15: Rapid Grammar Collection:
Language communities owning the orthography development process

Timothy Stirtz

Abstract: It is ideal when language communities decide their writing system, well-informed by factors affecting these decisions, and consistently use that writing system in producing reading materials that benefit the community. However, deciding and using a writing system are challenging in languages with linguistic complexities such as a high functional load of tone or morphemes that can legitimately be written either bound or free. This chapter describes a method to achieve these aims for languages with such complexities. From 2009-2017, the author facilitated workshops in South Sudan for ten language communities, working with one language at a time. In these, representatives of the language community collectively helped document the sound patterns and structures of their language to agree on a tentative writing system. In the weeks following each workshop, the word lists, narrative texts, documented grammar, and spelling rules agreed upon were made into a beginning dictionary and other reference tools for the language writers. Story books and other drafted reading materials were checked according to the agreed-upon writing system and revised as needed. This process trained writers to consistently use their writing system and empowered the language community to own the process of developing their written language.

Keywords: Nilo-Saharan languages, participatory method, morphosyntax, orthography development, consistent writing, language discovery, grammar book

1 Introduction

This paper describes the Rapid Grammar Collection (RGC) approach (Stirtz 2015) to language research. This participatory method typically involves several two-week workshops for representatives of a single language community in the beginning stages of language development, that is, deciding a writing system and producing reading materials. The aim is to quickly collect enough data to make informed decisions, not just regarding the alphabet, but also word breaks, tone representation, sound changes at morpheme boundaries, and natural constructions in the language. RGC is particularly useful in languages with complexities such as a high functional load of tone, numerous sound changes at morpheme boundaries, and morphemes that can arguably be written as either bound or free.

Among other activities, participants sort words on slips of paper according to vowel quality and tone melodies, discover the sound changes at morpheme boundaries using elicitation frames taken from texts, and decide how each morpheme functions in natural texts. Depending on the areas of complexity in a language, one or two workshops are needed for phonology, morphology and beginning syntax, and an additional workshop is needed to learn discourse analysis. After each workshop, the writing decisions and collected data can be described in reference tools for the language developers of the community, including a beginning dictionary and books presenting basic phonology and morphosyntax. These tools are useful reference works to remind various stakeholders of community decisions about writing but are often not enough for them to become consistent in following their writing decisions. For this, a training process can help them learn to apply writing rules as they draft each new written material.

Thus, the goal of the RGC approach is not just a collection of data or a way for language communities to make informed decisions about writing their language. It is also a way for

community representatives to learn to use those decisions in producing reading materials benefiting the community, and in this way more fully own the language development process.

I begin by describing the language contexts in which the RGC approach was used (§2), then explain the purpose for this approach (§3), the workshop activities and outputs specific to this approach (§§4,5), the checking process that trains language developers to use their writing choices (§6), a case study for one specific language community (§7), and finally the strengths and weaknesses of this approach (§8).

2 Language context

The RGC approach was used with ten language communities in South Sudan and Sudan from 2009-2017. These language communities are what many might consider small, their number of speakers ranging from 4,000 to 117,000.[1] The representatives who attended the workshops and became developers of their language were from a range of education backgrounds, some with secondary or post-secondary education, and others without having completed their secondary education.

Because of its history of language work in the region since 1977, SIL was approached by representatives of these ten communities, requesting support in community-based language development. SIL formed a partnership with each community through the local language committee. When a committee did not yet exist, communities were encouraged to establish one. For each of the ten languages, the language committee was involved in language planning and development decisions, such as which dialect to write, the alphabet choices, which materials to develop, which training workshops should be offered, and especially, who should attend. Working together with literacy experts, members of the language community then developed literacy materials, taught literacy classes, and took other steps towards language development.

Via SIL, I was assigned to assist with linguistic research to inform the writing system choices as well as identify natural constructions in the language. When I began working with these communities, most languages had a tentative alphabet under discussion which was inconsistently used, few if any beginning reading materials, and in some cases only a handful of people who could read them.

As is common in Nilo-Saharan languages, several of the ten languages mentioned[2] have extensive agglutinative or fusional morphology, as well as a high number of sound changes at morpheme boundaries, including many that are irregular (i.e. Table 1 of §5.3). Consistently representing these changes in the orthography can be extremely challenging. In several other languages, those more isolating,[3] there is often great confusion as to what should constitute a word. So, it is common for novice writers to sometimes attach morphemes, and at other times treat them as independent words. In addition, nine of the ten languages distinguish lexical or

---

[1] Eberhard et al. (2023) lists the populations and ISO 639-3 codes of the ten languages as follows. Pseudonyms are used for three languages because of the sensitive nature of the work described in this or another paper.

Belanda Bor [bxb]	26,000	Reel [atu]	116,000
Bongo [bot]	21,000	Kappa (pseudonym)	About 100,000
Beli [blm]	65,000	Omega (pseudonym)	About 30,000
Lopit [lpx]	117,000	Beta (pseudonym)	About 80,000
Mandari [mqu]	70,000	Narim [loh]	3,620 (1984)

[2] Those with extensive agglutinative morphology include Lopit, Mandari, Kappa, Omega, Beta and Narim. Reel mostly has fusional morphology.

[3] Those languages more isolating include Belanda Bor, Beli and Bongo.

grammatical meaning by tone, and eight of these have a significant enough functional load that tone needs to be at least partially represented in writing for meaning to be correctly understood.

3 Purpose

Among other factors, in order for the communities to be owners of their language development process, 1) they should be the ones to decide their writing system—including letter-sound choices, tone representation, and writing rules for word breaks and morphological sound changes, 2) their decisions should ideally be informed by the language patterns and constructions, and 3) there should ideally be community members who consistently use their writing system in producing reading materials. The purpose of the RGC approach is to achieve these three aims for languages with the above-mentioned complexities. §3.1 addresses how the workshops aimed to meet the first two. §3.2 addresses the third.

3.1 Informed writing decisions

As Lew (2019) claims, it is not enough for language communities to make writing decisions about their language; they must be empowered to make "*well-informed* writing decisions," and I would add, by examining the constructions of their language. In my experience, language communities often make inconsistent and conflicting choices when they consider the writing of one word or construction in isolation, but they usually make informed decisions when they see the big picture, having all relevant patterns in mind. And this is what can happen in RGC workshops.

   For example, the Bongo language community representatives first chose to connect the subject pronoun **bi** to the verb as in <**bi**lehi je> '**he**.worshiped us.' Later when they worked through many other constructions in their language in a workshop, they discovered that the conjunction **di** 'when' can separate this pronoun from the verb (<bi **di** lehi je> '**when** he worshiped us'). With growing awareness of this and other similar constructions learned with the wide view of the language (such as <**bi**lehi je> 'our worship' (*bi-* action verbal noun), <**'bi**lehi> 'worshiper' (*'bi-* person verbal noun)), they decided it was better to separate the subject as in <**bi** lehi je> '**he** worshiped us.'

   For workshop participants to be able to keep in mind the big picture of their language when making informed writing decisions, it helps to be led through many of the language constructions in a relatively short period of time—such as in two-week segments, rather than months or years. Over longer periods of analysis, there is a risk that certain relevant constructions will be forgotten when the time comes to make decisions. Besides, many participants find it challenging to be away from their regular responsibilities for more than two weeks at a time.

   An RGC approach can give community representatives a comprehensive view of the language from the beginning of language development. It typically covers about 80% of the phonology, morphology, syntax and discourse analysis[4] in just 4-6 weeks, divided into two-week segments over 8-18 months. Usually, enough data can be collected for beginning language development. It may take a linguist 4-6 months to write about the grammar learned in these workshops, but the time that community representatives need to be away from other responsibilities is relatively short in comparison. Workshop participants make well-informed

---

[4] 80% of the phonology, morphology, syntax and discourse analysis is a rough estimate. Often, even more than 80% of the language forms are collected, even if the functions are not fully understood.

writing choices as they learn the sound patterns and constructions throughout the language. After the workshops, they test these decisions for writability by producing written materials, which are tested for readability and acceptance within the community.

3.2 Consistent use of the writing system

The value of consistency in written language is increasingly debated in discussions of language variation, language and identity, code-switching, code-mixing, and use of language in digital space. However, these discussions often involve differences between language varieties or with a vernacular language in comparison with a dominant language, rather than consistency in the same language. In addition, some of these discussions involve creative alternatives used in social media rather than in publications and education materials (Deumert & Lexander 2013, Deumert et al. 2019). The writing inconsistencies that are discussed in this section are those that are unintentional, differing from the written standard previously agreed upon by the community for a particular language, and occurring in draft materials later to be published, such as for teaching reading or for use in schools or churches.

Even a well-chosen writing system by a language community can be worth little if they do not learn to consistently use that writing system. In a case study of the same language communities, Stirtz (2018) documents inconsistencies in the writing of nine of them after agreeing upon their writing systems in RGC workshops. In only one of the nine were there those who consistently used their writing system without further training. In the other eight, the language representatives either forgot their writing system, did not yet understand how to use it, or for other reasons did not consistently apply it when developing written materials. On average for these nine languages, there was one inconsistency for every ten words (Stirtz 2018:6-7). The most common types of inconsistencies involved similar vowels (such as in Bongo <afonj> vs. <afɔnj> 'try, test'), tone representation (Beli <dɔ-gbe⁵> [d̪ɔ̄ g͡bɛ̀] vs. <dɔ gbe> [d̪ɔ̀ g͡bɛ̀] 'on the son, head of the son'), word breaks (Bongo <bilehi ji> vs. <bi lehi ji> 'our worshiping, he worships us'), sound changes at morpheme boundaries (Kappa <wirí> vs. <wirrí⁶> 'slaughter, in order to slaughter'), or similar consonants (Belanda Bor <gbɛl> vs. <ngbɛl> 'only'). 21% of these inconsistencies, or 1 in 42 words, could not be found by computer tools since they involve similarly spelled words or constructions that are both correct in different sentences. 3.3% of these inconsistences, or 1 in 273 words, affect the meaning in that both words could equally make sense and be grammatical in the given context, and the meaning of the incorrectly written word is most often assumed. Thus, the extent of writing inconsistency in these languages was great and the potential for negative effects from this inconsistency was significant.

Inconsistent writing hinders reading fluency and comprehension, devalues the reading materials, and produces inconsistent writers as readers see what appears to be an acceptable example to follow (Smalley 1964, Karen 2006, Roberts & Walter 2021). For good reasons, there are few publishing companies willing to publish materials in any major world language without the materials first being edited according to the standards of that language. When language communities want their books to have the same quality as that of major world languages, they can be empowered to make well-informed writing decisions and be trained to use them.

---

[5] As explained in Stirtz (2014:24), the [blm] orthography uses a hyphen as alternative grammatical marking to distinguish prepositions (such as *dɔ-* 'on, above' with Mid tone) from the body-part nouns they are derived from (such as *dɔ* 'head' with Low tone).
[6] The incompletive verb with subject pronoun suffix /wir-i/ 'he slaughters' (which surfaces as [wīrí]) is often confused with the subjunctive form /wir-da-i/ 'in order for him to slaughter' (which surfaces as [wírí]).

The case study of Stirtz (2018) also demonstrates significant improvement in writing consistency by representatives of the nine language communities when given ongoing training. On average, the inconsistencies were reduced by 70% in two months of training over several years. By contrast, there was little if any improvement in writing consistency in a tenth language community that did not receive this ongoing training. The training given to the nine language teams merely involved an editor checking each word of each book in context, and for each inconsistency found, reminding the authors of their agreed-upon writing system. Through this process, there were occasional changes made to the initial writing decisions when the community was in agreement, and then all materials were revised accordingly. The initial writing decisions, changes to writing decisions, as well as all revisions to books in application of the writing decisions, were all at the discretion of the community members. Although the editor gave feedback and training, the communities had the final say in all language decisions.

4 Workshop prerequisites

Before the first workshop in the series, the language community representatives completed a translated word list of at least 1,000 words in a tentative orthography, often using the African word list of Snider and Roberts (2004) based on semantic domains. They also recorded, transcribed, and gave English glosses on eight to ten narrative texts according to a list of instructions (Stirtz 2015:54) so that those not speaking the language could understand the texts. Only one or two of the ten language communities needed help understanding the English word list or with glossing the texts. Although the texts often had inconsistent spelling and vague or inaccurate glossing, having this data from the beginning saved a great deal of time during the workshop, and gave the facilitator a head start in analysis. Only after receiving this data was the first workshop scheduled for a given community. Then, usually six representatives were invited to the workshop, all chosen by the community.

Because each workshop was only for a single language community, there was little need for training participants in language features in general (e.g. phonetics and typological linguistics), but only in an awareness of the features of their language. Thus, there were no lectures, only brief explanations of how to do each new activity and its purpose. In this way, the entire two weeks was reserved for mentally walking through displays of language patterns and constructions and then deciding the writing system based on what we found. Although much can be known about a language from the prepared word list and texts, or prior research of that language or related languages (if available), there were always surprises each day as the participants confirmed and rejected various insights. With no lesson plans and only a general list of features to analyze (noun and verb phonology; noun and verb morphology; syntax of particles, pronouns, conjunctions, and other parts of speech, etc.), the facilitator's biggest challenge was to continually engage the participants in activities producing data that would enlighten the choice of the next data to elicit, all the while not really knowing what would be discovered.

5 Workshop activities

There were three main activities used in RGC workshops: comparing the sounds of words on cards, learning the use of morphemes and constructions after marking them in texts, and determining various paradigms using elicitation frames taken from texts. Although each of these was more in focus at certain times during the workshop series, they can each be used at any time. In fact, regularly switching from one activity to another often helped participants

stay engaged. Besides the description in this section, Stirtz (2015) recommends other steps for how to facilitate RGC workshops.

5.1 Cards

Most of the first week of the first workshop was spent comparing the sounds of words on slips of paper or "cards," much the same as described by Kutsch Lojenga (1996; Chapter 9 this volume). The words were sorted according to word category, syllable patterns (CVCV, CVC, etc.), and other features potentially affecting the environment of the sound we wanted to focus on. We began with the largest sorted pile of nouns, and later the largest pile of verbs, to have the largest sets of words to compare in the beginning of the activity, before proceeding to smaller piles of nouns and verbs and eventually other parts of speech. Participants took turns reading a pile of cards and all decided how each word should be placed, according to whether the sound in focus was the same or different than the rest of the words in that particular pile. For example, there may be a pile of CVCV nouns, all initially written with the vowel <o> in both syllables, and the participants would decide if any should be placed in a separate pile to be written with a different vowel. We first compared the sounds of vowels, then any hard-to-distinguish consonants, and finally tone. After re-sorting certain words several times, the number of piles for vowels usually determined quite conclusively the number of vowel symbols that should be written in the orthography, and similarly for consonants. With tone, however, although the resulting piles represented the number of underlying tones (or tone melodies), there were additional factors to be considered before deciding what needed to be represented in writing. Often, many of the symbols used in the tentative orthography were already adequate for representing the sounds that distinguish meaning. However, when there was a need for something else, we discussed the advantages and drawbacks of various symbol options, usually those of related or neighboring languages.

One notable difference from the way Kutsch Lojenga (1996) describes this activity is that the facilitator wrote the words on cards ahead of time from the prepared word lists. This helped to save time during the workshop, and enabled participants to read the words more easily all in the same carefully written print. Another difference was that the words were re-sorted as needed before each new workshop session, again to save time. Often words were written on whiteboards in sorted order when comparing vowels[7] so that everyone could be engaged in the comparison process.

5.2 Texts

At least once a day during the first workshop, we also learned about the use of morphemes and constructions from the prepared narrative texts. Before that, however, the spelling, punctuation, glosses, translation, and other details of the texts needed to be revised. So, our first step was to read through a text several times, correcting it for any changes the participants suggested. Next, we began marking the text with different colors and shapes (underlines, overlines, circles, boxes, etc.) according to the various word categories, particles, bound morphemes, or constructions we wanted to learn more about on that day. Once a certain conjunction was marked throughout the text in a uniform way (i.e. with the same color and shape around it), it was much easier to find, read in context, and discuss what use it had in

---

[7] It was productive to write words sorted for vowels on the whiteboards for participants to read and compare because only a few changed from the way the vowel was originally written. However, this did not work as well with tone, since there were many changes even after the words were first sorted, and it was tedious to rewrite them on the board in a different group each time.

each occurrence. In a similar way we marked various verb morphemes, demonstratives, relative clauses, and many other challenging features.[8] Because there were usually 40 pages of interlinearized texts in total between the 8-10 narratives, most of the morphemes and constructions of the language were well-represented. As we discussed the use of each morpheme or construction in context, the facilitator wrote a few comments summarizing what was learned on whiteboards, along with an example sentence from the text, often using the participants' wording to describe it. Usually the participants also wrote down these summaries and examples in their personal notebooks.

During the workshop for discourse analysis, there was even more extensive marking and analyzing of texts with less focus on identifying morphosyntactic constructions. For example, various colors were used for tracking different participants, regardless of whether they were referenced by a noun phrase, bound morpheme, or null reference. Verb forms commonly used for mainline (storyline) clauses were marked with a different color than verb forms used for background information. In addition to marking and analyzing texts, the facilitator also developed tables to chart discourse features throughout the story, such as for paragraph/scene introducers or participant reference.

5.3 Elicitation frames

Often during the second week of the first workshop and throughout the second workshop, we elicited much data on whiteboards with the help of elicitation frames taken from the texts, that is, a fixed sentence, clause or phrase where one word or morpheme was substituted with others of that type. This was very helpful in discovering the remaining major constructions. Although generating a particular form of a noun or verb in isolation can be quite challenging, starting with a sentence from a familiar narrative containing one form often enabled us to collect variations of that form that may not have been discovered using other methods. Out of the three types of activities, these frame-based activities were used the most and also varied the most from one language to another.

For most languages, we began by writing pronoun paradigms with the same verb (I ate, you ate, he ate, etc.), followed by pronoun paradigms in various cases (subject, object, possessive, etc.). To help participants think of all the pronouns and to verify how each was used, we began with a sentence or clause from the texts with a known pronoun and grammatical case and replaced it with pronouns for other persons in that case. We used different clauses from the texts to generate pronouns in other grammatical cases. Writing the full set of pronouns together helped participants to quickly see the similarities and differences in pronominal forms and often the best way to write all of them, either bound or free, or with any other distinctions. When there was a disagreement as to how to write them, we discussed the benefits and drawbacks of each option, considering what might be easiest for the language community to accept, read, write or use with technology, usually in that order of priority. When all possible pronouns were documented in tables with agreed-upon spelling, we used the same process for demonstratives, prepositions and other parts of speech that have small sets of words, again relying on frames from texts to generate and confirm the use of each.

In several of the languages in the workshops, there are numerous morphemes with numerous sound changes (allomorphs) (i.e. the *-ong* demonstrative suffix of Table 1). These sometimes made it challenging to determine which morpheme was being used in any given sentence of a text until the possible sound changes were first determined. For these languages

---

[8] Although I always marked texts extensively before the workshops to have a better understanding of the language, I was often surprised by many of the participants' comments which confirmed, rejected, or brought new insights to the data.

with extensive agglutinative or fusional morphology, we used the above-mentioned process to generate and confirm all possible sound changes at the morpheme boundaries in nouns and verbs. Starting with a sentence taken from a text, such as in Table 1, we substituted different noun (or verb) lexemes, representing all known language features that might cause the sound change, such as nouns with different root-final consonants for a consonant-initial suffix, or nouns with various root tone melodies for a suffix with a change in tone.[9] Knowing the possible sound changes helped us determine which surface form belonged to which morpheme in the texts, and therefore how each might need to be represented adequately in writing for fluent reading.

Table 1: Example elicitation frame exercise for Caning [shj] with suffix –*ong* 'this' (Stirtz 2015:19)

Bonog, tong _____**ong**.			'Friend, bring **this** _____'	
Singular	**Plural**	**Singular-*this***	**Plural-*these***	
abad	abad**iny**	abad**ong**	abadiny**gong**	'stupid person'
säxäd	säxäd**iny**	säxäd**ong**	säxädiny**gong**	'egg yolk'
amkadad	amkadad**iny**	amkadat**song**	amkadadiny**gong**	'chisel'
led	le**tede**	led**ong**	letede**gong**	'dance type'
sud	su**tudi**	sud**ong**	sutudi**gong**	'mound, hump'
sad	sa**tu**	sa**tong**	satu**gong**	'bowl, calabash'

For other languages in the workshops, it is word order changes that indicate various functions (tense, aspect, mood, verb derivation, information flow, etc.), or particles combine with various clauses or with other particles to indicate these distinctions. For these languages that were more isolating, we used a similar process to test clauses with various possible word orders and combinations. Beginning with a sentence taken from a text, we changed the word order, substituted a different particle, or otherwise changed the sentence until we were satisfied that we had tested all relevant variations and combinations, especially those represented by other sentences in the texts. For each possible variation, we noted the implications for word breaks and tried to determine the construction's function, which would help us later when we came to analyze discourse for natural translation.

By the end of the first workshop, there were usually at least 1,000 words that participants had agreed upon for spelling and often 5-10 rules that participants agreed should be followed in writing their language. Before leaving, the participants practiced these writing conventions by revising portions of their texts accordingly.

6 Workshop reference tool outputs

There were three main written reference outputs of RGC workshops: a beginning dictionary, a consonant and vowel book, and a grammar book. All of these were completed by the workshop facilitator in the weeks following a workshop from the data collected, rather than by the participants during the workshop itself, although portions of these books were checked by speakers in following workshops. As language developers draft reading materials in the language, these reference tools can remind them of how they chose to write. Appendix A of Stirtz (2015) gives links to several examples of each of these books.

---

[9] Carefully studying the word list and texts ahead of time to some extent helped the facilitator find a representation of the possible sound changes in words and which features may be causing them.

The beginning dictionaries document 1,000 words from each workshop with spelling and word category agreed upon. They also include a brief introduction explaining how to read entries and a list of alphabet symbols and abbreviations, as well as select grammatical information. Each was compiled in Fieldworks (FLEx)[10] and uploaded to webonary.org, a website for dictionaries with searchable entries, so that communities may have easy access to look up entries.

The consonant and vowel books present the alphabet symbols, their combinations and restrictions in words, and especially a comparison of any symbols that could easily be confused, such as similar vowels, long and short consonants, and tonal contrasts. Lists of words with similar sounds are compared so that as language developers read each list, the sound in focus is reinforced with that symbol. The lists also provide a way to test new words against the relevant sound, helping language developers become consistent writers as they decide which letter should be written in each case.

The grammar books present the morphemes and constructions collected in the workshops, using brief, simple English explanation in short incremental lessons.[11] The writing rules decided upon during the workshops are listed at the beginning of these books, with page numbers showing where they are further explained with example data. Each lesson presents one concept, such as a part of speech or morpheme, first giving an example sentence taken from the texts to show its use, and then paradigms or data in tables to show how the relevant constructions and sound changes are written, such as in Figure 1. Each lesson has exercises with sentences taken from the texts where language developers can learn to recognize and

---

The suffix –**ong** with beginning vowel is the most common pointing near suffix on singular nouns. However, other singular nouns have pointing near suffixes with a beginning consonant, such as –**t-,-d- -s-,-c-,-z-,-n-**. Nouns with the same final letter in the root (such as **d**) sometimes attach suffixes with different beginning consonant (such as **Ø, -t-, -s-**, where **Ø** means without a beginning consonant).

Singular Noun	Suffix		Pointing near	
abad	**-ong**	*stupid person*	Abad**ong** kadeläng.	*This stupid person fell.*
sad	**-tong**	*bowl*	Sa**tong** kadeläng.	*This bowl fell.*
bul	**-dong**	*drum*	Bul**dong** kadeläng.	*This drum fell.*
amkadad	**-song**	*chisel*	Amkadat**song** kadeläng.	*This chisel fell.*
bac	**-cong**	*upper arm*	Bac**cong** kadeläng.	*This upper arm fell.*
nyang	**-zong**	*crocodile*	Nyang**zong** kadeläng.	*This crocodile fell.*
ya	**-nong**	*meat*	Ya**nong** kadeläng.	*This meat fell.*

Figure 1: Example partial lesson in the Caning Grammar Book (Alfira et al. 2017:55)

understand the use of each morpheme or construction. The full interlinearized texts are included at the back of these books and all language data is written in the agreed-upon writing system. In this way, the language developers not only learn the language patterns and constructions but also have numerous examples of how to write them.

---

[10] See https://software.sil.org/fieldworks/ for further information.

[11] Although language developers from among the communities can often read simple English, it would be ideal for these books to be written in their first language. The facilitator wrote them in simple English with the hope that someday the language developers would translate or rewrite them in their languages.

7 Checking materials

In the months following the first or second RGC workshops, community members, sometimes with assistance of technical experts, also produced beginning reading materials, such as an alphabet book, primer, story books, and translated materials. Although these books were drafted with the symbols chosen in the workshops, they contained many inconsistencies in following the writing rules—specifically those discussed in §3.2: similar vowels, tone representation, word breaks, sound changes at morpheme boundaries, and similar consonants.

It is not surprising that language community representatives initially had difficulty following their writing choices as they drafted new materials. Although many of them had completed their secondary education in English and were fluent speakers of their language, writing in one's own language is a high-level skill, usually only developed over years of training (Bear et al. 2012). Although the workshops and reference tools are a major step towards providing this training, the greatest challenge is learning to apply the writing rules in natural text, something that takes extensive guided practice and was successfully learned through a checking process.

For eight of the ten language communities,[12] the workshop facilitator checked each word of each book as it was first drafted, and later each revised portion. For each word or construction not written according to what was previously agreed upon, the facilitator made a written comment, often including the page in the reference tool where the writing rule or spelling could be found. The language developers, who had the final say in all writing issues of their language, revised these books according to about 90% of the suggestions. In a few cases, the rule or spelling was not intuitive enough or needed to be changed, giving the opportunity to discuss other options. Over time, language developers developed skills in consistency checking themselves, and eventually did not need assistance from the workshop facilitator. On average, this happened after about two months of full time checking for each team over several years. Stirtz (2018) gives other best practice steps for this checking process.

8 A case study

In 2010, representatives of the Beli (ISO 639-3 [blm]) language community of South Sudan finished translating the gospel of Luke at the request of Beli churches. Besides a brief spelling guide, this was the first published book in the language. There was little linguistic analysis done of Beli and the tentative orthography was untested. It is no surprise, therefore, that the community found it difficult to read. I then became involved with linguistic analysis in hopes of determining the source of the reading difficulty. Using the RGC approach we compared the sounds of hundreds of words and examined numerous grammatical constructions in several non-translated narrative texts. We learned that nearly one in five words were in a minimal pair, either for lexical or grammatical tone (Stirtz 2014), without being represented in writing. Our assumption was that this huge function load for tone was likely the main cause of the reading difficulty. By the end of the workshop, the participants chose four writing rules that distinguished 75% of these ambiguities. Next, they took a narrative text written according to these rules and tested it with the community. Although the writing rules were the simplest way to make reading possible, I suspected that the Beli would not easily learn them without

---

[12] One of the language development teams had an excellent native-speaker consistency checker from the beginning.

additional training. I therefore wrote a grammar book[13] to describe the language patterns and constructions that informed their writing choices. After several months of practice applying these agreed-upon choices to written materials over a period of years, the Beli writers for the translation project are now using these orthographic conventions with practically no errors, and they report that those in literacy classes and many in churches are reading with understanding.

9 Strength and weaknesses

In deciding whether to use the RGC approach, its strengths and weaknesses should first be carefully considered.

One major strength is that the approach has the potential for collecting extensive, quality data relatively quickly. Another is that it empowers language community representatives to be well-informed of language patterns and constructions so they can make choices for at least 80% of their writing system. The beginning dictionary, consonant and vowel book, and grammar book remind language developers how they chose to write, not just in spelling and writing rules, but through numerous data examples to model the writing system. The process of checking drafted reading materials can effectively train language developers to write their language consistently in about two months, rather than in several years, as might otherwise be needed. Finally, new language developers, instead of having to experience a workshop themselves, can often learn the skill of writing their language through the reference materials and checking process.

However, there are significant weaknesses of the RGC approach. The workshop sessions are intense, tiring, and monotonous, with many direct and repeated questions asked of participants. Because there are no workshop materials and only a general outline of topics to follow, the workshop management is only as good as the person leading it, so experience and being comfortable with ambiguity go a long way. Although all writing decisions are made by the language communities, the workshop agenda and activities are set by the facilitator. Although there are only 4-6 weeks needed for the workshops, 4-6 months are needed to complete the reference tools afterwards, and if time is not set aside for this extensive work, it will not happen. When the reference tools are produced by outsiders, the community representatives are less involved in the language description process and do not have as much ownership of them as is the case with other methods, such as the grammar sketches of Kroeger's method (Chapter 21 this volume).

Although I have not found ways to minimize these weaknesses while keeping these strengths, others using this method may be more successful. It is important to adapt the method as much as needed to one's own personality and language situation, so that the shared goals of researchers and language community members can be met.

10 Conclusion

The RGC approach, used with 10 language communities in recent years in South Sudan, provides a way for community representatives to make informed writing decisions and learn to use them. In a series of two-week workshops, participants sort words on cards according to vowel, consonant and tonal contrasts, learn the function of each word category and morpheme

---

[13] In addition to the *'Beli Grammar Book* www.sil.org/resources/archives/58712, there is also the *Reading and Writing 'Beli Book 2* www.sil.org/resources/archives/58704 and the *'Beli-English Dictionary* www.sil.org/resources/archives/54225.

in natural texts, and choose how to write tone, word breaks, and sound changes at word boundaries in elicitation frames taken from texts. After each workshop, the writing decisions, as well as the language patterns and constructions that inform these decisions, are described in reference books, written to be more accessible to language community members involved in language development—a beginning dictionary, consonant and vowel book, and grammar book. The paradigms and other data strategically collected in the workshops become tables and examples in these books, showing language developers the patterns that their writing choices are based upon. Not only are all the writing decisions listed, but also the agreed-upon way to write each construction. Finally, as each new reading material is drafted, each word is checked according to the agreed-upon way of writing, and language developers learn to write well in about two months rather than over several years. The approach has several weaknesses, including only minimal involvement of community representatives in the description of the language. However, community members become owners of their language development process, not just in making writing decisions for their language, but making *informed* decisions and using them to produce readable materials that benefit their communities.

References

Alfira, David Abi, Timothy Kuku Kafi, Hassan Kuwa Kaki, Ali Alaliim Hasan, Anjo Kuku Anjo, Dayan Kuku Jas and Sadik Kafi Sarukh. 2017. *Caning Grammar Book*. Caning Language Committee and Sudan Workshop Program. https://www.webonary.org/caning/files/CaningGrammarBookNov17.pdf (accessed 9 Apr 2023).

Bear, Donal R, Marcia Invernizzi, Shane Templeton and Francine Johnstone. 2012. *Words Their Way: Word Study for Phonics, Vocabulary, and Spelling Instruction*. Pearson.

Deumert, Ana and Kristin Vold Lexander. 2013. Texting Africa: Writing as Performance. *Journal of Sociolinguistics*. https://doi.org/10.1111/josl.12043 (accessed 9 Apr 2023).

Deumert, Ana, Ivan Panović, Dorothy Agyepong and David Barasa. 2019. African Languages and Mobile Communication – Between Constraint and Creativity. In H. Ekkehard Wolf (Ed.), *Cambridge Handbook of African Linguistics*, 555-574. Cambridge University Press. https://doi.org/10.1017/9781108283991.019 (accessed 9 Apr 2023).

Eberhard, David M., Gary F. Simons and Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. http://www.ethnologue.com (accessed 10 May 2023).

Karan, Elke. 2006. *Writing system development and reform: A process*. Theses and Dissertations. 3004. https://commons.und.edu/theses/3004 (accessed 9 Apr 2023).

Kutsch Lojenga, Constance. 1996. Participatory research in linguistics. *Notes on Linguistics* 73. 13-27. https://www.sil.org/resources/archives/4087 (accessed 7 May 2023).

Lew, Sigrid. 2019. Proposing a Facilitated Participatory Approach for Southeast Asian Minority Language Orthography Design. *Journal of the Southeast Asian Linguistics Society* 12 (1): ixvii–ixxv. http://hdl.handle.net/10524/52451 (accessed 14 May 2023).

Roberts, John and Stephen Walter. 2021. Tone Orthography and Literacy: The Voice of Evidence in 10 Niger-Congo Languages. *Studies in Written Language and Literacy* 18. John Benjamins Publishing Co.

Smalley, William. 1964. How Shall I Write This Language? Orthography Studies: Articles on New Writing Systems. *Helps for Translators* 6. London: United Bible Societies, 31–52.

Snider, Keith L. and James S. Roberts. 2006. SIL comparative African wordlist (SILCAWL). *SIL Electronic Working Papers* 65. https://www.sil.org/resources/archives/7882 (accessed 13 May 2023).

Stirtz, Timothy. 2014. 'Bëlï Phonology, Tone and Orthography. *SIL Electronic Working Papers* 111. https://www.sil.org/resources/archives/56015 (accessed 9 Apr 2023).

Stirtz, Timothy. 2015. Rapid Grammar Collection as an Approach to Language Development. SIL *Electronic Working Papers* 121. http://www.sil.org/resources/publications/entry/62483 (accessed 9 Apr 2023).

Stirtz, Timothy. 2018. Linguistic Checks: The Missing Link in Modern Language Development. *SIL Electronic Working Papers.* https://www.sil.org/resources/publications/entry/76614 (accessed 9 Apr 2023).