



Journal of the Linguistic Society of Papua New Guinea

ISSN: 0023-1959

Vol. 34 No. 1, 2016

Orthography as Social Practice: Lessons from Papua New Guinea*

John M. Clifton

SIL International and University of North Dakota

Sebba argues that orthography is a social, not linguistic, construct. In this paper I present four case studies from Papua New Guinea in which communities use orthography to encode the social relationships between dialects as well as between the language community and languages of wider communication. I then consider the question of how to best work with communities to develop an orthography. Drawing on workshops conducted in Papua New Guinea and in Bangladesh, I propose it is possible to develop an orthography based on community practices rather than on a phonological analysis. Rather than beginning with a phonological inventory, community members can directly analyze current or incipient attempts to write the language. Problems with conventions that are developing will become apparent during this process, and the community members can be given tools to solve these problems. By breaking the link between orthography and phonological analysis, the resulting orthography will more closely reflect the social dynamics that are important for wide-spread acceptance of any orthographic system.

1 Introduction

It is generally agreed that non-linguistic factors play an important role in developing orthographies. In an early discussion of the design of practical orthographies, Pike (1947) proposes that a “practical orthography should be chosen in such a way as to obtain an acceptable balance between phonemic principles and general sociological situations.” Similar considerations of both linguistic and non-linguistic factors are echoed in Nida (1953, 1954), Smalley (1959), Berry (1968, 1977) and Venezky (1970). More recently, Seifart (2006) discusses both linguistic and non-linguistic factors in orthography development within the context of language documentation, while Cahill and Rice’s (2014) collection of papers on developing orthographies begins with Cahill’s (2014) contribution cataloging non-linguistic factors. In addition to these more theoretical studies, there is a multiplicity of case studies in which the role of non-linguistic factors has been discussed.

While there is general agreement that it is important to consider non-linguistic factors when developing an orthography, there is disagreement about the relative importance of linguistic versus non-linguistic factors. After discussing the influence of the orthographic tradition of the national language, Pike (1947) comments that “It is sometimes necessary to write submembers of phonemes ... under this type of cultural pressure.” The phrases ‘sometimes necessary’ and ‘cultural pressure’ suggest a position that the best orthography is linguistically based, although sometimes it might be necessary to deviate from this ideal. I would summarize Pike’s approach to the relationship between linguistic and non-linguistic factors in the following three principles:

1. Some orthographies are linguistically better suited to a particular language than are others.
2. Orthography development should begin by designing an orthography that is linguistically-based.
3. This linguistically-based orthography should be modified only as necessary on the basis of non-linguistic factors.

*This is a revised version of a paper presented at the Fifth Austronesian and Papuan Language and Linguistics Conference in 2012 in London and at the Annual Meeting of the Linguistic Society of Papua New Guinea in 2015 in Port Moresby. I am grateful for the comments I received at both conferences and for general comments from Mike Cahill and Debbie Clifton. I am also grateful to Robbie Petterson for comments on §4, and to Karl Franklin, Rob Head, Dennis Malone and Al Stuckey for comments on §5. Any remaining mistakes are, of course, mine alone.

While linguistic theory has changed considerably in the past 70 years, adherence to these principles is still strong: Most current practitioners, even those who claim to value at least some non-linguistic factors more highly than linguistic factors,¹ follow these principles. For example, in a discussion of orthography reform in Eastern Tucanoan languages, Seifart (2006:289) states, “This example thus shows the importance of modeling orthographies as closely to the linguistic structure as possible, rather than taking over conventions of orthographies of better-known languages.” Similarly, Munro (2014:173) proposes “two basic technical rules for good orthographies,” the result of which is a linguistically-based orthography exhibiting a one-to-one correspondence between phoneme and representation which can then be modified as necessary.

Sebba (2009:37) presents an alternative view in which orthography is not seen as a linguistically-based technology suitably modified according to non-linguistic factors. Instead,

Writing systems involve conventional relationships at multiple levels. This is why they can be seen as social practices: entire communities of users, subgroups and even individual users first have to establish these conventional relationships and then, later on, to decide whether to follow or deviate from the established conventions.

In this model, non-linguistic factors are the basic building blocks of orthography.

In the rest of this paper, I investigate implications of the claim that orthography is basically a non-linguistic social practice rather than a linguistic technology. In §§2-5, I present four case studies from Papua New Guinea in which non-linguistic factors are central to orthography development, especially in the area of identity with related dialects and with languages of wider communication. Then, in §6, I outline a methodology for orthography development based on the claim that orthography is basically a social practice. This methodology breaks the tie between linguistics and orthography, helping communities develop orthographies without reference to phonological and syntactic analyses.

2 Bine

Bine [bon]² is a member of the Eastern Trans-Fly family (Wurm 1975) spoken south of the Fly River in Western Province in southwest Papua New Guinea. The language can be divided along village lines into seven closely-related dialects which exhibit a number of very regular sound correspondences: Boje-Giringarede, Irupi-Dragele, Kunini, Masingle, Sebe, Sogal and Täti (Fleischmann & Turpeinen 1976). Under the influence of the London Missionary Society (LMS), some Bine people were already writing Bine when SIL linguists Lillian Fleischmann and Sinnika Turpeinen began language development work in the early 1970s. These indigenous attempts to write Bine were based primarily on the orthography developed by the LMS for the neighboring Southern Kiwai [kjd] language, the language that the LMS used as a language of wider communication in the south Fly River region (Foley 1986:31). While no attempt had been made to standardize the orthography, the generally accepted alphabet was as given in (1).³

¹For example, Smalley (1959) explicitly values the non-linguistic factor of community attitudes towards the orthography of wider communication (**maximum motivation**) more highly than the linguistic factor of **maximum representation of sound**. This ranking, however, is purely pragmatic. In an analysis of competing orthographic traditions in Srê [kpm], Smalley (1954:172) states that, “...phonemic considerations are basic to any orthography...” Smalley’s analysis of the competing systems begins with determining the linguistic adequacy of each system. Non-linguistic considerations are brought to bear only after phonemic considerations are evaluated.

²The three-letter codes are ISO-639-3 identifiers.

³This is taken from Fleischmann & Turpeinen (1977:135-136), which focuses on the alphabet used for the Kunini dialect, supplemented by the addition of <mb> and <nd>, which would have been used in the dialects that included prenasalized voiced stops (Fleischmann, personal communication).

(1) <a b d e g i j k l m mb n nd ng o p r s t u w>

A number of the symbols were used to represent more than one sound, and at least one sound (the glottal stop) was generally ignored in indigenous attempts to write Bine. Therefore, Fleischmann & Turpeinen (1977) (F&T) proposed a number of changes they felt would make the orthography fit the Bine phonological system more adequately. A major consideration was choosing symbols which would encourage cross-dialectal intelligibility while still allowing for dialectal differences to be reflected in Bine literature (L. Fleischmann, personal communication). Thus, while separate primers were produced for Boje-Giringarede, Irupi, Kunini and Masingle, symbols were chosen that captured similarities between related words, and a given symbol had a consistent pronunciation in all the dialects. The goal was that beginning literacy would be conducted in the dialect of the learner before transitioning into the Kunini dialect. In the rest of this section I discuss four of the issues addressed by F&T: the use of <a>, the representation of the glottal stop, the use of <ng>, and the use of <u i> versus <w y>. While I discussed a number of these issues in Clifton (2014), that discussion focused almost exclusively on the dialectal issues involved. In this discussion, I deal both with dialectal identity and the relationship of the Bine orthography to three established orthographies with which the Bine community was familiar: Southern Kiwai (the church language of wider communication), Hiri Motu (the government language of wider communication), and English (the language of education).

2.1 Use of <a> in Bine

The first issue concerns the use of <a>. Like Southern Kiwai, Bine was initially written with the five vowels <a e i o u>. In Bine, however, <a> represented both [a] and [æ] in five of the dialects as illustrated in the Boje-Giringarede forms in (4).⁴

- (2) a. <maru> [maru] 'beach'
b. <mamle> [mæmle] 'afternoon'

In the other two dialects (Masingle and Irupi-Dragele), <a> consistently represents only [a], since [æ] merges with [a] in these dialects as shown in the Masingle forms in (3).

- (3) a. <maru> [maru] 'beach'
b. <mamle> [mamle] 'afternoon'

F&T introduced the symbol <ä> to symbolize [æ]. As shown in (4), this captures the similarity between parallel forms in the two dialect groups.

- (4) a. <mämle> 'afternoon' (Boje-Giringarede)
b. <mamle> 'afternoon' (Masingle)

The Bine community readily accepted <ä>. The fact that it was not present in Southern Kiwai, Hiri Motu, or English, was not problematic since it was being used to represent a sound that did not exist in Southern Kiwai or Hiri Motu, and that did not have a unique representation in English. Furthermore, a comparison of (2b) and (4a) shows that the new symbol <ä> was minimally different from the symbol <a> that had been used to represent [æ].

⁴The examples in this section are from Fleischmann & Turpeinen (1977), Fleischmann (1984a), Fleischmann (1984b), and Saari (1992).

2.2 Representation of glottal stop in Bine

The second issue was the representation of the glottal stop [ʔ]. The glottal stop is written as the apostrophe <'> in Southern Kiwai, but is quite rare in Southern Kiwai and was generally omitted in attempts to write in Bine. F&T proposed representing the glottal stop with <c>. A major reason for using a letter that can be capitalized is that there are contrasts between the presence and absence of initial glottal in a number of Bine dialects as illustrated in the Kunini forms <ime> 'hand' versus <cime> 'our (ex.)'. Using the <'> results in the need to capitalize the initial vowel when a glottal-initial word was also sentence initial. The choice of the specific letter <c> to represent the glottal is due to dialectal variation between <c> and <k>: Kunini <c> corresponds with Irupi <k> as shown in (5a), while Kunini <k> corresponds with Masingle <c> as shown in (5b).⁵

- (5) a. <cane> 'I' (Kunini)
 <kane> 'I' (Irupi)
 b. <keke> 'nose' (Kunini)
 <cece> 'nose' (Masingle)

Since the language of education was (and continues to be) English, most Bine speakers were aware of the fact that <c> and <k> are both used to symbolize [k] in English. The overlap in use in English made it natural to use <c> and <k> to represent related forms in Bine. And the prestige of English contributed to the acceptance of this use of <c>.

2.3 Use of <ng> in Bine

Another issue with dialectal ramifications was the use of the digraph <ng> to represent the velar nasal [ŋ]. As shown in the forms in (6) from Kunini, <ng> contrasts with <n>.

- (6) a. <nano> [nano] 'boy'
 <ngango> [ŋaŋo] 'uncle'
 b. <enenige> [enenige] 'he is drinking'
 <engenige> [eŋenige] 'he is whistling'

The velar nasal does not occur in either Southern Kiwai or Hiri Motu. Thus, most Bine writers followed English and used <ng>. The use of <ng>, however, led to problems in dialects like Boje-Giringarede that have a contrast between voiced and prenasalized stops. In these dialects, <ng> was used to represent both [ŋ] and [ŋg] as seen in the Boje-Giringarede forms in (7a) as compared with the Kunini forms in (7b) with no prenasalization.

- (7) a. <ngango> [ŋaŋo] 'uncle' (Boje-Giringarede)
 <drengo> [dreŋgo] 'dog' (Boje-Giringarede)
 b. <ngango> [ŋaŋo] 'uncle' (Kunini)
 <drego> [drego] 'dog' (Kunini)

While it would have been possible to account for this difference by using <ngg> to represent the velar prenasalized stop in the dialects like Boje-Giringarede, F&T felt it would be too easy for writers to leave out the second <g> (Fleischmann, personal communication). F&T had independently proposed using <ŋ> instead of <ng> in an attempt to emphasize its phonological status as a single sound, and to avoid sequences of three consonant symbols in words like <aŋwäri> 'crawl!' instead of *<angwari>. From a

⁵For more details regarding cross-dialectal patterns of [ʔ] and [k] see Fleischmann & Turpeinen (1976, 1977) and Clifton (2014).

cross-dialectal standpoint, use of <ŋ> had the additional advantage of the simpler representation of the prenasalized velar as <ŋg>.

The symbol <ŋ> was used in literacy materials and most Scripture portions for over ten years. In 1992, however, the people of Kunini asked my wife and me to coordinate a meeting to discuss the orthography. The manuscript for the New Testament was being finalized, and there was concern over the orthography that would be used. The issue was of particular concern to Kunini, since the translation was in the Kunini dialect. The participants at this meeting expressed unease at the use of <ä> and <ŋ>. Their concern was that the orthography should only use symbols available on any computer, including computers in the regional government offices. Assurances that appropriate fonts could be installed on governmental computers were not sufficient; people wanted assurance that all computers would have both symbols in their default inventories. Since, at that point in time, such assurances could be given for <ä> but not for <ŋ>, the community rejected further use of <ŋ>. While they all agreed that consistency between dialects was good, and acknowledged that rejecting <ŋ> would prove problematic for dialects like Boje-Giringarede, computer capabilities were considered to be of greater concern.

2.4 Use of <u i> versus <w y> in Bine

The final issue concerns the use of the symbols <i u w> for syllable onsets. The symbol <i> was consistently used for the front onset, as shown in (8), but both <w> and <u> were used for the back onset, as shown in (9), although the preference was for <w> word-initially (as in (9a)) and syllable-initially following a non-round vowel (as in (9b)).

- (8) a. <iungi> ‘light the fire’
 b. <iaia> ‘mama’
 c. <nie> ‘water’
- (9) a. <wapo> ~ <uapo> ‘rear of boat’
 b. <ewe> ~ <eue> ‘no!’
 c. <tuwe> ~ <tue> ‘dust’

This difference between front and back onsets probably arose from the fact that while the front onset is represented by <i> both Southern Kiwai and Hiri Motu, the back onset is represented by <w> in Southern Kiwai, but by <u> in Hiri Motu. Given the general preference for <w> over <u>, however, F&T proposed using <y> for the front onset, arguing that the use of <w> and <y> were similar. The Bine community did, in fact, adopt the use of <y>, but not because of the linguistic reason presented by F&T. Instead, the Bine argued that the use of <y> instead of <i> marked them as inland people as opposed to the coastal Kiwai people (Fleischmann & Turpeinen 1977:138). This resulted in the spellings shown in (10) (also incorporating other changes outlined above as well as long vowels).

- (10) a. <yuuŋi> ‘light the fire’
 <yääye> ‘mama’
 <niiye> ‘water’
 b. <waapo> ‘rear of boat’
 <ceewe> ‘no!’
 <tuuwe> ‘dust’

This use of <y> is an instance of branding, a situation in which an orthographic element becomes “emblematic of a group of people who use that element in their writing practices” (Sebba 2015). While

the Bine could have continued to use <i> in syllable onsets, they chose to use <y> to distinguish themselves from the Kiwai. The use of this orthographic element, then, holds social significance.⁶

3 Vanimo

Vanimo [vam] is a member of the Sko family (Donohue 2002) spoken on the north coast in Sandaun (West Sepik) Province. In 1993, my wife and I were asked to work with a self-selected group of Vanimo speakers to develop an orthography for the Vanimo language. The Sko languages have a number of features which distinguish them from surrounding languages and from Tok Pisin. These include a high number of monosyllabic roots (especially among the verbs), widespread compounding, nasalization, and a complex tone system (Ross 1980). As a result, while literacy was high in Tok Pisin, the language of wider communication, few people had tried to write in Vanimo.

An important aspect of the workshop was the presence of speakers of both the Vanimo and Waromo dialects of the language. The format for the workshop was an orthography workshop in conjunction with a writers' workshop. We began by having participants write stories. During the rest of the workshop, we alternated between resolving orthography problems encountered while writing and reading their stories, and applying the orthography decisions while continuing to write. Throughout the process, the participants worked to ensure that all decisions would be applicable to both dialects (Clifton & Clifton 1995).

As indicated above, the participants were all literate in Tok Pisin. Therefore, this background formed the basis for their orthography decisions. While we never formally wrote out an initial orthography, the actual practices followed while first writing stories involved the alphabet shown in (11).

(11) <a b d e g h i j l m n n g o p s t u v w y>

In the following subsections, I discuss four areas that we dealt with explicitly: representation of vowels; representation of the glottal stop; the use of <j> and <y>, and <v> and <w>, and compounds. While I discussed the first two issues in Clifton (2014), the presentation here is more comprehensive. I conclude this section with a summary highlighting the significant decisions that were made at the workshop.

3.1 Representation of vowels in Vanimo

The greatest problem faced was in the representation of the vowels. For example, the orthographic sequence <da> could represent any of the four pronunciations shown in (12).⁷

- (12) a. <da> [da] 'pig' (short, oral, level tone)
 b. <da> [dâ] 'west' (short, oral, falling tone)
 c. <da> [da:] 'water' (long, oral, level tone)
 d. <da> [dẫ] 'hair' (short, nasal, falling tone)

In fact, any given vowel could have six possible pronunciations: short oral, long oral, falling oral, short nasal, long nasal, and falling nasal.⁸ After a discussion of underdifferentiation, the group quickly decided not to represent tone, but to represent length as double vowels and nasality with <ng> following the

⁶It is possible that <c> and <ä> are also used for branding in Bine, since neither are used in either Southern Kiwai or Hiri Motu. I have no evidence, however, that either the Bine or the neighboring Southern Kiwai view either of these symbols as emblematic of the Bine. Differences are not sufficient for an element to have undergone branding. Branding entails that the element is seen by either the community itself or some outside community as a unique marker of the community.

⁷All examples are from Clifton (1995). The circumflex (e.g. [â]) indicates falling tone, while unmarked vowels take level tone.

⁸Long vowels in Vanimo can only take level tone (Ross 1980).

vowel. With these distinctions available, the words in (12) would be written as in (13). The sequence <da> is limited to representing ‘pig’ ([da]) or ‘west’ ([dâ]), while <daa> is ‘water’ ([da:]) and <dang> is ‘hair’ ([dâ]).

- (13) a. <da> [da] ‘pig’ (short, oral, level tone)
 b. <da> [dâ] ‘west’ (short, oral, falling tone)
 c. <daa> [da:] ‘water’ (long, oral, level tone)
 d. <dang> [dâ] ‘hair’ (short, nasal, falling tone)

Given these conventions, <da> still has more than one pronunciation ([da] or [dâ]), but that is more manageable than the original six potential pronunciations. In addition, it was decided that tone could be indicated by a following apostrophe in children’s books and primers. This would allow for differentiation between <da> ‘pig’ and <da> ‘west’ for beginning readers.

Another source of ambiguity in the vowels arose from the fact that two of the vowels were each used to represent two distinct vowel qualities. While there was no universal consensus on what vowels were used in this way, the most common practice was to use <i> and <e> for two vowels each as shown in (14).

- (14) a. <di> [di] ‘make’
 <ti> [ti] ‘sugar cane’
 b. <de> [de] ‘they’
 <pe> [pæ] ‘betelnut’

When the participants were told that <i> (barred-i) was commonly used in other Sepik languages to represent the sound in <ti> ‘sugar cane’, they readily adopted its use. However, there was no quick consensus regarding how to write the sound in <pe> ‘betelnut’. A majority did not want to use <e> for both forms in (14b), and so participants considered alternatives including a diacritic (e.g. <é>) or the digraph <ai>. None of the suggestions garnered wide-spread support, however. The resistance to the digraph <ai> was especially strong. Even though this sequence is not otherwise used in the Vanimo language, it is used in Tok Pisin for the diphthong [ai], and participants felt it would be confusing to have the sequence <ai> be pronounced differently in Vanimo than it is in Tok Pisin. After using <e> for both sounds for most of the workshop, the participants ultimately decided to use the digraph <ae> for the seventh vowel, with <ae> as its lengthened counterpart. This choice did not create transfer problems to or from Tok Pisin. Furthermore, it helped to maintain the visual similarity between related dialectal forms as shown in (15).

- (15) a. <pee> [pe:] ‘morning’ (Vanimo)
 b. <pae> [pæ:] ‘morning’ (Waromo)

3.2 Representation of glottal stop in Vanimo

Another problem was whether to write the glottal stop. It only occurs in the Waromo dialect, and was generally omitted in writing as shown in the Waromo forms in (16).

- (16) a. <ae> [ʔæ] ‘moon’
 b. <saae> [saʔæ] ‘smoke’

Omitting the glottal stop was especially problematic in forms like (16b) where this results in sequences of three vowels. While the apostrophe <'> is used in other Sepik languages to represent the glottal stop, this suggestion was rejected by the participants. Instead, they noted that the glottal stop in Waromo generally corresponded to either <g> or <h> in Vanimo. Based on their knowledge of both dialects, they decided to

write the glottal stop as <gh>. As shown in the examples in (17), this decision once again helped to maintain visual similarity between related dialectal forms.

- (17) a. <ghae> [ʔæ̃] ‘moon’ (Waromo)
 <gae> [gæ̃] ‘moon’ (Vanimo)
 b. <saghae> [saʔæ̃] ‘smoke’ (Waromo)
 <sagae> [sagæ̃] ‘smoke’ (Vanimo)
 c. <ghyaa> [ʔja:] ‘sea’ (Waromo)
 <hyaa> [hja:] ‘sea’ (Vanimo)

3.3 Use of <j> and <y>; <v> and <w> in Vanimo

As participants began to write, <j> and <y> were used interchangeably, as were <v> and <w>. After making lists of words for which there was consensus regarding the spelling, it was decided that <j> and <v> only occur before front vowels, while <y> and <w> only occur before back vowels, as illustrated in (18).

- (18) a. <ji> [dʒi] ‘sago’
 <yudi> [judi] ‘uncle’
 b. <vite> [vite] ‘reef’
 <wongbaba> [wõbaba] ‘sweet potato’

Thus, while the participants decided to underdifferentiate by not indicating tone, they also decided to overdifferentiate by writing both <j> and <y>, and <v> and <w>. All four symbols are used in Tok Pisin with the same phonetic values, although their phonological status is different in the two languages.

3.4 Representation of compounds in Vanimo

The final issue is tied to the fact that compounds are very common in Vanimo, and that the phonetic shape of morphemes varies depending on the environment. The question participants wrestled with is whether to write compounds to reflect the pronunciation of the components in isolation or within the compound. Examples involving long vowels and nasal vowels are given in (19a) and (19b), respectively.

- (19) a. <pii> [pi:] ‘breadfruit’
 <lu> [lu] ‘under’
 <piilu> ~ <pilu> [pilu] ‘under the breadfruit’
 b. <eng> [ẽ] ‘coconut’
 <lu> [lu] ‘under’
 <englu> ~ <englung> [ẽlũ] ‘under the coconut’

In (19a) ‘breadfruit’ has a long vowel in isolation, but a short vowel in the compound, while in (19b) the nasality of <eng> ‘coconut’ spreads to the vowel in <lu> ‘under’ in the compound. The question is whether to write each morpheme as it would be pronounced in isolation (<piilu> and <englu>) or as it is pronounced in the compound (<pilu> and <englung>). The participants felt that it would be easiest to read if the spelling reflected the pronunciation in isolation so that each morpheme had a consistent shape.

The decision was more difficult for nasal vowels followed by stops. As indicated above, <ng> following a vowel indicates the vowel is nasal. When a morpheme ending in a nasal vowel is followed by a morpheme beginning with a stop, the stop becomes prenasalized. As shown in (20), the question is whether to write the nasality on the vowel, for example <tangdongbi>, or as the homorganic prenasalization, for example <tandombi>.

(20) a.	<nyeng>	[nê]	‘octopus’
	<pang>	[pã]	‘banana’
	<nyengpang> ~ <nyempang>	[nêmpã]	‘snake’
b.	<tangdong> ~ <tandong>	[tândõ]	‘yesterday’ ⁹
	<tangdongbi> ~ <tandombi>	[tândõmbi]	‘day before yesterday’ ¹⁰

The participants initially decided to write the words as they sound: as <nyempang> and <tandombi>. These spellings follow the Tok Pisin conventions seen in words like <tambu> ‘taboo’, <bembe> ‘butterfly’, <kundu> ‘drum’ and <sanda> ‘perfume’. This decision was revisited, however, after the decision was made to spell <piilu> ‘under the breadfruit’ and <englu> ‘under the coconut’ to reflect how the individual morphemes were pronounced in isolation. At this point it was decided to write the compounds in (20) as <nyengpang> and <tangdongbi>, even though these forms looked less like parallel forms in Tok Pisin.

The decision to maintain the isolation spellings even in compounds was strengthened by a later decision regarding word breaks for compounds. The participants could not decide whether to write compounds as one word or two. While they recognized the component parts, they also realized that the meanings of many compounds were only loosely related to the meanings of the component parts (for example, ‘snake’ from ‘octopus’ and ‘banana’ in (20a)). The ultimate decision was to hyphenate compounds, resulting in the final spellings <pii-lu> ‘under the breadfruit’, <eng-lu> ‘under the coconut’, <nyeng-pang> ‘snake’ and <tangdong-bi> ‘day before yesterday’. The presence of the hyphen between the <ng> and the stop means the Tok Pisin convention that nasals agree in point with the following stop is less relevant.

3.5 Vanimo summary

Participants at the Vanimo orthography workshop made decisions based on a number of considerations. A paramount consideration was to maintain visual similarity between the Vanimo and Waromo dialects. This influenced the decision to use <ae> for [æ], and the decision to write the glottal as <gh> instead of the more common regionally-used <'>. In the absence of dialectal considerations, on the other hand, <i> was adopted for [i] because of its use in the region. The influence of Tok Pisin was felt in the decision not to use <ai> for [æ], and in the decision to write both <j> and <y>, and <v> and <w>, all of which exist in Tok Pisin. It is possible that the decision not to indicate tone except possibly for children was also due to the influence of Tok Pisin: The widespread use of the apostrophe for tone made the page look quite different from Tok Pisin.

The use of <ng> to mark nasal vowels is more complex, since it violates the Tok Pisin convention of using <ng> for the velar nasal [ŋ]. Participants indicated the use of <ng> to mark nasal vowels dated back to the work that Malcolm Ross did with two Waromo women at Goroka Teacher’s College from 1976-1978 (Clifton 1995), reported in Ross (1980). The subsequent decisions to write prenasalization indirectly as nasality on the preceding vowel, and to separate the mark of nasality from the stop by a hyphen in compounds puts even more distance between Tok Pisin and Vanimo.

Sebba (2007:112-118) argues that orthography plays a dual role in establishing group identity. On the one hand, similarities between the new orthography and the orthography upon which it is based link the identities of the two groups. On the other hand, differences between the two systems allow the group developing the new orthography to establish a unique identity. In the case of Vanimo, these two functions apply at both the language and the dialect levels. At the language level, the use of both <j> and <y>, and <v> and <w> ties Vanimo to Tok Pisin, while the use of the hyphen, <ae>, double vowels, and <ng> to mark nasal vowels (as well as prenasalization), mark Vanimo as unique. At the dialect level, the use of <gh> to indicate the glottal stop in Waromo simultaneously marks it as related to the Vanimo dialect

⁹There is no indication that this is a compound even though it is polysyllabic.

¹⁰The meaning of <bi> is unknown.

(which uses either <g> or <h> in related forms), but also as unique since the sequence never occurs in the Vanimo dialect.

4 Kope

Kope is one of five dialects that comprise Northeastern Kiwai [kiw] (Clifton 1987), a member of the Kiwai language family spoken in the Gulf Province (Wurm 1973, Foley 1986:233). Literacy in the languages of the Kiwai family began under the London Missionary Society (LMS), starting with Southern Kiwai [kjd], the southwesternmost member of the family. As much as possible, the LMS followed a uniform set of orthographic conventions in all the Kiwai languages. Their work in the closely-related language Kerewo [kxz] began in the 1920s, and Kerewo New Testament portions and a Kerewo hymnbook (Gido Buka 1969) were published by the LMS (Butcher 1963). While the only published material in Kope was a few songs included in the Kerewo hymnbook, Samoan pastors working with the LMS taught basic literacy skills in Kope schools. Because of this training, quite a few adults were literate in both Kope and Hiri Motu, the language of wider communication, when my wife and I began working with the Kope people on language development projects in the early 1980s. In addition, many people had basic literacy skills in English, since it was the language of education, and a large majority of the community had finished at least grade 6. The presence of Kope literacy was reflected in the fact it was not uncommon to see signs at the aid post, church and school in Kope. While there was a good deal of inconsistency in how the language was being written, it was generally agreed that the basic inventory of letters was as given in (21).

(21) <a b d e g h i k m n o p r s t u v ' >

Since, however, there were considerable inconsistencies in actual practice, we organized a number of writers' workshops as part of an ongoing discussion of orthography issues in an attempt to regularize the spelling system. The process and conclusions of these activities are presented in Clifton (1987). However, the emphasis of that presentation is to highlight the relationship between phonology and orthography. In this presentation, I examine the orthographic decisions on their own, compare the practices with the small amount of published material that predated our work, and present more recent changes to the orthography.

I examine four areas of orthography in this section: the representation of glottal stop and long vowels, the use of <n> and <r>, the use of <m> and <v>, and the use of vowel sequences. My overall claim is that while speakers continue to follow the conventions of Hiri Motu, they do not feel similarly constrained by the practices followed in Kerewo. In fact, many of the recent developments have served to differentiate Kope from Kerewo and from the other dialects that, with Kope, make up Northeastern Kiwai.

4.1 Representation of glottal stop and long vowels in Kope

In both Gido Buka (1969) and general community practice, each of the vowels represented two distinct sounds: a short vowel and a lengthened vowel. Examples are given in (22).¹¹

- (22) a. <abea> /ábèà/ 'father'
 <abea> /àbè:à/ 'sago squeezing bag'
 b. <nimo> /nímò/ 'lice'
 <nimo> /nì:mó/ 'we'
 c. <obo> /òbó/ 'water'
 <obo> /ó:bó/ 'woman'
 d. <tutu> /tútú/ 'handle'

¹¹Unless otherwise specified, all Kope examples are from Clifton (1991) or personal field notes. As indicated, each of the pairs in (22) are differentiated by tone in addition to vowel length, but tone has never been indicated in the orthography.

<tutu> /tùtù:/ ‘long’

While my initial concern over this underdifferentiation was based on my phonological analysis, discussion with community members revealed feelings more of resignation than of acceptance. While everyone recognized that different words were spelled the same in the system used at the time, they did not think there was anything they could do about it.

The most commonly used convention for long vowels in Papua New Guinea, doubled vowels, was intertwined with the representation of glottal stop in Kope. In Gido Buka (1969), the apostrophe <’> was consistently used to represent the glottal stop between both identical vowels, as in (23a), and dissimilar vowels, as in (23b).

- (23) a. <do’ou> ‘today’
 <oroi’io> ‘life’
 b. <o’u> ‘come’
 <a’o> ‘say’

By the early 1980s, however, the apostrophe was only used in words like those in (23b) in which the glottal occurred between dissimilar vowels. This allowed the system to differentiate between pairs of words that differed in the absence or presence of glottal stop as in <oruo> ‘climb down’ vs. <oru’o> ‘wash (oneself)’. The presence of the glottal stop between identical vowels was shown, on the other hand, by doubling the vowel so that the forms in (23a) were written as <doou> and <oroiio>. So if (22c) and (22d) were written as <oobo> and <tutuu>, respectively, they would be read as /oʔobo/ and /tutuʔu/.

A number of alternatives were tried and rejected including a colon (e.g. <o:bo> and <tutu:>), hyphens (e.g. <o-bo> and <tutu->), and macrons (e.g. <ōbo> and <tutū>) (Clifton 1987). After testing showed that ignoring the contrast between long and short vowels hindered reading, the decision was made to write long vowels as double vowels, and use <’> between identical as well as dissimilar vowels. More recently, Petterson (2014) reports that the use of both <’> and double vowels continues to be problematic. The apostrophe is sometimes used in long vowels (e.g. <pe’e> /pɛ:/ ‘canoe’) where it does not belong, while it is sometimes omitted even between dissimilar vowels (e.g. <ou> /oʔu/ ‘come’). This confusion affects the use of double vowels to represent long vowels. Readers who learn to use double vowels to indicate length like the convention, but readers who do not understand that double vowels are being used to indicate length sometimes actually cross one of them out to ensure they are not read as two vowels separated by a glottal stop. Furthermore, it is frequently difficult for speakers to identify which vowels are long, which leads to confusion in spelling even for those writers who do understand the use of double vowels for long vowels (R. Petterson, personal communication).

It appears that both the use of <’> to represent all occurrences of glottal stop and the use of double vowels for long vowels are problematic. This is not surprising, given that neither convention is used in either Hiri Motu (the language of wider communication) or in English (the language of education). It is of particular interest to me, however, that a significant number of speakers do not feel constrained to follow the convention used in Gido Buka (1969) of using <’> to mark all and only occurrences of the glottal stop.

4.2 Use of <n> and <r> in Kope

A second problem area is the use of <n> and <r>. In the Kope songs in Gido Buka (1969), the two are in near complementary distribution: <n> is used word-initially as shown in (24a) and <r> is used intervocalically as shown in (24b),¹² with the exception of a few forms in which <r> occurs word initially as shown in (24c).

¹²There are no phonetic consonant sequences in Kope, either within syllables or across syllable boundaries.

- (24) a. <nimo> ‘we’
 b. <mere> ‘person’
 c. <rautu> ‘with’

There was only one other word that began with <r> in Gido Buka (1969): <ro> ‘you (sg)’. One other word that we found native speakers frequently spelled with word-initial <r> was <rio> ‘desire’, generally used in possessive forms like <Mo rioka du> ‘My desire is sago/I want sago’. In Clifton (1987) I suggest that <rautu> and <rio> are pronounced with word-initial [r] because they occur almost exclusively in unstressed position within the phrase. But I also point out this will not account for word-initial <r> in the pronoun <ro> ‘you (sg)’ since it frequently occurs phrase-initially. In addition, there are four pronouns that begin with word-initial <n>: <nu> ‘he/she/it’, <nimo> ‘we’, <ni'o> ‘you (pl)’, <ni> ‘they’.¹³

There are also two affixes that are problematic with regard to <n> and <r>. First, the prefix <r-> ~ <n-> is used on verbs to mark first-person involvement. The only occurrence of this prefix in Gido Buka (1969) is in <na'ai>, the future tense clitic. It is spelled with an <n> because it is word-initial, even though it is normally pronounced as [r], for example, in the verb <emapui na'ai> ‘will lead me’. In actual practice, we found native speakers were divided as to whether to spell this prefix as <n> or <r>. The situation is complicated by the fact that the prefix can also occur after the far past prefix <pi->¹⁴ as in <pirodau> ‘I/we went’. Speakers consistently wrote the first-person prefix as <r> after the far past prefix. It can also occur after the certaintive marker <ai>. This marker is sometimes written as a prefix as in <airodau> ‘I/we definitely go’, in which case it was consistently written as <r>. But when <ai> was written as a separate word, there was again inconsistency in the representation of the first-person prefix: both <ai nodau> and <ai rodau> were found.

The other problematic affix was the ergative suffix <-ro> ~ <-no>. In Gido Buka (1969), this suffix was consistently joined to the previous word, and spelled <-ro> as in <nuro> ‘he (ERG)’ and <Iesuro> ‘Jesus (ERG)’. In actual practice, however, word breaks were not so consistent. While the ergative marker was almost always attached to pronouns (especially the one-syllable pronouns), it was less consistently attached to common nouns or adjectives, and most frequently written as a separate word after proper nouns. The consequences for orthography was that the ergative marker was consistently written as <-ro> when attached to the preceding word, but as either <ro> or <no> when written as a separate word.

At an orthography conference held in the mid-1980s it was decided to write both <r> and <n>, with <r> being used word-medially and in <rautu> ‘with’, <rio> ‘desire’, <ro> ‘you (sg)’, and the first-person prefix <r->; and <n> being used word-initially in all other words. It was further decided that the ergative marker would be consistently joined to the preceding word with a hyphen and spelled <-ro>, resulting in <Iesu-ro> ‘Jesus (ERG)’, <bomo-ro> ‘pig (ERG)’ and <ni-ro> ‘they (ERG)’. This decision was based on the fact that there is a phonetic contrast between word-initial [n] and [r] in the pronouns, and there is an orthographic contrast between <n> and <r> in Kerewo,¹⁵ the language of most of the songs in Gido Buka (1969), Hiri Motu and English.

The use of <n> and <r> continues to be problematic, however. Petterson (2014) reports that in 2005 most writers were consistently using <r> intervocally, but were inconsistent in word-initial position. They would sometimes use <r> in words like <raa> ‘fish’, even though it is always pronounced as [na:]. At the same time, they would sometimes use <n> in <nautu> ‘with’ even though it is pronounced as [rautu] and in the first-person involvement prefix in forms like <nodau> ‘I/we go’. In primers produced in

¹³The pronunciation of <ro> fluctuates between [ro] and [no], but this is still different from the pronunciation of the other pronouns cited, which consistently begin with [n].

¹⁴The far past prefix is realized as <pi-> before a consonant and as <p-> before a vowel, as in <podau> ‘went’.

¹⁵The distinction between [n] and [r] is phonemic in most Kiwaian languages (Wurm 1973), as seen in the difference between Kerewo [one] ‘sago grub’ vs. [oro] ‘thorn’(Kerewo examples from Petterson (2010)), neutralized in Kope as [ore] and [oro].

2007, <n> is consistently used in word-initial position except in a few suffixes that are sometimes written as separate words. A few years later, a reviewer/editor suggested using <r> in both medial and initial positions; this suggestion was implemented in post-primer materials printed in 2014. It is fair to say that there is continuing tension concerning the use of <n> and <r> (R. Petterson, personal communication).

Petterson (2014) suggests that the decision to write <rautu>, <rio>, and <ro> with word-initial <r> was due to influence from Kerewo where there is extensive contrast between <n> and <r>. This explanation is especially plausible in the case of <ro> ‘you (sg)’, since it is quite common in Gido Buka (1969). Whether or not this explanation is correct, however, the more recent decisions to either write these three words with word-initial <n> or to write <r> in all environments distinguishes Kope from Kerewo.

4.3 Use of <m> and <v> in Kope

A third problem area is the use of <m> and <v>. In Gido Buka (1969), <m> is always used in word-initial position and is almost always used word-medially. Only two words, <hivio> ‘sun, day’ and <orovidio> ‘to hear’, are written with <v> in Gido Buka (1969); the <v> never occurs in initial position. The <v> is also used in <ove> ‘moon’ giving rise to the orthographic minimal pair <ove> ‘moon’ versus <ome> ‘shark’. In careful speech, speakers would carefully differentiate between [β] and [m]. In normal speech, however, the two would merge into [β], so that both ‘moon’ and ‘shark’ are generally pronounced [oβɛ].¹⁶ There is also the near minimal pair <hivio> ‘sun’ versus <himia> ‘self’.

In the mid-1980s orthography conference, it was decided to write <v> in <hivio> ‘sun, day’, <ove> ‘moon’ and <orovidio> ‘to hear’, and <m> in all other words, both initially and medially. Since then, Petterson (2014) reports that the use of <m> is now limited to personal names; the three formerly exceptional words are spelled <himio> ‘sun, day’, <ome> ‘moon’ and <oromidio> ‘to hear’. He also suggests that the previous decision to write <v> in a handful of words in Kope might have been due to the fact that <m> contrasts extensively with <v> (before front vowels) and <w> (before back vowels) in the Kerewo songs in Gido Buka (1969). The words <hivio> ‘day’ and <orovidio> ‘hear’ are especially common in the Kerewo songbook. If this suggestion is correct, even the slow speech pronunciation of [β] in these words could be due to the influence of spelling. The de facto decision to use only <v> is, once again, an indication that the influence of Kerewo is no longer a driving force in Kope orthography decisions.

4.4 Use of vowel sequences in Kope

The final problem area is the use of sequences of vowel graphemes. In the Kiwaian orthographies developed by the London Missionary Society, a sequence of vowels can represent a sequence of syllabic vowels as in (25a), an offglide as in (25b), or an onglide as in (25c). Longer sequences can represent more complex syllabic structures, like the four vowels used to represent three syllables in (25d).

- | | | | | |
|------|----|----------------------|----------------|----------|
| (25) | a. | <tur ia ha> | [tu.ri.a.ha] | ‘all’ |
| | b. | <ra u tu> | [raʉ.tu] | ‘with’ |
| | c. | <ia ma gauri> | [ja.ma.gaʉ.ri] | ‘jump’ |
| | d. | <tu ia i> | [tu.i.aɨ] | ‘middle’ |

While the four-vowel sequence in <tuia*i*> is the longest found in Gido Buka (1969), I have found sequences of up to nine vowels in Kope: <oaeaoaeai> ‘wild okari tree’. My concern was that these long sequences of vowels would be difficult for readers to parse. In addition, there was at least one pair that seemed to differ in syllabification: <ue> could be pronounced either as [ue] ‘flat spear’ or [u.e] ‘sugar

¹⁶There is a tonal difference between these two words, but as noted in footnote 11, tone is not marked in the orthography. Therefore, if these two words were written as pronounced in normal speech, there would be no orthographic difference between the two.

cane'. Because of these perceived problems, I suggested using <y> and <w> to help readers identify syllables. This would result in <turiyaha> rather than <turiaha>, <yamagauri> rather than <iamagauri>, and <tuwiyai> rather than <tuiai>.

This suggestion was universally rejected. In an earlier analysis of Kope orthography (Clifton 1987), I propose a phonological reason for this rejection. Noting that readers do not seem to have difficulties reading strings of vowels, I show that syllabification of vowel strings is almost always predictable. Therefore, there is no pressure to adopt orthography conventions that make it easier to parse strings of vowels. I now feel, however, that the opposition to the use of <y> and <w> is more likely due to positive and negative pressures from neighboring orthographic systems.

The primary reason actually given by speakers for rejecting the use of <y> and <w> was that this made the language look like Kerewo and the closely related Gibaio dialect (Clifton 1987). Both of these speech communities have maintained a contrast between <m> and <w> before back vowels while, as noted in §4.3, only <m> is used in Kope. So <made> 'word, language' in Kope is <wade> in Kerewo and Gibaio. Use of <w> is perceived to make Kope look like Kerewo and Gibaio even though different words would be affected. This represents negative pressure against the use of <w> (and, by extension, <y>).

The positive pressure comes from the fact that the use of vowel sequences in Kope mirrors their use in Hiri Motu: Vowel sequences are used in Hiri Motu to represent a sequence of syllabic vowels as in (26a), an offglide as in (26b), or an onglide as in (26c).

- (26) a. <heatu> 'brawl'
 b. <hairaina> 'beauty'
 c. <uili> 'wheel'

Furthermore, longer sequences of vowels are also allowed in Hiri Motu as in <ataiai> 'above' and <haheaua> 'put into motion'. The use of <w> is restricted to the digraphs <kw> and <gw>; <y> is never used in Hiri Motu.

It is instructive to compare the use of <y w> in Kope and Bine (§2.4). The phonological analysis proposed in Clifton (1987) fits Bine with minor modifications. That is, it is possible to predict phonetic syllabification in both languages without positing either /w/ or /j/ in the phonological inventory. In both communities, there was a desire to differentiate their orthographic systems from Kiwaian systems that used <w>. But the solution was different in the two. Bine speakers decided to keep the <w> and differentiate their orthography from that of Southern Kiwai by adding <y>, which was not used in Southern Kiwai. Kope speakers, on the other hand, decided not to use either <w> or <y>. While the underlying motivation between the decisions was the same, a desire to differentiate themselves from a language with <w> but no <y>, the actual decisions were different.

5 Umbu-Ungu and Imbongu

Umbu-Ungu [ubu] and Imbongu [imo] are members of the Chimbu family, which is likely a part of the Trans-New Guinea family of languages (Foley 2000).¹⁷ Umbu-Ungu is spoken primarily in the Western Highlands province, while Imbongu is spoken in the Southern Highlands province. They are either closely related languages (as indicated by separate ISO-639-3 identifiers) or dialects of a single language (generally referred to as Kaugel).

Prior to 1983, two orthographic systems were in use in Umbu-Ungu: one developed by Dieter Klemm and used by the Lutheran church, and one developed by Bruce and Ruth Blowers (Blowers & Blowers

¹⁷The Chimbu family is called the Chimbu-Wahgi family by Lewis et al. (2015) and the Central East New Guinea Highlands family by Capell (1962) (as referenced by Hammarström et al. (2015)).

1970)¹⁸ and used by the Evangelical Bible Mission, the Catholic church and SIL International (R. Head, personal communication). I refer to these as the K orthography and the B orthography. The two orthographies both drew from a common set of letters given in (27).

(27) <a b d dl e g gl i j k l ll m mb nd ng ni nj ŋ o p r s t w y>

In 1983, an orthography conference was held for Umbu-Ungu speakers to agree on a single orthography to be used by all Umbu-Ungu communities (Malone & Malone 1983). This was followed in 1985 by a similar conference for Imbongu speakers (Malone et al. 1985), in which I participated. I examine five topics that were discussed at these meetings: the use of <mb>, <nj>, <nd>, and <ng>; the representation of the laterals; the use of <s>; the use of <ŋ>; and the use of <y> and <w>. As I show, the end result of the two conferences was to bring the Umbu-Ungu and Imbongu orthographies closer to each other, and, for the most part, closer to English, the language of education, and to Tok Pisin, the language of wider communication.

5.1 Use of <mb>, <nj>, <nd>, <ng> in Umbu-Ungu and Imbongu

One major difference between the two orthographies was the use of the digraphs <mb>, <nj>, <nd>, and <ng>. (The diagraph <nj> represents the dental affricate [ndʒ].) Klemm uses , <j>, <d>, and <g> in word-initial position, and <mb>, <nj>, <nd>, and <ng> in medial position, although both sets of consonants represent prenasalized stops. While Blowers & Blowers (1970) agree there is a phonetic difference between the two sets of stops,¹⁹ they claim the difference is not phonemic and write all prenasalized stops as , <j>, <d>, and <g> regardless of position. These differences are shown in (28).²⁰

(28)	Klemm	Blowers	
	<bana>	<bana>	/mba.na/ ‘sweet potato variety’
	<kumbe>	<kube>	/ku.mbe/ ‘large grasshopper’
	<dumu>	<dumu>	/ndu.mu/ ‘shell decoration’
	<mundu>	<mudu>	/mu.ndu/ ‘mound’

While it is possible that Klemm was reacting to the phonetic difference in prenasalization between initial and medial positions, it is also possible that as a non-linguistically trained speaker of German, Klemm is reflecting German (and English) orthographic conventions: Both German and English allow the sequences <mb>, <nd> and <ng> in medial (cf. German <Amboß> ‘anvil’, <Stunde> ‘hour’, <Singular> ‘singular’) but not initial positions.

The participants at the Umbu-Ungu conference in 1983 decided not to write prenasalization in either position, following the practice of the B orthography. The participants at the Imbongu conference in 1985, however, decided to write intervocalic prenasalization, following the practice of the K orthography. The ultimate decision has been to write intervocalic prenasalization in both Umbu-Ungu and Imbongu. This standardizes the orthographic practice for prenasalized stops for both communities. In addition, it aligns itself with the practice of English and Tok Pisin, even though the phonemic status and syllabification of /mb/ and /nd/ is different in Umbu-Ungu and Imbongu from that in English or Tok Pisin. Head (2011:3)

¹⁸While Blowers & Blowers (1970) is a traditional phonemic statement with no recommended orthography, the symbols chosen to represent each phoneme were used as the de facto practical orthography.

¹⁹Blowers & Blowers (1970) claim that prenasalized stops are lenis in word-initial position.

²⁰Umbu-Ungu examples in this section are from Malone & Malone (1983) and Umbu-Ungu OPD (2011); Imbongu examples are from Malone et al. (1985) and Imbongu OPD (2011).

notes that while writing intervocalic prenasalization is more difficult for illiterates, it is preferred by those educated in English or Tok Pisin.

5.2 Representation of laterals in Umbu-Ungu and Imbongu

A second issue was how to write the laterals. The basic distinction in the B orthography was between <ll> and <l>, while the basic distinction in the K orthography was between <l>, <dl> and <gl>. As shown in the Umbu-Ungu forms in (29), <ll> in the B orthography corresponds with <l> in the K orthography as in (29a), while <dl> and <gl> in the K orthography correspond with <l> in the B orthography as in (29b-c). Both systems use <l> before a consonant as in (29d).

(29)	Blowers	Klemm		
a.	<llapo>	<lapo>	[lapo]	‘lizard sp.’
	<wallo>	<walo>	[walo]	‘small’
b.	<lemenu>	<dlemenu>	[dʒemenu]	‘daughter’
	<akilio>	<akidlio>	[akidʒio]	‘last born’
c.	<laporale>	<glaporadle>	[glaporadʒe]	‘wealth indicator’
	<olo>	<oglo>	[oglo]	‘stomach’
d.	<ltebo>	<ltembo>	[ldembo]	‘I will throw away’
	<ulke>	<ulke>	[ulge]	‘house’

In general, the K orthography used symbols which reflected the orthographic conventions of English and Tok Pisin more closely than did the B orthography. K <l> in (29a) is used for the flap /l/ which is reasonably similar to <l> in English and Tok Pisin,²¹ while B <ll> does not correlate with anything in either English or Tok Pisin. Similarly, K <dl> and <gl> in (29b-c) reflect the pronunciation of the lateral affricates [dʒ] and [gʟ] much more closely than does B <l>.

The B orthography is based on the observation that there is no contrast between <dl> and <gl>: <dl> occurs before <i> and <e>, while <gl> occurs before all other vowels. Therefore, both sounds can be represented by the same letter. Given this fact, it would be possible to use <l> for the flap and some other symbol (for example, <ll>, <dl> or <ɫ>) for the affricate. The flap, however, is quite rare while the affricates are quite common. Therefore, Blowers & Blowers (1970) decided to use <l> for the more common affricate and <ll> for the rarer flap even though the pronunciations did not match as well with the English and Tok Pisin. The fact that the K orthography aligns more closely with the English and Tok Pisin systems than does the B orthography can be seen in the spelling of loans like Tok Pisin <balus> ‘airplane’. This would be written as <balusi> in the K system, but as <ballusi> in the B system (Head 2011:3).

In spite of the fact that the representation of laterals is closer to English and Tok Pisin in the K orthography than in the B orthography, the decision at both conferences was to use <l> for the affricates and <ll> for the less common flap /l/. This decision was based on maximizing the use of the simpler grapheme rather than using a system that was closer to English or Tok Pisin.

²¹The flap /l/ was sometimes written as <r> in the K orthography, probably because it also represents a flap, and the <l> does not represent a flap in either English or German. But there is a contrast between <l> and <r> in Umbu-Ungu and Imbongu, so the <r> should not be used in either the B or K orthography for the lateral flap.

5.3 Use of <s> in Umbu-Ungu and Imbongu

A third issue was the use of <s> to represent the voiceless dental phone [t̪]. Blowers & Blowers (1970) observe that [t̪] and [t̪̚] apparently contrast in some speech varieties²² but not in others. While Klemm consistently used both <t> and <s> to represent this phonetic difference in all dialects, Blowers & Blowers suggest that both <t> and <s> should be used only in varieties which show contrasts, and that only <t> should be used in varieties where there is no contrast. In Umbu-Ungu the two do not contrast: <s> only occurs contiguous to a high vowel, while <t> occurs elsewhere (Head 2011:3). However, both <s> and <t> were used in early SIL materials in Umbu-Ungu as shown in (30) (Head 2011:3).

- (30) <tawe> [tawe] ‘laugh’
 <simu> [simu] ‘he gave’

Head notes that the use of both helped readers educated in English or Tok Pisin. On the other hand, since the distribution of <s> was not uniform between varieties, using <s> meant the same word had different spellings in different varieties which, in turn, caused pressure to produce materials in more than one variety. Because of this, the use of <s> was abandoned in SIL materials beginning in mid-1976. Writing only <t> allowed greater homogeneity across the varieties (Head 2011:3). Both <t> and <s> continued to be differentiated in the K orthography, however. At the Umbu-Ungu conference it was decided to write both <t> and <s>, following the practice of the K orthography. While this results in differences between varieties, it aligns the conventions with those of English and Tok Pisin.

The <s> does not occur in any native words in Imbongu, but it does occur in some loanwords. These loanwords were sometimes written with <t> and sometimes with <s>, as shown in (31).

- (31) <ballusi> ~ <balluti> ‘airplane’ (from Tok Pisin <balus>)
 <kusa> ~ <kuta> ‘salt’ (from Enga <kosa> (Draper & Draper 2002))²³

At the Imbongu conference it was decided that both <t> and <s> should be used in loanwords, with the spelling reflecting the more common pronunciation. For example, the loanword ‘airplane’ would be written <ballusi>, since the word is most generally pronounced with the [s] from Tok Pisin. This creates a contrast between <ti> in <tipe> ‘fire’ and <si> in <ballusi> ‘airplane’. On the other hand, the loanword ‘salt’ would be written <kota>, since the Enga [s] is normally changed to the [t] found in Imbongu.

5.4 Use of <ŋ> in Umbu-Ungu and Imbongu

The fourth issue was whether to use <ŋ> in some subdialects of Umbu-Ungu (as well as in the neighboring variety Bo-Ung) in which the velar nasal corresponds to the prenasalized velar stop in the standard dialect, as shown in (32).

- (32) <ŋ> variety <ng> variety
 <nuŋa> <nunga> ‘your’
 <aŋenu> <angenu> ‘sibling, same sex’
 <laŋi> <langi> ‘food’
 <kaŋo> <kango> ‘boy’

²²I am intentionally using the term ‘variety’ instead of ‘dialect’ since the differences can occur at language, dialect and subdialect levels.

²³Even though <kosa> refers to commercial salt, the traditional trade in native (locally-produced) salt came from Enga (K. Franklin, R. Head, personal communication).

The use of <ŋ> to represent the velar nasal was common in many of the orthographies developed by German missionaries, including those in some of the neighboring languages. In the Umbu-Ungu communities, it contrasted with the prenasalized velar stop <ng>. The use of <ŋ> made it simple to differentiate between the nasal and prenasalized stop.

The dialectal difference illustrated in (32) was considered to be an important distinction: Umbu-Ungu speakers divide themselves into those who say [nuŋga]/[naŋga] ‘your’/‘my’ and those who say [nuŋa]/[naŋa] (D. Malone, personal communication). In Umbu-Ungu, then, the <ŋ> marked a division that was recognized as significant by the local community.

As noted in §5.1, the participants at the Umbu-Ungu conference decided that prenasalized stops should be written as , <j>, <d>, <g> in all positions. At the same time, they agreed that the <ŋ> would be retained in the alphabet for the use of the dialects which differentiated between the velar nasal and the prenasalized velar stop. At the Imbongu conference, however, it was decided that medial prenasalized stops should be written as <mb>, <nj>, <nd>, <ng>, a decision that brought the orthography into agreement with the conventions of English and Tok Pisin. After this decision was made, the [nuŋa]/[naŋa] community agreed to use <ng> for the velar nasal. Given this decision, <ng> represents both the velar nasal and the velar prenasalized stop. In this respect, the decisions mirror the conventions followed in English and Tok Pisin: The sequence <ng> represents both the velar nasal as in English <singer> and Tok Pisin <sangana> [saŋana] ‘groin’ and the nasal-stop sequence in English <finger> and Tok Pisin <sanguma> [saŋguma] ‘type of murder’.

In summary, the use of <ŋ> was an important marker of dialectal differences, as well as allowing for the straightforward differentiation of velar nasals and velar prenasalized stops. Ultimately, however, the use of <ŋ> was abandoned in favor of <ng>, following the conventions of both English and Tok Pisin.

5.5 Use of <y w> in Umbu-Ungu and Imbongu

The fifth issue, the use of <y> and <w>, was discussed during the Imbongu conference (Malone et al. 1985), but not during the Umbu-Ungu conference. It was decided to use <y w> word-initially, as shown in (33a), but <i u> word-finally and before a consonant, as shown in (33b).

- (33) a. <yokoli> [jokoli] ‘blood’
 <wale> [uale] ‘netbag’
 b. <kou> [kou] ‘stone’
 <koipe> [koipe] ‘club’

It was further decided to use <y w> for intervocalic glides as shown in (34a), even glides that are not necessarily perceptible, since they agree in backness and roundness with the preceding vowel, as shown in (34b).

- (34) a. <kamaye> [kamaje] ‘type of bush cane’
 <awili> [aʷili] ‘large’
 b. <iye> *<ie> [i.e] ‘man’
 <owa> *<oa> [o.a] ‘dog’

While these decisions were made specifically for Imbongu, they have also been adopted for Umbu-Ungu.

These decisions follow Tok Pisin conventions for word-initial and word-final glides: Word-initial glides are written with <y w> in Tok Pisin, as seen in words like <yake> ‘teakwood’ and <wasket> ‘chin’, while word-final glides are written with <i u> as seen in words like <flai> ‘fly’ and <dinau> ‘debt’. Tok Pisin conventions are less clear word medially. While there are a few Tok Pisin words like <pawa> ‘power’ and <mowa> ‘mower’ that are written with an intervocalic <w>, most Tok Pisin words with these sequences are written with <i u> as in <paia> ‘fire’ and <plaua> ‘flower’ or with no glide as in <spia> ‘spear’ and <foa> ‘four’. Word-medial <y w> are common in English, however, even though

medial <y> is most frequently found morpheme-finally (as in <player> and <playing>). The occurrence of medial <y w> in English seemed to be a major factor in the decision to use these letters in Imbongu (and, by extension, in Umbu-Ungu).

It is instructive to compare the use of <y w> in Imbongu and Umbu-Ungu with that in Bine (§2.4) and Kope (§3.5). In both Bine and Kope, the primary motivation for decisions to use (Bine) or not use (Kope) <y w> was a desire to differentiate themselves from some other group. In the case of Imbongu, however, the primary motivation was a desire to look more like English and, to a lesser extent, Tok Pisin. So different decisions in Bine and Kope arose from similar motivations, while the same decision in Bine and Imbongu (and Umbu-Ungu) arose from different motivations. Group identity is a strong factor in these orthography decisions.

6 Methodology

The claim that orthography is a social practice rather than a linguistically-based technology raises the question as to how orthographies should be developed. My proposal is that the Vanimo model outlined in §3 is a good starting point. In this model orthography is not developed in a vacuum. Instead, it is developed in tandem with an activity in which orthography decisions are applied to a real world problem. In Papua New Guinea, the Vanimo workshop united an orthography workshop and a writers' workshop. This basic pattern was further developed in Papua New Guinea and used in more than 100 language groups there (Easton 2003). More recently, I used a similar model in Bangladesh, working with representatives of eight language groups simultaneously involved in orthography development and literacy material development (Clifton 2013).

In these workshops, we begin with participants who are already literate in some language, whether or not they have ever tried to write their own language. Some of the participants may write their own language, but inconsistently. There is frequently disagreement related to the spelling of particular words. If possible, participants should represent the range of orthographic practice found in the community as a whole. Participants begin by writing down a working alphabet, and then make lists of words for each of the letters (or combinations of letters in the case of digraphs and trigraphs). These word lists are made on the basis of their orthographic practices, not on the basis of the 'actual' sounds. Participants are encouraged to make separate lists for each letter in different positions: the beginning of the word, the end of the word, between vowels, and so on.

As part of this, it is important to raise awareness of basic orthographic concepts. One of the most important concepts is digraphs. This can frequently be accomplished by referring to sound/symbol correspondences in the language of wider communication. For example, the sequence <ng> represents the single sound [ŋ] in English, Tok Pisin and Hiri Motu. Once participants get the idea that a sound can be represented by a sequences of symbols, they can identify digraphs (and trigraphs) in their own language, and make lists of words for each of the digraphs (and trigraphs).

As participants make these lists of words, they inevitably encounter problems. They may find that a given letter has two or three different pronunciations depending on the word, or that some letters only occur in a small number of words, or that different letters or sequences can be used to write the same sound. In some cases, they only discover the problems as they try to apply decisions while writing stories; in other cases they discover problems when they try unsuccessfully to read their own or someone else's story. Regardless of how the problems are identified, they form the basis for further discussions of the orthography.

The role of the linguist is to be a facilitator, not a decision-maker. Even during the discussion of problems, participants should be encouraged to take the lead. Rather than identify the source of the problems, the linguist should ask the participants to verbalize why they think a particular practice is problematic. Participants are trained in possible ways to handle what they see as problems, but only when they see the situation as a problem.

Patterns of underdifferentiation generally arise as participants realize that the same sequence can be pronounced in more than one way. For example, during the Bangladesh workshop, participants from the

Tripura communities asked what they should do when they found that word-final <n> had two different pronunciations, as illustrated in (35).

- (35) a. <awan> [awan] ‘rice cake’
 b. <san> [sã] ‘bush sp.’

During the ensuing discussion, participants indicated that they felt it was better to write the two forms differently, even though writing both forms with <an> would almost never cause problems with reading. At that point, we discussed various ways of differentiating between the two, including digraphs and diacritics. In the end, the group decided on their own solution, writing the words with nasalized vowels with a diacritic, but still keeping the (now redundant) final <n>: (35b) would be represented as <sãn>. Although the circumflex on the vowel rendered the final <n> unnecessary, the group decided it was important to keep it to maintain visual unity with closely related dialects that might still pronounce the <n> in these forms (Clifton 2014).

Patterns of overdifferentiation generally arise as participants realize that a particular letter or sequence has a limited range of use. A limited range of use is not, however, a marker of a definite problem. For example, participants from Umbu-Ungu recognized that <s> only occurred before <i> and <u>, but they were not concerned about it. Even though a linguist would see this (along with the fact that <t> never occurs in this position) to be a case of overdifferentiation, it was not a problem needing a resolution. If, on the other hand, the participants saw it as a problem (perhaps because some write <s> while others write <t>) it would become the basis for discussion. Again, the focus of the discussion should be whether there is a consensus that this is a problem, and if so, why it is a problem. At that point we can discuss the possibilities of writing both <s> and <t>, versus choosing just one of the two.

By keeping the focus of the discussions on what the participants see as problems, and why they are perceived as problems, the group can weigh various factors involved in orthography decisions.²⁴ As indicated above, there are competing pressures involved in orthography decisions. Groups use orthography to simultaneously show solidarity with a language of wider communication, as well as their independence from that language. They use orthography to show their unity with related dialects as well as their distinctiveness. Furthermore, these relationships change over time, and groups may well want to change their orthographic conventions over time to show these changes. It is important to emphasize during discussions that decisions made at the workshop are meant to be tentative. Groups should be encouraged to resist making decisions in areas in which there is no consensus. During the Bangladesh workshop, for example, the Khyang group could not even agree on a script: While some wanted to retain the Roman-based script that had been in use, a significant minority wanted an entirely new indigenous Khyang script. Even among those who wanted to retain the Roman-based script, there was no consensus on whether to use <oy> or <oi>, and <ei> or <ey> (Clifton 2013). So, many of the materials were produced in both scripts, and no attempt was made to standardize the spelling of diphthongs in the Roman-based script. The participants agreed to put off these decisions until some unspecified time in the future.

A major focus of the orthography workshop is an alphabet guide which includes a list of the letters and sequences, an indication of how each is to be pronounced, and a list of words in which the letter or sequence appears. As indicated above, it is important to emphasize that this guide is a work in progress, to be changed and modified as conventions are changed. But it provides the community with a touchstone as the process continues. Related to this, participants need to be able to identify symbols which modify the basic meaning of a letter or digraph. These would include indicators of nasality in vowels, length and tone. All these need to be documented in the spelling guide.

²⁴I do not mean to imply that the group needs an overt list of factors to consider. While the consideration of multiple factors might be conscious, it could just as well be subconscious.

Finally, the alphabet guide needs to be accompanied by a body of literature. This might be a collection of stories written at a writers' workshop, or it might be short stories included in a primer. Whatever the form, these stories will allow speakers to get an idea of the overall look of the alphabet in use. They will also allow speakers to practice reading their own language.

7 Conclusion

Orthography development has generally been closely tied to linguistic analyses, modified only as necessary to take into consideration non-linguistic factors. Following Sebba's claim that orthography is primarily a social practice, not a linguistically-based technology, I have presented four case studies from Papua New Guinea in which orthography functions to show the relationship between dialects and between a language community and languages of wider communication. Given the central role of non-linguistic factors in orthography design, I have proposed that orthography workshops should focus on the orthography itself, not on how it is related to a linguistic analysis. By breaking the direct link between orthography and linguistic analysis, language communities will be freer to exploit orthography to express their identity.

References

- Berry, Jack. 1968. The making of alphabets. In Fishman, Joshua A., ed. *Readings in the sociology of language*, 737-747. The Hague: Mouton & Co.
- Berry, Jack. 1977. 'The making of alphabets' revisited. In Fishman, Joshua A., ed. *Advances in the creation and revision of writing systems*, 3-16. The Hague: Mouton & Co.
- Blowers, Bruce L. & Blowers, Ruth. 1970. Kaugel phonemic statement. In *Papers in New Guinea linguistics*, no. 13, 1-12. Pacific Linguistics A-26. Canberra: Australian National University.
- Butcher, Benjamin T. 1963. *We lived with headhunters*. London: Hodder and Stoughton.
- Cahill, Michael & Rice, Keren, eds. 2014. *Developing orthographies for unwritten languages*. Dallas: SIL International.
- Cahill, Michael. 2014. Non-linguistic factors in orthographies. In Cahill, Michael & Rice, Keren, eds. *Developing orthographies for unwritten languages*, 9-25. Dallas: SIL International.
- Capell, Arthur. 1962. *Linguistic survey of the South-Western Pacific*. New and revised Edition. Noumea: South Pacific Commission.
- Clifton, John M. 1987. Orthographic reform in Kope. In Clifton, John M., ed. *Studies in Melanesian orthographies*, 1-15. Datapapers on Papua New Guinea Languages 33. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics. <http://www-01.sil.org/pacific/png/abstract.asp?id=19867> (17 Dec 2015)
- Clifton, John M. 1991. Organized phonology data: Kope dialect of Northeast Kiwai language [kiw]. Ukarumpa, Papua New Guinea: SIL International. ms. http://www.sil.org/pacific/png/pubs/928474542498/Kope_OPD.pdf (18 Apr 2012)
- Clifton, John M. 1995. Organized phonology data: Vanimo language [vam]. Ukarumpa, Papua New Guinea: SIL International. ms. http://www.sil.org/pacific/png/pubs/928474542501/Vanimo_OPD.pdf (18 Apr 2012)
- Clifton, John M. 2013. Orthography as social practice: Is there a role for the linguist? In Norris, Marry Jane & Anonby, Erik & Junker, Marie-Odile & Ostler, Nicholas & Patrick, Donna, eds. *Endangered languages beyond boundaries: community connections, collaborative approaches and cross-disciplinary research = Langues en péril au delà frontières: connexions communautaires, approches collaboratives et recherche interdisciplinaire*, 189-190. Bath, U.K: Foundation for Endangered Languages.
- Clifton, John M. 2014. Orthography as a marker of group identity. In Dyck, Carrie & Granadillo, Tania & Rice, Keren & Labrada, Jorge Emilio Rosés, eds. *Dialogue on dialect standardization*, 81-97. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Clifton, John M. & Clifton, Deborah A. 1995. Vanimo orthography workshop. *READ Magazine* 30(2):3-8.
- Donohue, Mark. 2002. Which sounds change: Descent and borrowing in the Skou family. *Oceanic Linguistics* 41:171-221.
- Draper, Norm & Draper, Sheila. 2002. *Dictionary of Kyaka Enga, Papua New Guinea*. Pacific Linguistics 532. Canberra: Australian National University.
- Easton, Catherine. 2003. Alphabet design workshops in Papua New Guinea: A community-based approach to orthography development. Ukarumpa, Papua New Guinea: SIL International. ms.
- Fleischmann, Lillian, ed. 1984a. *Bine acnema buka: Kunini tage pogogo ca*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Fleischmann, Lillian, ed. 1984b. *Bine acnoma buka: Boje-Giringarede tago pogogo ca*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

- Fleischmann, Lillian & Turpeinen, Sinikka. 1976. A dialect survey of the Eastern Trans-Fly languages. In *Papers in New Guinea linguistics*, no. 19, 39-76. Pacific Linguistics A-45. Canberra: Australian National University.
- Fleischmann, Lillian & Turpeinen, Sinikka. 1977. Bine phonemes. In *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> (29 Nov 2011)
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, William A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29:357-404.
- Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2015. *Glottolog 2.6*. Jena: Max Planck Institute for the Science of Human History. <http://glottolog.org>
- Gido Buka. 1969. *Gido buka: Hymnbook of the Kerewo language (Kikori Delta)*. Hong Kong: Chinese Christian Literature Council.
- Head, June. 2011. A grammar of Umbu-Ungu. Ukarumpa: SIL Papua New Guinea. ms. <http://www.sil.org/pacific/png/abstract.asp?id=928474542774> (14 Dec 2014)
- Imbongu OPD. 2011. Organized phonology data: Imbongu (Imbo-Ungu) language [IMO]. Ukarumpa: SIL Papua New Guinea. ms. <http://www.sil.org/pacific/png/abstract.asp?id=928474542332> (14 Dec 2014)
- Lewis, M. Paul & Simons, Gary F. & Fennig, Charles D., eds. 2015. *Ethnologue: Languages of the world*. 18th Edition. Dallas: SIL International. <http://www.ethnologue.org>
- Malone, Dennis L. & Malone, Susan E. 1983. Kaugel (Umbu-Ungu dialect) orthography conference. Ukarumpa: SIL Papua New Guinea. ms.
- Malone, Dennis L. & Malone, Susan E. & Keruwa, Rambai. 1985. Imbu Ungu orthography conference. Ukarumpa: SIL Papua New Guinea. ms.
- Munro, Pamela. 2014. Breaking rules for orthography development. In Cahill, Michael & Rice, Keren, eds. *Developing orthographies for unwritten languages*, 169-189. Dallas: SIL International.
- Nida, Eugene A. 1953. What is phonemics? *The Bible Translator* 4:152-156. <http://www.ubs-translations.org/tbt/1953/04/TBT195304.html?num=&num1=152&x=15&y=11> (2015 Dec 04)
- Nida, Eugene A. 1954. Practical limitations to a phonemic alphabet. *The Bible Translator* 5:35-39, 58-62. <http://www.ubs-translations.org/tbt/1954/01/TBT195401.html?num=&num1=35&x=12&y=12> (2015 Dec 04)
- Pettersen, Robert. 2010. *Kerewo wade buka, urio ramu*. Ukarumpa, Papua New Guinea: SIL Press.
- Pettersen, Robert. 2014. Kope phonology, orthography, and literacy notes. Ukarumpa, Papua New Guinea: SIL International. ms.
- Pike, Kenneth L. 1947. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan Press.
- Ross, Malcolm. 1980. Some elements of Vanimo, a New Guinea tone language. In *Papers in New Guinea linguistics*, no. 20, 77-109. Pacific Linguistics A-56. Canberra: Australian National University.
- Saari, Sinikka. 1992. Organized phonology data: Bine language [ORM]. Ukarumpa, Papua New Guinea: SIL International. ms. <http://www-01.sil.org/pacific/png/pubs/928474541693/Bine.pdf> (5 Feb 2015)
- Sebba, Mark. 2007. *Spelling and society: The culture and politics of orthography around the world*. Cambridge: Cambridge University Press.
- Sebba, Mark. 2009. Sociolinguistic approaches to writing systems research. *Writing Systems Research* 1:35-49.
- Sebba, Mark. 2015. Iconisation, attribution and branding in orthography. *Written Language and Literacy* 18:208-227.
- Seifart, Frank. 2006. Orthography development. In Gippert, Jost & Himmelmann, Nikolaus P. & Mosel, Ulrike, eds. *Essentials of language documentation*, 275-299. Berlin: Mouton de Gruyter.
- Smalley, William A. 1954. A problem in orthography preparation. *The Bible Translator* 5:170-176. <http://www.ubs-translations.org/tbt/1954/04/TBT195404.html?num=170&x=10&y=15&num1=> (2015 Dec 04)
- Smalley, William A. 1959. How shall I write this language? *The Bible translator* 10:49-69. <http://www.ubs-translations.org/tbt/1959/02/TBT195902.html?num=49&x=16&y=16&num1=> (2015 Dec 04)
- Umbu-Ungu OPD. 2011. Organized phonology data: Umbu-Ungu language [UMB]. Ukarumpa: SIL Papua New Guinea. ms. <http://www.sil.org/pacific/png/abstract.asp?id=928474542287> (14 Dec 2014)
- Venezky, Richard. 1970. Principles for the design of practical writing systems. *Anthropological Linguistics* 12:256-270.
- Wurm, Stephen A. 1973. The Kiwian language family. In Franklin, Karl, ed. *The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea*, 219-260, 565-571. Pacific Linguistics C-26. Canberra: Australian National University.
- Wurm, Stephen A. 1975. The central and western areas of the Trans-New Guinea Phylum: The Trans-Fly (sub-phylum level) Stock. In Wurm, Stephen A., ed. *New Guinea area languages and language study*, vol. 1: *Papuan languages and the New Guinea scene*, 323-344. Pacific Linguistics C-38. Canberra: Australian National University.