Single-event Rapid Word Collection workshops: Efficient, effective, empowering

Brenda H. Boerger
SIL International

Verna Stutzman
SIL International

In this paper we describe single-event Rapid Word Collection (RWC) workshop results in 12 languages, and compare these results to fieldwork lexicons collected by other means. We show that this methodology of collecting words by semantic domain by community engagement leads to obtaining more words in less time than conventional collection methods. Factors contributing to high and low net word senses are summarized, addressed, and suggestions given for increasing effectiveness of the RWC procedures. Relevant points are illustrated in detail using a 2015 Natügu [ntu] RWC workshop in the Solomon Islands. We conclude that the advantages of the single-event RWC workshop strategy warrant recommending it as best practice in lexicographic fieldwork for minority languages.

1. Introduction[1] Descriptive linguists have long had the tripartite goal of producing a grammar, text collection, and dictionary in the languages in which they work. While the focus of this article is on the word collection phase of dictionary production, it is good to review briefly why dictionary work is important, especially in minority and endangered languages. First of all, besides merely preserving language and culture data for posterity, dictionary work can also be a conduit for beginning to understand other cultures. This is particularly true when producing a bilingual or trilingual dictionary with the glosses in a language of wider communication (LWC). Similarly, a dictionary can also serve as a language development tool, contributing to growing competencies in literacy and taking its place in the schools of the community.

It has been observed that concomitant with language development comes an increase in the status and value of the language, on the one hand, and the self-image of its speakers on the other (Ostler 2003:176; Boerger 2015:152). For the scholar-fieldworker, an extensive wordlist like the one collected in a Rapid Word Collection (RWC) workshop helps inform the texts that are collected, giving a head start on text glossing and comprehension at the outset of a project.

---

[1]The authors are thankful for the encouragement of Gary Simons, SIL's CRO, to pursue this publication and for the constructive comments of Paul Unger, who read an earlier draft. All errors are the responsibility of the authors. This research was conducted in part through funding provided by a Documenting Endangered Languages Fellowship from the US National Endowment for the Humanities and the National Science Foundation during 2015–2016.

No matter what purpose one might have for collecting words in the field, there is evidence that single-event Rapid Word Collection workshops facilitate collecting more words in less time than more conventional methods, thereby increasing the effectiveness of fieldwork, with anecdotal accounts claiming to double or triple the size of lexicons collected by other means. In this article, we describe the single-event RWC workshop methodology and report in detail about results for a 2015 Natügu [ntu][2] RWC workshop in the Solomon Islands. These are compared with results from single-event RWC workshops held for 11 other languages[3] over the past five years – seven in Africa, three in Asia, and one in the Pacific. RWC workshops in seven other languages are not part of the study due to incomplete data for our purposes.

Our findings demonstrate that the concept of collecting words by semantic domain has several advantages over traditional corpus-based methods, during which a linguist collects oral texts, transcribes them, and adds them to a database. For example, the RWC method addresses semantic relations, such as synonymy, by providing better evidence through disambiguated senses. These advantages hold with regard to time invested, numbers of words collected, and consensus about the data through group collaboration and a system for checking the words collected. We also show that the single-event RWC workshop method results are more effective than intermittent collection by semantic domain. Using these findings, we address several criticisms of RWC in §10. We suggest ways to improve RWC workshop results, and conclude by summarizing the approach's advantages and their implications.

2. A brief history of Rapid Word Collection The processes and procedures which have become Rapid Word Collection (RWC) have their origins in the work of Ron Moe, especially (2001, http://www.sil.org/resources/publications/search/contributor-/moe-ronald), who developed a technique he called the Dictionary Development Process (DDP). The underlying assumption for elicitation by semantic groupings is that we humans organize words in our minds based on semantic relationships.[4] Assuming

---

[2]Natügu [ntu] is spelled 'Natqgu' in the local orthography in which c, q, r, x, and z are vowels. That common local orthography will be used from here forward, since it is the one used by scribes in the Natqgu RWC workshop. We gratefully acknowledge the assistance of 27 Natqgu speakers who gave more than three weeks of their time for the Rapid Word Collection (RWC) workshop and the seven interns who assisted in the research, namely Jeremiah Aviel, Alexander Boerger, Donald Furnival, Adam Walker, Kim Wells, and two others.

[3]The authors would also like to thank the following individuals for sharing information about their experiences with Rapid Word Collection workshops in the languages listed here, with ISO 639-3 codes in square brackets. *AFRICA*: Cameron and Valerie Hamm for Bambalang [bmo] of Cameroon, Pamela Morris for Bissa Barka [bib] of Burkina Faso, Doug Higby for Buli [bwu] of Ghana, Stuart Showalter for Kaansa [gna] of Burkina Faso, John Walker for Kabwa [cwa] of Tanzania, Leoma Gilley for Shilluk [shk] of South Sudan, and Bep Langhout for two languages of Cameroon: Chungmboko [cug] and Kemedzung [dmo]. *ASIA*: Mari-Sisko Khadgi with regard to these three languages of Nepal: Madhya-Purbiya Tharu [thq], Syuba [syw], and Thulung [tdh]. *PACIFIC*: Steve and Kim Blewett for Rapoisi [kyx] of Papua New Guinea. We were also assisted by Kevin Warfel who connected us with some of the individuals and information above.

[4]Clearly words may also be arranged based on their phonological similarity, and it is this combination of phonological similarity and semantic intersection that forms the basis for most puns. Similarly, words can also be arranged by what one associates with them. For example, the word *mom* might elicit words which are associated with her, but which are not necessarily otherwise in the same or similar semantic domains, such as *apple pie, hugs, "clean your room,"* and *PTA*.

that words with similar meanings or words related to a single activity will often cluster together means that it is cognitively easier for speakers to think of these words all at the same time, rather than coming up with them alphabetically or from a wordlist elicitation. For example, if we ask for a word for 'rebuke' it is likely that only one word will be given. However, if instead we ask for all the Natqgu verbs for correcting someone, from a gentle chiding to a harsh tongue lashing, then we might end up with a list like the following (Boerger et al. 2016). More words could be elicited by broadening the category to think about punishment and reward.

Table 1. Natqgu words for rebuke

Natqgu	Natügu	English	Notes
adcpx	adâpä	notify, make aware of	
rngidr	öngidö	urge, correct, caution	
apqti	apüti	reprove, reprimand, chide, scold	not shared by all dialects
ycmne-apqbz	yâmne-apübë	admonish, exhort, warn	*yâ*mne 'speak'
lc-apqbz	lâ-apübë	admonish, exhort, warn	*lâ* 'complain'
lclvz	lâlvë	rebuke	
ycmne-kilvz	yâmne-kilvë	strong rebuke	
ycmne-plzti	yâmne-plëti	very strong rebuke	
lcki	lâki	rebuke repeatedly	
pnana	pnana	castigate, tongue lashing	

The set of words which are closely related in this way is called a semantic domain. Starting in 2000, Moe organized a set of nearly 1800 hierarchically arranged semantic domains, under nine major headings. That list is available at http://semdom.org/.

The resulting database from semantic domain elicitation would form the skeleton of a dictionary. Moe also developed elicitation questions and lists of English words within each domain to serve as examples. Even though his list is based on English and therefore has an inherent Western cultural bias, human cultures have striking similarities in the semantic domains we use. In fact, similarities in cultural categories have long been recognized, as evidenced by Yale University's *Outline of Cultural Materials* which categorizes aspects of culture and assigns them numerical abbreviations (http://www.yale.edu/hraf). The Yale categories are discussed further below and continue to be used by scholars today.

Both Moe's semantic domains and the Yale cultural domains are necessarily etic categories[5] which reach an emic arrangement in a particular language or culture. Given that languages and cultures divide the world differently, there is not a one-to-one correspondence between the domain labels in RWC and available lexical items in a particular language. Often it is difficult to find terms for the higher level nodes. In particular, we discuss in §6.10 how the arrangement of the animal kingdom semantic domain presented problems to Natqgu workshop participants in several instances.

---

[5]While the intention is to form etic categories, clearly the culture and experience of the designers is influenced by the language(s) they speak and the culture(s) they have experienced. Thus, there is no purely etic categorization possible.

The advantage of doing elicitation by semantic domains along with the sample domain words is that neither the speakers nor the fieldworkers need to generate target words or categories spontaneously, because a thorough set has been outlined in advance.

While elicitation by semantic domain is not new (Boerger et al. 2016:140, 206ff), what is innovative is the focused elicitation by groups of speakers during a single-event Rapid Word Collection workshop, which we will show later to be more effective than dividing RWC elicitation into separate elicitation sessions over weeks or months or years, as well as surpassing more conventional elicitation by wordlists and mining words from texts.

RWC workshop procedures have been highly systematized by SIL International's Dictionary and Lexicography Services (DLS), now led by co-author Stutzman. It is this format that we will be describing, whose procedures we will be evaluating, and whose results we will be documenting. At the heart of an RWC workshop is community engagement, because it ideally involves 40–45 individuals who contribute directly to the word collection effort in some way, as well as those who prepare and serve break refreshments and midday meals. Over and over again, Stutzman and RWC consultants trained by her team have witnessed growth in the community's interest in its own language and culture as a result of the participants sharing about what they do and learn during the RWC workshop. And since the language experts, one of the categories of participant in the collection groups, do not need to be literate in any language, it is possible to engage the full demographics of a community, as long as participants are physically and mentally able to endure the full-time pace of the workshop. Of course, it is also possible to let some participants join as they are able, as long as there is a basic core contingent coming on a daily basis.

To date, the standardized RWC workshop format – as carried out by DLS-trained personnel – has involved six groups of four or five people collecting words full-time over a period of ten days, during two consecutive work weeks. In this context, full time will generally run around six or seven hours, including breaks and lunch. This is preceded by three days of training for the collection group leaders, scribes, and glossers, and it is followed by one week of editing of the words and glosses collected during the workshop. In that format, a single-event RWC workshop takes about one month. Turning the resulting lexical database into a bilingual dictionary will take considerably longer.

However, since the number of speakers available may vary, the claims made throughout will be clearer if we specify the time investment parameters more clearly. To do this let us introduce the concept of "*60 consecutive, full-time group days*" as the reasonable target for completing all the semantic domains. This is more specific than just saying two weeks or ten days and can be manipulated more easily. This time frame was calculated by six groups working for 10 days equaling 60 group days. The adjective "consecutive" means two things: first, that the group members are all trained together for two or three days in advance of the collection times; and second, that they then work through the domains for as many days as it takes to complete them, over a consecutive period, with weekends off. The adjective "full-time" indi-

cates that the groups will need to work from 8:00 a.m. to 5:00 p.m., or some similar interval, during these days, in order to finish the task. This clarifies expectations and makes time calculations easier depending on the numbers of speakers available and for how long.

3. Defined roles in a formal RWC workshop   In order to report on results of RWC workshops, our focus will be on 12 languages which used the single-event format with advance training and distinct, formal roles for participants. The personnel involved in an RWC workshop include the following: workshop coordinator, logistics manager, RWC consultant, record keeper, group leader, group scribe, group language experts, glossers, and typists. These are discussed in more detail below in chronological order leading up to the RWC workshop event.

3.1 RWC workshop coordinator and logistics manager   Before the workshop takes place, it is critical that two roles be filled. The *workshop coordinator* has responsibility for the overall organization of the workshop, including advertising, recruitment of participants, oversight of training and feedback to participants, and acting as the primary interface with the outside consultant. The coordinator is assisted by a *logistics manager* who takes responsibility for organizing venue, meals, and transport and who may act as pay master depending on local and funder expectations. Together they are responsible for planning and administration.

Photo 1. Workshop coordinator

3.2 RWC workshop consultant and record keeper   Two further workshop roles are the RWC linguist-consultant and the record keeper. The workshop *consultant* is normally an outside expert who has been trained first as an apprentice consultant in an earlier RWC workshop. The consultant leads the preliminary training sessions over a three-day period and is encouraged to mentor someone who is learning to be a

consultant. The consultant assists the coordinator in giving helpful feedback to the participants. A *record keeper* organizes and helps collect and file informed consent documents from the participants and records attendance. At the start of each day or session, the record keeper distributes semantic domain folders. Then, as collection groups turn in their folders, the record keeper serves as the hub to assure that the pages are in numerical order and then enters data into a preset, 10-day spreadsheet, which tracks the number of words collected in each domain. The record keeper then passes the folders to the glossers and typists, and distributes a new domain folder to the collection group which has turned in their completed folders. When folders return after having been glossed and typed, the record keeper performs another check to be sure all the words are glossed, puts the pages in numerical order again, enters the statistics in the spreadsheet, and files the folder in a box for completed work. On a daily basis the record keeper calculates and conveys two totals: the number of words collected for a particular day, and the total words collected so far. A visual representation of these totals is maintained to help keep the word collection groups encouraged and to motivate them to find more words. The consultant and record keeper support the word collection groups.

3.3 RWC word collection groups: leaders, scribes, and other language experts The word collection groups form the heart of the RWC workshop, with each group having its own structure. Groups are comprised of four to six fluent speakers of the language, including a team leader, a scribe, and from two to four language experts. The *team leader* helps the group work through the Questionnaire domains in the folder assigned to that group.

Photo 2. Word collection group

Therefore, the leader needs to be fluently bilingual in both the vernacular and the LWC used in the Questionnaire. The leader also needs people skills and general leadership skills to keep the group moving forward. The *scribe* in each group writes down the words the team members call out. Therefore, the scribe needs to be able to spell accurately, write clearly and quickly using the standard orthography, and to

help the team by serving as time keeper, documenting the start and end times for processing each domain. The remaining group members are called *language experts*. They are speakers who know the language best, who have spent most of their lives in the language and culture area, and who are able to cooperatively work in groups. The language experts need not be literate in any language and they should be representative of the entire community, including males and females, as well as multiple generations. Membership in groups can be somewhat fluid, as long as there is a leader and scribe in each group. Ideally, each group itself would also be a microcosm of the community.

The scribes in each group receive printed response sheets with space for 30 words and their glosses. Each sheet also collects metadata about the semantic domain name and number, the scribe's name, the initials of the group's participants, the date they worked, and the time the group started. At the top are three bookkeeping boxes for a) when the words have been counted by the record keeper, b) when the words have been glossed and the glosser's initials, and c) when the words have been typed into the FLEx database and the typist's initials. This data can be combined with basic participant metadata collected along with informed consent to explore dialect variation and to track certain tendencies a scribe (misspellings and handwriting idiosyncrasies), glosser (lack of precision), or typist (repeated misinterpretation of handwriting of a particular scribe) might have. A completed response sheet is included below in the discussion about the Natqgu RWC workshop.

3.4 RWC glossers   As the word collection teams submit their words to the record keeper, they get passed on to the glossers, who are normally located in a separate room. The role of the *glossers* is to write the meaning of the vernacular word in the LWC being used for the workshop. In addition, they are instructed to check spelling, to be sure words are in the correct citation form, to be sure that the words submitted actually belong to the domain given, and to add additional words to any of the domains as they think of them. The glossers also check back with collection groups if they have questions about words the groups have given them. For example, one collection group misunderstood the scope of a category, as we describe below in the section on etic and emic animal kingdom challenges in Natqgu.

So glossers serve as checks and balances to the word collection groups. This is one of the most difficult jobs in the workshop because glossers not only need to be fluent in both the vernacular and the LWC, they also need to be able to recognize homonyms and multiple senses of the same word in order to give the appropriate gloss for the domain they are working on at the time.

Some scholars might find it profitable to preserve the original spellings as containing information about native intuitions about phonology and even to include them in a finder list which would lead to the correct spelling of the target word. This could be done in FLEx or through other ways of preserving the data for later analysis.

For example, look at the following Natqgu entry for *temz*. Its primary meaning is 'moon', while the associated meanings relate to 'month', as the cycle of the moon, and to 'things that have the appearance of a full moon', such as round, white forms.

The glosser must not just write 'moon' every time he encounters *temz*, but needs to record the gloss that corresponds most closely to the domain being glossed at that time. Given that many cultures function best through working by consensus, it was effective for Natqgu to have two rooms, with two or three glossers working in parallel in each one, so that they could consult each other, which happened often.

Photo 3. Dancer with breastplate
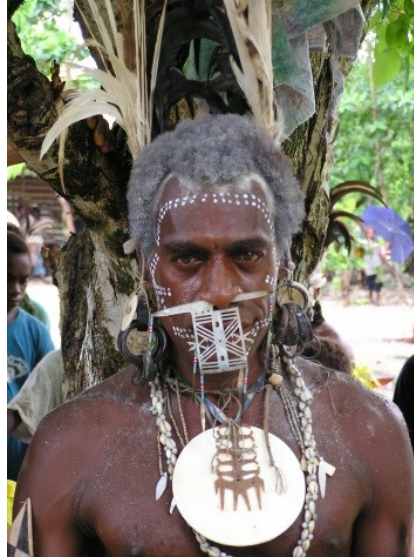
temz N **1**) moon *Temz trnapnapxu.* The moon will not shine. sem. dom.: 1.1.1.1 – Moon, 8.3.3.1.1 - Light source.) der. nctq temz, temz kxpo, temz ngipe ncte, temz nqngibenyz, temz yc-atwrngr **2**) month *Ycbep la temz pwx, x leplz kc tqrkqlzpe.* There are still four months, and then a person harvests. (sem. dom.: 8.4.1.4 - Month.) **3**) menstruation *Kzdq kc olvz kx tresakiu nzrpwx-krde temz.* There was a woman whose monthly flow did not cease. (sem. dom.: 2.2 – Body functions.) **4**) breastplate; round white, chest protector made from a shell; undecorated version historically used during warfare; decorated version used by dancers in *nelc* dance; smaller versions of decorated shell sold as a symbol of Santa Cruz Island. *Ncblo ngr vea nzyrlqbzlr temz ngr lomrdr kc tqyrputrngr nipna, murde ma tzluplxtrpz enqmi rdrng lomrdr.* The warriors put on their breastplates which arrows bounce off, lest their enemies pierce their chests. (sem. dom.: 4.8.3.7 – Weapon, shoot, 4.2.4 – Dance, 5.4.1 – Jewelry, 5.3.5 – Clothes for special people, 8.3.1.8 – Pattern, design.) der. temz ngr lomr kx nzwzngr mz aian. **5**) operculum; the small, flat, coin-sized shell which acts as the closure to a turban shell. *Meya mcle kc temz mrkc drta'.* Meya saw the operculum shell on the beach. *wh:* nqbli. (sem. dom.: 1.6.1.5 – Fish.)

Photo 4. *Nqbli* shell with operculum

We see from the semantic domains in this entry that *temz* was collected for ten different domains, leading to five distinct senses. Therefore, unlike wordlist elicitation which is normally satisfied with one word per prompt, RWC is likely to capture multiple senses of a word, since the domains for the individual senses are elicited separately. This is valuable in light of recent work in historical linguistics which finds that co-lexification provides another piece of evidence of language relationships (Koptevskaja-Tamm et al. 2016). That is, by having a list of senses associated with a root, there are more chances to find cognates in languages presumed to be related and more possible hypotheses can be made about semantic shift, polysemy, and other semantic and cognitive concepts.

3.5 RWC typists   When a folder containing multiple related domains has been glossed, it is passed to the *typists*, who are optimally located in the same room as the glossers. The computer-literate typists use the "collect word" tool in FLEx, described in "software and tools" below, to enter both the vernacular word and its gloss into the computer. This is done by selecting the correct semantic domain for each wordlist in a folder from the domain drop down menu. These personnel must be experienced, competent typists, with familiarity regarding how to access any special characters used for the language. It is also recommended that the typists be bilingual so that they can catch their own typing errors and are able to decipher handwriting which is unclear. Using typists who only know the LWC can lead to problems, as we show in the Natqgu case-study below, although more experienced RWC consultants have found ways to make this successful. Having a minimum of two typists means that all the words are entered into the computer as soon as the glossers make them available.

4. Task and event flow in a formal RWC workshop

4.1 Pre-workshop training for language roles: Two to three days During the two or three days of training before the word collection groups start working, the participants with technical roles – the team leaders, scribes, glossers, and typists – are given special training to help them learn the tasks required for each role. After there has been a description of the tasks for each role, there is a role-playing game, with individuals selected to fill the particular roles. Each of them wears a headband with the particular role written on it, with evaluation by others regarding how they did the tasks assigned to that role. Then a demo group works through a practice domain together, while others observe and give feedback on things they did well and the things they overlooked.

Photo 5. Role-playing during training

Another aspect of the training involves thinking through the multiple levels of embedding in the semantic domain hierarchy, which as noted previously is necessarily etic, in order to cover all domains and to be relevant to multiple languages and cultures. Clearly the addition of emic information will need to happen outside the hierarchical organization of the domains.

Hierarchy training is done in at least two ways. First, participants are shown a genealogical chart drawn by the consultant or consultant trainee. It shows the levels of embedding from great grandparents, down through a family line. Another way this is done is by having the participants practice a domain together. The Natqgu RWC workshop used the domain *6.5.2 Parts of a building*, to illustrate levels of embedding. It is part of the parent domain *6.5 Working with buildings* and part of the grandparent domain *6 Work and occupations*. We then showed how the parts of the building can be seen as its eight children – wall, roof, floor, door, window, foundation, room, story – each with a separate domain number. After brainstorming in the large group about

any words related to house-building, participants took turns coming to the board and underlining the ones in different colors to match the smallest domain level, such as 'wall' words being underlined in blue chalk.

Photo 6. Parts of a building

Other topics during training were citation forms, orthography and spelling, idioms, and whether or not to record borrowed terms which are used daily. Training for the scribes asked them to use a black pen. Then glossers learned how to check that the gloss they choose for a word is from the correct domain, how to make corrections using a red pen, and how to check and correct spelling from the collection groups.

4.2 Word collection for 60 consecutive full-time group days   Immediately following the training, word collection groups, glossers, and typists begin their assigned tasks, working through the folders of domains until the domains are complete. The groups start with concrete domains with primarily nouns, then move to action domains, and conclude with the more abstract domains. This makes the learning curve gradual and increases effectiveness. The amount of time required will vary depending on how many groups are working.

4.3 Post-workshop on site FLEx database editing   Following the workshop, one week has conventionally been allotted to general editing of the words collected. This involves four or five of the most gifted participants who among them command the typist, scribe, and glosser skill sets – the more skills per person, the more effective they can be. It is critical that the editing team be comprised of those who can speak, read and write both languages and who have not recently been away from the language area for an extended time. It helps, too, if at least one of them is computer literate, though this is not essential if they are working with someone who is.

The main purpose of this editing group is to serve as a second revision/approval loop in the process, in support of such checking already done by the glossers. At this point, the editors would have an opportunity to discuss what should be treated as homonyms and what as a single word with different senses. Again, consensus decision-making would come into play. That is, the editors work in the FLEx database to examine all the words collected, correct spellings, purge duplicates, and – if time allows – also decide on merging entries and senses, adding example sentences, and other higher-level lexicographical tasks as time allows. In actuality, one week was completely inadequate for this task in Natqgu, as will be discussed below.

This means that at the end of the RWC workshop process, there will be a database of raw data which still needs extensive editing. That is, we have rapidly collected words, but we have not fully organized or processed them. That processing and editing requires its own work flow, as described in the next section.

4.4 Post-workshop off site dictionary preparation   We see that the fruit of a single-event RWC workshop then is a lengthy, somewhat clean, glossed wordlist. For this wordlist to become a bilingual dictionary, more concerted lexicography needs to take place, such as adding fully developed definitions, information about pronunciation, synonyms and antonyms, example sentences, lexical relations, dialect notes, parts of speech, grammatical appendices, and any further information desired by the stakeholders, both the community and any linguists they relate to (Bartholomew & Schoenhals 1983).

Progress in these tasks will necessarily vary considerably, based on who and how many personnel are editing the database and how many hours they are able to give to it. A linguist who has no means of consulting speakers will be able to do some of the work independently. For example, a linguist working alone will be able to add some example sentences to the lexical database by finding them in texts which have already been collected and glossed. But there will also probably be a list of flagged items requiring native speaker input. The work can be facilitated through internet contact with speakers who are able to spell their own language well and to respond to the researcher's questions. But realistically, to obtain answers to these questions could require at least one or two further extended fieldwork trips to complete the editing and to consult with the community about what kinds of products they desire to see from the database.

5. RWC software and tools   Moe's semantic domain categories have been incorporated into the Fieldworks Language Explorer (FLEx) lexical database software (http://software.sil.org/fieldworks/), with dictionary production and RWC workshops in mind.[6] Its "Collect Words" tool is set up for inputting all words collected for a particular semantic domain, making workshop data entry straight forward. The initial word collection automatically logs the semantic domain for each word. Pawley

---

[6]There is a FLEx Google group at flex-list@googlegroups.com where FLEx users and developers assist each other in answering questions about using FLEx. There is often a quick turn around and it is recommended for anyone using FLEx.

(2006:188) comments that "a close study of different lexical domains would reveal interesting variations." If the domains are made available in electronic media the lexemes can be sorted by domain, with an amazing array of study options available. Another advantage of tagging each word with its related domains is that it makes it straight-forward to compile specialty dictionaries. For example, we could create a dictionary of fish names by using just the domain *1.6.1.5 Fish* or we could expand it to include other sea creatures, including also *1.6.1.6 Sharks and rays*. This specialty dictionary could then be made more valuable by the addition of photos and illustrations to create a picture dictionary for use in schools.

An additional feature of the FLEx software is the anthropological "Notebook" tab which allows users to annotate words with their appropriate anthropological category from the *Outline of Cultural Materials*. This involves using another dropdown menu to find a major category and then to tag terms appropriately. It is intended to replace the anthropological field data notebook, and includes places for extended notes. Then, items in the lexicon can be linked to the notebook labels, thereby connecting the language and culture sides of the database.

For example, one of the interns during the 2015 fieldwork trip studied the Santa Cruz Island banana fiber weaving done on a back strap loom. The process is used to make the breechclout worn by the hired dancers in the Santa Cruz *nelc* dance. The weaving is done by only one clan in one village on the island, and the weaving craft master agreed to teach the intern the entire process. On the lexical side of the FLEx database, the student used the semantic domain *6.6.1 Working with cloth*, primarily its subdomain of *6.6.1.4 Weaving cloth*. But stages in the process and uses of the cloth also involved the domains of *6.6.6.1 Cloth*, *6.6.1.2 Spinning thread*, *6.6.4.2 Weaving baskets and mats*, *8.3.1.6 Pattern, design*, *8.3.3.3 Color*, as well as *1.5.1 Trees* and *5.2.3.1.2 Food from fruit*, when discussing the varieties of banana tree from which the weaving fibers are made. Other semantic domains relate to the uses of the woven fabric, such as *2.6.1.2 Wedding*, *5.3.1 Men's clothing*, *5.3.2 Women's clothing*, *5.3.3 Traditional clothing*, *5.3.4 Clothing for special occasions*, and *5.3.5 Clothes for special people*. These categories can be compared to the anthropological categories accessed from the Notebook tab. There, the weaving vocabulary was categorized under the primary category of Material Culture (MC), then the category labels used were primarily *285 Mats and Basketry* and *286 Woven Fabrics*, as well as *280 Leather, Textiles, and Fabrics*, *290 Clothing* and *292 Special Garments*. The intern is a graduate student who plans to pursue this research further for her MA, expanding the anthropological aspects of the significance of the weaving craft. The usefulness of the semantic domains and anthropological categories is that one can sort the database using anything which has been tagged in this way.

A few additional websites warrant being mentioned here, all of which relate to making effective use of FLEx for dictionary production. First, http://rapidwords.net/ is the official website for the Rapid Words Collection methodology, where many of the materials used in a RWC workshop can be downloaded. It also contains reports about RWC workshop results from around the world and news of interest to those engaging in RWC.

Collaboration on a FLEx database is made easy at https://languageforge.org/, where individuals can collaborate on Language Forge then sync their changes back into FLEx. Alternatively, one can make edits in FLEx and then use "Send/Receive" to share them at Language Depot (https://public.languagedepot.org). Then members of the project can use "Send/Receive" to get changes others in the same project have made.

For data management, we recommend using SayMore software (http://saymore-.palaso.org/) which allows all project data files to be stored in one place with the ability to link between them. This is particularly useful in linking informed consent documents with individual photos of the RWC participants or other language consultants in a fieldwork project.

6. Natügu RWC in-depth discussion   Boerger lived with her husband and two sons on Santa Cruz Island, Solomon Islands for nearly 20 years between 1988 and 2008, doing Natügu/Natqgu [ntu] language development tasks. The RWC workshop there was conducted as a research component for a 12-month Documenting Endangered Languages (DEL) Fellowship[7] awarded to Boerger for a project entitled, "Natqgu Dictionary and Legacy Texts," during 2015–2016.

Boerger had previously compiled a Natqgu lexicon based primarily on the input of several individual Natqgu consultants and discussions in team meetings. This data was augmented by dialect information which arose during checking of written materials produced by her and her colleagues. The Natqgu language has several varieties in a dialect chain which wraps around the island, eventually encompassing all four Santa Cruz languages (Boerger & Zimmerman 2012:98; Boerger et al. 2012:117–118). The intention was to include Natqgu varieties in the database, but to exclude the other Santa Cruz languages: Nalrgo [nlz], Noipx [npx], and Engdewu [ngr].

The October 2015 Natqgu RWC workshop came at a time when there was a growing felt need for it. In fact, at a March 2015 workshop for teachers sponsored by the Temotu Province Ministry of Education on Santa Cruz, Solomon Islands, the teachers were asked what the next step would be in moving toward multilingual education (MLE). Their response was that there needed to be a bilingual dictionary[8] for Natügu [ntu]. While the community did not request the RWC workshop, there was a general acceptance that this would help them move forward in language development.

Since Boerger served as the RWC consultant for the Natqgu workshop, in this section of our paper about Natqgu, there is a shift to first person singular references referring to her. With that given, this section describes how the parameters of the single-event RWC workshop played out during September–October 2015 for Natqgu [ntu].

I led a team of seven interns to conduct the research. This included nearly three weeks of prefield training plus 10 weeks in the field, with eight weeks in the language

---

[7]I, Boerger, gratefully acknowledge NEH/NSF support through a 12-month DEL fellowship during 2015-2016. Any views, findings, conclusions, or recommendations expressed in this article do not necessarily reflect those of the National Endowment for the Humanities and the National Science Foundation.
[8]Personal communication, Elizabeth Ilovz, teacher at Kati Primary School, Graciosa Bay, Santa Cruz, Solomon Islands.

area and a week either side of that in the national capital. The team was together from August through mid-November 2015. While this was my first time to lead an RWC workshop, my interns and I had been trained for three days in RWC procedures by Verna Stutzman, who has been consultant for multiple RWC workshops around the world and has mentored others who have become RWC consultants.

6.1 US team demographics As principal investigator, I recruited from acquaintances and from advertisements placed at three linguistics schools in the Dallas-Fort Worth area: the University of North Texas at Denton, the University of Texas at Arlington, and the Graduate Institute of Applied Linguistics. Team members' education levels ranged from several years of post-secondary education through an MA in linguistics. This resulted in the following team of seven interns, all of whom assisted with the RWC workshop and also had responsibility for other areas of research, several of which allowed deeper exploration of culturally significant semantic domains:

- Jeremiah A.–has completed all MA linguistics coursework at GIAL, and has done language and culture documentation fieldwork in Papua New Guinea. He helped with prefield training in FLEx and led the research on **kinship** terms, as well as valence of Natqgu verbs.

- Alex B.–the PI's son, grew up on the island and commands the language and culture. He helped with prefield training, computer and AV technical expertise, cross-cultural insights, served as team cook, and led the research in **ethnobotany**.

- A couple–came to me from GIAL. They were anticipating work elsewhere in the Pacific and desired this experience to help prepare them cross-culturally. They were assigned to lead processing of legacy oral texts using Basic Oral Language Documentation (BOLD, as in Boerger 2011).

- Donald F.–had just completed an MA in linguistics at UNT Denton and was interested in exploring language and culture documentation. He helped train the team in SayMore software and was responsible for data management. He was assigned the **house building** domain, and played a major role in helping the other researchers be successful by contributing to their research objectives.

- Adam W.–I knew from the Dallas area community. He has done some graduate work in linguistics and has additional research interests in ethnic dance and sign languages. He led the research about the Santa Cruz *nelc* **dance** (Walker 2016).

- Kim W.–I also knew from the Dallas area community as an elementary music teacher. She had two previous courses in linguistics for bilingual education. In conjunction with this trip she resigned from her teaching position and started MA coursework in anthropology at GIAL. She had previous experience in loom weaving and was therefore assigned to research the island's **banana fiber weaving** craft, which she did as an independent study course (Wells 2016; Wells & Balq 2017).

- Valentina A.–was not part of our Natqgu [ntu] team, but she was a PhD student working on Nalrgo [nlz], the most closely related language on the other side of the island. There were several **collaborations** between her and some of our team members.

6.2 Village location The norm for RWC workshops has been to host them at a regional center where there are greater resources than in a village setting. However, for the Natqgu RWC workshop I felt it was critical for it to be held in the center of the language area for several reasons. First, I thought we would be more likely to receive natural Natqgu data with participants returning home each night, rather than housing them in an artificial environment outside their home area. Second, in the language area we were better able to get enough participants for the workshop to be a success. Staying in the language area is likely to be even more critical for highly endangered languages. Thirdly, the financial and logistical cost of sending everyone by ship to Honiara was prohibitive and there was no advantage of everyone being housed and fed in Lata, the provincial capital, just a few miles from the language center.

6.3 Basic statistics

ISO 639-3 code:	ntu
Location:	Bznwz, Santa Cruz Is, Temotu Province, Solomon Islands
Word collection dates:	September–October 2015
Total US facilitators:	8
Total training days:	3
Total FT collection days:	15
Total Natqgu participants:	27
Average daily participants:	20
Attended 10 days or more:	19
Total domains treated:	1792 (all)
Starting total senses in FLEx:	4,820
Total raw words collected:	14,270
Current FLEx total senses:	11,997 with editing ongoing
Estimated unique senses:	11,700
Signed consent forms:	27
Photos:	Each leaf represents 500 words, days 02, 05 and 09

Photo 7. Word count, day 2

Photo 8. Word count, day 5

Photo 9. Word count, day 9

6.4 Participant demographics The workshop participants included young, middle-aged, and older individuals, including both men and women. This meant it was possible to include at least one woman and one older man in each word collection group, thereby giving the opportunity to engage with a broad spectrum of the community. It also meant that broader areas of competence and specialized vocabulary could be accessed. For example, on the day when private body parts, dating, sexuality, rude terms, marriage, and childbirth were discussed, and knowing some taboos of the society, I asked the participants whether we should revise our groups for that day. In consultation with each other they suggested that it would be better to form single-sex groups for these domains, which led to two groups of single guys taking the domains for male body parts, dating, flirtation, sexuality and rude terms; a group of married guys discussing marriage, terms of endearment, and sexuality; and a group of single and married women discussing female body parts, sexuality, and child birth. There were no older women participating regularly in word collection. But for this day, the women's group recruited an older woman to join them, who had married into the village where the workshop was being held. A note about marriage practices is relevant here.

> Santa Cruz marriages are arranged. The woman's adult male relatives meet several times with the man's adult male relatives to set an agreed upon bride price…It is forbidden or *tambu* to pronounce the given name or 'home' name, as opposed to baptized name of one's adult in-laws–all of one's in-laws. In fact, Pijin *tambu* is used to mean 'in-law,' and Natügu has two words referring to in-laws: *këdo* is a male in-law and *lväbü* is the female in-law of a female, such as a woman's daughter-in-law or a bride's

mother-in-law. Since villages are usually inhabited by extended families, when the new bride moves to her husband's village, everyone there is a new in-law of some kind.

A new bride's primary relationships are with her mother-in-law and other female relatives of the groom. They are the people with whom she works in the gardens and cooks and does other chores, so it is critical that the bride and her mother-in-law get along. After a marriage, if a girl's birth family feels that she is not being treated well by her new relatives, they can return the bride price and bring the girl home, in effect having the marriage annulled.

In Davenport's era, (Davenport 1964) the bride price would have been composed in part of coiled red feather money, and carried on the heads of the groom's female relatives. But in Santa Cruz of the early 21st century, the bride price is paid in Solomon dollars…

> (Excerpted from Boerger, in progress)

In that context, then, the grandmother was able to share with middle-aged and younger women about childbirth practices before the establishment of the provincial hospital just a few miles away. This relates to what Sperlich & Pawley (2013:6) say about the importance of the woman's viewpoint in compiling dictionaries:

It is noteworthy that the chief author (http://hdl.handle.net/1885/107217) of the Lakalai dictionary is a woman. A survey of the major dictionaries of Oceanic languages reveals that almost all have been compiled by men. The question arises, does the gender of the compiler influence the character of a dictionary? Insofar as a dictionary is a kind of ethnographic record, to ask this question is essentially to ask whether the gender of the ethnographer influences the character of the ethnographic description. We think that in the case of the Lakalai dictionary, it does, to some extent–that Chowning's interaction with Lakalai women yielded linguistic and cultural data that a male counterpart would probably not have obtained.

In relation to Natqgu, it was not so much the gender of the ethnographer which allowed for the collection of cultural information relating to childbirth, for example, but rather that the group was composed entirely of women and that they were given the freedom and encouragement to talk about things which are not normally part of their day-to-day discussions. Likewise, the men were laughing together when they realized that for some of the rude terms, it was a former Anglican priest who said a particular word, the church catechist who wrote it down, and Alex B. who glossed it in English. The factors which allowed exploration of these domains are that the community made the choice of how to divide the groups and the thoroughness of the domain list meant that certain topics were not avoided because they might have been too uncomfortable for them to bring up themselves.

6.5 Equipment and data management   The team had combined resources of six personal computers and one tablet. By charging them nightly with mains power at the rest house where half of us were staying, it was possible to use two computers in the morning and two in the afternoon in the two glossing-typing rooms. The tablet was used primarily by the record keeper, and two computers served as backup in case battery life on one computer proved insufficient for its assigned period of operation. This meant that team members could use their own computers for their assigned tasks. As principal investigator, I kept the master files for the FLEx database, as well as backing up documents generated by the record keeper on my own computer. All files generated daily were additionally backed up each evening on two external hard drives. To keep all the computers up to date, a flash drive was circulated at the end of the morning and afternoon sessions so that all the words collected were uploaded and synced and everyone was using the same starting database.

The Natqgu team used FLEx for the database and SayMore for data management, as recommended above.

6.6 "60 consecutive, full-time group days" for Natqgu RWC   The average number of Natqgu participants on a daily basis was 20, generally allowing for four word collection groups of four people each, and four glossers per day. This was two fewer groups than recommended for the standardized workshop to be conducted in two weeks. To reach "60 group days" for Natqgu meant adding a third week of five working days. But, rather than completing the glossing of all the domains, the group elected instead to quit at midday for a closing feast, party, and speeches. That being said, for planning purposes, the Natqgu experience shows that "60 group days" is a reasonable, approximate target for a single-event RWC workshop, as described above.

6.7 Logistical notes

6.7.1 Delays   The main hindrance to the workshop was that there was a death every week for the first six weeks of the ten weeks the US facilitators were on the island. For one death, it meant delaying the start of the workshop for a full week following the initial training. The other deaths meant that key personnel were missing for a few days at various times. Each interruption meant that momentum was lost and that fewer groups were possible. But the participants themselves decided to extend the workshop in order to complete the domains.

In fact, they even insisted on doing all the grammar domain questions in folder nine, because they said, "We want to break our heads on it." In other words, they wanted to learn about their language in ways they had not previously looked at it. Since I had already worked extensively on a draft grammar of Natqgu, there were no new words or categories gathered during this time. The two interns with MA-level linguistics training rotated amongst the groups to help them unpack the linguistic concepts. So while it was less necessary for the data gathered, it was satisfying to

the speakers to explore these categories and to have completed the domains. For the interns it was useful for them to attempt to explain the concepts simply.

Another delay was that on rainy days, people came considerably later than usual. And until we rented a vehicle for our exclusive use, we found we could not rely on hired transportation to get the US team to the workshop on time each day.

6.7.2 Informed consent We read and discussed aloud the informed consent form, explaining to participants that this was required by our educational institutions in the US. We explained the options for what to put in the various blanks and the implications of saying "yes" to various items. After a form was signed, an intern would take a photo of both the form and the person and tag them with their name, so that the document and the photo could be paired in SayMore. This worked well.

The participants are not considered co-authors of the dictionary, per se, because it is impossible to include all of them in the day-to-day decision-making that has to happen. However, the plan is to list all of their names in the front matter of the dictionary so that they are acknowledged historically as having been the speakers who contributed in this way.

6.7.3 Intern responsibilities We found that eight outsiders were too many once the participants were comfortable with their roles. So five worked and three rested each day. However, Alex B. and I worked full-time all of the days, because our combined skills in English and Natqgu were needed daily in the two glossing rooms to help glossers find specific rather than generic glosses for words.

6.7.4 Infrastructure lacks We had no internet or data projector, so we did not show videos of previous workshops for the training phase, as is the convention established for standardized RWC workshops. Similarly, we also had no printer, so the daily or weekly words were not printed out for people to take home and talk about. These may have affected initial output and community enthusiasm, but it was made up for as momentum built during the workshop and results were shared person to person.

The workshop participants were the first group to use a recently constructed local community hall and we had no tables for the collection groups to sit at. Each group sat on the floor and had a clipboard for the scribe to use. The three available tables were assigned one each to the two glossing rooms and one for the record keeper in the main room where all the groups sat. Sitting on the floor is the norm for Melanesia, but not for RWC workshops which have customarily been held in regional centers with tables, chairs, and other advantages like reliable electricity and internet access. While it may have become uncomfortable after a period of time, with morning and afternoon tea breaks and another break for lunch, participants had the opportunity to move and stretch. No one commented on it or complained, and the older participants were healthy enough to participate full-time. There is no evidence that sitting on the floor either helped or hindered the process. My impression, though, is that it may have made responses more natural since the sitting postures reflected norms for conversations held on the porches of local houses.

6.8 Linguistic challenges

6.8.1 Borrowings   There was no consensus among the participants regarding the status of borrowed terms. One natural leader and several others expressed resistance to including borrowed terms in the eventual dictionary, even though most of the indigenous languages in the country have borrowed extensively from English via Solomon Islands Pijin (Boerger et al. 2012). This meant that one creative individual and his group insisted on inventing new terms on the spot. For example, instead of saying *lada* from English *ladder*, for 'stairs' or 'steps' they wanted to use a term for a tree that is leaned against another tree for climbing and harvesting in the bush. And instead of saying *poket* from English *pocket*, they said 'basket of clothing', which is ambiguous in Natqgu just as it is in English. It was surprising that there was so much resistance to these borrowings, since they are so much a part of the language as it is used daily. If I had anticipated it, I would have taken more time during the training phase for them to discuss the purposes and uses of a dictionary, as well as the pros and cons of including borrowings in it. However, since we did want all the groups to be processing things similarly, I took a session mid-week of the first week of word collection to do a presentation on the naturalness of borrowing between languages in contact, with data about words English has borrowed through time and the languages from which the words were borrowed.

6.8.2 Dialects   In the word collection groups we had speakers from the dominant central dialect, as well as at least one speaker from each of four other dialects in the chain that winds around the island. The groups noted dialectal variants by writing the word with a single capital letter code after it for one of the four other dialects: B for Balo, also called "Bottom Bay", L for Lvepx village, M for Mzlo village, and V for Vxngr village. All lexical items collected, including those from different dialects, were included in the word count. The intention is to include these as separate headwords for inclusion in the dictionary once it is published, for ease in looking them up, as well as for it being a way to show value to all the variants.

   For example, the target dialect for the dictionary uses the word *beningi* 'edge', while in the dialect of the area around Vxngr village people say *benigqnyr* 'edge'. Both dialects combine 'edge' with 'of boat' to mean the gunnels of any water craft, either *beningi r bot* or *benigqnyr r bot*. The plan is that all four forms will be in the dictionary, with the Vxngr forms flagged as dialectal variants. Part of the editing process includes deleting the capital letter from the dialectal citation form and moving it to an appropriate field in FLEx, which will allow finding and sorting by dialect variant, if any.

6.8.3 Citation forms   Unfortunately, some groups did not grasp the idea of citation forms, and perhaps the pre-workshop training could be improved to give more focus and feedback on this. Instead of using citation forms, some groups included conjugations of verbs in multiple (but not all) persons. Likewise, a verb would occur with

multiple optional particles and affixes, but the basic meaning was still the same. Or similarly, we got a noun possessed by multiple persons. Clearly, even if the elicitations are not ideal, these kinds of forms can be dealt with as part of the editing or even the typing process. In fact, see below about glossers in the section on skill set challenges, where we added a step prior to typing to eliminate unwanted forms.

6.9 Skill set challenges   In this section, I report on problems we faced in each of the conventional RWC roles, due to there being insufficient "ideal" people with the skill sets needed for the workshop. The average levels of education in Melanesia, for example, are significantly lower than those of many in African nations.

6.9.1 Group leader   The group leaders did quite a good job, overall. But early in the workshop they started bringing English dictionaries to look up every single word in the English examples. It didn't seem to matter that I said to only use the example words if they were helpful, because some groups made the task a translation exercise and not a brainstorming event. I eventually used a marker to black out all the sample words in English, except a very few. Plans for how to improve this are included below, but basically include using a few words from an LWC rather than all possible English words.

The full questionnaire is available at: http://rapidwords.net/resources. But for discussion purposes, here is the questionnaire section for *1.3.1.3 River*. Our suggestion is that it might be less obstructive and more helpful to just use 'river' and 'creek' in (1) or similarly to use 'flood' in (3) and 'swell' or 'swollen' in (2). Most minority language populations are not going to have the competency in English to distinguish amongst the various options provided in the questionnaire or to benefit from them.

6.9.2 Scribe   Within each group, the scribes had the most difficult task. For our former co-workers it had been nine years since we'd worked together, and they had not done much writing in Natqgu since then. We had to urge some of them to be willing to try to do it again.

1.3.1.3 River
Use this domain for words referring to bodies of flowing water.

(1) What words refer to rivers of different sizes?
   • *river, stream, streamlet, creek, brook, brooklet, rill, rivulet*
(2) What words refer to a river when it has a lot of water?
   • *flood, torrent, flash flood, freshet, swell, surge*
(3) What words refer to a river overflowing its banks?
   • *overflow its banks, flood, deluge*
(4) What words refer to two rivers coming together?
   • *branch, confluence, fork, tributary*

---

(5) What words refer to a part of a river that is steep?
- *waterfall, falls, rapids, cascade, cataract, spillway, race*

(6) What are the parts of a river?
- *source, head, headwaters, mouth, upstream, downstream, bend, whirlpool, eddy, river bottom, riverbed, backwater, narrows, oxbow*

(7) What words refer to a river when it has no water?
- *overflow its banks, flood, deluge*

(8) What words refer to a man-made river?
- *canal*

(9) What words refer to a place to cross a river?
- *bridge, ford*

(10) What words refer to something floating down a river or caught in it?
- *flotsam, snag, logjam*

(11) What words refer to dirt, sand, and rocks washed down by a river?
- *sandbar, delta, alluvium, alluvial, silt*

(12) What words refer to the area of land drained by a river?
- *sandbar, delta, alluvium, alluvial, silt*

(13) What words refer to the movement of a river?
- *current, flow*

6.9.3 Language expert It was easiest to be a language expert because they were not required to command English fluently or to know how to read or write Natqgu. This made participation possible by some older men and some unmarried young men, nearly all of whom could read English through exposure to the English language Melanesian Prayer Book of the Church of Melanesia (Anglican). However, by the end of our time together, many in this group could read Natqgu and some had begun to write it. This was a positive side effect.

6.9.4 Record keeper One of the interns was our primary record keeper and she shared that role with another team member so that there were two of them who knew that task. Besides the RWC data, they also kept track of daily attendance so that we could pay participants for the time they worked.

6.9.5 Glossers and typists We decided to have the glossers and typists in the same room so that they could consult with one another as needed. This gave the Natqgu speakers access to English speakers for help with glossing and the English speakers/typists could check on spelling and handwriting with the Natqgu speakers. There were two glossing/typing rooms.

Photo 10. A glossing and typing room

6.9.6 Glossers It soon became clear, however, that Alex and I needed to each be in one of the rooms because we knew both English and Natqgu. Otherwise, some glossers were putting the same gloss on five or six different Natqgu words. Our final procedure was for two glossers working together to get consensus about words and their meanings. Alex and I gave input, as needed. Once a sheet was glossed it was handed to the typist who sat at the same table and who heard the glossing discussion.

Then, part way through the first week, I noticed another way to improve the quality. As it turns out, a large number of misspelled words were getting past the group with Alex helping with glossing, because although he is a native speaker of both languages he did not know the Natqgu spelling conventions. So I inserted another step in the process and asked them to bring me their glossed words before they were typed. I consulted and then corrected the Natqgu spellings and passed them back to the typist in the other room. Other RWC coordinator-consultants have also found this helpful. I did the same kind of proofing in our room before passing each page to the typist. This helps minimize the amount of database spelling clean up needed following the workshop.

6.9.7 Typists As it turned out, the Solomon Islands handwriting conventions proved challenging to the US typists. They saw the letter <ɑ> as <q> because the tail of the <ɑ> was a little overlong. They saw the letter <Y> as <T> because the left stem was nearly perpendicular to the long stem. This resulted in the word *lue* 'water' in cursive being seen as 'we' because the <l> was a little short and it was connected to the <u>. For example, here is the response sheet for the same domain as above for the questionnaire, 1.3.1.3 river.

Sometimes my command of Natqgu was adequate to correct such mistakes after they had been entered into FLEx without referring to the handwritten response sheets, but at other times I did not know what the word was or why it looked wrong, and I had to consult the response sheets. I was very glad that I decided to take all the response sheets back to the US.

Photo 11. Natqgu response sheet for 'river'

6.10 Animal kingdom challenges The Natqgu speakers also had some difficulties with the animal kingdom, since animals were divided differently from the way that they categorize them. For example, the group that worked on the fish domain only had 17 words, in spite of living along the coast of an island in the middle of the Pacific Ocean. Of these, instead of names of fish they listed six categories of fish, such as river fish, sea fish, reef fish, deep sea fish, surface fish, and coral fish. And even more surprisingly, of the remaining 11 words, three were salt water fish and eight were fresh water fish. We had expected that speakers who had spent their lives in the area would easily come up with at least as many salt water fish as Alex knew from his years of spear fishing while growing up there.

My hypothesis is that the questionnaire prompts for this domain were unclear. For example, the main prompt said, "Use this domain for fish (phylum Chordata, class Osteichthyes)," and few people know what 'phylum' means, nor the Latin words used. Then the next questions asked for species of freshwater and saltwater fish, but none of the freshwater fish are known to them. Similarly, while they do have many of the species of saltwater fish listed, only a few Natqgu speakers would also command the English word for those fish.

Apparently the group thought that "species of fish" meant categories of fish rather than "names of (kinds of) fish." They also did not include many mollusks or crustaceans in the words collected, because it was not clear which category of the domains they should be part of. Mostly, if entered, they were included in "small animals," with no way to distinguish whether they are sea or land animals.

(2) What species of freshwater fish are there?
- *trout, minnow, catfish, mudfish, perch*

(3) What species of saltwater fish are there?
- *anchovy, barracuda, bass, blackfish, bonefish, bream, carp, cod, codfish, crappie, darter, flatfish, flounder, gar, garfish, garpike, goldfish, grouper, guppy, halibut, herring, lungfish, mackerel, marlin, pickerel, pike, piranha, pollack, pompano, sailfish, salmon, sardine, sea horse, shad, sole, spearfish, sturgeon, sucker, sunfish, swordfish, tarpon, tuna, turbot, whitefish, whiting, wrasse*

Alex caught this mismatch of concepts during glossing, and as a result, he and one of the Natqgu glossers worked on adding fish names for the next couple of days. In fact, the Natqgu man borrowed all the fish resource books so he could consult with other speakers over a weekend. The current total of "fish" plus "sharks and rays" is 256, far surpassing the initial 17 terms. This includes terms collected previously, but which had not been tagged by semantic domain. For most of these I have also been able to add the scientific names due to follow up fieldwork in 2017 with access to Natqgu speakers and with help from WorldFish personnel and resources in Solomon Islands (https://www.worldfishcenter.org/). The Natqgu numbers are well short of those found in the language Wayan [fyy], which has 400 fish names, 140 mollusk names, and 45 crustacean names (Pawley 2006:188). Further suggestions for ways to improve this category are summarized later.

6.11 Highlights

6.11.1 Targeted in-depth domain research

6.11.1.1 Ethnobotany   We just mentioned the research regarding fish. Following the workshop, a similar effort took place with regard to trees, tree names, and functions. As part of this, the US team, led by Alex B., and their local guides collected two sets of identical botanical samples and archived them with Myknee Sirikolo, Director of the Solomon Islands Herbarium and of the Solomon Islands Botanical Gardens. One set was sent by him to the Botanical Research Institute of Texas, where the US team had trained. In spite of this focus, the total number of plant-related terms is only 406 to date. This number was reached by including four plant domains (tree, bush, grass, moss), as well as the subdomains of "food from plants." Some of the foods were not tagged as plants, so this needs to be done as part of the database editing. This number is far below the 1200 terms for Kalam [kmh] in inland Papua New Guinea, but understandably more comparable to the 500–600 attributed to Wayan, spoken on a small island in the central Pacific (Pawley 2006:187).

6.11.1.2 Weaving   Also following the RWC workshop, Kim Wells did more in-depth research into the use of a back strap loom for weaving banana fiber fabric. The breech-clout made by this process is worn by the dancers in the *nelc* dance. The process is also used to make carrying bags. She learned the parts of the loom, the processing of the banana fibers from harvesting to weaving, the weaving process, and the origin story about weaving. Eventually, 60 weaving-related terms, some of them endangered vocabulary, were added to the FLEx lexical database. Cultural notes were added in the anthropological "Notebook" tool of FLEx, as described previously. This endangered craft had nearly died out until one man recognized that if he did not take it up the craft would be lost. He became the master craftsman who has since taught it to seven other men in his clan (Wells 2016).

Photo 12. Santa Cruz dancer with woven breechclout

6.11.1.3 Dancing   Another intern, Adam W., studied the Santa Cruz Island *nelc* dance, its history, customs, and costuming, contributing to a total of 25 dance terms in the Natqgu database. He photographed all of the dance rings in the Natqgu-speaking area and many of them elsewhere on the island and helped video record two demonstration *nelc* dances (Walker 2016). The dance is shared by the four related Santa Cruz languages there: Natügu, Nalögo [nlz], Engdewu [ngr], and Noipä [npx].

6.11.1.4 Santa Cruz birds   The island is not known to have many bird species, and the current Natqgu FLEx lexicon has only 18 bird names. This is about half of those available in the Tarburton PDF (no date), which records 36 Santa Cruz birds, with sightings as late as 1999. It gives English and Latin names, and our task would have been to ask speakers for the Natqgu names. While there, we heard about a bird book which workshop participants knew about, but it had been loaned out to

someone who took it to the Reef Islands, 40 miles away. Following that, we heard that conservation scientist, Ray Pierce (http://www.raypiercepacific.com/), would be coming to Santa Cruz to study bird populations on the volcanic island of Tinakula. He had his own copy of the relevant bird book and agreed to take a day to help with the Santa Cruz Island bird name survey. So we did not pursue that domain during the RWC workshop. However, Dr. Pierce became sick, cancelled his trip, and the names were not collected or refined. In surveys he conducted on the island from 2014 to 2016, his website reports that, "the endangered Santa Cruz Ground-dove was still present in good numbers on rat-free Tinakula. [...] Proposed commercial logging looms as a huge threat for the endangered Santa Cruz Shrikebill and Santa Cruz Flying-fox on Nendo."[9] The bird names remain to be explored.

6.11.1.5 House-building The team, led by Donald F., had also planned to study house-building. This would have been facilitated by a line drawing of a house from Malaysia, with parts labeled. The plan was to discover whether there were parallel Natqgu terms for each of the terms listed. His work was derailed by health issues in his host family, making the man who had agreed to help him unexpectedly unavailable for a significant portion of our time in the language area. We expect that there are still numerous terms in that domain that have not been collected. There are currently 120 words tagged for the domain *6.5 Working with buildings*.

6.11.1.6 Nelc dance song lyrics Another anticipated focus of our work was to collect song lyrics sung during nelc dances and to transcribe them in order to identify endangered and archaic vocabulary and to add it to the Natqgu FLEx database. This work also had to be cancelled for several reasons. First, I was unaware that one of the lead dancers we expected to work with had died. Then the person who had stepped into his role was off the island working in his forestry role and unavailable for recording.

I hope to conduct future fieldwork on Santa Cruz Island with other interns to collect bird names, house building terms, and to audio and video record *nelc* dance songs for glossing and documenting expected endangered vocabulary there. Prior to any print publication, all new terms will be added to the lexical database in FLEx and will help it be more comprehensive.

6.11.2 Community feedback The RWC workshop also had significant positive impacts on community well-being (Boerger 2017). In closing speeches and in private comments the participants made statements about:

(a) Feeling inadequate to work on a dictionary but coming to the workshop anyway because of the relationships we'd built here previously or because they knew it was important;

---

[9]Nendo, Nendr, Nedö, Ndeni, and multiple other variants are other names for the island of Santa Cruz, and that is the name given to it by Äiwoo speakers of the nearby Reef Islands.

(b) Recognizing that they'd made a contribution to the history of the island, the vitality of the language, and the future of their children's children's children;

(c) Having come to the workshop without knowing how to read or write in Natqgu (some of them) and leaving it with the ability to read (and even to write);

(d) Thankfulness for the US team for investing the time and money to come and to be willing to live for two and a half months at a lower lifestyle than what we have in the US; and

(e) Appreciation for Alex's essential role as the insider-outsider who understands Natqgu, and them, and English better than anyone else they know.

6.11.3 Two languages discovered In the section above on linguistic challenges, we discussed making note of dialect differences during the workshop. But after some weeks on the island and in discussion with workshop participants and others outside the workshop, it soon became clear that the so-called dialect in Noipä village was not mutually intelligible with any of the other three Santa Cruz languages, but a separate language itself. There were no Noipä speakers at the workshop and no Noipä words were included in the Natqgu database. As it turns out, while Noipä speakers have a passive knowledge of one or two of the other Santa Cruz languages, speakers of the other three such languages all claim to not be able to understand Noipä. Two interns recorded and transcribed a wordlist in Noipä, documenting its uniqueness. A change request was submitted and a new ISO 639-3 code [npx] was granted to Noipä in late 2016.

During RWC workshop domain discussions about deafness and signed languages, we were told that the nephew of one of the participants attends a school for the deaf on Guadalcanal, the main island of the Solomon Islands. Since intern Adam W. also commands multiple signed languages, once we returned to Honiara, the national capital, he made a trip to the school for the deaf and was able to video record a wordlist. It appears that students, faculty, and others at the school are borrowing signs from other signed languages and finger spelling systems, and then using them as a basis for creating their own signs. We thought therefore that Solomon Islands sign language also warranted its own ISO 639-3 code, but on our return we found it had already been identified as unique, with the code [szs].

6.12 Natqgu FLEx database editing In addition to giving some supervision to interns, I spent approximately five weeks editing the words collected during RWC. This included adding glosses for the words not completed on the last day, checking the words, spellings, and meanings collected in the workshop. About half of the alphabet was completed for a first pass and further domains have been edited since then for use in special research topics.

The Natqgu FLEx lexicon work is ongoing. Anyone who has worked in lexicography knows that dictionary-making is an endless task. There was a time when I was somewhat satisfied with having collected 4,820 senses in the process of doing other

tasks. Then following the RWC workshop, I was even more pleased to have over 11,000 senses to edit for publication. But as discussed earlier, there are still some remaining domains to explore in more depth. In addition, I expect to add a few vocabulary items after annotating all the legacy texts. At the end, my hope is that the published Natqgu lexicon comes somewhat higher up the scale of Oceanic languages noted by Pawley (2006).

7. Average of 12,862 raw words collected in "60 group days" We look now at how the Natqgu results compare to RWC workshops in other languages. But, in order to evaluate how effective the RWC workshop process is, we need to recall that all of the words collected are not unique words or senses, as seen above for *temz* 'moon.' Rather, the same word may be collected multiple times during a workshop because it is relevant to multiple domains. Every time that word is collected it is counted and added to the daily total by the record keeper. Each collection group submits its folder with words from a set of related domains and another collection group may have listed the same word. These are reflected in the raw totals for a workshop before any editing has taken place. In the Natqgu RWC workshop, for example, the word *toki* 'knife', which has only one sense, was collected in the following 12 domains:

Table 2. Domains for 'knife'

4.8.3.7 – Weapon, shoot	6.2.4.1 – Cut grass	6.6.3.1 – Lumbering
5.2.1.3 – Cooking utensil	6.2.4.4 – Trim plants	6.6.5.3 – Sculpture
5.2.2.8 – Eating utensil	6.2.5 – Harvest	6.7.1 – Cutting tool
5.4.7 – Fingernail care	6.2.8 – Agricultural tool	7.8.3 – Cut

This means that in the total word count for the workshop *toki* was processed 12 times. This does not mean that there are 12 senses for *toki*, since the domain list is etic and the senses assigned to words are emic. The same thing happened, to a lesser degree, with many other words. Similarly, scribes may also just write the form of a word that a group member calls out, rather than trying to think of the citation form and then writing that. If not corrected by the glossers to conform to a citation form, this yields sets of words like, *eat, eating, he ate, we ate,* etc., such that different forms of what will become a single headword are actually collected multiple times. Furthermore, if there is already a lexical database at the start of an RWC workshop, the groups inevitably duplicate words already collected, since it is easiest to think of the more common words first. These three factors contribute to high numbers of non-unique words being collected during RWC workshops.

That being said, the statistical data discussed next reflects results for twelve single-event RWC workshops held since 2012.[10] The workshops are represented by the two following tables. Table 3 shows totals for six languages with *no lexical database prior to the RWC workshop* and Table 4 shows totals for six languages which *already had*

---

[10] Basic RWC reports are found at http://rapidwords.net/reports, but to augment data there the authors corresponded with those involved to request further information regarding the ongoing progress of dictionary development occurring in the relevant languages.

lexical databases prior to RWC. Word collection groups in these twelve languages gathered from 8,800 to 23,600 total raw words in workshops targeting "60 consecutive, full-time group days", normally conducted within a ten-day period. The average number of raw words collected was 13,762 words. But we assume that one language had an atypically high figure of 23,600 words, in part due to inexperience in writing the language as well as having participants from six areas with slightly different pronunciations for some words, all of which were collected. If we eliminate that artificially high total, the average for the remaining eleven languages is **12,862**, giving us a conservative working estimate for the raw total words one might expect to collect covering all the semantic domains in "60 consecutive, full-time group days".

Table 3. Languages with no lexical database before their RWC workshop

Language	ISO 639-3	Country	# RWC words collected	# post-RWC total	estimated # net senses	% unique senses
Bissa Barka	bib	Burkina Faso	13,802	14,354	9,580	67%
Buli	bwu	Ghana	14,747	–	10,114	69%
Kaansa	gna	Burkina Faso	11,500	15,000	10,000	67%
Madhya-Purbiya Tharu	thq	Nepal	23,600	–	12,500	53%
*Shilluk	shk	South Sudan	15,117	–	8,000	53%
*Syuba	syw	Nepal	12,608	–	3,723	29%

While one week for preliminary editing is recommended following the workshop, in practice this is not always done. Sometimes the editing is started but not completed. And other times it is delayed or incorporated into general dictionary construction, making it difficult to say exactly when the workshop-related editing was complete. One workshop coordinator did significant editing during the workshop, by printing each day's pages and sending them home each night with selected participants who further edited them with a red pen. Then the next day the coordinator input corrections into the database. Adding this step would also save considerable time, and doing it when in the language community means it is easy to ask questions.

For RWC workshops elsewhere, due to time constraints, lack of availability of necessary language speakers, and other obstacles, not all of the language teams have completed the post-workshop editing to their satisfaction. Given this factor of incomplete editing, the authors asked the linguists working in each language to provide either the net total senses or estimates of the net number of senses that will have been processed following the post-workshop editing. The number of entries can be determined by sorting the database by headword, with the total shown in the bottom right-hand corner of the FLEx display screen in the Lexicon Edit tab. The number of senses can be determined by sorting the lexicon by glosses, since each sense receives a separate gloss. Counting in this manner using FLEx is much easier than counting entries, subentries, and senses on representative pages of a print dictionary and then multiplying for estimates, as Pawley (2006:175) did in his comparison of lexicon sizes in a number of published print dictionaries.

But as we wrote this paper, word counts in FLEx lexicons have continued to change, even since we solicited feedback from the linguists working in the 12 languages. Certainly the Natqgu totals have changed. Four of these languages, marked with an asterisk in the tables, have published dictionaries on Webonary (http://www-.webonary.org/). This allows others to view the work, observing whether definitions have been added or whether lexical relationships such as synonyms have been included. The input to Webonary is uploaded from one's FLEx database and the template used for publishing lexical data on Webonary allows the user to indicate whether the data is a rough draft or a formal publication format, with various stages in between. This allows for publishing early and often. Currently, Webonary hosts over 100 dictionaries from 37 countries.

Based on actual figures or estimates given, the percentage of unique senses generated by each of the RWC workshops ranges from 29% to 82% of the words collected during the workshops for languages with no lexical database prior to the workshop, plus those with a lexical database in progress. The average percent of unique senses is 65%. However, the 29% figure in Syuba is quite low, considering that 12,608 words were collected. This was due to several factors. Apparently the number of words collected was artificially high because the scribes and glossers are not yet familiar with the orthography and spelling rules. The more literate the participants are, the fewer spelling discrepancies will be found. While all the words will be counted by the record keeper before they are glossed, it is possible to avoid putting all of them into the FLEx database by having the glossers and typists act as checkers for spelling and citation form. If necessary, one could also make spelling and citation form checking a separate role prior to typing, as Boerger did for Natqgu, as described above about the glossers in "skill set challenges." The database resulting from Syuba RWC also contained quite a few multi-word expressions, most of which were eliminated by the speakers during the editing phase following the workshop, because the phrases were predictable. Another factor in the low senses collected for Syuba were the different inflections of the same verb which were collected when only the citation form (head word) was desired.[11] If we eliminate the Syuba data's outlier 29% and average the remaining results, it pushes the percent of unique senses to 69%.

Table 4. Languages with a lexical database before their RWC workshop

Language	ISO 639-3	Country	# pre-workshop senses	RWC # words collected	estimated # net senses	% unique senses
Bamblang	bmo	Cameroon	2000	11,000	8,000	73%
Chungmboko	cug	Cameroon	1484	12,264	10,000	82%
*Kemedzung	dmo	Cameroon	2751	13,685	11,050	81%
*Kabwa	cwa	Tanzania	1900	8,880	6,000	68%
Rapoisi	kyx	Papua New Guinea	3000	13,091	8,000	61%
Natügu	ntu	Solomon Islands	4820	14,270	11,700	80–82%

Calculations for the languages which had a lexicon prior to the RWC workshop were processed as if they did not. That is, to reach the percent of unique senses, we

---

[11]Mari-Sisko Khadgi, personal communication.

divided the estimated net senses by the total number of RWC words collected. The reason for this is two-fold: a) the early words in most dictionaries are the most common ones, and b) these are exactly the words which are most likely to have been duplicated in the RWC workshop. A quick look at the Natqgu database shows that the 325 headwords untouched by RWC input are those which are dialect variants or individual variants or otherwise controversial or pending in some way.[12] Subtracting these 325 previously existing senses from 11,700 estimated total unique senses reduces the Natqgu percent of unique senses from 82% to 80% and the overall 11-language average from 69% (with Syuba excluded) to 68%.

8. Unique senses average 65-70% of raw total or 8,746 tokens   With the average raw words collected and the average percent of unique tokens based on the 12 languages in the study, we have the ability to predict the effectiveness of a single-event RWC workshop for a community working for "60 consecutive, full-time group days." That is, assuming an average collection rate of 12,862 words, a typical language could anticipate 8,746 unique senses, using the 68% figure. These expectations would need to be adjusted downward where the language in question is highly endangered, such that it had already lost a considerable amount of vocabulary or there were no longer sufficient speakers to form even one word collection group. But in a small, but vibrant language, with a longer history of literacy, the totals might very well be higher.

These projected results of RWC can be compared to lexicons of published dictionaries, as described by Pawley (2006:175) regarding the size of the lexicon in Australian, Austronesian, and Papuan languages. Our own study has one Austronesian language, namely Natqgu [ntu] and one Papuan language, namely Rapoisi [kyx]. The Austronesian dictionaries discussed by Pawley are all from the Oceanic sub-branch of Austronesian, and so is Natqgu. He finds that the Oceanic published dictionaries range in estimated senses from 9,500 to 31,100. In his discussion he notes that few Oceanic dictionaries beyond Polynesia and Fiji have more than 10,000 senses, what he calls "lexical units." Given these figures, the results for Natqgu are expected to surpass the lower range of 9,500 senses and may even exceed the 10,000-sense threshold. For published Papuan dictionaries, Pawley finds that they generally have fewer senses than the Oceanic ones. Their range in senses is from 5,800 to 15,400. Here again, the estimated RWC results of 8,000 senses place Rapoisi's lexicon firmly within the number expected for a Papuan language. Pawley (2006:186) attributed the differences to some combination of three factors: "1) the completeness of the lexical record for each language; 2) the number and size of different vocabularies representing different cultural domains; and 3) the number and productivity of derivational processes, i.e. in the ways of forming new words from roots." The Natqgu and Rapoisi results demonstrate that the single-event RWC workshop is effective in mining a significant, representative lexicon in a relatively short amount of time. The words and senses collected can be filled out later, but the basic skeleton of the dictionary is present.

---

[12]Natqgu also included 681 proper names from the Bible, which were deleted from the pre-workshop Natqgu totals to better reflect actual core vernacular words. These are not part of the calculations in this paper.

9. RWC more effective than "corpus-based approach" for minority languages How does one create a dictionary? One conventionally accepted means is the corpus-based approach. According to this method, the lexicographer collects words from written textual material, which has the advantage of providing words with a particular context revealing a meaning of the word. In fact, we are aware of one minority language database which is being supplemented using the corpus-based approach with input from the translated New Testament. While this is certainly not ideal, since translated material is not produced naturally, it is an effective way to assure that all the terms in a scripture publication are also in the lexicon. The Bible software Paratext (http://paratext.org/) can be linked with FLEx to allow the lexicographer to create an alphabetical list of unique words in the scripture text, which can then be entered from Paratext into FLEx.

9.1 Problems with corpus-based approach Even so, there are at least three problems inherent in the corpus-based approach to lexicography as applied to minority languages. First, while major language corpora contain millions of words, minority languages often lack written corpora of any significant size. Therefore, to create a corpus the linguist-lexicographer is required to collect wordlists and oral texts in multiple genres, then to transcribe them, to translate them, and perhaps even to write them in a community-approved orthography. Since such annotation is exceedingly time consuming, and cannot be accomplished without also doing fairly serious grammatical analysis, many linguists – perhaps even most linguists – collect considerably more texts than they actually annotate, and the typical minority language grammar and dictionary are based on only a subset of the data collected. The second weakness is that, given a minority language's comparatively small corpus size, the corpus will inevitably not reflect the full scope of semantic domains in use in a language. A third, related drawback is that fieldworkers cannot invest sufficient time into text collection and analysis to encompass all the semantic domains and thereby produce a representative corpus for input into a dictionary. The scholar has the pressure to publish and move on to something else. And while it is good to publish from the texts one has collected so that the lexicon being produced is more widely available, the result is that corpus-based minority language dictionaries tend to be less robust, with the number of senses being less than or toward the lower end of what Pawley (2006) reports for the published dictionaries in the three language families he examined.

9.2 Some Oceanic corpus-based results Initially we wanted to compare the results of RWC, i.e., "8,746 unique senses in 60 group days", to results from corpus-based approaches, but we lack the data for a systematic study of the major variables. Instead, we cite anecdotal evidence by looking at five further Oceanic languages,[13] not otherwise part of our statistical study, each with somewhat more semantic domain work than the preceding one.

---

[13]Word and sense counts were provided through personal communications by Catherine McGuckin, Åshild Næss, Greg Mellow and Paul Unger, respectively, for these languages.

Lou [loj] is spoken on Lou Island in Papua New Guinea. When Stutzman arrived there to undertake language work, a retired school teacher had been collecting words from his own language for 20 years. Whenever he went into the villages, he took his notebook with him and wrote down every new word he heard. In 20 years he had noted only a little over 3,000 words, even though he was a mother-tongue speaker. The online Lou dictionary currently has 4,261 entries in its rough draft stage at http://loudictionary.webonary.org. No RWC workshop has been done there.

Gapapaiwa [pwg] is also spoken in Papua New Guinea. The linguists had a FLEx database with 2,397 entries, giving 2,678 senses after 20 years of fieldwork. They are slowly adding more words using the corpus-based strategy and plan a future publication. No RWC has been done to date.

Äiwoo [nfl], a language related to Natqgu, is spoken in the Reef Islands, Temotu Province, Solomon Islands. The recently published dictionary (Næss 2017) was based primarily on corpus-based elicitation plus a short wordlist, and no RWC. It contains 3,845 headwords and 4,294 senses. The headwords include all parts of speech, including affixes and clitics.

Owa [stn] is spoken in eastern Makira Province of the Solomon Islands. While a small RWC collection was attempted, the semantic domain list was primarily used by Mellow and a language consultant to check their work. The published dictionary (Mellow 2013) has just under 4,000 main entries, with 7,122 subentries and senses, including nouns, verbs, number words, pronouns, prepositions, complex nouns, and complex verbs, plus 143 idioms.

Doku [lgr], also called Lengo, is spoken in Guadalcanal Province of the Solomon Islands. After four years' work by linguist Paul Unger, the lexicon had around 1,200 entries. They were obtained by mining words from a limited text corpus and *ad hoc* interactions with Doku speakers. But Unger realized there were many words he had not yet encountered in conversations and other words that he could *never* hope to hear in normal conversation. That is when he heard about RWC elicitation by semantic domain and began using it. There are now over 5,000 head words in the Doku dictionary, after completing about half of the semantic domains using RWC. We give more details about that process in one of the later sections.

Note that all of these results are well below the average 8,746 unique senses expected through the use of a single-event RWC workshop. They are also considerably less than the 9,500 to 31,100 senses found by Pawley in the Oceanic dictionaries he examined (Pawley 2006).

9.3 Before and after RWC workshops The same kinds of low numbers of senses are found in the six languages of our study which had started lexicons prior to their RWC workshops. Four of the languages are from Africa – Cameroon and Tanzania, specifically; and one of the languages, like Natqgu, is spoken by Melanesian peoples of the South Pacific. Table 4 shows those statistics.

The pre-workshop numbers in the African languages all started with around 2,000 senses; the numbers almost certainly reflect the use of the SIL comparative African wordlist (Roberts & Snider 2006), which contains 1,700 semantically arranged items.

The increased effectiveness of RWC is seen in the post-cleanup estimates from these languages. Nearly all of them estimate that the total number of unique senses following cleanup to be triple the size of the starting database.

The two languages from Melanesia show similar results, except that these projects had been in existence longer than the African ones. The Rapoisi workshop on Bougainville in Papua New Guinea was led by an experienced team with many years of experience in the language. Work there had been hampered during a decade of unrest. But even so, they started with somewhat more senses than the African languages, and their estimated net senses more than double the starting number, but do not quite triple it. Natqgu had parallel results to Rapoisi, with 4,820 senses in the Toolbox database, which were later imported to FLEx. We worked primarily with texts, interactions with Natqgu speakers, and direct elicitations related to translation terminology, but were not focused on lexicography or enrichment of the database per se. The estimated 11,700 unique senses following the RWC workshop more than doubled in 18 days what it had previously taken 18 years to accomplish. Editing is ongoing, as well as fleshing out the entries. Remember, the workshop just targets the skeleton of the dictionary.

For all of these, the results are noteworthy because many more words were collected in a short amount of time compared with the lower numbers during the preceding longer period of time. Results support the claim that it is beneficial to do systematic collection by semantic domain during a single-event RWC workshop.

10. Three critiques of RWC unwarranted   There have been three main criticisms of RWC and its predecessor DDP conveyed informally by colleagues who may or may not have tried it. We want to specifically address these here, even though they have been discussed in passing elsewhere in the article. First, it has been offered as a criticism that the domains are not emic. However, as we said in the second section above, this is intentionally and necessarily so. The same is true of nearly any wordlist instrument in use today. Part of what these instruments help us do is to start with etic categories to identify emic concepts through interactions with speakers. The etic categories mean that the same list can be used by multiple languages and it is not skewed toward any particular language, except English, for which dictionaries already exist, so there is no favoritism at work. Perhaps an attempt to make wordlists more emic is reflected in wordlists targeting particular areas of the world, such as the Comparative African Wordlist mentioned previously,[14] or others for Indian and Austronesian languages (Boerger et al. 2016:268).

A second word-of-mouth objection has been that RWC collects too much unusable data that you are still dealing with years later. We did find in Natqgu that there tended to be quite a bit of "junk" and we suggest ideas for making results more streamlined, less cluttered, and more useable in the section on ways to improve RWC results. In general, the suggestion is to go more slowly at the outset, checking the early work so that later work is more refined. However, considering that all the domains are

---

[14]http://www-01.sil.org/silewp/2006/silewp2006-005.pdf.

covered and a great degree of new data is generated, it is not surprising that the great bulk of new data creates considerable post-workshop labor to refine and annotate it. That work is worth it in order to have a more comprehensive dictionary.

The common practice for compiling a dictionary has been to use a short wordlist followed by corpus-based mining, as was done by the first two linguists in the discussion on Oceanic corpus-based results above. Some colleagues have expressed reservations that RWC is not corpus-based. However, as we also showed above, the conventional practice gleans fewer senses than a single-event RWC workshop. That is, in the Natqgu work, there were many semantic domains covered in the workshop for which the Natqgu data contained no representative text, nor the likelihood of obtaining one. However, by covering all semantic domains during the RWC workshop, words were discovered which would have otherwise been missed. Most languages with large corpora already have extant dictionaries. Given this, the RWC workshop method is more thorough and more suitable than corpus-based methods for documentation of small or endangered languages which lack a large, balanced corpus. It is also more efficient for such languages, since the broad span of semantic domains are all collected during one fieldwork trip while the language is still viable and the words collected are immediately available for parsing any texts collected.

11. RWC outside the standardized, single-event workshop format   Three colleagues from the Solomon Islands shared with us about their experiences collecting words by the RWC semantic domains outside the "60 consecutive, full-time group days" workshop structure.[15] Greg Mellow's work was mentioned previously. He explored the vocabulary of the natural world (sea shells, fish, and birds) Then he organised a couple of RWC sessions which yielded a small number of words. But, a greater number of Owa words were collected from natural and translated texts. However, the greatest number of words were collected in collaboration with David Taaru, an Owa translator turned lexicographer, who used words already in the database and considered synonyms, different parts of speech of the word, other affixed forms, different senses, idiomatic expressions, collocations, and antonyms for them. In providing one or two example sentences for each word form, he also discovered new roots to add to the dictionary. Taaru and Mellow subsequently used the RWC semantic domain list to check that all likely productive concepts had been explored. This resulted in a 639-page dictionary (Mellow 2013).

Paul Unger organized four one- to two-week long RWC workshops for Doku [lgr] in various dialect areas over a span of five years. Despite only completing just over half of the semantic domains, about 4,000 words were added using the RWC method, as reported above. Clearly RWC was more productive than previous strategies for adding words to the lexicon. Some participants were able to attend more than one workshop, but none of them was able to attend all the workshops, and none of the workshops had all the same participants. There were three to five word collection groups each time, with four or five people in each group. In the first workshop, three

---

[15]Personal communications from Greg Mellow, Paul Unger, and Andrew VanAndel contributed to this section.

groups worked for eight days, or "24 group days." This means that about half the domains were covered in an estimated "120 group days." The decrease in efficiency could be partially attributed to a combination of having to train and retrain people at the four workshops, losing momentum from not doing the days consecutively, and perhaps also doing more editing while the participants were still present, giving cleaner results.

In addition to the four workshops, Unger also took the novel approach of engaging prisoners in RWC following the ethnic tension in the country. His hopes were: a) that by participating the prisoner(s) would feel a sense of validation in that they could still contribute to their language and cultural community; and b) that the community would receive the prisoner(s)'s contribution and in so doing affirm their ongoing membership in the community. The result was that even though only one inmate participated, the hoped-for outcomes were realized.

Andrew VanAndel for Malango [mln] has met intermittently with groups of speakers from various Malango-speaking communities, covering a few domains at a time. He organized three separate RWC workshops during 2014–2015, each one collecting lexical items from a different dialect area within Malango boundaries. The two-week workshops averaged 4,500 items. Although most domains were repeated and thereby produced many duplicate forms, each successive workshop helped to confirm and correct items collected initially, as well as give some insights into dialectal variation within the language community. The database continues to be reviewed and refined.

Results from the ways these three languages used RWC and the semantic domain list suggest that it is more fruitful to carry out a standardized, single-event RWC workshop. Its advantages would mean training people only once, maintaining momentum by working consecutive days, and involving more of the community.

12. Ways to improve RWC results   Given our experiences in multiple RWC workshops, we suggest the following possible adjustments and improvements for making single event RWC workshops more effective.

Translate questionnaire into LWC– In addition to English, it could have been helpful to also include Pijin prompts for the Natqgu work. In fact, for the South Pacific it could be constructive to create a Pacific-focused list with an adjustment in domains, plus a sample few words each from English, French, Tok Pisin, Pijin, and Bislama. While pidgins and creoles are inherently less rich than most vernaculars, including these could enhance domain comprehension for speakers who are less comfortable in English and French. To mitigate this kind of problem, the questionnaire has already been translated into major languages such as Chinese, French, Hindi, Indonesian, Russian, Spanish, Swahili, and Urdu, and can be downloaded from http://rapidwords.net/resources.

Review domains twice daily before group elicitation– An RWC consultant in Cameroon started going over each Semantic Domain set with the large group in the morning and again in the afternoon to help the groups get an overview of the domains they would be working on during those half-day sessions. It took more time

at the outset, but probably saved time and avoided some of the misunderstandings described above in Natqgu and other RWC workshops.

Provide more training on multi-level hierarchy– Part of the standardized RWC pre-workshop training involves helping key participants understand the hierarchical organization of the wordlists. One way that is done is by using a family tree diagram showing who is related to whom. Genealogies are familiar in most parts of the world and participants seemed to grasp the hierarchy following that lesson. Those who have conducted RWC elicitations without this introduction to the multiple levels have found that speakers are confused by the complexity of the lists. It might be constructive to review this again with all participants in the middle of the first week, once they have interacted with the levels for a few days.

Eliminate irrelevant questionnaire domains in advance– Consultants or participants who are familiar with a culture or a region could eliminate consideration of domains which are essentially irrelevant for that situation. For example, in Natqgu, we were able to eliminate most of section *6.6.2 Working with minerals*, because there is no mining, or work with metals, clay, glass, bricks, or chemicals. Oil and gas were relevant only as products purchased for running equipment. That left *6.6.2.7 Working with stone*. Similarly, in the entire domain of *6.9 Business organization*, there were only 11 total records, with none for *6.9.6 Insurance*. Likewise, *6.6.5.2 Photography* only yielded words for camera, image, photograph (English borrowing), and picture.

Expand record-keeping spreadsheet– We also suggest expanding the tabs in the conventional record-keeping spreadsheet to allow workshops to distribute the work over any number of days or weeks, as needed, while maintaining the formulas that calculate the statistics for the columns and rows. This would allow for projects to incorporate the "60 full-time group day" concept without being limited to six groups working for ten days.

Rearrange glossers and typists for optimal results– Glossers are normally in one room with typists in another. Instead, for Natqgu, we split into two rooms, but had one team of glossers and a typist in each of them, which allowed for more refined results. Other RWC consultants have also found it helpful to rearrange the work flow at this point.

Assign two glossers per sheet– Many cultures like to work by consensus, and this is also true in Melanesia, where Natqgu is spoken. Therefore, we recommend assigning two glossers per sheet or having two glossers working in the same room, allowing them to consult and get consensus. This increases the confidence they have in their work.

Give immediate and frequent early feedback to scribes, glossers, and typists– The RWC consultant should be sure spelling, headword, and data entry conventions are followed. Desire to let the group synergy give momentum to the Natqgu workshop meant that Boerger failed to give adequate feedback and correction at several points in the process. Evaluation times for scribes, glossers, and typists should be built into the training and the first few days of the workshop so that such checks are expected, with spot checks on an on-going basis. The US interns for the Natqgu RWC needed

to ask more questions than they did and not assume they were getting it right or that it was "good enough." They also made English typing mistakes from typing too quickly and not double-checking the work.

Gloss in the collection group and have an outside glosser check it– In Cameroon, they found it effective to make the scribe and glosser the same person, so the glosser was part of the collection group. This could save time, but it also eliminates the check provided by glossers not being part of the collection group. The advantage is that the group still remembers their discussion and the gloss reflects the relevant sense of a word.

Create roles of gloss checker and spelling checker– If glossing is done in a collection group, then it could be profitable to create the role of a "glossing checker", who would check what has been done in the groups before passing it on to typists. Similarly, a spelling checker could go over the spelling of the vernacular words and their glosses before they are given to the typists.

Scan or photograph all collection sheets– In order to bring home an entire box full of Natqgu word collection sheets in their domain folders, Boerger made room in her suitcase by leaving other things behind. Instead, it would have been rather simple, if tedious, to have two interns photograph each page and name it according to the domain number and name. That would have made better use of the interns' time; Boerger would have had less weight to bring back; all the pages of data would be available to file and access electronically and even to archive with the Natqgu corpus. This could be considered a pending assignment for a future intern.

Plan to use a portable solar system for village workshops– It would also be good to plan to have a portable solar system or generator plus a printer, so that RWC videos can be shown, computers charged, and sheets printed out daily for participants to take home. This would increase community exposure to the work and expand the participants' conception of what has been done elsewhere.

Community meetings– Plan a pre-workshop meeting with community representatives so that they can give input into what domains should be included in a general print dictionary or in a school dictionary. For example, how much of the rude vocabulary should be included? Meet with them again post-workshop to confirm decisions.

Schedule workshops during school holidays so teachers can participate– Several teachers joined the Natqgu RWC for the third week and helped give the momentum needed to finish all the domains. Teachers or other more highly educated speakers may better understand the concept of a citation form as opposed to other forms of a word. They can also be assigned tasks of scribe and glosser due to their better understanding of the LWC. Furthermore, teachers are likely to have other input on the dictionary as its future users in their classrooms.

Design a RWC workshop readiness profile– It would be beneficial to establish a RWC readiness profile for communities to complete which has all the expectations of what they need to do, and to provide, clearly delineated. This could be done in consultation with a RWC staff person, preferably face-to-face, in the months before a workshop is scheduled. A draft for such a readiness profile is included as an appendix to this article.

Routinely hold RWC workshops in communities rather than regional centers–One of the consultants who responded to our questions about their RWC experiences said that when she held RWC workshops in community settings, there was more community ownership than when they are held in regional centers. In the community, the people themselves donated food and labor for the meals. Two years after the Natqgu RWC, it is reported that speakers are overtly more interested in their language and await the eventual publication of the dictionary with anticipation.

13. Conclusion We have presented several advantages or benefits of using single-event Rapid Word Collection workshops as a foundation for dictionaries and language development. One such advantage is that the full breadth of the lexicon is available from the outset of a language development project. We also showed that the standardized RWC workshop method of populating a lexicon in "60 consecutive, full-time group days" yields more comprehensive results and is more time-efficient than populating it from texts. Furthermore, it is more effective as a single event than other methods of RWC which split up the elicitation times into multiple workshops. Also, for linguists wanting to make a written record of an endangered language, the RWC workshop method provides the skeleton of a representative dictionary approaching the number of expected unique senses for particular language families. The capturing of all domains also shows promise for delving into culturally significant domains – like the botanical, weaving, and dance domains in Natqgu – and contributes to culture documentation and vitality (Boerger et al. 2017). A further benefit is that involvement by the community in their own language development improves self-image and well-being, as well as supporting language vitality (Boerger 2017). That is, single-event RWC workshops were found to be efficient, effective, and empowering. Given these considerable advantages, the implementation of single-event RWC workshops in field linguistics should become best practice for minority language lexicography.

References

Bartholomew, Doris A. & Louise C. Schoenhals. 1983. *Bilingual dictionaries for indigenous languages*. México: Instituto Lingüístico de Verano. http://www.mexico.sil.org/resources/archives/2384. (9 May 2016).

Boerger, Brenda H. In progress. *Natügu Grammar*.

Boerger, Brenda H. 2017. Rapid word Collection, dictionary engagement, and community well-being. E-poster presented at the 5th International Conference on Language Documentation and Conservation (ICLDC5), Honolulu, March 2–5, 2017. http://hdl.handle.net/10125/41988.

Boerger, Brenda H. 2015. Bible translation as Natqgu language and culture advocacy. In Beerle-Moor, Marianne & Vitaly Voinov (eds.), *Language vitality through Bible translation*, 145–176. Peter Lang: Berkeley Insights in Linguistics and Semiotics.

Boerger, Brenda H. 2011. To BOLDly go where no one has gone before. *Language Documentation & Conservation* 5. 208–233. http://hdl.handle.net/10125/4499.

Boerger, Brenda H., Alexander Boerger, Leonard Menrlwz & Myknee Sirikolo. 2017. On integrating ethnobotany with field linguistics. Paper presented at the 10th International Conference on Oceanic Linguistics (COOL10), Honiara, Solomon Islands, July 10–15, 2017.

Boerger, Brenda H., Stephen N. Self, Sarah Ruth Moeller & D. Will Reiman. 2016. *Language and culture documentation manual*. LeanPub. https://leanpub.com/languageandculturedocumentationmanual.

Boerger, Brenda H., Åshild Næss, Anders Vaa, Rachel Emerine & Angela Hoover. 2012. Sociological factors in Reefs-Santa Cruz language vitality: A 40 year retrospective. *International Journal of the Sociology of Language* 214. 111–152. http://dx.doi.org/10.1515/ijsl-2012-0023.

Boerger, Brenda H. & Gabrielle Zimmerman. 2012. Recognizing Nalögo and Natügu as separate languages: Code-splitting in ISO 639-3. *Language and Linguistics in Melanesia* 30(1). 96–133.

Chowning, Ann & Ward H. Goodenough. 2016. *A dictionary of the Lakalai (Nakanai) language of New Britain, Papua New Guinea*. Canberra, ACT: Asia-Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, The Australian National University. http://hdl.handle.net/1885/107217.

Davenport, William Hunt. 1964. Social structure of Santa Cruz island. In Goodenough, Ward H. (ed.), *Explorations in cultural anthropology*, 57–93. New York: McGraw-Hill.

Koptevskaja-Tamm, Maria, Ekaterina Rakhilina & Martine Vanhove. 2016. The semantics of lexical typology. In Riemer, Nick (ed.), *The Routledge handbook of semantics*, 533-604.

Mellow, Greg. 2013. *A dictionary of Owa: A language of the Solomon Islands*. Pacific Linguistics, DeGruyter Mouton.

Moe, Ron. 2001. Lexicography and mass production. *Notes on Linguistics* 4(3). 150–156.

Ostler, Nicholas. 2003. Desperate straits for languages: How to survive. *Language Documentation and Description* 1. 168–178.

Pawley, Andrew. 2006. On the size of the lexicon in preliterate language communities: Comparing dictionaries of Australian, Austronesian, and Papuan languages. In Genzor, Jozef & Martina Bucková (eds.), *Studies in honour of Victor Krupa*, 171-191. Institute of Oriental Studies, Slovak Academy of Sciences.

Roberts, James S. & Keith L. Snider. 2006. SIL comparative African wordlist (SIL-CAWL). SIL *Electronic Working Papers* 2006-005. http://www.sil.org/resources/publications/entry/7882. (9 May 2016).

Sperlich, Wolfgang B. & Andrew Pawley. 2013. The making of the Lakalai dictionary. Paper presented at the 13th International Conference on Austronesian Linguistics (13ICAL), Taipei, July 18–23, 2015.

Stutzman, Verna P. & Kevin Warfel. 2015. *Rapid word collection coordinator's manual*. SIL Dictionary and Lexicography Services. (In-house publication. Available on request.)

Stutzman, Verna P. & Kevin Warfel. 2017. *Rapid word collection workshop consultant's training manual*. SIL Dictionary and Lexicography Services. (In-house publication. Available on request.)

Tarburton, Michael K. *Nendo (Santa Cruz, Ndendi, Ndeni) island bird checklist, Santa Cruz group, Solomon Islands*. Pacific Adventist University, PNG. http://birdsofmelanesia.net/solomons8.html/santacruz/santacruzproper/nendo.pdf.(28 July 2015).

Walker, Adam. 2016. *The nelc dance of Nedr island: Some initial thoughts on the dance, its cultural import, and spiritual significance*. Dallas, TX: GIAL Academic Forum.

Wells, Kim Beebe. 2016. *Rescued from extinction: The resurgence of Santa Cruz banana fiber weaving*. Dallas, TX: GIAL Term Paper.

Wells, Kim Beebe & Selwyn Balq. 2017. Santa Cruz island banana fiber weaving and its endangered technical vocabulary. Paper presented at the 10th International Conference on Oceanic Linguistics (COOL10), Honiara, Solomon Islands, July 10–15, 2017.

Brenda H. Boerger
Coordinator for Language &
Culture Documentation Services
SIL International
brenda_boerger@sil.org

Verna Stutzman
Coordinator for Dictionary &
Lexicography Services
SIL International
verna_stutzman@sil.org

Appendix: RWC Workshop Readiness Checklists[16]

There are two major stages to RWC readiness: 1) workshop feasibility and 2) workshop logistics.

Stage 1 – Determining Feasibility

Linguist Checklist:
- ☐ **Community support** – Members of the language community are motivated to support a RWC workshop leading to a dictionary.
- ☐ **Orthography** – The vernacular language has a provisional orthography with some level of acceptance by the community, as evidenced by at least a small corpus of texts, which could include native-authored, transcribed from audio recordings, or translated materials.
- ☐ **Good writers and spellers exist** – Several speakers know how to use the orthography to read, write, and spell with reasonable ease, and are available for the workshop.
- ☐ **RWC Questionnaire** – The RWC Questionnaire exists in or has been translated into a language in which at least some members of the language community are bilingual.
- ☐ **RWC consulant** available or linguist has resources to train speakers independently.

Community Checklist:
- ☐ **Personnel to fill roles** – There are enough people with the needed leadership skill sets to fill the roles of Coordinator (often the linguist working in an area), Group Leader, Scribe, Glosser, Typist, Record Keeper, and Logistics Manager for the duration of the workshop.

General Checklist:
- ☐ **Financial** – There is adequate financial backing from within the language community or from the linguist's resources to undertake the workshop, including, as necessary: a) transportation, food, housing, and payment for participants; b) equipment and supplies. Budgets will differ depending on multiple variables.
- ☐ **Workshop dates** – The dates for an RWC workshop have been set and agreed upon by linguist and community organizers at least six months in advance.

---

[16]These are synthesized from Stutzman and Warfel (2015; 2017) and further amended for this publication.

Stage 2 – Logistics in place

3 months in advance

Community Checklist:
- ☐ **Admin roles** of Coordinator and Logistics Manager have been filled.
- ☐ **Language leader roles** of Group Leader, Scribe, Glosser, Typist, and Record Keeper have been filled by qualified people who can serve for the necessary length of time. No individual should perform multiple roles.
- ☐ **Venue** has been identified, rental rates, if any, agreed upon, and reservations made for the workshop's duration.

General Checklist:
- ☐ **Electricity** is available at the workshop site, with a backup plan in place in case primary power fails.

1 month in advance

Linguist Checklist:
- ☐ **Equipment** has been acquired.
  - ☐ One computer for each typist
  - ☐ One computer for record keeper
  - ☐ One computer for linguist
  - ☐ One printer and several reams of paper for printing each day's output
  - ☐ One high capacity flash drive plus backup for sharing changes to FLEx database daily
- ☐ **Resource books** such as vernacular dictionaries, pictures of flora and fauna, etc. have been collected for transport to workshop venue.

Community Checklist:
- ☐ **Transportation**, if needed, has been arranged for the participants to and from venue.
- ☐ **Meals** for participants have been arranged for the duration of the workshop, as appropriate; minimally two tea breaks and lunch daily for weekdays of workshop.
- ☐ **Housing** arrangements have been made, if needed, for the participants in workshop.
- ☐ **Language experts** have been identified, invited, and agreed to come.
- ☐ **Daily payment rates** have been identified, invited, and agreed to come.
- ☐ **Glossing language** has been decided on, if not a given.
- ☐ **Citation form** conventions have been agreed upon by all stakeholders.

<u>**1 week in advance**</u>

Linguist Checklist:
- ☐ **Printing –** Necessary documents have been printed:
    - ☐ 2200 copies of response sheet;
    - ☐ Questionnaire printed, divided up, and placed in the domain folders.
    - ☐ Semantic domain hierarchy lists. One per group. One for admin.
- ☐ **Computers configured –** Computers have been configured to enter data the same way with same version of FLEx and same configuration of Collect Words tool.

Community Checklist:
- ☐ **Supplies** have been acquired: folders, 30 dark pens, 15 red pens, 3 boxes for folders, stapler & staples, scissors, marker, paperclips & clamps.
- ☐ **Training capacity** requirements such as chalkboard, whiteboard, or large paper and appropriate writing materials for group instruction are in place.
- ☐ **Roles** of Group Leader, Scribe, Glosser, Typist, Record keeper, and Language Experts are reconfirmed.
- ☐ **Venue has been prepared** to facilitate workshop activities: tables & chairs or writing surfaces, lunch and break venue, cooking and serving utensils.
- ☐ **Food** responsibilities and coordination is confirmed and menus planned for each day.
- ☐ **Closing event**, if any, has been planned in conjunction with wider community.

General Checklist:
- ☐ **Funds** are in hand to pay the workshop expenses.