihood estimation procedure to guard against erroneous standard errors generated by possible model misspecification.

[6]There are many examples of local literary traditions that develop on their own, without the use of mass printing technology. Besnier (1995) reports on the use of a local script for writing letters on the Pacific island of Nukulalae. Cherokee is written with a syllabary invented independently by Sequoyah in the 1830s (Walker 1981, 1984). The Vai of Liberia use a local script, in addition to Arabic and English, and some Vai men write English and Arabic with Vai script (Scribner and Cole 1981; see also Dalby 1967). Early in the 20th century, Silas John created a writing system for Apache (Basso and Anderson 1973), as did King Njoya, in Cameroon, for Bamun. Several of the Munda tribal languages of central India developed their own writing systems in the 20th century (Zide 1996), and the Pahawh script for writing Hmong (Smalley, Koua Vang, and Yee Yang 1990) was developed in 1959 by Shong Lue Yang, a previously nonliterate peasant.

References Cited

Awedoba, Albert K.
1990 Kasem Tones and Orthography. Journal of West African Languages 20:31-46.

Basso, Keith H., and Ned Anderson
1973 A Western Apache Writing System: The Symbols of John Silas. Science 180 (4090):1013-1022.

Belsley, David A., Edwin Kuh, and Roy E. Welsch
1980 Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Bernard, H. Russell
1974 Otomí Tones in Discourse. International Journal of American Linguistics 40:141-150.
1980 Orthography for Whom? International Journal of American Linguistics 46:33-36.
1985 The Power of Print: The Role of Literacy in Preserving Native Cultures. Human Organization 44:88-92.
1997 Language Preservation and Publishing. *In* Indigenous Literacies in the Americas: Language Planning from the Bottom Up. Nancy H. Hornberger, ed. Pp. 139-156. Berlin: Mouton de Gruyter.

Bernard, H. Russell, and Gery Ryan
1998 Text Analysis: Qualitative and Quantitative Methods. *In* Handbook of Methods in Cultural Anthropology. H. Russell Bernard, ed. Pp. 595-646. Walnut Creek, Calif.: AltaMira.

Bernard, H. Russell, George N. Mbeh, and W. Penn Handwerker
1995 The Tone Problem. *In* The Complete Linguist. Anthony Traill, Rainer Vossen, and Megan Biesele, eds. Pp. 27-44. Papers in Memory of Patrick J. Dickens. Cologne: Rüdiger Köppe Verlag.

Besnier, Niko
1995 Literacy, Emotion, and Authority: Reading and Writing on a Polynesian Atoll. New York: Cambridge University Press.

Bird, Steven
1999a Strategies for Representing Tone in African Writing Systems. Written Language and Literacy 2:1-44.
1999b When Marking Tone Reduces Fluency: An Orthography Experiment in Cameroon. Language and Speech 42:83-115.

Chia, Emmanuel
1984 Guide to the Kom Alphabet. Yaounde, Cameroon: Summer Institute of Linguistics.

Chia, Emmanuel, and Joseph C. Kimbi
1992 Guide to the Kom Alphabet. Rev. ed. Yaounde, Cameroon: Societé Internationale de Linguistique, B.P. 1299.

Crofts, Marjorie
1976 Must Tone Always Be Written in a Tonal Language? Technical Papers for the Bible Translator 27(1):127-134.

Crystal, David
2000 Language Death. New York: Cambridge University Press.

Dalby, David
1967 A Survey of the Indigenous Scripts of Liberia and Sierra Leone: Vai, Mende, Loma, Kpelle, and Bassa. African Language Studies 8:1-51.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Donaghy, Keola
1998 'Olelo Hawai'i: A Rich History, a Bright Digital Future. Cultural Survival Quarterly 22(4):45-49.

Dorian, Nancy C.
1981 Language Death: The Life Cycle of a Scottish Gaelic Dialect. Philadelphia: University of Pennsylvania Press.

Essien, Udo E.
1977 To End Ambiguity in a Tone Language. *In* Language and Linguistic Problems in Africa: Proceedings of the VII Conference on African Linguistics. Paul F. A. Kotey and Haig Der-Houssikian, eds. Pp. 155-167. Columbia, S.C.: Hornbeam Press.

Fagborun, J. G.
1989 Problems in Yoruba Orthography. Journal of West African Languages 19:74-92.

Fishman, Joshua, ed.
2001 Can Threatened Languages Be Saved? Reversing Language Shift, Revisited: A 21st Century Perspective. Buffalo, N.Y.: Multilingual Matters.

Gil-White, Francisco J.
2002 The Cognition of Ethnicity: Native Category Systems under the Field-Experimental Microscope. Field Methods 14:161-189.

González Ventura, Josefa
1997 Experiences in the Development of a Writing System for Ñuu Savi. *In* Indigenous Literacies in the Americas: Language Planning from the Bottom Up. Nancy H. Hornberger, ed. Pp. 157-170. Berlin: Mouton de Gruyter.

Grimes, Barbara F., ed.
2000 Ethnologue. Languages of the World. Vol. 1. Dallas: Summer Institute of Linguistics.

Gudschinsky, Sarah C.
1970 More on Formulating Efficient Orthographies. The Bible Translator 21:21-25.

Gujarati, Damodar N.
1988 Basic Econometrics. 2nd ed. New York: McGraw-Hill.
1995 Basic Econometrics. 3rd ed. New York: McGraw-Hill.

Harris, Marvin, Josildeth G. Consorte, Joseph Lang, and Bryan Byrne
1993 Who Are the Whites? Imposed Census Categories and the Racial Demography of Brazil. Social Forces 72:451-462.

Henrich, Joseph
2000 Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon. American Economic Review 90:973-979.

Hinton, Leanne
1998 Language Loss and Revitalization in California: Overview. International Journal of the Sociology of Language 132:83-93.

Hosmer, David W., and Stanley Lemeshow
1989 Applied Logistic Regression. New York: John Wiley.

King, Kendall
1996 Indigenous Politics and Native Language Literacies: Recent Shifts in Bilingual Education Policy and Practice in Ecuador. *In* Indigenous Literacies in the Americas. Nancy H. Hornberger, ed. Pp. 267-284. Berlin: Mouton de Gruyter.

Kroskrity, Paul V., ed.
2000 Regimes of Language: Ideologies, Polities, and Identities. Santa Fe, N.M.: School of American Research Press.

Lojenga, Constance K.
1993 The Writing and Reading of Tone in Bantu Languages. Notes on Literacy 19:1-19.

Longacre, Robert
1964 A Tone Orthography for Trique. The Bible Translator 4:8-13. (First published in 1953.)

Mbah, Gabriel
1990 Langues maternelles au niveau du primaire (Model 2). *In* Le Defi de Babel au Cameroun. Maurice Tadadjeu, ed. Pp. 81-130. Collection PROPELCA No.53. Yaounde: University of Yaounde.

Newman, Barry
1997 Quizz: How do u spel milenium? Not That Way, Actually. Wall Street Journal, June 25:A1, A12.

Pike, Kenneth
1948 Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion. Ann Arbor: University of Michigan Publications in Linguistics.
1964 Phonemics: A Technique for Reducing Languages to Writing. Ann Arbor: University of Michigan Press. Vols. 1 and 2. Glendale, Calif.: Summer Institute of Linguistics. (First published in 1947.)

Salinas Pedraza, Jesús
1997 Saving and Strengthening Indigenous Mexican Languages. *In* Indigenous Literacies in the Americas: Language Planning from the Bottom Up. Nancy H. Hornberger, ed. Pp. 171-188. Berlin: Mouton de Gruyter.

Scribner, Sylvia, and Michael Cole
1981 The Psychology of Literacy. Cambridge, Mass.: Harvard University Press.

Seligmann, Charles G., and Brenda Z. S. Seligmann, with Charles S. Myers, and Abraham M. Gunesekara
1911 The Veddas. Cambridge: The University Press.

Smalley, William A., Chia Koua Vang, and Gnia Yee Yang
1990 Mother of Writing: The Origin and Development of a Hmong Messianic Script. Chicago: University of Chicago Press.

Steinberg, Dan, and Phillip Colla
1991 LOGIT: A Supplementary Module for SYSTAT. Evanston, Ill.: SYSTAT, Inc.

Tadadjeu, Maurice
1990 Le Defi de Babel au Cameroon. Collection PROPELCA No. 53. Yaounde: University of Yaounde.

Tadadjeu, Maurice, and Emmanuel Chia
1982 Projet de recherche operationnelle pour l'enseignement des langues au Cameroun (PROPELCA). Centre de Recherches et d'Etudes Anthropologiques, Department des Langues et de Linguistique, University of Yaounde.

Tadadjeu, Maurice, E. Gfeller, and G. Mba
1991 Training Manual for the Teaching of National Languages in Primary Schools. PROPELCA No. 56. Younde: University of Yaounde.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Voorhoeve, Jan
1962 Some Problems in Writing Tone. The Bible Translator 13:34-38.

Walker, Willard
1981 Native American Writing Systems. *In* Language in the USA. Charles A. Ferguson and Shirley Brice Heath, eds. Pp. 144-174. Cambridge: Cambridge University Press.
1984 The Design of Native Literacy Programs and How Literacy Came to the Cherokee. Anthropological Linguistics 26:161-169.

Watahomigie, Lucille J., and Akira Y. Yamamoto
1988 Linguistics in Action: The Hualapai Bilingual/Bicultural Education Program. *In* Collaborative Research and Social Change: Applied Anthropology in Action. Donald D. Stull and Jean J. Schensul, eds. Pp. 77-98. Boulder, Colo.: Westview Press.

Watson, Richard L.
1991 Moru-Ma'di Orthographies. *In* Proceedings of the Fourth Nilo-Saharan Linguistics Colloquium. Nilo-Saharan: Linguistic Analysis and Documentation. Vol. 7. M. Lionel Bender, ed. Pp. 273-282. Hamburg: Helmut Buske Verlag.

White, Halbert
1982 Maximum Likelihood Estimation of Misspecified Models. Econometrica 50:1-25.

Wilkinson, Leland
1992 SYSTAT: The System for Statistics. Evanston, Ill.: SYSTAT, Inc.

Wiessmann, Ursula
1989 Orthography Matters. Notes on Literacy 57:14-21.
1995 Tone Orthography and Pedagogy. Notes on Literacy 21:25-31.

Zide, Norman
1996 Scripts for Munda Languages. *In* The World's Writing Systems. William Bright and Peter T. Daniels, eds. Pp. 612-618. New York: Oxford University Press.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.