Orthographies in Papua New Guinea through the Years

Ray Stegeman

Abstract. I would like to present the use of non-English graphemes by many other SIL-PNG colleagues over the years. SIL-PNG has been at work documenting little-known languages in Papua New Guinea for over 60 years. The motivation for using certain graphemes, diacritics and other writing strategies has changed through the years, particularly related to choices made in nearby languages, related languages and the choices available through the dominant, colonial language of English.

It is particularly interesting how in recent years, linguists and translators have moved away from using diacritics and other unique graphemes, especially with the advent of cell phone use, so underdifferentiation in the alphabet has become commonplace, and its relative acceptance and efficacy will be the focus of my presentation.

I have done a systematic analysis of graphemic choices made for most SIL projects in PNG for the past 60 years. Unusual, non-English graphemes are the focus of that research, and a questionnaire was sent to current SIL members asking about their motivations for using or not using certain unusual graphemes. I wish to compare which uncommon graphemes are chosen to represent which phonemes and gain insight into their efficacy as well as their general acceptance among the people who use the orthographies.

1. Introduction

SIL has been at work for over 60 years in Papua New Guinea, a land teeming with hundreds of languages. My colleagues and I have worked in over 300 of the over 800 languages that exist in this most linguistically diverse of countries/regions of the world.

As most everybody in the field of linguistics knows, Papua New Guinea is the most linguistically diverse nation on earth. The Ethnologue currently lists PNG as having 841 living languages. Of these, 164 are either "in trouble" or "dying". Those are not particularly disparaging numbers, given that Wikipedia (https://en.wikipedia.org/

Ray Stegeman
SIL International
Dallas, TX, United States

Y. Haralambous (Ed.), *Graphemics in the 21st Century. Brest, June 13-15, 2018. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2534-5192), Vol. 1.
Fluxus Editions, Brest, 2019, p. 269–292. https://doi.org/10.36824/2018-graf-steg
ISBN: 978-2-9570549-0-9, e-ISBN: 978-2-9570549-1-6

wiki/Endangered_language) reports that 50–90% of the world's languages will be gone by 2100. It seems that language identity and loyalty in PNG is still a very strong social force, and people here have a binding, common identity with their wantoks (people who share "one talk" with them) that overcomes many of the obvious forces of nation building and national identity through promotion of a national language or languages elsewhere. Western culture and psyche focus on "being on the same page" and "speaking the same language". This is a strong cultural goal for us, particularly in the internet age, so that any linguistic individuality we may have falls to these preeminent pressures. In Papua New Guinea, not so much.

SIL International (formerly The Summer Institute of Linguistics) has been at work in PNG since 1956. By 2006, we had served in 337 language communities and we are currently serving in 187 of those same communities. Much linguistic work has been done on almost all of these languages (http://www.pnglanguages.sil.org/resources) including many grammars, phonologies and literacy materials. Along with those, there is a formidable amount of material on the design and reform of the orthographies that would best serve these many groups in reducing their languages to writing. You might imagine that when small communities pride themselves in preserving their own languages, they might also be interested in having their orthographies stand out as unique—different from nearby neighbors, if not necessarily different from the national languages of English, Tok Pisin and Hiri Motu. So, in surveying these orthographic materials, one finds a good number of strategies used for writing the myriad of languages in this country.

Many of our personnel have been and continue to be committed to seeing the indigenous languages of PNG written so that the speakers can feel a sense of pride at seeing their mother tongue on paper, in books, words in a dictionary, and so forth, and realize more completely the value of their language. Of course, this is not the only way to recognize the value and prestige of a language, but it is one very important way. And so, all those hundreds of languages must have orthographies.

Before being allocated to one of these language groups, my colleagues all had linguistic training, and we might admit that some of us were overzealous in trying to make plain all the esoteric phonological qualities of any one language in its orthography. While making it easier for an outsider to read and write in their language, this concept may actually make it more difficult for the local speakers to write and/or read their first languages—some of which are arguably among the most morphologically complex in the world. One colleague reports that in the language he studied, they can have more than 10,000 forms of any one verb (Menya, Whitehead, pc).

So, among these challenges arises the unique challenge of writing these languages—allowing them to learn from and share with each other

in ways that weren't available to them before—for parents to write a letter to their child, far away at a national high school; for someone to write to a brother working in the mines across the island. And, of course, with written Bible translation, people can check for themselves the Bible stories so often talked about. "Look, it's that story I heard about; it's written down right here, and I can read it for myself—and I understand it clearly!" Whether it be the Gospel of Mark or a community meeting notice at the local shop or checking up on what their children are learning at the local tokples preschool, they have a new-found power and prestige of seeing their language, not only talking it. They are not only preserving their language, they are expanding the use of their languages into new milieus.

As an introduction, I want to discuss some of the literature on the subject of orthographies in general and move toward more specific ideas coming out of PNG and SIL's work in PNG, since I have most access to our colleagues' work. After that, I'd like to present the data and some results I have gleaned from the data. I want to include some evidence from local testing of graphemes and especially subjective feedback from the SIL teams and the communities they worked with on this alphabet enterprise—the very people who are left to use the alphabet either given them, or the alphabet they choose, and more ideally, the orthography that represents the best of both efforts.

In following sections, I will present the use of non-English graphemes by many other SIL-PNG colleagues over the years. The motivation for using certain graphemes, diacritics and other writing strategies has changed through the years, particularly related to choices made in nearby languages, related languages and the choices available through the dominant, national language of English and trade language Tok Pisin. It is particularly interesting how in recent years, especially with the advent of cell phone use, underdifferentiation in the orthography has become more common.

I have done a systematic analysis of graphemic choices made in many of the SIL projects in PNG for the past 60 years. Use of non-English graphemes are the focus of that research, and a questionnaire was sent to current SIL members asking about their motivations for using or not using certain unusual graphemes and the graphemes they chose for sounds not described in the English alphabet. I wish to compare which uncommon graphemes are chosen to represent which phonemes and gain insight into their efficacy as well as their general acceptance among the language communities that use the orthographies.

2. Literature on Orthography Development

The idea that local language writing systems should resemble official and/or national languages is a well-established goal for any local orthography (Grenoble & Whaley, 2004:158 and others). For that reason, I won't investigate the possibility of languages in Papua New Guinea using any other kind of writing system, such as abjads or logographies, etc. It seems obvious they would not be helpful in allowing the language communities of PNG to bridge[1] between their local languages and the national languages of English and Tok Pisin. There has been some recent success in teaching the reading and writing in Uniskript (R. Petterson and D. Petterson 1994; also `http://uniskript.org/`), especially as a precursor to learning to read and write using a more regular alphabet. The Uniskript alphabet encourages its users to adapt the symbols to how a particular language community views the sounds they make, so although the "common" way of representing a t sound is with a triangle (showing the tip of the tongue touching the roof of the mouth), it can be adapted in different ways to suit the perceptions of sounds in each language community.[2]

Going back to the formative years of SIL, Pike (1947) wrote a book that included a chapter titled "The Formation of Practical Alphabets" which included linguistic and non-linguistic motivation for designing alphabets for local communities. He also addressed important topics such as dialects, knowledge transfer of first language reading and writing to languages of wider communication and advice against excessive use of diacritics, among other things.

A book by Smalley et al. (1964) called *Orthography Studies* addresses many topics including a chapter called *Practical limitations to a phonemic alphabet*. He suggests that many other sociolinguistic factors be considered when forming a writing system among a language community.

Gudschinsky's book, A manual of literacy for preliterate peoples (1973, has many helpful chapters on such literacy ideas as functional load, underdifferentiation, orthography testing and phonemic as well as morphophonemic representation. This book was adapted to language work in PNG and published by SIL in PNG. The sources from Pike and Gudschinsky both are very practical in their information on implementing orthography development and reform.

---

1. Bridging, here, is referring to the idea that mother-tongue orthography and literacy be designed as a reflection of the national/official language as much as possible, so that letters and spelling rules chosen for the mother-tongue help a student move more easily from mother tongue literacy to national/official language literacy and vice versa.

2. Unrelated to Uniskript, but on the same topic, one language community chose the tilde over their vowel letters to symbolize nasalization, and they call it *titi*—'wave,' and the symbol seems to them to represent a wave (Kala, Mitchell Michie, pc).

Simons (1977) suggests viable solutions for orthographies for multiple dialects, which is a highly salient topic in many areas in PNG. His multidialectal approach to orthography design can help closely related dialects to use the same orthography following seven principles that include social acceptability, minimal potential ambiguity, simplicity and phonemic contrast and neutralization between dialects. He also discusses four levels of psycholinguistic reality—phonetic, phonemic, morphophonemic and fast speech. He stresses that it is possible for dialects to have the same reality of a word in one of these realities, even when in other ones they would be different.

UNESCO has sponsored many articles and books on literacy and orthography development, one of which is *The manual for developing literacy and adult education programs in minority language communities* (Malone, 2004). It has a chapter on devising alphabets for previously unwritten languages, giving practical guidelines and helpful case studies from around the world.

Also from UNESCO is *Writing unwritten languages: A guide to the process* (Robinson and Gadelii, 2003), which has very practical help for grassroots level workers. It focuses on the stakeholders—those who would be using the orthography—and the issues that most often affect acceptability of the orthography. They also mention non-linguistic factors that can most often affect—positively or negatively—the acceptance of an alphabetic orthography. These include social, political and cultural considerations. Most of these are written in the form of questions, to get the language communities thinking for themselves, such as:

- Socially, what are the relationships between the language community and its own dialect areas? How do they view their language and possible orthography choices? What about the relationships between the language community and other languages within the greater area?
- Politically, what degree of autonomy does the language community have in making orthography decisions? What about colonizing effects on national as well as local literacy?
- Regarding preserving a cultural heritage, what does writing do to an exclusively oral culture? Who are the gatekeepers for preserving this newly written base of knowledge, when it has up to now been held by those who preserve it orally? (ibid., Section 3, p. 11–13)

In writing about language structure, Robinson and Gadelii speak about these important ideas for consideration:

- Representing distinctions in sounds to avoid confusion (minimize underdifferentiation).
- The phonemic ideal that one symbol represent one sound, always and only.
- How does the grammatical structure of the language influence how it is written? (e.g., word breaks, elision, verb morphology, etc.)
- Can the structures of related languages give ideas for developing the orthography? (ibid., Section 3)

Cahill (2014) mentions that in addition to sound scientific study and community involvement, effective teaching materials and practices are crucial to having a lively, sustainable literate public. He also talks about promoting the following benefits to a language community that is seeking to own and utilize mother-tongue literacy:

Tools to deal with the larger world which is unavoidably coming (to all language communities) in:

- interacting with computers and the internet;
- thwarting attempted land grabs and other legal tangles;
- more math awareness, to deal with others, especially regarding money;
- more access to beneficial materials;
- gaining knowledge about health (water, AIDS, nutrition, etc.);
- knowing the contents of government and NGO documents (e.g., UN Declaration of Human Rights);
- not losing almost-forgotten folk tales and other local lore;
- preservation of a community's cultural resources;
- strengthening one's cultural identity and having a higher view of one's own language;
- strengthening the language (Malone (2004) lists "materials for language education and literacy" as one of the nine factors that most affect language vitality);
- for women especially, and sometimes whole language groups, increased self-esteem;
- letter-writing can be more private than conversations;
- an aid when traveling (e.g., following directions, finding destinations, etc.). (Cahill, 2014, Appendix, p. 6)

Other literature was also consulted while writing this paper; please see the bibliography for details.

3. PNG Milieu

There are over 800 living languages in PNG—a nation of less than 8 million people, occupying an area less than the size of France. There are obviously many challenges to nation building and communication among this ethnically and linguistically diverse people. Still, many of the language communities have been and continue to stay connected by trading and using common trading languages like Tok Pisin and Hiri Motu. In the more recent past, with roads connecting communities and better infrastructure, more and more people are moving to the bigger cities for economic and educational opportunities. These people, while desiring to keep their cultural and linguistic heritage, are also realizing the opportunities for advancement by way of modernization and living in multi-linguistic and multi-cultural communities.

Many of the languages are related to many others, so even though one can walk down the road just 5–10 km and likely find another group

of people speaking a different language, many times, the languages are related. Many of the same cultural ideas are preserved in multilingual communities (e.g., trade, the cultural custom for men to marry outside their clans, etc.) These ideas feed the concept that many people are multilingual, depending on the environment and the audience. And so, many Papua New Guineans have a knack for learning languages, appreciating sound and grammar patterns that those of us with fewer multilingual opportunities could easily miss or overlook.

The official languages of Papua New Guinea are English, Tok Pisin and Hiri Motu. Tok Pisin is an English-based pidgin/creole and used largely in the north and highlands areas, while Hiri Motu is an adaptation of Motu, an Austronesian language, developed through trade and by colonial efforts mostly in the south of the country. All three use similar orthographies and spelling rules, and these are taught in schools, so any literate person working on indigenous alphabet development will be strongly influenced by these traditions when it comes to creating an alphabet in one's own language for the first time.

The two biggest language families in Papua New Guinea are Trans New Guinea (483, Ethnologue.com)[3] and Austronesian (1,256—not all in PNG, Ethnologue.com). Austronesian languages are mainly found along the coast and in the outlying islands of PNG, while Trans New Guinea languages are largely found in the highlands. There are some exceptions to this, and one can find largely Austronesian languages displaying some obvious Trans New Guinea features (similar pronouns and counting systems, word order, grammatical features, etc.) and vice versa. This might seem natural in a place like PNG, since over time there has been so much language contact among the different communities. It would be and indeed is, hard to tease apart the indigenous features from the borrowed features of any individual language.

One feature of interest for this study is that the number of graphemes is largely related to the number of phonemes needing representation in the orthography. While Austronesian languages are said to be less phonologically complex than Trans New Guinea languages (`https://www.britannica.com/topic/Austronesian-languages/Structural-characteristics-of-Austronesian-languages`), I have found phoneme numbers to be widely divergent and seemingly not related to the family of any one language. One famous PNG language, Rotokas (N Bougainville language family), is known for having (arguably) the fewest phonemes of any language on earth—11. The Trans New Guinea language of Melpa has 26 phonemes. Alekano, also a TNG language, has only 16 phonemes while Sudest, an Austronesian language, has 40 phonemes. These are offered

3. Trans New Guinea is actually a grouping of many language families that are known not to be Austronesian.

as a small sample that confirms languages from either major language family can have a big difference in the numbers of phonemes.

One objective of this study was to look at how many non-English letters and diacritics are needed/used in a local orthography, so the number of phonemes, while far from the only consideration, is a significant consideration—possibly the best starting point for engaging the local communities in making alphabet choices.

4. SIL's Work in Papua New Guinea

SIL International began work in PNG as the Summer Institute of Linguistics in 1956. Following what Pike (1947) and others promoted, many of our SIL teams worked hard to discover the basic building blocks of each language they encountered—sounds, phonemes, morphemes, morphophonemic sound changes, and other linguistic concerns that are directly related to any orthography developed for the languages. Most of their work is documented and can be found on www.pnglanguages.sil.org.

While many of the teams were careful in their linguistic studies, and in accordance with their linguistic training, they also consulted members of the language communities in which they worked, eliciting feedback from local leaders as to the best ideas for alphabet choices. In the questionnaire I gave to 39 current SIL colleagues, 37 of them said they had ongoing consultations with local leaders—often school teachers and/or school administrators, respected village and church leaders—in making initial and ongoing decisions about alphabet letter choices and spelling rules. They worked together to address linguistic concerns as well as many sociological concerns with respect to the orthography. Many SIL colleagues spent many long years doing literacy and translation work among the people, living in their villages for months at a time. They have documentation of testing different alphabet choices made by them, in conjunction with the greater language community as to how they "liked" what they saw and/or were successful in learning to read and write. Many of the SIL teams were following the extensive testing procedures covered in Gudschinsky's book (1973, ch. 13) along with other resources.

One of the issues that seemed to come up most frequently was how similar or different one's language "looks" on paper compared with neighboring languages. Some language communities desired to follow what they saw being used around them and sought to use letters and diacritics similar to nearby languages. Other communities sought to "show off" the uniqueness of their languages by doing things differently from those nearby. So, as a non-linguistic factor, one of these two opposing

preferences could be mainly responsible for how particular items in a single orthography come to be used.

More recently, with the advent of large-scale cell phone use in PNG, it has become a concern of many language communities how relatively easy or difficult it is to text in their mother tongues, without the aid of apps that need to be downloaded and often set up (InKey, Keyman, etc.) This makes it a priority to use letters and diacritics available on simpler, lower-priced cell phones, to allow those who wish to text in tokples ('talk place'—the local language). This was the topic addressed most in the Questionnaire section of my research for this paper.

5. Grapheme Choices

In looking at SIL phonologies and orthographies of PNG languages through the years, I was interested in charting strategies that teams used to describe sounds that required non-English graphemes. These included diacritics, digraphs and trigraphs (and one tetragraph) that are not used in English, which I group together to call multigraphs from here on. I was also interested in the use of English letters in the orthography used for sounds other than regular English sounds, as in <b> for the phoneme /β/, or <c> for the glottal stop. I also looked at the strategies of overdifferentiation and underdifferentiation, to see if there were any trends developing through the many years SIL has been at work among PNG languages.

I had access to three different sources of this orthography data, based on the relative dates of the information. The oldest material comes from charts of phonemes mapped with graphemes of language projects from before 1990. The second group of data comes from OPDs (Orthographic and Phonology Data) which are on file in the Linguistics Office at Ukarumpa. Many of these OPDs have been written and/or updated up to 2010. The third group of data available to me came from answers to a questionnaire given to current language teams and the current graphemic data they provided. The time frames overlap from the standpoint of when the orthographies were developed, but the endpoints are fixed and exclusive, so that orthography strategies from the first group (before 1990 would not have been concerned with, say, relative ease of texting in the mother tongue, whereas data from answers to the questionnaire (that is, data from a currently active language project) would be concerned about strategies for employing graphemes in direct relation to that issue or others. So, I believe the three groups of data correspond with different philosophies of literacy, reading and writing, and communication tools available to the local speakers/readers/writers in separate time frames.

Along with their graphemic data, those who completed the questionnaire gave answers to questions regarding who was involved in the language program and the grapheme choices as well as changes made to their graphemes throughout the life of the program and the reasons they felt they needed to make changes. A list of the questions from the questionnaire is together with the answers in section 8 below.

Regarding the grapheme data, I counted the number of instances of each strategy ("strategy" here refers to things like using a tilde to show nasalization, or the digraph <ng> for the velar nasal phoneme /ŋ/) and kept track of them on a chart. Then I put these individual strategies into major groups ("major groups" here refers to diacritics, multigraphs and letters not used elsewhere). Some interesting results emerged as I put the data into six groups and compared their numbers in relation to the 3 historical eras. I needed to give each section a weighted average, since the number of languages researched was different from those in the other two groups.

One caveat: The languages in this survey are from a number of language families (charted below in section 6), and the percentages from the different language families are not constant in the 3 time-related sections (the middle three columns). It is possible that the increased number of Austronesian languages from the newer data (Questionnaire) section of research was the main reason for, say, the increase in diacritic use from the older data, which more heavily favors Trans New Guinea languages. This skewing factor could be eliminated in any future study of a similar nature.

6. Data Collection

Table 1 shows the number of languages in my research and in what province they are mainly located.

Table 2 shows the language families to which the researched languages belong. Again, Trans New Guinea is a convenient designation for several large language families traditionally recognized as non-Austronesian.

And finally, Table 3 shows the actual research data. The numbers in columns 4–7 represent the raw data (RD) of the research and the second number in each cell represents the weighted mean (WM), based on the different number of languages in each section divided by the total number of languages, so the second number in each cell are the actual numbers for comparison across columns.

Table 1. Languages from research and in what province they are mostly spoken

	older data ⟵		⟶ newer data	
PNG Province names	Languages from before 1990	Languages from before 2010	Language teams responding to the Questionnaire (current projects)	Total languages from provinces
Morobe	5	9	9	23
East Sepik	9	7	3	19
Madang	2	15	2	19
Milne Bay	5	9	4	18
Gulf	7	6	2	15
Western	5	5	4	14
Autonomous Region of Bougainville	7	2	3	12
East New Britain	8	1	3	12
Eastern Highlands	7	4	1	12
Central	10	1	1	12
Sandaun	4	7		11
New Ireland	7	1	3	11
Manus	6	2	2	10
Oro	6	2	1	9
Simbu	6	1	1	8
Southern Highlands	5	1		6
West New Britain	3			3
Western Highlands	3			3
Enga	2			2
Hela			1	1
Jiwaka	1			1
Total	108	73	40	221

7. Data Results

7.1. Increase in Use of Diacritics

Moving from the past to the present (moving across Chart 3, from left to right), the data show an increasing use of diacritics among the orthogra-

TABLE 2. The number of researched languages and the language families to which they belong

	older data ⟵		⟶ newer data	
PNG language families represented	Languages from before 1990	Languages from before 2010	Language teams responding to the Questionnaire (current projects)	Total in language families
Trans New Guinea	53	37	11	101
Austronesian	35	15	21	71
Sepik	6	4	0	10
Torricelli	1	4	3	8
South-Central Papuan	0	3	3	6
Isolate	3	1	0	4
Ramu-Lower Sepik	1	3	0	4
South Bougainville	2	1	0	3
Border	1	1	0	2
Eastern Trans-Fly	1	0	1	2
Senagi	1	1	0	2
Yele-Western New Britain	0	2	0	2
Arai	1	0	0	1
East New Britain, Baining	0	0	1	1
East New Britain, Taulil	1	0	0	1
Fas	1	0	0	1
North Bougainville	1	0	0	1
Skou	0	1	0	1
Total	108	73	40	221
Divide raw data below by this number to get a weighted mean	.489	.330	.181	1.000

TABLE 3. Comparing use of each orthography strategy across the 3 time periods of research

			older data ⟵		⟶ newer data	
	Strategy used	Examples	Languages from before 1990	Languages from before 2010	Language teams responding to the Questionnaire (current projects)	Change in use of each strategy
			RD/WM	RD/WM	RD/WM	
1.	diacritic	ë, ã, ú	78/160	65/197	54/298	increase in use
2.	multigraph	th, mp, ndr	264/540	263/797	153/845	increase in use
3.	underdifferentiation (including phonemes not written)	<e> for both /e/ and /ə/	40/81.8	35/106	33/182	significant increase in use
4.	overdifferentiation	<b> and <mb> for /b/	114/233	89/270	32/177	eventual decrease in use
5.	English letter not used elsewhere	c, q, x	125/256	72/218	33/182	decrease in use
6.	non-English letter	′, ʔ, ŋ	36/73.6 (6 ŋ, 17%)	31/93.9 (13 ŋ, 42%)	8/44.2 (5 ŋ, 62%)	eventual decrease in use (but increase in use of ŋ)

phies SIL language teams employ. The weighted mean number (WM, the second number in each cell) increases. (Row 1 has the numbers 160, 197, 298, from left to right.) Some examples of diacritic use include certain marking on vowels to show nasalization, so that the languages have a set of non-nasal vowel letters, say <a, e, i, o, u> along with a nasalized set, say <ã, ẽ, ĩ, õ, ũ>, or possibly a complementary set of long vowels

marked with diaresis, say <ä, ë, ï, ö, ü>. Of course, if a language team decide to represent nasality or length with diacritics, this adds significantly to the overall number of diacritics used—5 or more, as opposed to using a single diacritic to show, say, a dental t phoneme different from alveolar t. I chose to document the marking of nasalization and length (the two most common need for graphemic adjustment) by the diacritic used on each letter, as opposed to documenting use of only one diacritic. This is because some languages documented only some vowels as having nasalized counterparts and not the whole set of vowels.

Use of diacritics in orthographies employed by SIL-PNG language projects were mostly used in the vowel systems, either showing nasalization, as above, or a similar place of articulation on the vowel chart. An example of this would be <u> for the close, back, rounded vowel /u/ and <ü> (with a diaresis) for the close, middle vowel /ɨ/. Gizrra does this, along with using an acute mark on the o <ó> for the schwa phoneme /ə/ while also having the <o> letter for the middle, back rounded vowel /o/ phoneme. Thus, for seven vowel phonemes, they use the regular 5 vowel letters of English and two of the same vowel letters with two distinct diacritics.[4]

7.2. Increase in Use of Multigraphs

There was also an increase in the number of multigraphs used in the alphabets chosen by SIL-PNG language projects in more recent years. Many PNG languages have phonemic systems that involve prenasalization /^{m}b, ^{n}d, ^{ŋ}g/ and labialization /p^{w}, t^{w}, k^{w}/ of plosive phonemes. These can have some phonemic alterations, such as word-initial, word-medial or word-final forms. This can mean that it is not phonemically critical to show the prenasalization or labialization in the alphabet as a matter of linguistic description, but it may be preferred by the language communities. Often, this is the case, due to the influence of English as an official language. and the letters they see when they read English words like co*mb*ine, co*nd*ition, *tw*in, *qu*ick, etc. and they see both sounds represented, they feel like their languages should be written the same way, with both letters used, even though the phonemic reality for the two languages can be quite different. Some PNG languages have both prenasalised *and* labialized consonants /^{m}b^{w}, ^{n}d^{w}, ^{ŋ}g^{w}/, which are sometimes represented as trigraphs, with one tetragraph used in one language for the prenasalized, labialized velar plosive, /^{ŋ}g^{w}/, spelled

4. It is interesting to note that this team felt it was easier to recognize different diacritics for the extra two vowel phonemes, rather than using the same diacritic. Both the letter (o and u) and the diacritic show the difference, to emphasize recognition of the difference in reading more quickly (vanBodegraven, pc).

<nggw> (Khehek, Manus). As with nasalization on vowels, representing these phoneme series as multigraphs can have a multiplying effect on the number of counted strategies used in orthography design. It also tends to make certain words (more words in some languages than in others) unduly long, and the length is multiplied since the condition of prenasalization and/or labialization usually occurs over the whole range of plosives, and not just a single phoneme.

Use of digraphs is also a common strategy in vowel phonemes. Double vowels such as <aa, ee, ii, oo, uu> are often employed in orthographies to identify length. Use of double letters for vowel length, and digraph use in general, can lead to disproportionately long words, depending on the actual language and how many syllables have nasalization or length (or both) on the vowels. In fact, digraphs representing vowel length is a common strategy among SIL-PNG language projects.

Although it isn't phonemically necessary, language communities in a situation where English is the national language feel compelled to use nasal letters together with the prenasalized consonant phonemes. This is a valid consideration in relation to the official languages of English and Tok Pisin, especially when the concept of bridging between languages is a serious concern.[5] Local speakers have learned about spelling rules in English, and they have learned to read English from being taught in school. When they hear the same sounds in their language, they tend to want to use the same letters and spelling rules that they know from another language, particularly the prestigious official language. This can be alright in some cases, and it can be more helpful in a multi-lingual environment like PNG, where people move easily from language to language based on the social situation, but such uninformed orthographic transfer can ignore the unique phonemic and morphophonemic tendencies of any one language, which could help in more quickly and completely acquiring fluency in reading and writing one's mother tongue. Languages, after all, are so very different from each other; it seems natural (to an outsider/linguist) that the orthographies representing these languages would also be very different from each other. It is probably best from a bridging perspective to utilize some of the same letters for the same phonemes (*not* sounds) while also showcasing other unique phonemic qualities of a language by making some unique grapheme choices, which could include the writing of prenasalization.

5. Bridging, here, is referring to the idea that mother-tongue orthography and literacy be designed as a reflection of the national/official language as much as possible, so that letters and spelling rules chosen for the mother-tongue help a student move more easily from mother tongue literacy to national/official language literacy and vice versa.

7.3. Increase in Use of Underdifferentiation

This seems like a surprising result, considering the previously mentioned two trends. One might think that using more diacritics and multigraphs would correspond with (adequate) differentiation or even overdifferentiation, as I have been talking about above.

The increase in use of underdifferentiation could be the result of the increased use of technology having a direct effect on language and literacy development, and the felt need for communicating in one's mother tongue using different electronic devices, including cell phone use. Cell phone use has skyrocketed in PNG in the recent past, and it is a felt need, at least in some language communities, to use one's mother tongue in calling and texting each other. While the phonology of a particular language may be complex enough to need many diacritics and/or multigraphs for a more phonemic representation in the alphabet, it may be even more desirable by the community to reduce the number of "untextable" letters in the alphabet, to make it easier to communicate with each other by using today's technology. This is mentioned in a few of the questionnaire responses, as can be seen in section 8.4.

This underdifferentiation has a trade-off in that while it is easier to text/write, it is often much more difficult to read. The onus of communication falls to the reader in deciphering a message that could have more than one meaning based on the lack of sufficient letters for the meaningful sounds of a language. Many of the questionnaire respondents mentioned that speakers usually text without the diacritics in the official orthography, and for a few diacritics (depending on how many and how often they are left out) they can make themselves easily understood. This would obviously have a limit, so that by leaving off 5 or 8 or 12 special characters or digraphs, one's texting would certainly become more of a deciphering challenge than actual effective communication.

7.4. Decrease in Use of Overdifferentiation

Both increasing underdifferentiation and decreasing overdifferentiation could be tied to a growing interest in using modern technology in one's mother tongue without the need for special apps. More and more Papua New Guineans have access to computers and smart phones, but they don't necessarily have the ability or knowledge to adapt them for mother tongue language use. To have an orthography that is simple enough to use a regular cell phone to communicate in one's mother tongue seems to be more of an interest than before.

Some teams also mentioned the desire for shorter words. A lot of PNG languages can have complex morphology, especially on the verb, and this can make words unwieldy in their length. Together with multi-

graphs, written words become difficult to read. One way to counteract this problem is to use fewer multigraphs, which might be preferred for other reasons (like bridging) but would decrease the use of overdifferentiation.

7.5. Decrease in Use of English Letters Not Used Elsewhere

Over the decades, there has been a noticeable decrease in the use of English letters not used elsewhere in the mother tongue grapheme inventory for SIL-PNG language teams. These letters often include v, z, q, c, x, w, and/or y. In the past, using one of these letters was often the strategy for indicating the glottal stop phoneme, /ʔ/, for which some languages use <q> or <c>. Some languages use <x> for uvular phonemes like /x/ or /ɣ/. This seems like a good strategy to use, especially based on current texting concerns. These letters are immediately available on regular computer keyboards and texting devices, and they don't require special apps or set-up. But the data in this survey shows they are not used as often as they were in the past. This could be due to the bridging concerns mentioned earlier, where letters used in one's mother tongue are expected to reflect the alphabet and sound patterns used in the official language(s). So, for example, using a <c> for the glottal stop doesn't "feel" natural, when one has a strong association that the c letter should/must represent the [k] sound as in <cat> or [kæt] and not the glottal sound. Of course, the major difference is that the glottal stop is not a phonemic sound in English, while it is a meaningful sound in many PNG languages, and necessary to include in the orthography for that reason.

7.6. Decrease in Use of Non-English Letters

The use of ŋ as a grapheme has increased over time, which contrasts with the overall decrease in using other non-English letters. These include the apostrophe or question mark (both mostly for the glottal stop), certain IPA graphemes (usually for similar sounds/phonemes) including letters like <ɬ, æ, ə> etc. The decreasing use of letters like these in PNG orthographies is perhaps to be expected, again considering the spread of technology and the texting phenomenon. These characters for use in an alphabet are not standard on computer keyboards or phone touchpads. Some special, non-English letters are found on smartphones by pressing and holding certain letters, which reveals a choice of alternate characters, but this feature is only available on higher-end smart phones and often only with the special characters used in European languages; that is, mostly English letters with certain diacritics and no non-English letters like above.

It may seem contradictory that with an increase in the availability of technology, we would have more choices in the alphabet letters available to us. But in fact, it seems like we are still bound to using only a standard computer keyboard (like we were bound to using typewriter keyboards before computers) and bound to using only a cell phone keypad with the only options being European diacritics favored by the phone makers or stakeholders other than those people with emerging writing systems for their languages. While Unicode has increased the availability of a vast array of characters for use, the vast majority of them are still not available to the average language speaker who would like to text/write in his/her first language, but who doesn't have or know about the options for using all those characters on simple technology.

7.7. Overall Impressions

Based on the questionnaire responses (newer data) versus older orthographic data, there seem to be forces at work causing language projects to use strategies for crafting orthographies that are less purely based on phonemics, more based on mirroring official languages, and more a reflection of the strong felt need for simpler orthographies for use on technology currently available to language communities. These sociolinguistic forces are made evident in the increasing use of digraphs, the increasing use of underdifferentiation and the decreasing use of overdifferentiation.

8. Questionnaire Feedback

8.1. I asked the currently active SIL teams to share with me some general ideas about the specific situation they found among the local speakers of the languages with whom they work. The idea they mentioned most was the challenges they face trying to develop an alphabet for multiple dialects. One team mentioned the people all had strong dialect loyalty, which meant it was hard for the speakers of various dialects to utilize a common alphabet and spelling rules that didn't reflect their particular pronunciations.

A couple teams mentioned that they had no significant dialect challenges and were able to realize a unified orthography across minor dialect boundaries.

Another challenge to teams was that of having had multiple SIL teams at work in the same language communities through the years. There were also instances of other mission agencies working in the same area previously, in particular, German missionaries who made alphabet choices based on German sound-symbol correspondences (e.g., <ch> for

/x/), which make it difficult for transferring reading skills from mother tongue to the official language now that it is English. Another example was the influence of certain Fijian missionaries who chose <g> to represent the velar nasal /ŋ/. Some choices made by earlier teams are often difficult to overturn, especially as the older generation owns and appreciates the earlier choices, but the younger generations would like to see something based more on current realities.

8.2. I also asked current teams about the stakeholders that were involved in the alphabet-making enterprise. Who gave input into the process? How were they chosen?

Those who became involved (other than the SIL team themselves) were mainly school teachers, local speakers of the language (through informal, occasional meetings), and local church and other community leaders. Many of the SIL teams formed either language committees or translation committees who were responsible for making orthography decisions, often meeting on a regular basis and conducting surveys or tests related to orthographic choices and/or changes.

The SIL teams have many tools at their disposal to help make or force decisions about the orthography. Some of the more common tools mentioned as being instrumental in the process are: Alphabet Development Workshops (Nukna, Blafe, Middle Kodut and others), developing tokples prep school literacy materials (Nek, Seimat, Nehan and others), orthography testing methods (Edolo, Mato, Seimat, Iyo and others), writers' workshops (Misima, Solos, Lote, Arop-Lokep, Kanja, Nehan) and distributing copies of the trial orthography along with locally authored stories using the orthography and then getting feedback from fellow speakers about their alphabet preferences.

8.3. The next question had to do with orthographic strategies they tried early on and decided to change. There were not many common answers in this section, as a testimony to the linguistic diversity and language communities' unique preferences. The idea of overdifferentiation was mentioned most often, particularly for prenasalization, pre-clusivized nasals and nasalized vowels. It was determined in most of these cases that readers benefit from "seeing" the overdifferentiation, likely because of what they had learned from literacy in the official language in their school education.

There were some decisions made in one language community for which another community made a contrary decision. The two strategies mentioned in the answers to the questionnaire were marking nasalization on vowels and the orthograph chosen for the velar nasal. One team hadn't marked nasalization on their vowels initially but finally decided they preferred marking them. The reason given was that the speakers preferred seeing/reading the difference. Another team, while at first marking nasalization changed so as not to mark it. They made

this decision as a concession to writers, to make their job easier, but it no doubt gave readers more work. It was mentioned elsewhere in the questionnaire replies that writers often leave out certain letters—often those representing overdifferentiation—and that the readers are often able to read the materials anyway (Gizrra, among others). This was mentioned most often when talking about texting. It is also common among Westerners (at least in my home country, USA, and for many PNG people, messaging in Tok Pisin) texting and Facebooking to take shortcuts in their writing. Maybe these language communities in PNG have become familiar with others' texting and posting habits in other languages, and they learn to be brief (including underdifferentiation) based on other people doing the same in English. It could be a unique way of expressing oneself, even if others must work harder at deciphering the content.

A lot of angst was expressed when dealing with a language that is found to have more than 5 vowel phonemes. In some of these cases, it was decided that the vowel phonemes should be underdifferentiated. Dadibi has phonemic nasalization on all 5 vowels, but they choose not to write the nasalization. In other language communities, digraphs of vowel combinations or of the same vowel were used for a vowel phoneme of a similar quality. Ambulas, for example, uses <a> for the vowel phoneme /ɐ/ and <aa> for the /ɑ/, while Gapapaiwa uses <i> for /i/ and <ii> for /ɨ/. In an effort to keep words as short as possible, it is a common practice when using double letters to use them for the less common phoneme.

It was mentioned in one team that a previous orthographic influence was from Fiji, and they found it necessary to move from orthographs common in Fijian languages to others, since that social influence was no longer in effect here in PNG (Mussau-Emira). It was determined that it is more important to have letters that help speakers bridge from their mother tongue to the official language of English, so the letter <g> for the phoneme /ŋ/ and the letter <q> for the phoneme /ɣ/ were not helpful in today's linguistic climate.

8.4. The next question in the questionnaire for current language teams was about technology and texting using the orthographies they had supervised. The most common answer was that there are no special characters or diacritics in the orthography, so texting using one's mother tongue was no special challenge in that regard. The next most common response was that the speakers were experimenting with texting by not using any of the diacritics and "getting by" with the lack of differentiation. This includes using <ng> for texting the velar nasal phoneme, even when their orthograph for this phoneme is <ŋ>. Whether or not the local speakers were feeling successful about this enterprise was not mentioned. One community (Gizrra) that had only two diacritics (an acute and a diaeresis) were successfully texting without these diacritics, and

they were making themselves understood. I suppose the relative number of diacritics and special characters in the official orthography would determine the relative success of this endeavor in any language.

One enterprising community was using numbers to reduce the effort texting takes, especially when the words are long due to reduplication. For example, for the word <waiwaisana>, they were texting <wai2sana>.[6]

8.5. The final question in the questionnaire for current language projects was about the major influences on the orthography, whether mostly phonemic or more a result of language community input. It seemed that of all the stated responses in current teams, the influence was about equal between the linguistic/phonemic influence and community input. This pairs well with the literature on the issue, that often states that linguistic and non-linguistic forces are at work in language development in general. Of course, the orthography has everything to do with how a language looks to its readers and their perception of how their language looks to the outside world—through their writing system. A language community must feel confident that the orthography they are using to showcase their language is adequate and practical but also a personal expression of themselves through language. It's not just a string of sounds—it's my language.

Questionnaire responses also talked about wanting to have reading easier and/or the teaching of reading and writing to be easier. This seemed to be important among language communities that also wanted an easier bridge to reading and writing in the official language. This desire seems to point to a need for more of a phonemic influence on the orthography, which would be a pull in the other direction from many of the other influences mentioned in this paper.

8.1. Conclusion

Grenoble and Whaley (2004, p. 158) list five recommendations at the end of their chapter on orthographies that I think are representative of the

6. *Note by the Editor.* The convention of using a <2> to reduplicate the graphemes preceding it has been attested in Latin-script Malay: according to Haji Omar (1989, §10), "There are three types of reduplication in Malay: the reduplication of the first syllable of the root, the reduplication of the stem of a complex word, and the reduplication of the whole word, be it a simple or complex word. In the old spelling systems both in Malaysia and Indonesia, the first type of reduplication was spelt in toto, but the character <2> was used to indicate the reduplication of the second and third types. In the reduplication of the whole word, the character <2> was placed at the end of the word, for example, <rumah2> was read as *rumah-rumah* 'houses,' <makan2> as *makan-makan* 'to while away the time eating'."

SIL experience in PNG. The first item is that orthographies devised to be used in a language revitalization project should be focused primarily on utilizing an *alphabetic* system. This goes without saying in PNG, where bridging concerns are primary in the thoughts of language describers/documenters, the government and the language communities themselves. This was obvious in the responses to my questionnaire. The orthography of a language in this situation will be used by many semi-literate people, so it needs to be instructive, teaching and reinforcing a speaker's knowledge of the sounds of one's language, and s/he will pick that up most easily from an alphabetic orthography. This point is well-accepted in PNG, so it needs no further discussion here. All our SIL-PNG projects have adopted this stance.

The second characteristic of a successful orthography they mention is *learnability*, where the orthography helps and encourages the learner in any way possible. Motivation can easily be discouraged if the enterprise of reading seems too difficult. Languages in PNG being revitalized through the use of a new orthography do not have the advantage of a well-established national or official language, where learning to read and write offers its own rewards of being in touch with the greater world through books, movies, internet, etc. Local language learners must be encouraged through any means possible to learn to read and write their own mother tongue, and the orthography design can aid in this process. Our SIL-PNG teams have shown this concern in their answers to my questionnaire by showing a real desire for successful local level literacy, working together with the language communities to make sure alphabet choices reflect the community's wishes, including different and often competing ideologies.

The third point they make is that orthographies should be *phonemic* as much as possible. The meaningful sounds of the language should be evident in the orthography, particularly those items with a high functional load. Depending on the language, application of this principle could be in tension with Grenoble & Whaley's second point above. A language in which there are relatively more phonemes doesn't allow for a simple, easily learned alphabet and spelling rules. But a phonemic understanding and basis is a good starting point, and most of the simpler issues can be easily addressed with this point in mind from the beginning.

Point four speaks of *transparency* in that "spelling conventions should coincide with those of the language of wider communication wherever possible" (p. 159). This is the same "bridging" concern mentioned by SIL-PNG teams in responses to the questionnaire. Any benefit to having a unique system of reading and writing (in competition with the national language) is offset by the limitations on how it helps or hinders the literate person. If the orthography is transparent, the language learners can become literate in both their own language and the official language(s). The skill of reading and writing can transfer more easily

to another language that shares a mostly common alphabet. Many SIL-PNG teams have made this a priority in their orthography designs. They are using many of the digraphs we use in English, such as <ng> for the velar nasal.

Their last point refers to the *acceptability* of the orthography. A writing system is only successful if it is acceptable to the language communities that are motivated to learn it and use it. Our SIL-PNG teams showed this to be a constant concern in their orthographies and in their literacy programs. We need to continue consulting with the language communities, so that the factors that concern them get incorporated into the orthography and allow them to realize the language revitalization they desire.

References

Cahill, Michael (2014). "Non-linguistic Factors in Orthographies". In: *Publications in Language Use and Education*. Vol. 6: *Developing Orthographies for Unwritten Languages*. Ed. by Michael Cahill and Keren Rice. Dallas, TX: SIL International, pp. 9–25.

Cahill, Michael and Keren Rice (2014). *Developing Orthographies for Unwritten Languages*. Dallas, TX: SIL International.

Clifton, John M., ed. (1987). *Data Papers on Papua New Guinea Languages*. Vol. 33: *Studies in Melanesian Orthographies*. Ukarumpa: SIL.

Grenoble, Lenore A. and Lindsay J. Whaley (2004). "Orthography". In: *Saving Languages: An Introduction to Language Revitalization*. Cambridge: Cambridge University Press.

Gudschinsky, Sarah C. (1973). *A Manual of Literacy for Preliterate Peoples*. Ukarumpa: SIL.

Haji Omar, Asmah (1989). "The Malay Spelling Reform". In: *Journal of Simplified Spelling Society* 11, pp. 9–13.

Karan, Elke (2006). "Writing System Development and Reform: A Process". Master's thesis. Univ. of North Dakota.

———— (2014). *The ABD of Orthography Testing: Practical Guidelines*. Vol. 54. Dallas, TX: SIL.

Larsen, Robert E. (1977). *Multidialectal Orthographic and Lexical Adjustments for Orokaiva*. Vol. 21. Ukarumpa: SIL, pp. 343–348.

Litteral, R. and Susan Malone (1991). *The Sounds of Your Language*. Port Moresby: Department of Education.

Malone, Susan (2004). *Manual for Developing Literacy and Adult Education Programmes in Minority Language Communities*. Bangkok: UNESCO.

Petterson, Robbie and Debby Petterson (1994). "Failures and Successes in Literacy in Gulf Province Schools". In: *Conference of the Linguistic Society of Papua New Guinea, Madang*.

Pike, Kenneth L. (1947). *Phonemics: A Technique for Reducing Languages to Writing*. Vol. 3. Ann Arbor: University of Michigan Publications.
Roberts, John R. (2002). *Orthography Reform in Amele*. Ukarumpa: SIL.
Robinson, Clinton and Karl Gadelii (2003). *Writing Unwritten Languages: A Guide to the Process*. Paris: UNESCO. https://pdfs.semanticscholar.org/bda5/26059b80037af03b0eaf4fec84ab696bf114.pdf.
Sarvasy, Hannah and Diana Forker, eds. (2018). *Word Hunters*. Amsterdam: John Benjamins.
Schreyer, Christine (2017). "Reflections on the Kala Biŋatuwã, a Three-Year-Old Alphabet from Papua New Guinea". In: *Creating Orthographies for Endangered Languages*. Ed. by Mari C. Jones and Damien Mooney. Cambridge: Cambridge University Press.
"SIL Papua New Guinea. Annual Report" (2017).
Simons, Gary (1977). *Principles of Multidialectal Orthography Design*. Vol. 21. Ukarumpa: SIL, pp. 325–342.
Smalley, William A. et al. (1964). *Orthography Studies: Articles on New Writing Systems*. London: United Bible Societies.
Snyder, David (1994). "Orthographic Symbols in Papua New Guinea Languages". In: *Conference of the Linguistic Society of Papua New Guinea, Madang*.
Spilioti, Tereza (2009). "Graphemic Representation of Text Messaging: Alphabet Choice and Code Switches in Greek SMS". In: *Pragmatics* 19, pp. 393–412.
Tomokiyo, Laura Mayfield (2018). "Orthography Development". Presentation https://slideplayer.com/slide/13082277/.
Venezky, Richard L. (2003). "In Search of the Perfect Orthography". In: *Written Language and Literacy* 7, pp. 139–163.
Whitehead, Carl R. (2004). "A Reference Grammar of Menya, an Angan Language of Papua New Guinea". PhD thesis. University of Manitoba.