

Dialects, Orthography and Society

John M. Clifton

SIL International and University of North Dakota

*While there is broad consensus that both linguistic and non-linguistic factors play a role in orthography, there is disagreement as to the relative weight played by the two sets of factors. In this paper, I present case studies from four languages, two from Papua New Guinea and two from Bangladesh, in which orthography interacts with dialectal differences. In two of the studies, orthography is used to show the unity of the communities, while in the other two it is used to establish separate identities. This interaction of orthography and group identity follows naturally from the claim that orthography is primarily a social practice, not a linguistic construct.**

1. Introduction

It is generally accepted that both linguistic and non-linguistic factors are relevant to orthography design. There is less agreement, however, on which factors are more important. The dominant view is that orthography design should begin with linguistic analysis and then be modified on the basis of non-linguistic factors. For example, while Pike (1947:208–215) lists phonemic and social goals to be considered in orthography development, the phonemic goals are clearly seen to be more basic in statements like: “It is sometimes necessary to write submembers of phonemes ... under this type of cultural pressure (Pike 1947:214).” For Pike, a good orthography is based first on the phonemic analysis, and is only modified by social goals when forced by social considerations. More recently, Seifart (2006:288) made a similar point in arguing for:

... the need for a thorough linguistic analysis as a basis for orthography development and for the advantage of an orthography that respects the particular structure of the languages as opposed to one that uses conventions of orthographies of surrounding languages with established orthographies.

Working within the framework of New Literacy Studies, Sebba (2009) argues that this view is incorrect. For Sebba, orthography is primarily a social practice, not the embodiment of the linguistic structure. Non-linguistic factors, rather than occupying a peripheral position, are central. Discussing the relationship of linguistic and non-linguistic factors, Sebba (2009:46) says:

While no doubt the primary work of writing systems is to provide written medium for linguistic messages, in virtually every case, sociolinguistic work is done as well. Frequently this sociolinguistic work has to do with the construction of identities.

In this paper I show through four case studies that the construction of identities is a crucial component in understanding how orthographies are used in the context of dialect differences. In the first two case studies, one from Papua New Guinea and one from Bangladesh, the communities speaking different dialects use orthography to show they are a unitary group. In the other two case studies, again from Papua New Guinea and Bangladesh, orthography is used by communities to show they are different from related communities.

* This paper was presented at Methods in Dialectology 14, University of Western Ontario, August 2-6, 2011.

2. Case Study 1: Vanimo

Vanimo [vam] is a non-Austronesian language spoken by almost 2700 people in three villages on the north coast of Papua New Guinea. Its location is shown in Figure 1.

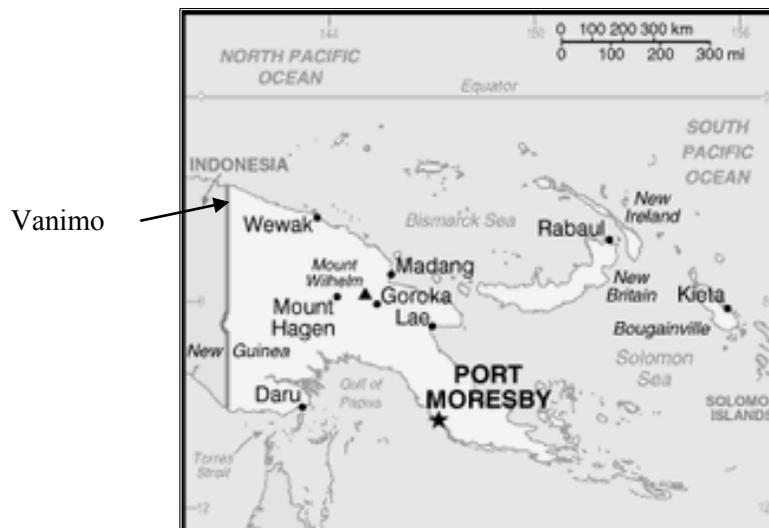


Figure 1. Location of Vanimo Language¹

In 1993 my wife and I were asked to work together with members of the Vanimo community to develop an orthography. Over a period of seven days, we guided a self-selected group who decided how to write the language and then applied their decisions as they wrote and edited stories (John M. Clifton & Deborah A. Clifton 1995).

The group contained speakers of both dialects of Vanimo: Vanimo and Waromo. A major concern of the group was to ensure that whatever decisions were made would be suitable for both dialects. One of the main differences between the two dialects is that the contrast between /h/ and /g/ in the Vanimo dialect is neutralized as /ʔ/ in the Waromo dialect. Contrastive pairs are shown in (1).

(1)	Vanimo	Waromo	
	/gâ/ ²	/ʔâ/	‘moon’
	/sagæ/	/saʔæ/	‘smoke’
	/hja:/	/ʔja:/	‘sea’

We made a number of suggestions for how to handle this contrast: Write <g> and <h> in Vanimo and <ʔ> in Waromo; overdifferentiate in Waromo by using the Vanimo forms in both dialects; underdifferentiate in Vanimo by using the Waromo forms in both dialects. The workshop participants proposed a fourth alternative: Write <g> and <h> in Vanimo, and <gh> in Waromo. By using <gh> instead of the <ʔ> used in neighboring languages for /ʔ/, visual unity could be maintained between the two dialects without having to over- or underdifferentiate in either dialect.

3. Case Study 2: Kok Borok and Kau Brung

Kok Borok [tp] and Kau Brung [usi] are Tibeto-Burman languages. Kok Borok is spoken by 5,000 people in the Chittagong Hill Tracts and by 691,000 people in India, while Kau Brung is spoken by almost 22,500 people in the Chittagong Hill Tracts. The locations of both languages are shown in Figure 2.

¹ Maps are from *The world factbook* (2011) and are in the public domain.

² Vanimo is tonal. The circumflex indicates a falling tone.

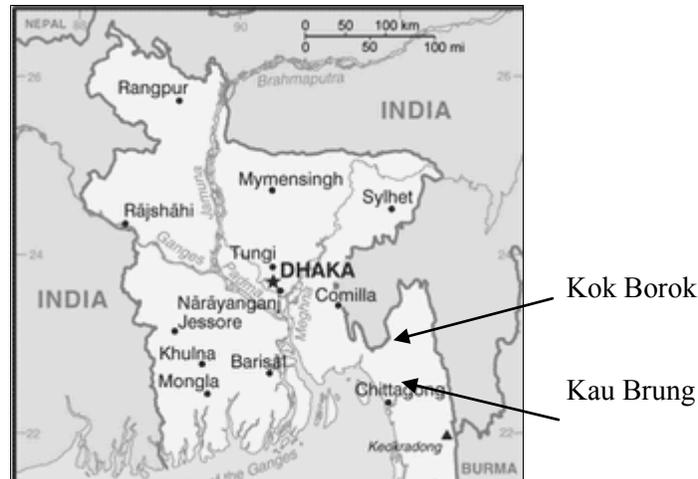


Figure 2. Location of Kok Borok and Kau Brung Languages

On the basis of word list comparisons and self-reported intelligibility, Kim et al. (2011) conclude that the mutual intelligibility between Kok Borok and Kau Brung is low; they should be considered separate languages. The communities speaking these two languages, however, are part of the larger Tripura ethnic community which consists of 36 clans, the majority of whom live in India. Kau Brung is spoken by the Usoi clan, while Kok Borok is spoken by a number of Tripura clans. Given the linguistic differences but ethnic unity, Kim et al. (2011) propose that while different materials are needed for the two languages, development should occur within the context of the overall Tripura community.

Historically the languages of the Tripura communities were written using both the Bangla and Roman scripts. In the early 2000s, the Tripura community in India decided to use only the Roman script (Debbarma 2006). By 2008, when I began consulting on orthography with language committees from eight language groups in the Chittagong Hill Tracts, both Tripura communities in Bangladesh had decided to also use only Roman script. The major reason for this decision was a desire to maintain the unity of the Tripura community.

During a workshop concentrating on orthography and literacy material production, both the Kok Borok and Kau Brung committees noticed a problem in the orthography: The letter <n> was realized as /n/ in some words, but as nasalization on the preceding vowel in others. A contrastive pair is given in (2).

- | | | | |
|-----|-----------------------|------------------------|--------------|
| (2) | /punsa/ | <punsa> | ‘(goat) kid’ |
| | /nũk ^h uŋ/ | <nun ^h kuŋ> | ‘roof’ |

We discussed a number of options for dealing with this difference including diacritics and underdifferentiation. The committees for the Kok Borok and Kau Brung speaking communities ultimately decided to mark the nasalized vowels with a circumflex followed by the nasal. Following this decision, the words in (2) would be written as <punsa> and <nũn^hkuŋ>. We discussed the pros and cons of writing the ‘silent’ nasal. The decision to keep it was made to maintain visual unity between varieties with nasalized vowels and varieties with a full nasal consonant.

4. Case Study 3: Bine

Bine [bon] is a non-Austronesian language spoken by approximately 2000 people in nine villages on the south coast of Papua New Guinea. Its location is shown in Figure 3.

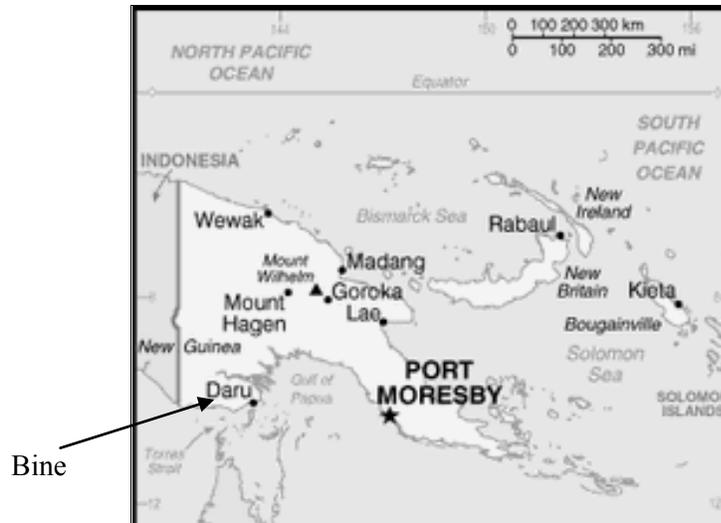


Figure 3. Location of Bine Language

Bine can be divided into seven dialects along village lines (Fleischmann & Turpeinen 1976). In the Boje-Giringarede dialect voiced stops /b d g/ contrast with prenasalized stops /mb nd ŋg/. Examples are given in (3) at the alveolar and velar points of articulation.³

- (3) /dade/ ‘platform’ vs. /gændæ/ ‘tree kangaroo’
 /tago/ ‘mouth’ vs. /dreŋgo/ ‘dog’

In the Kunini dialect, however, there are no prenasalized stops. Examples from the Kunini dialect corresponding to those in Boje-Giringarede are given in (4).

- (4) /dade/ ‘platform’ vs. /gædæ/ ‘tree kangaroo’
 /tage/ ‘mouth’ vs. /drego/ ‘dog’

There is also contrast between /n/ and /ŋ/ in both dialects, as shown in example (5).

- (5) /nano/ ‘boy’ vs. /ŋupo/ ‘rain’ Boje-Giringarede
 /nani/ ‘drink’ vs. /ŋupe/ ‘rain’ Kunini

Fleischmann and Turpeinen, linguists working with the Bine people, felt it was important to have consistent grapheme/phoneme correspondences in the Bine dialects (Fleischmann, personal correspondence). The crucial question was how to represent the velar nasal and prenasalized stop in Boje-Giringarede. The three options are given in (6).

- (6) a. /ŋ/ <ng> /ŋg/ <ng>
 b. /ŋ/ <ng> /ŋg/ <nng>
 c. /ŋ/ <ŋ> /ŋg/ <ŋg>

Option (6a) was rejected as leading to ambiguity. Option (6b) was rejected since it would result in sequences of three consonants, and there was fear that the second <n> in <nng> would be deleted in casual use. Therefore, option (6c) was chosen for Boje-Giringarede.

Writing /ŋ/ as <ng> would not have been problematic in the Kunini dialect since there is no /ŋg/. To maintain consistency between the dialects, however, the decision was made to write /ŋ/ as <ŋ> in Kunini as well. This decision was reflected in primers used in literacy classes at least back to 1979, and when

³ Boje-Giringarede examples are from Fleischmann (Fleischmann 1984a); Kunini examples are from Fleischmann (Fleischmann 1984b). Fleischmann and Turpeinen (1977) present a phonemic analysis of the Kunini dialect.

Debbie Clifton was asked to produce materials for use in vernacular preschools in 1992 she used the <ŋ> for /ŋ/.

In spite of this use of <ŋ>, the people of Kunini village asked for a meeting to discuss a number of orthography decisions before the New Testament was published in the Kunini dialect. The most contentious issue was the use of <ŋ>. We explained that by using <ŋ> the dialects could all use the same orthographic conventions. But the people of Kunini were not swayed by this argument: They decided that /ŋ/ should be written as <ng> in Kunini, even though this meant that /ŋ/ would not be realized uniformly in all the dialects.

5. Case Study 4: Tanchangya and Chakma

Tanchangya [tnv] and Chakma [ccp] are Indo-Iranian languages. Tanchangya is spoken by 21,600 people in the Chittagong Hill Tracts of Bangladesh, while Chakma is spoken by 150,000 people in the Chittagong Hill Tracts of Bangladesh and by 400,000 people in India. The locations of both languages in Bangladesh are shown in Figure 4.

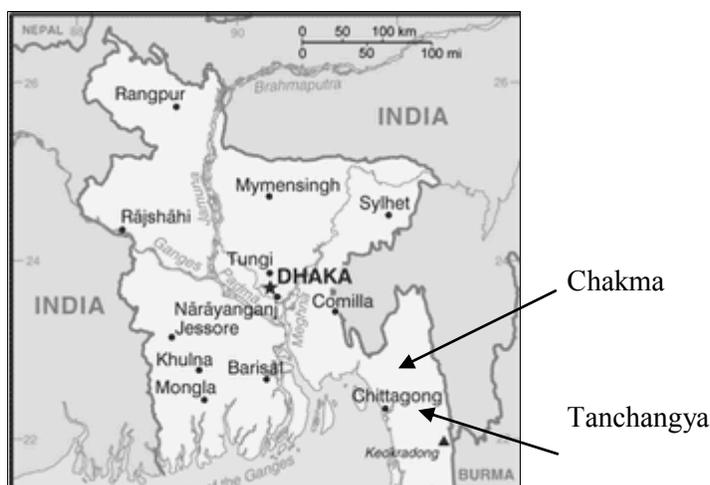


Figure 4. Location of Tanchangya and Chakma Languages

On the basis of wordlist comparison, recorded text tests,⁴ and self-reported dialect perceptions, Maggard et al. (2007) suggest that Tanchangya and Chakma are closely-enough related that it might be possible for the two communities to use a common literature. They also note that there are already materials published in Chakma. At the same time, they indicate that there is an asymmetrical relationship between the two groups. While the Chakma feel the Tanchangya are a part of the Chakma community, the Tanchangya consider themselves to be separate.

The ambivalent relationship between the Tanchangya and Chakma communities is reflected in orthographic decisions. The decision was recently made to write Chakma with the traditional Burmese-based Chakma alphabet documented by Grierson (1903) instead of the Bangla alphabet. The graphemes for the five basic points of articulation in this alphabet are shown in Table 1.

⁴ In a recorded text test, subjects listen to a story from another speech variety, and answer question about the story. A combination of the average number of correct responses and the standard deviation can be used to determine levels of mutual and learned intelligibility (Casad 1974).

Table 1. Chakma Alphabet

	vls unasp	vls asp	vd unasp	vd asp	nasal
velar	ᱠ	ᱡ	ᱢ	ᱣ	ᱤ
alveopalatal	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ
retroflex	ᱪ	ᱫ	ᱬ	ᱭ	ᱮ
dental	ᱯ	ᱰ	ᱱ	ᱲ	ᱳ
labial	ᱵ	ᱶ	ᱷ	ᱸ	ᱹ

The Tanchangya could have used the Chakma, Burmese,⁵ or Bangla alphabets. However, they decided to design a new alphabet based on, but different from, the Chakma. The graphemes for the five basic points of articulation in this new Tanchangya alphabet are shown in Table 2.

Table 2. Tanchangya Alphabet

	vls unasp	vls asp	vd unasp	vd asp	nasal
velar	ᱠ	ᱡ	ᱢ	ᱣ	ᱤ
alveopalatal	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ
retroflex	ᱪ	ᱫ	ᱬ	ᱭ	ᱮ
dental	ᱯ	ᱰ	ᱱ	ᱲ	ᱳ
labial	ᱵ	ᱶ	ᱷ	ᱸ	ᱹ

A majority of the graphemes are identical in the two alphabets. Where differences occur in the Tanchangya alphabet, they are variations on the basic Chakma form. For example, there is a change in orientation for the voiceless aspirated velar (C ᱠ vs. T ᱡ). The points for the voiced aspirated dental and voiced unaspirated bilabial are rounded in Tanchangya (C ᱯ ᱵ vs. T ᱰ ᱶ). The outer loops are extended for the retroflex nasal (C ᱮ vs. T ᱮ), but shorted for the dental nasal (C ᱳ vs. T ᱳ). Even the graphemes showing the greatest differences, the voiceless aspirated and unaspirated retroflex, have a horizontal bar on top, a loop, and a difference between a secondary loop and the lack of a loop (C ᱪ ᱫ vs. T ᱪ ᱫ).

Sebba (2007) argues that in their orthographies, communities exhibit a tension between identification with, and differentiation from, the language on which the orthography is based. This is exactly what we see in the case of Tanchangya. The Tanchangya community decided to base their alphabet on the Chakma to show they were related to the Chakma. However, they systematically changed the alphabet to show that they were different from the Chakma – they were not simply a part of the larger Chakma community.

6. Conclusions

There is no linguistic explanation as to why some communities speaking different dialects want to maintain orthographic unity while others do not. Considering only linguistic factors we would expect that Chakma and Tanchangya could be written with a single orthography, since the two varieties are close enough that they could probably share a common literature. On the other hand, there is no obvious linguistic reason to predict that speakers of Kok Borok and Kau Brung would want to maintain a common orthography. In fact, while the two groups initially attempted to develop a single set of literacy materials to meet the needs of both groups, they have since decided the varieties are too different for that to be a viable option. The question remains why they want to maintain a common orthography.

Referring back to the claims of Sebba (2009) outlined in the introduction, I claim that the differences can only be accounted for in terms of social identity. Speakers of Tanchangya see their identity as separate from the Chakma, while speakers of Kok Borok and Kau Brung both see their identity as members of the larger Tripura community. Group ethnonyms highlight this difference in the Bangladeshi case studies: Speakers of Kok Borok and Kau Brung have a single ethnonym and see themselves as

⁵ They did decide to use the Burmese numerals, unchanged.

members of the larger Tripura community which straddles India and Bangladesh, while there is no ethnonym that covers both Tanchangya and Chakma communities.

Group ethnonyms do not help in the case of the two case studies from Papua New Guinea. The villages of Kunini, Boje and Giringarede all see themselves as part of the larger Bine community. But this identity is evidently not strong enough to convince them to continue to use the <η> and maintain a visual unity with the Boje-Giringarede community.

There are several lessons to be learned from these case studies. First, social factors, especially issues of identity, are often more important than linguistic factors when designing orthographies to meet the needs of multi-dialectal communities. This suggests that we need to spend a significant amount of time studying social factors. Second, even if we do study social factors, it is likely that we will miss crucial elements that will affect final orthography decisions.

These two lessons lead to a third lesson: It is important to include the community from the beginning when making orthography decisions. This may well involve challenging some of our deeply felt convictions. On the one hand, we may find a push from the local communities to unify orthographies for speech varieties that will cause problems in reading for significant numbers of users. On the other hand, we may find that the wish of a community to distinguish itself from related communities may result in a body of literature that will have a smaller pool of potential readers. In the end, the most important issue is whether the communities involved will accept the orthography, not whether the orthography is linguistically correct or whether it will be usable by the greatest number of readers.

References

- Casad, Eugene. 1974. *Dialect intelligibility testing*. Norman, OK: Summer Institute of Linguistics.
- Clifton, John M. & Deborah A. Clifton. 1995. Vanimo orthography workshop. *READ* 30(2). 3-8.
- Debbarma, Binoy (ed.). 2006. *The national seminar on Kokborok script - with special reference to its tonal representation*. Agartala, India: Kokborok Tei Hukumu Mission.
- Fleischmann, Lillian (ed.). 1984a. *Bine acnoma buka: Boje-Giringarede tago pogogo ca*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Fleischmann, Lillian (ed.). 1984b. *Bine acnema buka: Kunini tage pogogo ca*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Fleischmann, Lillian & Sinikka Turpeinen. 1976. A dialect survey of Eastern Trans-Fly languages. *Papers in New Guinea linguistics, no. 19*, 39-76. (Pacific Linguistics A-45). Canberra: Australian National University.
- Fleischmann, Lillian & Sinikka Turpeinen. 1977. Bine phonemes. *Miscellaneous papers in P.N.G. linguistics*, 99-144. (Workpapers in Papua New Guinea Languages 22). Ukarumpa, Papua New Guinea: Summer Institute of Linguistics. <http://www.sil.org/pacific/png/abstract.asp?id=15669>.
- Grierson, George Abraham Sir 1851-1941. 1903. *Linguistic survey of India*. Calcutta: Govt. of India, Central Publication Branch.
- Kim, Amy, Seung Kim, Palash Roy & Mridul Sangma. 2011. *The Tripura of Bangladesh: A sociolinguistic survey*. (SIL Electronic Survey Reports 2011-038). Dallas: SIL International. <http://sil.org/silesr/abstract.asp?ref=2011-038>; accessed Dec 23, 2011.
- Maggard, Loren, Mridul Sangma & Sayed Ahmad. 2007. A sociolinguistic survey among the Chakma and Tanchangya communities. Dhaka, Bangladesh, ms.
- Pike, Kenneth L. 1947. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan Press.
- Sebba, Mark. 2007. *Spelling and society: The culture and politics of orthography*. Cambridge: Cambridge University Press.

- Sebba, Mark. 2009. Sociolinguistic approaches to writing systems research. *Writing Systems Research* 1. 35-49.
- Seifart, Frank. 2006. Orthography development. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*. New York: Mouton de Gruyter.
- The world factbook*. 2011. Washington, D.C.: Central Intelligence Agency.
<https://www.cia.gov/library/publications/the-world-factbook/index.html>; accessed Nov 30, 2011.