Testing mutual intelligibility between closely related languages in an oral society

Gooskens, Charlotte; Schneider, Cindy

Published in:
Language Documentation & Conservation

Publication date:
2016

Testing mutual intelligibility between closely related languages in an oral society

Charlotte Gooskens
University of Groningen

Cindy Schneider
University of New England

This paper describes a new methodology for testing intelligibility across closely related languages and dialects in a traditional oral society in Vanuatu. There are many reasons why it could be useful to establish how well speakers of related varieties can understand one another: such knowledge is relevant to language planning and policy making, and it can shed light on the dynamics of language contact. However, conventional approaches to intelligibility testing, such as 'recorded text testing' (Hickerton et al. 1952; Pierce 1952; Voegelin & Harris 1951), are time consuming to score, and difficult to implement consistently. In Europe, fast and efficient intelligibility testing has been successfully carried out across closely related varieties (cf. Vanhove 2014; Gooskens forthcoming; Schüppert & Gooskens 2011a, 2011b, *inter alia*). However, these methods assume that test subjects are literate and computer-savvy. The methodology discussed in the present paper adapts European methods to conventional 'fieldwork' conditions. In Vanuatu we piloted a picture task and a translation task. Although some words had to be removed from the final analysis, the experiment was successful overall and we anticipate that this method can be fruitfully applied in other oral language communities.

1. Introduction[1] For various reasons it may be interesting to establish the degree to which a speaker of one language variety understands the speakers of another closely related variety. For instance, such information can be used to resolve issues that concern language planning and policies, second-language learning, and language contact. In the past, various methods have been used to establish the mutual intelligibility of small and large languages and dialects all over the world. A major division can be made between investigations in which subjects are asked how well they think they understand the other language (opinion testing) and investigations testing how well subjects actually understand the other language (functional testing). Gooskens (2013)

---

[1] We thank Andrew Gray for his invaluable guidance and advice as we were compiling our initial wordlists in the Raga, Suru Kavian, and Suru Mwerani varieties. We also thank the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands for generously providing us with access to their picture database. Finally we are grateful for useful comments given by two anonymous reviewers to an earlier version of this paper.

provides an overview of various intelligibility tests and discusses their advantages and disadvantages for various purposes.

Opinion testing is an easy and efficient way to get a quick impression of the intelligibility of a language. Subjects are simply asked to rate along scale(s) how well they think they understand the language at hand. It may provide a shortcut to functional intelligibility tests, and in addition, it provides information about people's subjective ideas about the intelligibility of languages. The results should be interpreted with some care, however, as a person's reported language behavior may not be in line with his or her actual language behavior. It may, for example, be influenced by his attitude towards the test language.

Doubting the validity of intelligibility judgments, most researchers prefer to test actual speech comprehension. By means of tests, the degree of intelligibility can be expressed in a single number, often the percentage of input that was correctly recognized by the subject. The methods for testing and measuring are becoming increasingly sophisticated. For example, web-based experiments now provide new avenues for collecting large amounts of data in a relatively short period of time.

However, field linguists working in indigenous communities who may be interested in testing intelligibility across closely related varieties often face challenges that their peers working with more developed communities do not have to contend with. Special problems may arise where subjects have limited or no literacy skills; where the use of modern technology is necessarily constrained; or where the researcher does not speak the test languages fluently. For such situations the recorded text testing (RTT) method has been developed. This method was first used in the 1950s to establish the mutual intelligibility of American Indian languages (Hickerton et al. 1952; Pierce 1952; Voegelin & Harris 1951). The methodology has been standardized and is still being used, for example in the context of literacy programs where a single orthography is developed to serve multiple closely related language varieties (Casad 1974; Nahhas 2006).

The standard RTT method uses a short text recorded from a speaker of the speech variety to be tested. The subject hears the text, with questions about the text in his or her own mother tongue interspersed following the portion that contains the answer to the question. The subjects are required to answer these questions. An alternative approach to the standard question format is the RTT retelling method, which requires subjects to listen to a narrative that has been broken down into natural segments of one or two sentences each, and to retell the recorded text, segment by segment, in their L1 (see Kluge 2006). For each segment, the number of correctly retold core elements are counted, and the segment scores are added up to obtain the overall score for a given RTT text. The main advantage of the RTT retelling method, when compared to the standard RTT question method, is that comprehension of an entire text is tested, rather than selected sections only. A second major advantage is that in many more traditional societies, retelling a story is more appropriate and less threatening than answering questions. An additional advantage is that this method does not require the design of comprehension questions and the translation of these questions into the speech varieties of the communities under investigation. A significant disadvantage

of both versions of the RTT method is that it is very time consuming both to develop the tests and to count the number of correctly retold segments. Furthermore, speakers of various varieties tell different stories, which makes it difficult to make tests of the same level of difficulty for each language variety to be tested.

In the research that we carried out ourselves, we were interested in establishing the mutual intelligibility of various language varieties spoken in the northern half of the island of Pentecost in Vanuatu (see Figure 1). This is a traditional oral, rural society, which makes it difficult to use some of the testing methods that have recently been developed and that depend on modern technology. The Raga language is widely spoken in north Pentecost, and Apma varieties are spoken to the south of the Raga area, in the central part of Pentecost. Suru Kavian and Suru Mwerani are both classified as Apma dialects, along with the Suru Rabwanga dialect. Gray (2012:14) calculated cognacy rates on the basis of a list of 247 words. He found Suru Mwerani and Suru Rabwanga to have 99% cognacy while there is only 90% cognacy between Suru Kavian and the other two Apma varieties. Raga is considered a different language and shares only 60% cognacy with Apma varieties (Gray 2012:14). In addition to these indigenous languages, the creole lingua franca, Bislama, is widely used across the country and most children master this language well before their teens.[2]

We wanted to test how well speakers of the Raga and Apma varieties understood Raga, Suru Kavian, Suru Mwerani and Bislama.[3] Apma speakers (including Suru Kavian speakers) often claim that the Suru Kavian dialect is incomprehensible to Suru Mwerani and Suru Rabwanga speakers (Schneider & Gray 2015), while Suru Kavian speakers seem to have less difficulty understanding the other Apma varieties.[4] Due to the closer relationship, we expect the mutual intelligibility between Raga and Apma speakers to be lower than mutual intelligibility between Apma varieties. By carrying out intelligibility tests in the area, we sought to investigate the relationship between the dialects. Should Suru Kavian be considered an Apma dialect or a separate language? What role does exposure play in the mutual intelligibility between speakers of various language varieties in the area? How well do children understand the test languages in comparison with adults? In addition, by comparing the intelligibility of the three Pentecost language varieties to that of Bislama, we could put the results into perspective. Is it, for example, easier for an Apma speaker to understand Raga than Bislama?

The aim of the present paper is to present and evaluate a new methodology for testing intelligibility under 'fieldwork' conditions in oral language communities. In the next section we present the method in detail. In §3 we present and compare the

---

[2]Despite Bislama's importance in national life, the authors have only observed its usage in North and Central Pentecost as a language of necessity. People prefer using an indigenous variety and only use Bislama when this is not possible. Some younger people who have lived in Port Vila are more inclined to use Bislama nouns instead of adopted indigenous nouns for introduced western concepts such as 'church' and 'school,' and business terminology such as 'export' and 'country.'

[3]We excluded Suru Rabwanga as a test language because it is so similar to Suru Mwerani.

[4]The comparative distinctiveness of Suru Kavian relative to the other two varieties has resulted in Suru Kavian speakers code-switching to Suru Rabwanga or Suru Mwerani in the presence of speakers of these varieties as a form of accommodation. See Schneider (under review) for detail on language practices in North-Central Pentecost.

overall results of the two tests we developed, and evaluate the suitability of the tests for our purpose. More detailed analyses of the results will be presented in future publications.
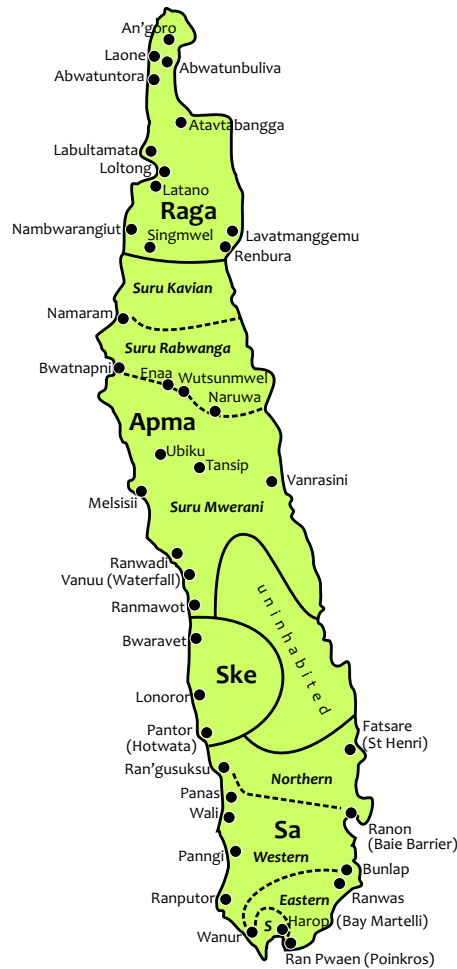
Figure 1. Present-day languages of Pentecost Island (source: personal communication, Andrew Gray, September 1, 2015). Purported language boundaries are drawn in solid lines; purported dialect boundaries are drawn in dotted lines.

2. Method We aimed to test a large number of participants of different ages from different places in north and central Pentecost. Knowing that the RTT method is time consuming, we decided to try to adapt tests that have recently been used for

intelligibility testing, for example in Europe (Vanhove 2014; Gooskens forthcoming; Schüppert & Gooskens 2011a, 2011b). In doing so we had to take a number of circumstances into consideration. For example, we wanted participants of all ages who could not read or write to be able to take the test. We had a limited amount of time to collect our data, so we wanted short, efficient tests that could easily be carried out in the field and test a large number of participants in a short time. Also, it should be taken into consideration that the authors are not native speakers of the test languages.

We opted for word intelligibility tests rather than testing intelligibility of whole texts. An advantage of testing isolated words is that the influence from the context on the understanding of a word can be excluded. This allows us to draw conclusions about the role of individual word characteristics for intelligibility. For example, we were interested in the role of lexical differences for the mutual intelligibility between our test varieties. We also wanted to have a closer look at the intelligibility of cognates to see whether the degree of similarity between the test words and the corresponding words in the participant's native could predict intelligibility of individual words. By analyzing non-cognates separately we could draw conclusions about the role of exposure, since non-cognates would only be understood by participants who have heard the words before. At sentence or higher levels, poor intelligibility is difficult to trace back to specific sources. If the words are presented in a sentence, the context or the situational redundancy is likely to make up for poor intelligibility. We are aware of the fact that a word test is ecologically less valid than a test involving whole sentences or texts. However, a recent investigation (Gooskens forthcoming) comparing the results of three spoken intelligibility tests used to test mutual intelligibility between 16 different languages in Europe show that the results of a word translation task correlated highly with the results of a cloze test set up to test the intelligibility of a text of 200 words ($r =.73$). This seems logical, since to understand a text a listener has to be able to understand individual words. Previous research has shown that, in general, morphosyntactic differences play a small role for intelligibility compared to lexical and phonological differences (Hilton et al. 2013). Gooskens (forthcoming) also asked the participants in the intelligibility test to indicate how well they thought that they understand the test language on a scale from 1 (not at all) to 10 (very well). She correlated these perceived intelligibility scores with the results of the spoken word translation task and got a correlation of $r =.78$. This shows that the results of word translation tasks are a good reflection of the overall intelligibility of the test language as perceived by the participants. On the basis of these results, we conclude that word intelligibility tests can be assumed to be good reflections of the overall intelligibility of a language.

We decided to test spoken word intelligibility with two different tasks: a picture-pointing task and a word translation task. In the picture-pointing task, the participants listened to recorded words in the test languages, and for each test word they were shown a card with four pictures, of which one depicted the test word. The researchers noted down whether the participants selected the picture that correctly depicted the test word. The responses could easily be corrected on the spot, even by

the first author who did not speak any of the test languages. A similar test was used by Schüppert & Gooskens (2011a, 2011b) to test mutual intelligibility of Danish and Swedish pre-schoolers and adults. Their results show the distribution of the scores of the children were near-normal with mean scores of 63% for the Danish and 65% for the Swedish participants, while the scores of the adults had a ceiling effect (percentage of correct recognitions of the test words around 90%). The ceiling effect suggests that adults used more cues than their native language to recognize the stimuli. These cues could be foreign language knowledge, dialect knowledge, or the orthography of their native language.

We deemed the picture-pointing task suitable for our purposes since there is not a strong tradition of literacy in this community. Still, we were concerned that the picture-pointing task would be too easy for our adult Pentecost participants and that a ceiling effect result would not discriminate sufficiently between the intelligibility of the four test languages. We therefore decided to carry out a word translation task in addition to the picture-pointing task. We were fortunate that, except for young children and some elderly women, all people we tested could speak Bislama and could therefore translate the test words into this language. We checked that this was indeed the case by testing subjects' knowledge of Bislama in the picture-pointing task, as well as in that part of the translation task where the participants translated their own variety (and therefore would be expected to translate consistently from their own language into Bislama). The second author, a Bislama speaker, noted down whether or not the participants translated each test word correctly during the test. In this way we were able to calculate individual intelligibility scores immediately after each test.

In the next section we describe how we developed the testing material. In Appendix A we provide a checklist of items to do and to prepare beforehand when developing the material for the picture task to help other researchers interested in conducting similar testing. Except for the preparation of the picture cards, the same steps also apply to the translation task.

2.1 Testing material

2.1.1 Word lists Since we wanted the results of our tests to reflect real-life intelligibility to the greatest extent possible, it was important that the selected words formed a representative sample of the language varieties as a whole. High-frequency words are normally selected from frequency lists to make sure that all words are commonly known and frequently used.

In our case we did not have a frequency list to base our selection of words on. We therefore based our selection on Gray's compilation of 247 words (personal communication, Andrew Gray, 30th May 2012). This list is based on one that Tryon (1976) compiled with the aim to produce a classification of the approximately 105 languages of the New Hebrides, the colonial name for the island group in the South Pacific Ocean that is now known as Vanuatu. Tryon collected wordlists of 309 items from 330 villages across Vanuatu and used them for the computation of cognate

percentages. On Pentecost, lists were collected in five villages, two of which represent Raga- and Apma-speaking villages. However, the location of the villages is not indicated.

The 309 lexical tokens include the Swadesh 100 and 200 wordlists as modified by Samarin (1967:220–223), together with additional items considered suitable for the Island Melanesia area. The 309-item list was reduced to 258 items for the purpose of the computation of the cognate densities because some items were difficult to elicit accurately in the Island Melanesian region. Gray (2012) collected additional material on Pentecost to be able to establish cognacy rates between 10 varieties indigenous to the island. He adapted Tryon's word list in a number of ways. He excluded 49 words because they were derived from similar words in the list, showed lexical overlap or because no simple word existed for the term. He also changed 25 words to simplified or less ambiguous terms.

Since the test would become too long if we were to test all words in this list, we made a selection of 80 words: 40 nouns and 40 verbs. The 40 nouns were randomly selected, with the constraint that we needed to use nouns that could be clearly portrayed in a line drawing. Therefore if a noun that was 'unpicturable' was selected, we would throw it away and select another noun. The 40 verbs were randomly selected.

A few transitive verbs in our word translation test could not occur in isolation; these were presented together with an object. For example, the verb 'hit' required an object, so we used the generic noun 'someone.' Along similar lines, two of the intransitive verbs, 'sit down' and 'lie down,' sounded more 'natural' when followed by another word (see Appendix B, verbs number 17 and 28; also see §2.1.3).

The nouns were tested in both the picture-pointing task and the word translation task. Verbs are generally more difficult to depict pictorially than nouns are, and therefore we decided to test them in the word translation task only. A list of all nouns and verbs in the 4 language varieties is given in Appendix B, together with their English translations.

2.1.2 Picture cards For the picture-pointing task we created 40 cards, each with four pictures (one in each corner of the card; see Figure 2). One of the pictures depicts the stimulus word (the target picture). The target pictures were randomly placed in the top left corner, the top right corner, the lower left corner or the lower right corner to make sure that the position of the target picture on the card would not influence the overall results.

The pictures were all simple black and white drawings depicting the stimulus words as clearly and unambiguously as possible. A majority of the pictures depicting nouns were selected from the picture database developed at the Max Planck Institute for Psycholinguistics.[5] For words that were not found in this database, we added pictures ourselves that were as similar as possible in their drawing style. In order to

---

[5] We would like to thank Anne Cutler for generously providing us with access to the picture database developed at the Max Planck Institute for Psycholinguistics in Nijmegen.

check whether all pictures were unambiguous and recognizable to the target group, we consulted native speakers from Pentecost while developing the material.[6]

Three of the pictures on each card were distractors with the same drawing style as for the target pictures. We took great care to select distractors that showed pictures of concepts that would be just as well-known to the participants as the target pictures. Ninety percent of the words depicted in the distractor pictures are also in Gray (2012), so we can be sure that they are commonly known words. Those words not in Gray that we included as distractors were mostly loan words and represented objects that people are very familiar with, such as 'book,' 'cow,' 'hammer,' 'ladder,' and 'airplane.'[7] Also, we took care not to place homonyms on the same page, or words with high neighborhood density—that is, none of the distractor pictures should have depicted words that sounded too similar to the target word. Nor did pictures with similar meanings appear on the same card.

Each card with four pictures was printed one-sided in A5 format. All cards were laminated and bound into a booklet of 20 test cards. In this way two booklets were made, one with pictures 1 to 20 to be used to test one half of the words, and the other with pictures 21 to 40 to be used to test the other half of the words (see details about the design of the experiment in §2.3). The 20 test cards in each booklet were preceded by four practice cards, to give the participants an opportunity to get used to the task before embarking on the 'real' task.

2.1.3 Morphology across the three indigenous varieties

Since the three indigenous languages (Raga, Suru Kavian, Suru Mwerani) are related to each other, they share similar morphological structures. Many of the nouns in our list are in fact mono-morphemic across the three varieties. Where the nouns are bi- or multi-morphemic, this pattern tends to be shared, as with the nouns 'old person,' 'egg,' 'grass,' 'centipede,' and 'back.' There are inconsistencies in this pattern, however. The words for 'forest' differ in origin and morphological composition in the three varieties: *ute-vono* in Raga is glossed as 'place-blocked/thick' (Andrew Gray, personal communication, February 11, 2016). In contrast, *lee-wa-kina* in Suru Kavian is glossed as 'at-PIECE.OF.PLANT-thing' and *katraba* in Suru Mwerani is mono-morphemic but with *ka* 'TREE.GENERIC' probably an old prefix frozen to the beginning of the word. Along similar lines, 'nose' consists of a single bound noun suffixed with the third singular possessive marker *-n* in both Suru Kavian and Suru Mwerani. Raga, however, uses the compound *hala-n davi-na* 'road-3SG.POSS snot-3SG.POSS,' literally 'path of his snot' (Andrew Gray, personal communication, 11th February 2016). And while the Suru Kavian and Suru Mwerani words for 'urine' require a possessive suffix (third

---

[6]We considered using full-color photos rather than line drawings to avoid cross-cultural difficulties. However, it was difficult to find enough existing photos that did not contain distracting elements. A solution would have been to take the photos ourselves but this would have been too time consuming. Still, we encourage researchers to consider using photos in future work if possible.

[7]There were two non-borrowed distractor words represented in our picture book, which are not in Gray (2012); these were 'land-dive' and 'spider web.' Both of these are well-known concepts in Pentecost society. 'Land-dive' refers to the 'original bungy jump' that is undertaken in South Pentecost between April and June, to the great attraction of foreign tourists.
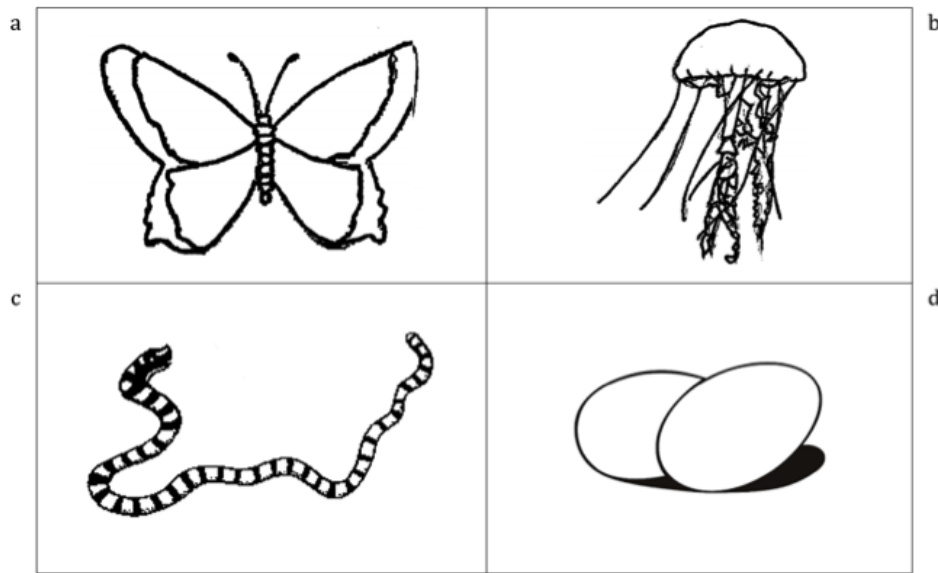
Figure 2. Example of a card used for the picture-pointing task. The target word is 'egg' and the correct answer is picture 'd'.

person singular possessive *-n* is used in our list), Raga *mere* can stand alone as an independent noun and does not require suffixation.

The verb phrase has the same structure in Raga, Suru Kavian and Suru Mwerani. All three varieties express the 40 verbs in our list in the following way:

{IMPERFECTIVE/PERFECTIVE} VERB ({NP OBJECT/COMPLEMENT})

Normally the subject pronoun precedes the perfective/imperfective marker. We chose to express verbs in our list in the third person singular because in all three varieties, third person singular is zero marked. In this way we could maximize participants' focus on the lexical roots themselves and minimize any distraction caused by differences in the form of compulsory grammatical morphemes that co-occur with verbs.

However, there are differences across varieties in the way transitivity is morphologically encoded in the verb root. Raga intransitive forms are indistinguishable from transitive verbs taking an object NP (Andrew Gray, personal communication, 11th February 2016). Similarly, in Suru Kavian, transitivity is not productively marked on the verb. On the other hand, Suru Mwerani verbs tend to take a transitive suffix (which is instantiated in various forms). For example, the Suru Mwerani verb expressing 'he/she drinks (something),' *mwa=mn-i* '3SG.IPFV=drink-TR' has an intransitive counterpart, *min*. Note that the transitive form of drink in Suru Mwerani is a bound root; *mni* requires prefixation. Bound roots are a common feature of Suru Mwerani verbs. The other Suru Mwerani verbs in our list that are also bound are *mtatsi* 'afraid,' *tbo* 'lie down,' *mdeptep* 'light,' *mkan* 'sharp,' *mtsuu* 'sleep,' *mtsiltsil* 'thick,' and *mnipnip* 'thin.' By contrast, Raga and Suru Kavian verb roots are free.

It is not uncommon in all three varieties for simple roots to be reduplicated in the derivation of intransitivity, as for 'dig.' However, many stative verb roots that would appear to be reduplicated are actually frozen forms, for example 'dry' in Raga and 'thick' and 'yellow' in Suru Kavian and Suru Mwerani.

2.2 Speakers and recordings To prepare the test material, we needed to make recordings of native speakers of the three indigenous varieties (Raga, Suru Kavian and Suru Mwerani) plus Bislama. To ensure consistency across the four recordings, we decided to record only adult females. We needed people who were literate to the degree that they could read a list of words in their native language (or in Bislama). Furthermore, the reader needed to articulate words clearly and correctly, and have a strong speaking voice.

Our time in Vanuatu was limited, so we decided to seek out native speakers of these varieties who were living in Port Vila, Vanuatu's capital, rather than travel to each of the areas of Pentecost Island where Raga, Suru Kavian and Suru Mwerani are spoken. In addition, we wanted to record a Bislama speaker who was <u>not</u> from Pentecost but from another island; if the speaker came from within Pentecost, then there was a small chance that her accent would be unduly favored by test participants who came from the same area, and that other participants would be at a disadvantage.

Given our time constraints, it proved impossible to successfully record four women all of the same generation. The second author had only four days to locate and arrange meetings with women in various neighborhoods of Port Vila. Appointments were cancelled or rearranged. 'Good' recordings ended up having to be scratched due to irreconcilable background noise. Ultimately, the four women who made it onto the 'final cut' for the recordings ranged in age from their mid-twenties (Bislama and Suru Mwerani) to their mid-forties (Suru Kavian) and mid-fifties (Raga).

Before making the recordings, speakers were shown the two lists of 40 nouns and 40 verbs and were given as much time as they needed to familiarize themselves with the words. Indeed, reading aloud is a challenging task even for literate people, as schools in Vanuatu have only begun teaching indigenous language literacy relatively recently. Therefore, even women in their twenties would never have been formally instructed in how to read their native language. The readers were therefore being asked to extrapolate from their knowledge of reading French or English (the two official languages of Vanuatu that are taught in school) or Bislama (which is not formally taught in school, but is widely used in the community as a written lingua franca).

Speakers were asked to read each word individually, and to pause between words. They were also instructed to read each word with falling intonation (rather than as a 'list' where only the last word in the list has falling intonation). If they made a mistake, the recording continued and they were asked to correct the mistake. Of course all mistakes were edited out for the final version.

The recordings were made with a Sennheiser 100-P series lavalier microphone attached to a Zoom Q3HD recorder.[8] As mentioned above, environmental noise was an issue for the recordings, but ultimately not an insurmountable one. The Raga speaker worked in a women's craft market and as there was nowhere else to go, she was recorded at her workplace. The whirring noise of sewing machines entered into the recordings, but fortunately we managed to edit most of this from the final cut. Similarly, the Bislama speaker (from the island of Epi) was recorded in the women's market; the same issues were present, and adequately dealt with in our editing process. A positive aspect of recording in a craft market was that the space was filled with textiles—clothing, sarongs, and woven material—and this enhanced the acoustics of the recordings. The Suru Kavian speaker was recorded at the Port Vila Public Library, which was generally quiet—the overhead fans did create some background noise, but we were able to filter this out from the final recording. The Suru Mwerani speaker was recorded in a private office, in generally good recording conditions; there was an echo, but this was filtered from the final recording.

2.3 Design For the nouns, we used a semi-crossed design. Each participant listened to 40 nouns, 10 in each language, and never listened to the same noun twice. First, half of the nouns were presented in the picture-pointing task, and next, the other half of the nouns were presented in the translation task. An exception was made for Bislama, since it would not be meaningful to translate stimulus words from Bislama into Bislama. Therefore, each test participant only translated 15 nouns, 5 from each of the three Pentecost varieties. However, the participants still listened to Bislama words in the picture-pointing task. In this way we could test whether the participants had sufficient Bislama proficiency to carry out the translation task. The results of participants who recognized fewer than 4 of the 5 Bislama words in the picture-pointing task were excluded from the analysis of the word translation task. If we knew in advance that an adult participant did not know Bislama, he or she was only offered the picture-pointing task. Children under the age of 16 only got the picture-pointing task, since we assumed that some of the children would not know Bislama well enough to do the translation task. In addition, we were concerned that many children would lose concentration if they were tested for too long.

We used a design with 8 different versions as shown in Table 1. The languages in versions 1 to 4 were presented in the mirrored order of the languages in versions 5 to 8. Also, for each version, the order of the languages in the translation task is the mirrored order of the picture-pointing task. In this way we made sure that the potential effect of fatigue was the same for all language varieties and all test words in our investigation.

The verbs were only tested during the translation task. Half of the verbs were translated in versions 1 to 4 and the other half in versions 5 to 8. The order of the languages in version 1 to 4 was mirrored in version 5 to 8.

---

[8]In retrospect, a directional microphone (either unidirectional or cardioid-shaped) would have been a better option for making recordings.

Table 1. Design of the experiment. The numbers refer to word number. SK=Suru Kavian, SM= Suru Mwerani, RA=Raga, BI=Bislama. Bislama is not tested in the translation task, which results in empty cells in the table.

Picture-pointing task									
	version					version			
	1	2	3	4		5	6	7	8
nouns 21–25	BI	RA	SK	SM	nouns 1–5	RA	SK	SM	BI
nouns 26–30	SM	BI	RA	SK	nouns 6–10	SK	SM	BI	RA
nouns 31–35	SK	SM	BI	RA	nouns 11–15	SM	BI	RA	SK
nouns 36–40	RA	SK	SM	BI	nouns 16–20	BI	RA	SK	SM

Translation task									
	version					version			
	1	2	3	4		5	6	7	8
nouns 1–5	RA	SK	SM	–	nouns 21–25	–	RA	SK	SM
verbs 1–5	RA	SK	SM	–	verbs 21–25	–	RA	SK	SM
nouns 6–10	SK	SM	–	RA	nouns 26–30	SM	–	RA	SK
verbs 6–10	SK	SM	–	RA	verbs 26–30	SM	–	RA	SK
nouns 11–15	SM	–	RA	SK	nouns 31–35	SK	SM	–	RA
verbs 11–15	SM	–	RA	SK	verbs 31–35	SK	SM	–	RA
nouns 16–20	–	RA	SK	SM	nouns 36–40	RA	SK	SM	–
verbs 16–20	–	RA	SK	SM	verbs 36–40	RA	SK	SM	–

2.4 Background questionnaire Before the listening experiment, the participants were asked a number of questions about their personal background and their language background. Each participant was asked at least the following information: gender, age, current village of residence, native dialect, mother's and father's native dialect, and educational attainment. Adults were also asked what church they were affiliated with. While questions about religion may raise eyebrows in a western interview context, in Vanuatu, as in the rest of the Pacific, religion (almost always some variety of Christianity) forms the basis of one's social life. Therefore, religion may prove to be a factor in explaining individuals' abilities to understand one another. We also asked adults whether they had lived outside Pentecost and if so, for how long, as long periods away from Pentecost may affect individuals' language abilities. In addition, we asked the adult participants whether they spoke and understood each of the test languages (Raga, Suru Kavian, Suru Mwerani, and Bislama).[9]

For each village we also elicited information about the number of inhabitants and which languages were spoken there.

The full questionnaire can be found in Appendix C.

---

[9]We limited the questionnaire to a minimum to keep the testing session short. This means that we had to leave out questions that could be helpful in the interpretation of the results. For example, it would have been useful to know how long participants from other villages had lived in the present village and whether the participant was married to a person from another village.

2.5 Procedure Each testing session consisted of three parts in the case of adults and two parts in the case of children:

1. Background questionnaire

2. Picture-pointing task

3. Translation task (adults only)

The session could be completed within 10 minutes for adults and within 5 minutes for children. First, the participant answered the questions from the background questionnaire (see §2.4). The second author asked the questions orally (in Bislama or, in the case of Suru Kavian or Suru Mwerani speakers, in the native variety of the speaker) and noted down the answers.

Next, the testing session began. It consisted of a picture-pointing task and, for the adults, a translation task. Before the tests, the participants were told which languages they would hear. The participants started with the picture-pointing task, and after a short pause, the adults proceeded to the translation task. The participants from each village were divided equally across the 8 versions. All participants listened to the test words via headphones connected to an MP3 player. A second set of headphones was also hooked up to the MP3 through a splitter. In this way the second author was able to listen to the test words as well and therefore keep track of which word a participant was listening to at a particular moment. An advantage of the use of headphones was that participants were not disturbed by other villagers gathering around them as they took part in the test. At the same time, the physical presence of their peers prevented them from feeling isolated from their group. In this way, the test was less threatening than it would have been if participants were made to go to a separate location, away from others. The testing situation also made it easier to recruit more participants because onlookers could see that the task was not unpleasant, and that in fact it was even enjoyable.

In the picture-pointing task the participants listened to a set of 20 words (5 in each language). The recording of each word was followed by a pause of 5 seconds, during which time the participants were shown a card with a picture of the target word and three distractors (see Figure 2). The participants had to point to the picture that depicted the word that they had just listened to. One of the authors noted down whether the participants pointed to the correct picture. After the pause, there was a beep and the next word was played. The second author turned the pages in the picture book after each beep.

In the translation task, the participants listened to the recordings of 15 nouns and 15 verbs in the three Pentecost varieties. Each word was followed by a pause of 5 seconds, during which the participant gave a spoken translation of the word into Bislama. The second author listened to the test words together with the participant via headphones and noted down for each word whether the participant translated it correctly. If the word was translated incorrectly, the wrong translation was noted down. After each pause, a beep signalled the next word.

Both tests were preceded by a test session of 4 words so that the participants could get used to the task. These trial sessions were repeated when it was occasionally necessary. The participants were given the opportunity to ask questions before starting the real test.

2.6 Participants

2.6.1 Finding participants

We wanted to target speakers of all of the varieties of north and central Pentecost to see how well they understood our three test languages (Raga, Suru Kavian, and Suru Mwerani, with Bislama as a mediating language). We chose to conduct our testing in villages located in the 'heart' of the territories of our target speakers. Thus our chosen test villages were Loltong (for Raga speakers), Namaram and surrounds (for Suru Kavian speakers), Waterfall (for Suru Mwerani speakers), and Tanbok (for Suru Rabwanga speakers). Note that although Suru Rabwanga was not included as a test language because it is so similar to Suru Mwerani, we still wanted to test Suru Rabwanga speakers to see how well they could understand the languages spoken in their neighborhood.

We assumed (correctly, as it turned out, except for the Suru Kavian area) that we would likely find many native speakers of our targeted languages in our chosen locations. Figure 1 shows the Raga, Suru Kavian, Suru Rabwanga, and Suru Mwerani areas within Pentecost Island. Raga and Suru Mwerani / Suru Rabwanga are large varieties, with a 2001 estimate of 6500 and 7800 speakers, respectively (Lynch & Crowley 2001)—this figure would be higher today due to natural population growth. Suru Kavian is reported to have only about 250 speakers (Emil Tawal, personal communication, January 2010), and due to intermarriage and the sale and leasing of customary land, the traditional territory of Suru Kavian speakers has diminished considerably in recent decades. This has placed great pressure on the variety itself, as we shall see below.

Once in a test village, we tested anyone who resided in that village, regardless of whether or not the participant was a native speaker of the language of that area. Women frequently marry into new areas; furthermore, teachers often hail from faraway places. In these situations, the newcomers bring their linguistic backgrounds with them. Some successfully learn the language of their new village; others tend to rely more on Bislama. This linguistic complexity is a reality of village life not only in Pentecost, but across Vanuatu. We therefore decided against 'cherry-picking' our test participants to weed out non-native speakers. We feel that our resulting test sample is a reasonably accurate reflection of the linguistic composition of the overall population of the villages we visited.

We started our testing in Loltong, which is in the centre of the Raga area, on the west coast. Neither of us had ever set foot in Loltong before. People had no idea who we were: two western strangers conspicuously wandering around with iPods, headphones and clipboards in hand undoubtedly made a very unusual spectacle for locals. What became apparent fairly quickly was that we needed a local person, preferably someone with some influence in the community, to accompany us in our

Table 2. Number of participants 16 years and older, males (M), females (F), number of speakers of language native to given area, their mean age and mean number of years of education per village. In parentheses the ranges are given. RA=Raga, SK=Suru Kavian, SR=Suru Rabwanga, SM= Suru Mwerani

Place (target language)	Adults			
	N	Number of native speakers of target language	Mean age	Mean years at school
Loltong (RA)	27 17 M 10 F	22 RA	42.2 (18–69)	8.3 (3–14)
Namaram (SK)	45 31 M 14 F	26 SK (14 SR)	32.8 (16–64)	7.1 (0–14)
Tanbok (SR)	36 14 M 22 F	31 SR	32.1 (16–59)	5.1 (0–12)
Waterfall (SM)	39 23 M 16 F	32 SM	38.3 (18–68)	7.8 (0–15)

quest to find test subjects. After a Raga man offered to lead us around the village and introduce us to people, we had little difficulty in finding participants who were agreeable to doing the task. Our experience in Loltong prepared us well for our visits to subsequent villages. In each place we found a local person to lead us around to different people in the community for testing. Generally, we were taken to public places where there were many people about. For example, in Loltong we were taken to people's homes, but we were also taken up to the fields around the Catholic Mission, which are public access. While it is possible that our test participants were to some extent 'pre-selected' by whoever our guide happened to be friendly with, this potential weakness is diluted somewhat by the fact that we also conducted interviews in public places. That said, the fact that one of our two guides in Tanbok was female, whereas all of our other guides were male, may help to explain why we had more female participants in Tanbok than in other places (see Table 2).

After Loltong, we visited Namaram in search of Suru Kavian speakers. Figure 1 shows that Namaram is located at the border of the Suru Kavian and Suru Rabwanga varieties. Indeed, many residents of Namaram do not speak Suru Kavian. Table 2 shows that only 26 of the 45 people we interviewed there were speakers of Suru Kavian. Of the remaining interviewees, fourteen were Suru Rabwanga speakers, two were Raga speakers, and three were Suru Mwerani speakers. In an effort to find more Suru Kavian speakers, we spent three days visiting small villages in the hinterland of Namaram. This turned out to be a somewhat thankless task. Clearly and very sadly, Suru Kavian is a variety in a state of significant decline.

Tanbok is home to Suru Rabwanga speakers and there were few outsiders. Similarly, Waterfall is home to Suru Mwerani speakers, and we had no difficulty in locating an ample supply of native speakers. Although Waterfall is far enough south for residents to have regular contact with Ske speakers, this was not a significant concern for us as the density of Ske speakers in our dataset was not high.

2.6.2 Statistics on participants Children under 16 years of age did not do the word translation task and, since the main aim of the present paper is to compare the picture task with the word translation task, we only present the results of the adult participants here. Six adults did not do the translation task, or were excluded afterwards, because they did not recognize at least 4 of the Bislama words in the picture-pointing task (see §2.3). In Table 2 we summarize the characteristics of the remaining adult participants in each location. We see that in general there are no large differences between the participants in the four places. The adult participants in Loltong are a bit older and have attended school for a bit longer than the participants from the other places. The Tanbok speakers are on average the youngest and have had the least amount of schooling. Most of the participants are native speakers of the local language spoken in the test area. An exception is in Namaram, where only 26 of the 45 participants speak Suru Kavian. In three places, approximately two-thirds of the participants were male, but in Tanbok, more females than males took part in the investigation. As previously mentioned, this may be due to the fact that our guide for one of our two days in Tanbok was a woman.

3. Data processing For practical reasons, we were not able to put our testing material through a 'pilot run.' Preferably we would have tested whether all pictures for the picture-pointing task were indeed easily recognized by people with the background of our participants, whether all words in the lists were the best tokens of the intended concepts, and whether the recordings of all nouns and verbs were of good quality. Since we were not able to do pre-testing, we decided to test all participants in their own variety in addition to the three other varieties. Words that were misunderstood by more than one participant listening to their native variety were excluded after the experiment. This resulted in the exclusion from the picture task of two Suru Kavian nouns (numbers 29 and 34), five Suru Mwerani nouns (9, 26, 29, 34 and 36) and three Bislama nouns (13, 24 and 35). From the translation task, one Raga verb (number 1), three Suru Kavian nouns (9, 21 and 25) and three verbs (1, 10 and 33), five Suru Mwerani nouns (25, 29, 34, 35 and 40) and four verbs (1, 4, 26 and 27) were excluded.

There are various reasons why words were misunderstood by participants. In many cases the listeners' failure to comprehend their own language resulted from errors and deviations from the methodological objectives explained in section 2. For example, while we strove to pair verbs with generic NPs, we were remiss on one occasion. In the case of the first verb for example, 'he/she bends (bamboo leaf),' the collocation of noun with its object was too unusual and without any context to

facilitate understanding, most listeners simply could not pick this phrase out of thin air.

We tried to produce high quality recordings but *dakbus* 'forest' in Bislama was read much too slowly, resulting in a distorted recording and subsequent comprehension problems. The Suru Kavian verb *nggak* 'fly' is monosyllabic; we believe that this word was simply too short for listeners to comprehend without the aid of context. In the case of *tsivii* 'conch shell,' the Suru Mwerani reader misread the word as *sivii*, which means 'rainbow lorikeet' in Suru Mwerani. This confused test takers who had to listen to the word and point at the corresponding picture.

In some cases, the reason was a complete mystery. For example, in Suru Kavian and Suru Mwerani, many listeners simply did not understand the word for 'nose' in their own language; we have no idea why.

4. Results   As previously mentioned, the objective of the present paper is to compare the picture task with the translation task. Participants younger than 16 years were only tested by means of the picture-pointing task, so we will only focus on adults here. Also, Bislama was only tested in the picture-pointing task and for this reason we do not present the intelligibility results of Bislama.

In Figure 3, the mean results of the two word intelligibility tests in the three test varieties, Suru Mwerani, Raga and Suru Kavian, are presented for each of the four places in Pentecost where we administered the test. We included all participants from each village in the analysis, both native and non-native speakers; see Table 2. In each of the places (except for Tanbok, which is home to Suru Rabwanga speakers), most of the participants were native speakers of one of the test languages. As we expected, participants got high scores when tested in their native variety. This shows that the tasks were not too difficult to undertake. It also confirms that participants spoke Bislama well enough to translate into Bislama.

In Namaram participants displayed great linguistic sophistication. They of course scored well in Suru Kavian, the indigenous variety of the area. However, participants even scored marginally higher for Suru Mwerani than for Suru Kavian. We cannot be sure whether we are dealing with a ceiling effect here, or whether the Namaram participants actually understand Suru Mwerani just as well as Suru Kavian. They also got rather high scores even for the Raga language, but in this case there is no ceiling effect. We assume that these high scores must be attributed to the very high degree of exposure that Namaram citizens have to all three varieties (see Schneider & Gooskens (under review) for a closer analysis of the role of exposure in mutual intelligibility in Pentecost).

The participants in Loltong have comparatively less exposure to Suru Kavian than Suru Kavian speakers have to Raga. Furthermore, since the two varieties share only 60% cognacy (cf. Gray 2012:14), Suru Kavian is a difficult language for Raga speakers to understand. The Suru Mwerani speakers in Waterfall also have difficulty understanding Suru Kavian, but less so than do the Raga speakers, probably due to the closer relationship between the two Apma varieties Suru Kavian and Suru Mwerani. The Loltong and the Waterfall participants also have problems understanding
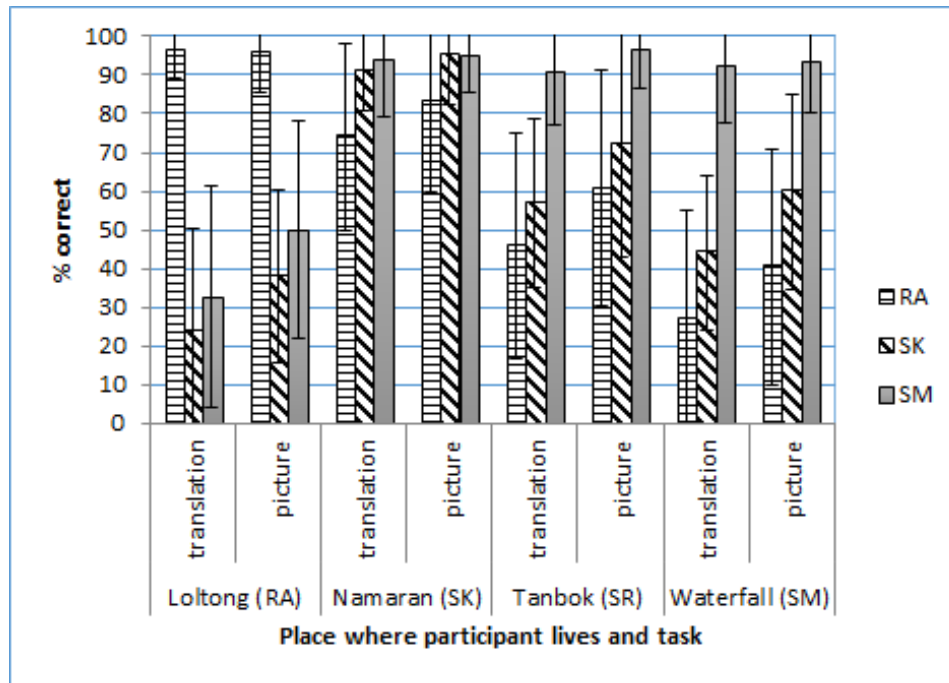
Figure 3. Percentages of correct word recognitions per task (translation task or picture-pointing task) and standard deviation for each of the places and test languages. RA=Raga, SK=Suru Kavian, SR=Suru Rabwanga, SM=Suru Mwerani

each other's varieties. The varieties are not very closely related (again, 60% cognacy; cf. Gray 2012:14) and speakers from the two regions have little contact. Finally, most of the participants in Tanbok are native speakers of Suru Rabwanga, an Apma dialect closely related to Suru Mwerani. Suru Rabwanga speakers understand a fair amount of all three varieties, probably due to a mixture of exposure and linguistic similarities between their variety and the test varieties.

As expected, the results for the picture-pointing task are higher than for the translation task. The picture task was easier than the word translation task, where the participants had to come up with the correct answer themselves. In the picture-pointing task there was a 25% chance of guessing the correct picture that corresponded to the test word. However, a visual comparison of the results of the two tasks for each test location makes it clear that the relative order of the three test languages from easiest to most difficult is always the same for each place. When correlating the results of the translation task and the picture task per language combination (in total 4 places times 3 test varieties = 12 combinations), we get a very high correlation (r =.99, p <.01). Each participant did a word translation task and a picture-pointing task, so we can also correlate at the participant level. Here we get a lower correlation (r =.70, p <.01), which could be explained by the fact that the statistics are based upon each participant only being tested with a few words. If they had been tested with a larger number of words, the results per participant would probably have been more stable

and shown a higher correlation. However, we assume that the fact that we tested a large group of participants in each place in a total of 40 nouns (and 40 verbs for the translation task) make the overall results stable enough to draw solid conclusions. The high correlations show that the picture task and the translation task both test intelligibility in the same manner. If the aim of an intelligibility investigation is to get a quick impression of the mutual intelligibility between language varieties in a given area, it therefore seems that a picture-pointing task is sufficient. This test is easily administered to a large group of participants of different ages and backgrounds and the researcher does not have to master the varieties to be tested. However, as described in §2.1, a lot of care still has to be taken when preparing the stimulus material. For this part of the investigation the researcher is dependent on his or her own knowledge or the knowledge of another language expert about the varieties to be tested to select suitable words and pictures for the stimulus material.

The results of the intelligibility experiments reflect the perception of the participants to a large extent. On the basis of the answers that the participants gave to question 3 in the questionnaire before doing the test (see Appendix C), we calculated mean perceived intelligibility scores. Participants who said that they understood the test language got 1 point; if they answered that they understood a little, they got half a point; and they got zero points if they said that they did not understand the language. The results are presented in Figure 4. We see that overall the scores show a similar pattern as in Figure 3. The correlations between perceived intelligibility scores and the functional test scores are high: $r =.84$, $p <.01$ for translation task and $r =.80$, $p <.01$ for the picture task). The correlations would probably have been even higher if we had used a more elaborate scale than the very coarse one we used. The Suru Kavian speakers claim to understand all varieties well, while the other Apma speakers (and the Raga speakers) are much less positive about their ability to understand Suru Kavian, and even claim to understand Raga better than Suru Kavian. In fact, this is the only discrepancy between perceived and actual intelligibility: in Tanbok and Waterfall, participants have lower confidence in their ability to understand Suru Kavian than Raga, but the results from intelligibility testing find that in fact Apma speakers actually understand Suru Kavian better than Raga. These results confirm the impression Schneider & Gray (2015) got from interviews with Apma speakers that the other Apma speakers consider Suru Kavian difficult to understand. It also shows that on the one hand, the inhabitants of Pentecost were in general well able to judge the intelligibility of the neighbouring language varieties and, on the other hand, that the results of the intelligibility tests reflect the perceptions of the inhabitants reasonably well.

5. Conclusions and future work   The main aim of the present paper is to present a method for testing mutual intelligibility between closely related language varieties under circumstances where the test participants may have no literacy skills and where it is not possible to use modern technologies for testing. We developed two tests, a picture-pointing task and a word translation task. Both tests turned out to be suitable and efficient for testing both child and adult populations. Our results show a high
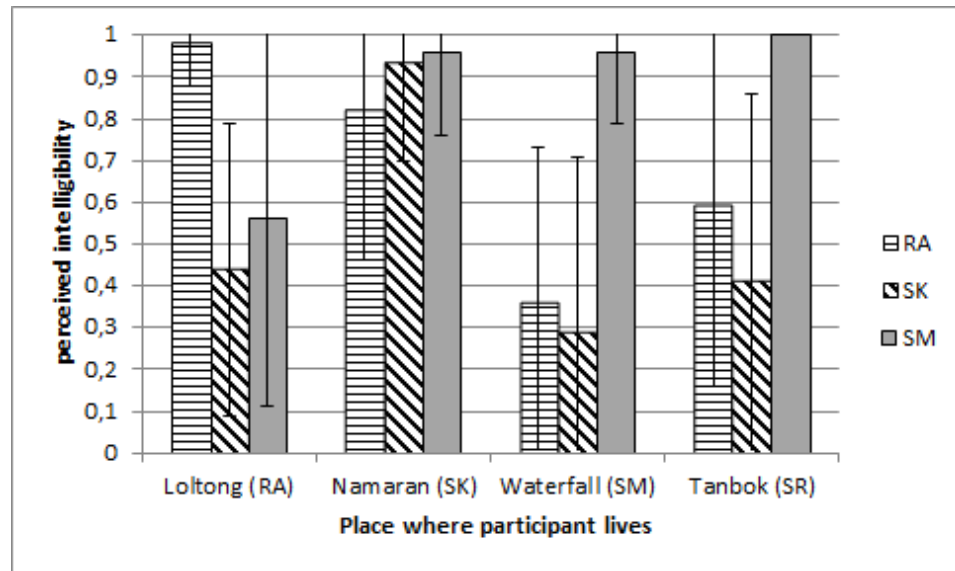
Figure 4. Mean perceived intelligibility scores and standard deviation. RA=Raga, SK=Suru Kavian, SR=Suru Rabwanga, SM=Suru Mwerani

correlation between the results of the two tests, and there was also a high correlation between the two tests on the one hand and intelligibility as perceived by the participants themselves on the other hand. This shows that the tests can be assumed to reflect real-life intelligibility to a large extent and that it is possible to get an impression of mutual intelligibility with a simple and quick test even under circumstances where the researcher has limited knowledge of the test languages. We therefore expect that our tests can be useful to other field workers wishing to test intelligibility in oral communities. For researchers interested in similar testing we added an Appendix with a checklist of steps to take when using a picture-pointing task for testing intelligibility. Most of the steps also apply to the word translation task. In addition, we make the picture books available via http://www.let.rug.nl/gooskens/picture_cards.

Even though intelligibility testing at the word level seems to reflect overall intelligibility well, we realize that other linguistic levels may also play a role in intelligibility. In future work we therefore intend to develop a test that can easily be used under fieldwork conditions to test intelligibility at sentence level and at text level. For example, a task completion task where the subjects listen to a recording and then complete a simple task in relation to what they hear may be a good way to improve the ecological validity of our testing. It may even be possible to develop apps that can be easily used on tablets for our purpose. This would make it possible to register reaction times and correct answers by means of a touch screen.

We presented the overall results of the adult participants in our investigation mainly to show that the results of the two functional intelligibility tests give the same overall picture and that the results of the functional tests reflect the perceived intelligibility scores well. In future publications, we will analyze the data in more detail.

For example, we are interested in comparing the results of adults to the results of the children. The Vanuatu government has recently implemented a policy of vernacular literacy. Children are now to receive the first three years of schooling in a vernacular language. However, due to the large number of languages spoken in Vanuatu, many classrooms have more than one L1. The new policy therefore makes it important to collect more information about how well children can understand the various language varieties, including Bislama, that they may be confronted with at school. The results also allow us to come closer to an answer to the question of whether Suru Kavian should be considered an Apma dialect or a separate language (see Gooskens in press for a discussion of criteria for distinguishing between dialects and languages).

To explain our intelligibility results, we plan to correlate the results with quantitative measures of linguistic and non-linguistic factors that may play a role in mutual intelligibility. For example, exposure is likely to be an important factor, and may explain the asymmetric intelligibility between Suru Kavian and the other Apma dialects. We did not ask the participants questions about the amount of exposure they had with the test languages. However, we are still able to quantify exposure by calculating the percentage of non-cognates that each participant understood. If a participant has had little exposure to a variety, he or she is not likely to understand the non-cognates in our testing material. The participant will probably understand more non-cognates if he or she has been exposed to the variety more often. Also, geographical distance to the place where the test language is spoken might give an indication of the amount of exposure a test participant is likely to have had. To quantify linguistic factors, we will use methods for measuring phonetic and lexical distance that have recently been successful in explaining the results of a project carried out to test and explain the level of mutual intelligibility between closely related European languages (Gooskens & van Heuven forthcoming).

References

Casad, Eugene H. 1974. *Dialect intelligibility testing*. Dallas, TX: Summer Institute of Linguistics.

Gooskens, Charlotte. Forthcoming. The relationship between judged and functionally tested intelligibility.

Gooskens, Charlotte. In press. Dialect intelligibility. In John Nerbonne, Charles Boberg & Dom Watt (eds.), *Handbook of dialectology*. Boston: Wiley.

Gooskens, Charlotte. 2013. Experimental methods for measuring intelligibility of closely related language varieties. In Robert Bayley, Richard Cameron & Ceil Lucas (eds.), *The Oxford handbook of sociolinguistics*, 195–213. Oxford: Oxford University Press.

Gooskens, Charlotte & Vincent van Heuven. Forthcoming. Mutual intelligibility between closely related languages in Europe.

Gray, Andrew. 2012. *The languages of Pentecost Island*. Ashford, Middlesex: Manples (British Friends of Vanuatu) Publishing.

Hickerson Harold, Glen D. Turner & Nancy P. Hickerson. 1952. Testing procedures for estimating transfer of information among Iroquois dialects and languages. *International Journal of American Linguistics* 18(1). 1–8.

Hilton, Nanna H., Charlotte Gooskens & Anja Schüppert. 2013. The influence of non-native morphosyntax on the intelligibility of a closely related language. *Lingua* 137. 1–18. doi:10.1016/j.lingua.2013.07.007

Kluge Angela. 2006. RTT retelling method: An alternative approach to intelligibility testing. *SIL Electronic Working Papers* 2007-006.

Lynch, John & Terry Crowley. 2001. *Languages of Vanuatu: A new survey and bibliography*. Canberra: Pacific Linguistics.

Nahhas, Ramzi W. 2006. The steps of recorded text testing: A practical guide. Unpublished manuscript.

Pierce, Joe E. 1952. Dialect distance testing in Algonquian. *International Journal of American Linguistics* 18(4). 208–218.

Samarin, William J. 1967. *Field linguistics*. New York: Holt, Rinehart & Winston.

Schneider, Cynthia. Under review. Language ideologies, language practices: A case study of dialect idealisation and loss in North Central Pentecost, Vanuatu.

Schneider, Cynthia & Charlotte Gooskens. Under review. The roles of cognacy and exposure in explaining mutual intelligibility between closely related varieties of Pentecost island, Vanuatu.

Schneider, Cynthia & Andrew Gray. 2015. Is it worth documenting "just a dialect"? Making the case for Suru Kavian (Pentecost Island). In Alexandre François, Sébastien Lacrampe, Michael Franjieh & Stefan Schnell (eds.), *The languages of Vanuatu: Unity and diversity*, 197–216. Canberra: Asia Pacific Linguistics.

Schüppert, Anja & Charlotte Gooskens. 2011a. The role of extra-linguistic factors in receptive bilingualism: Evidence from Danish and Swedish pre-schoolers. *International Journal of Bilingualism* 16(3). 332-347. doi:10.1177/1367006911426389.

Schüppert, Anja & Charlotte Gooskens. 2011b. Investigating the role of language attitudes for perception abilities using reaction time. *Dialectologia*. Special Issue II. 119–140.

Tryon, Darrell T. 1976. *New Hebrides languages: An internal classification*. Pacific linguistics series C, no. 50. Canberra: Pacific Linguistics.

Vanhove, Jan. 2014. *Receptive multilingualism across the lifespan. Cognitive and linguistic factors in cognate guessing*. Fribourg (Switzerland): University of Fribourg dissertation.

Voegelin, Charles F. & Zellig S. Harris. 1951. Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society* 95(3). 322–329.

Appendix A: checklist steps to take for the picture task

Step	Considerations
Select test words	- Concepts known to participant - Words common in test language - 'Picturable' - Limit number of words to keep the length of experiment acceptable for participants - Select extra words for trial session
Translate words into all test languages	- Check with native speakers whether all words are suitable tokens of intended concepts
Select target and distractor pictures	- Recognizable to target group - Unambiguous - Uniform style
Make crossed design (see example in Table 1)	- Each participant listens to all languages - Participant should not listen to the same word more than once - All words should be tested in all test languages - Potential effect of fatigue must be the same for all languages and test words (use mirrored order in half of the versions) - Include own language as a control
Prepare cards with pictures of test words and distractors	- Randomize distractors and target words - Randomize place of test words on cards - Replace distractors in case of homonyms, high neighborhood density or similar meanings as target words
Collect picture cards into booklets	- One booklet per version in design - Cards in order of the words in the design - Add extra cards for a trial session at the beginning
Prepare questionnaire	- Collect information necessary to make a selection of comparable groups of participants per language variety - Collect information that may help to explain the results afterwards
Prepare scoring forms	- Print on one page for easy scoring in the field - Arrange optimally for data entry
Make recordings	- Ideally several per test language - Literate speakers from same gender and age group - Same recording equipment (directional microphone) - Quiet surroundings with good acoustics - Give reader time to read through word list beforehand

Continued on next page

Continued from previous page

Step	Considerations
	- Instruct speakers to speak clearly and not too fast with a falling intonation and with a short pause between each word - Ask speaker to repeat words if he/she makes reading errors
Selection of recordings	- Same sound quality - Same reading speed - Same voice quality
Improve recordings with speech processing program	- Adjust level of recording - Suppress noise
Cut words from recordings	- Check that all words are still recognizable by native speakers
Prepare one file per version in design	- Assemble words in correct order - Insert pauses and beeps between words - Save one file per version onto MP3 player
Run a pilot with target groups	- Check that the pauses between words are not too long or too short - Control for ceiling or floor effects
Testing session	- An equal number of participants per version - Sufficient number of participants for statistical analysis - Researcher explains purpose and nature of experiment - Participants listen through headphones - Researcher listens via a second set of headphones connected via a splitter - After the trial session participants are given opportunity to ask questions - Researcher notes down answers

Appendix B: word lists

Nouns

	English	Bislama	Raga	Suru Kavian	Suru Mwerani
1	*old person*	wan olfala	bwatmetua	asi tabwas	atsi tebwet
2	*fly*	flae	lango	lang	leng
3	*village*	vilej	vanua	veni	vini
4	*dog*	dog	vwiriu	kuli	kuli
5	*men's house*	nakamal	gamali	wunis	kamel
6	*egg*	eg	idolin manu	undelin bweeil	dulun bwihil
7	*shark*	sak	bageo	beke	beke
8	*ant*	anis	butubutu	busbus	butbut
9	*conch shell*	pupu sel	taiva	tevi	tsivi
10	*taro*	taro	bweta	bwet	bwet
11	*grass*	gras	bwatbwanea	wavelih	wavilih
12	*water*	wota	wai	sileng	sileng
13	*hill*	hil	vusi	rep	rep
14	*rope*	rop	gao	kaawa	kaawa
15	*Malay apple*	nakavika	gaviga	kavik	kavik
16	*rat*	rat	garivi	koup	temwa
17	*flying fox*	flaengfokis	bwaratu	bwaras	bweret
18	*pig*	pig	boe	kavi	bo
19	*kava*	kava	malogu	seni	sini
20	*snake*	snek	teltele	teltel	teltel
21	*fence*	fenis	ara	kooh	koo
22	*mosquito*	moskito	namu	tabwaken	tabwaken
23	*pawpaw*	popo	uhi	bwarurit	bwarus
24	*forest*	dakbus	utevono	leewakina	katraba
25	*centipede*	milpat	bwanseresere	bwanseesee	bwanseesee
26	*banyan*	nambangga	ramute	wale	baga
27	*bamboo*	bambu	bua	vaskubu	vatkubu
28	*bird*	pijin	manu	bweeil	bwihil
29	*nose*	nus	halan davina	ngosin	ngusun
30	*knife*	naef	bua	bu	bu
31	*people*	pipol	sinombu	taris	tarut
32	*chicken*	faol	toa	madeede	mwateete
33	*stone*	ston	vatu	vas	vet
34	*wave*	wef	navo	nap	nap
35	*cliff*	klif	bahara	bas	bas
36	*sweat*	swet	mamaono	datuwan	tatsuwan
37	*palm*	kokonas	niu	ni	kul
38	*teacher*	tija	vagahi	sasarakan	sesesrakan
39	*back*	bak	gatiguna	tekun	tsukun
40	*urine*	pispis	mere	mesin	misin

Verbs

	English	Bislama	Raga	Suru Kavian	Suru Mwerani
1	*he/she bends (bamboo leaf)*	hemi benem (lif bambu)	mwa mbigi (varasi)	bwii (karas)	bwiri (karas)
2	*it is black*	i blak	nu meto	te met	te mee
3	*it is cold*	i kolkol	mwa masisi	te ilili	mamamdidi
4	*he/she digs*	hemi dig	mwa nggelgeli	nggilkil	mwigilkil
5	*he/she drinks (water)*	hemi dring (wota)	mwa min (wai)	min (sileng)	mwamni (sileng)
6	*it is dry*	i drae	nu ngalangala	te mamah	te ragah
7	*he/she falls down*	hemi foldaon	mwa hovi	mweiah	mwaiah
8	*he/she is afraid*	hemi fraet	mwa matagu	mataasi	mwamtatsi
9	*it is finished*	i finis	nu nogo	te nok	te nok
10	*he/she/it flies*	hemi flae	mwa nggaga	nggak	mwegak
11	*it is good*	i gud	nu tavuha	te kabis	te gabis
12	*he/she hears something*	hemi harem	mwa rongoe	mworonga	mworongo
13	*he/she hits (someone)*	hemi kilim (man)	mwa wehi (atatu)	mwih (asi)	mwahi (atsi)
14	*he/she holds (the rope)*	hemi holem (rop)	mwa hanggo (gao)	buh (kaawa)	mwabuhu (kaawa)
15	*it is hot*	i hot	mwa aruaru	mwapmwap	mwapmwap
16	*he/she laughs*	hemi laf	mwa mana	man	man
17	*he/she lies down (on the bed)*	hemi ledaon (long bed)	mwa eno (la bata)	mween (ili kabwal)	mwatbo (li kabwal)
18	*it is light*	i laet	nu mamara	te mendepdep	temdeptep
19	*he/she plays*	hemi pleiplei	mwa mwo-somwoso	mamaplel	mamaplel
20	*he/she pulls (the rope)*	hemi pulum (rop)	mwa rav (gao)	mworap (kaawa)	mwerava (kaawa)

Continued from previous page

	English	Bislama	Raga	Suru Kavian	Suru Mwerani
21	*he/she pushes (someone)*	hemi pusum (man)	mwa sogai (atatu)	dambesok (asi)	mwedaba-soni (atsi)
22	*he/she rests*	hemi spel	mwa mambu	membe	mebe
23	*he/she runs*	hemi ron	mwa rovo	mworop	mworop
24	*he/she sees (someone)*	hemi lukim (man)	mwa nggita (atatu)	nggida (asi)	mwigita (atsi)
25	*he/she sews*	hemi somap	mwa ligoligo	solsol	mwosolsol
26	*it is sharp*	i sap	nu mamagani	matan dok	temkan
27	*he/she sings*	hemi singsing	mwa lol iboi	sasah	mwesaasaa
28	*he/she (still) sits*	hemi sidaon (yet)	mwa ndogo (radu)	dok (ngamwa)	mwesadok (ngamwa)
29	*he/she sleeps*	hemi slip	mwa maturu	mesuu	mwamtsuu
30	*he/she spits*	hemi spit	mwa lodo	gasuu	mwegasuu
31	*he/she stands up*	hemi stanap	mwa ndomare	si boswos	mwidi boswos
32	*he/she takes (the ship)*	hemi tekem (sip)	mwa lai (wangga)	lap (angga)	mwali (aga)
33	*it is thick*	i tiktik	nu vonoga	te metiltil	temtsiltsil
34	*it is thin*	i tintin	nu manevnevi	te menipnip	temnipnip
35	*he/she throws (a stone)*	hemi sakem (ston)	mwa mbohai (vatu)	bohni (vas)	mwabohni (vet)
36	*he/she vomits*	hemi traot	mwa lua	liait	mwiliaut
37	*it is wet*	i wetwet	nu meho	te mes	te mes
38	*it is white*	i waet	nu maita	te ndap	te dap
39	*he/she works*	hemi wok	mwa rorovoga	muum	muum
40	*it is yellow*	i yelo	nu angoga	te sesende	te sesede

Appendix C: questionnaire

1) Village Survey Questions (ask once per village)

a. How many people live in this village?_____

b. How many people speak as a first language

Suru Kavian	Suru Rabwanga / Suru Mwerani	Raga	other

2) Background Questions (ask all participants)

a. Name_____

b. Gender_____

c. Current village _____

d. Birthdate _____

e. Native dialect_____

f. Father's dialect_____

g. Mother's dialect_____

h. Education _____

i. Religion _____

j. Lived outside Pentecost, and if so, doing what_____

3) Language Usage Patterns (ask only adults)

	Speak	Understand
Raga		
Suru Kavian		
Suru Mwerani / Suru Rabwanga		
Bislama		

Charlotte Gooskens
c.s.gooskens@rug.nl

Cindy Schneider
Cindy.Schneider@une.edu.au