# Implementing free-listing: possessive classifiers in Oceanic *

Michael Franjieh, Alexandra Grandison, Anne-Laure Dotte and Greville G. Corbett
University of Surrey, University of Surrey, Université de la Nouvelle-Calédonie, University of Surrey

**Abstract:** We evaluate the design and implementation of the free-list experiment, which is a relatively easy method for exploring the membership of semantic categories (Weller and Romney 1988). Different factors that could affect the replicability and validity of the experiment are explored, and these are balanced with the need to work sensitively with speakers of endangered and minority language communities. By including aspects of a Participatory Research approach (van der Riet and Boettiger 2009), such as building rapport and respecting participants' knowledge, the experimenter can extend the free-list experiment to include wider discussions around the linguistic categories under study. We include a case study from our research on Oceanic possessive classifiers to show that a free-list experiment results in a wealth of data, and offers up opportunities for discussing and valorising different speakers' understanding of linguistic categories.

**Keywords:** free-listing, Oceanic, possessive classifiers, Participatory Research, rapport

## 1 Introduction

Free-listing is a technique used to isolate and define a domain (Weller and Romney 1988) and to explore salient members of a domain (Stausberg 2011). Quite simply, participants are asked to list members of a domain or category specified by the researcher. Lists collated from multiple participants can be combined to uncover salient members of a category. Frequency of an item across lists produced by different participants and order of an item within individual participant's lists have been shown to be indicators of the salience of a category member. These metrics are often combined to give an overall numeric value indicating salience, such as Smith's Salience (Smith et al. 1995) or the Cognitive Salience Index (Sutrop 2001). Though we do not address salience directly in this paper, it should be noted that identifying salient members of a domain is one of the main types of analyses that free-listing is used for.

Our main focus in this paper is the design and implementation of the free-listing technique, rather than the analysis of the data collected. At each stage of the design and implementation, decisions must be made that can affect the replicability of the experiment, and may lead to either priming of participant responses or possible deviations from pre-determined experimental protocols. Working with minority and endangered language communities is a complex matter in itself (c.f., Grinevald 2007), but also offers challenges to an experimental approach, as speakers are unlikely to be accustomed to strict psycholinguistic protocols and must be carefully guided through the process. We

* The support of the ESRC under grant: ES/R00837X/1 'Optimal categorization: the origin and nature of gender from a psycholinguistic perspective' is gratefully acknowledged. We thank Penny Everson and Lisa Mack for their help in the preparation of the manuscript. We wish to thank the 121 participants from the six different language communities in Vanuatu and New Caledonia. In particular, we wish to thank Willie Salong from the North Ambrym community; Elder Simeon Ben from the Vatlongos community; Korau Melio from the Lewo community; Melkio Wulmele and Adam and Hester Pike from the Merei community; Willion Phadom from the Nêlêmwa community; and Wejë Bae from the Iaai community. This research has undergone a full ethical review and been granted ethical approval by the University of Surrey ethics committee.

argue that following aspects of a Participatory Research approach (van der Riet and Boettiger 2009) treats experiment participants more like traditional language consultants and can help build important relationships between the experimenter and members of a language community, while still retaining the replicability of the experiment. This interaction between researcher and participant is important for all experimental and semantic research with minority and endangered language communities as research is more often than not conducted outside the laboratory environment and within the community and homes of the speakers.

Free-lists have been used in a wide range of academic fields including anthropology (e.g., Gatewood 1983), the study of religion (e.g., Stausberg 2011), ethnobotany (e.g., Vogl et al. 2004) and psychology (e.g., Henley 1969; Rosch et al. 1976). The main focus within psychology has been the study of colour terms (e.g., Grandison et al. 2014; Davies and Corbett 1995; Smith et al. 1995). With the exception of psycholinguistics, there is a noticeable lack of research in linguistics that uses the free-listing technique.

Free-listing tasks are often used as an initial starting point to aid in the definition of a particular domain for further study (Weller and Romney 1988) and to give an up-to-date understanding of a particular category. For example, Grandison et al. (2014) used a free-list task to elicit colour terms used by Otjiherero speakers in north-western Namibia. The free-listing task uncovered a new borrowed term, *girine* 'green', which was used frequently by respondents. This new term was the basis for further experimental investigation into how this newly acquired term interacted with another older borrowed colour term *burou* 'blue-green' and whether this change in the language impacted speakers' perception.

There is a distinction in the methodology of data collection and in relationships between the researcher and participant depending on the type of research being undertaken. In quantitative studies the researcher aims for objectivity and attempts to limit bias between the research subject and experimenter. An objective research approach implies limited interaction between researcher and participant in order to eliminate treating participants differently (Sandoval 2017). With qualitative research there is often an assumption that the interaction between researcher and participant is non-neutral and that the building of rapport with participants is an important research methodology (Sandoval 2017). Undertaking quantitative/experimental research in minority and endangered language communities should aim for objectivity and replicability, while respecting participants' individual needs and expectations. Participants are often not familiar with psychological experimental techniques and so a great deal of interaction between researcher and participant can be required. These interactions should be guided by clear experimental protocols wherever possible to ensure that researchers interact in a consistent way and explain the experimental procedures clearly so that participants feel confident and comfortable. Rapport is the level of ease or comfort between a researcher and participant and can be built in a relatively short time, for example in data collection through interviews (e.g., Given 2008) or, in our case, during psycholinguistic experiments. Building rapport in a short space of time is aided by respecting participants' knowledge, beliefs, well-being and wishes (c.f., Rice 2006).

Participatory Research is a broad term used to refer to a range of research methods that make use of participatory inquiry. It originally emerged in response to the dominance of the positivist paradigm (the notion that human behaviour is best understood through observation and experimentation) and as a critique of its application within research conducted in the developing world (van Vlaenderen and Neves 2004). This was largely due to the perceived imbalance of power between the observers and the observed within this context. One of the fundamental assumptions of the Par-

ticipatory Research process rests on the importance of engaging the people on whom the research is focused (Chambers 1994). Furthermore, all people have a right to contribute to processes that claim to generate knowledge about them, irrespective of age, gender, or level of education (e.g., Gaventa and Cornwall 2006). Indeed, as Wallerstein (1999: 43) suggests, Participatory Research aims to "produce knowledge that clarifies and seeks to change the maldistribution of power and resources and can probe the power relationship between researchers and research participants."

## 2 Current study

Our study investigated the possessive classifiers found in many Oceanic languages, a subgroup of the Austronesian language family, and we give the essential background for readers to appreciate the issues involved. This is necessary since possession is "usually one of the more complex aspects of the grammar of Oceanic languages" (Lynch et al. 2002: 40).

There is a distinction between direct and indirect possessive constructions, which maps onto the difference between inalienable and alienable possessions, respectively. Our study focusses solely on the indirect marking strategy, which utilises a range of possessive classifiers that carve up the semantic space of alienable possession. We do not investigate direct possessive constructions and therefore the discussion of possession is purely for background information and to provide information on potential grammaticalisation pathways from directly possessed nouns to possessive classifiers (see section 5 for further discussion).

Importantly for our interests, Oceanic languages vary in the number of possessive classifiers, from two[1] in Merei, spoken in Vanuatu (Chung 2005), to well over 30, such as in the Micronesian language Chuukese (Benton 1968). Proto Oceanic has been reconstructed as having three possessive classifiers, one for possessions that are eaten, one for possessions that are drunk, and a residual, general classifier for everything else. This earlier three-way distinction is still represented by Lewo (Central Vanuatu), one of the languages in our study:

(1) a. ka-la kinan-ena
FOOD.CLF-3PL eat-NMLZ
'their food'

b. ma-na wi
DRINK.CLF-3SG water
'her (drinking) water'

c. sa-u sisi la lua
GENERAL.CLF-1SG child PL two
'my two children' (Early 1994: 210, 214)

An aspect of Oceanic possessive classifiers which is often discussed is the ability of a possessed noun to occur with different classifiers dependent upon the real-world relationship between the possessor and possessed (Lichtenberk 1983). This ability of a noun to occur with different classifiers is

[1] If a language has only one way of marking indirect possession, then the language shows a split between direct and indirect possession. Only in a language whose indirect possessive constructions show a further split among semantic lines can we say that it has a system of possessive classifiers, thus the minimum amount of classifiers is two (see Lynch et al. 2002: 41, for a discussion on this).

termed overlap or fluidity (Lichtenberk 2009; Lynch 1973). Compare the use of the DRINK classifier in (1b) to the use of the GENERAL classifier in (2) in Lewo.

(2) sa-na wi
GENERAL.CLF-3SG water
'her (washing) water' (Early 1994: 214)

Oceanic languages differ in how free or rigid the association is between a possessed noun and possessive classifier. In North Ambrym (Central Vanuatu), another language in our study, this overlap is much more restricted, as shown in (3).

(3) a. mwe-ng we
GENERAL.CLF-1SG water
'my water (for drinking, washing, etc.)'

b. *mwene-ng we
GENERAL.CLF-1SG water
intended: 'my water' (Franjieh 2016: 95)

An additional dimension is that possessive classifiers need not be semantically homogenous. Some Oceanic languages often allow non-edible or non-drinkable possessions to occur with the FOOD and DRINK classifiers respectively. Again, we use Lewo as our example:

(4) a. ka-la piaki
FOOD.CLF-3PL pot
'their pot'

b. ma-miu yuma
DRINK.CLF-2PL house
'your house'
(Early 1994: 210, 212)

Direct possession patterns with semantically inalienable possessions, such as body parts, parts of wholes and kinship terms, as shown in (5) from Lewo. In direct possessive constructions possessor suffixes attach directly to the possessed noun.

(5) a. la-la lala
leg-3PL PL
'their legs'

b. vine-u
sister.of.male-1SG
'my sister (male speaking)'
(Early 1994: 123)

Our study focussed on a sample of six Oceanic languages spoken in Vanuatu and New Caledonia (see figure 1), which vary in the size of their possessive classifier inventory and in their semantic homogeneity: Merei (2), Lewo (3), Vatlongos (4), North Ambrym (5), Nêlêmwa[2] (20) and Iaai (23).

More broadly, comparing across languages allows us to compare categories that on the surface appear similar, but in fact differ in intriguing ways. For example, we can identify sets of related

[2]Bril (2002) lists 13 possessive classifiers for Nêlêmwa. However, we also tested the bi-morphemic compound classifiers, such as *shâ-pwee-* 'CATCH-LINE' and *shâ-doo-* 'CATCH-SPEAR' which indicate different methods of catching fish; incipient classifiers which occur with modified possessed nouns; and the general prepositional possessive construction, as there is no GENERAL classifier in Nêlêmwa.

lexical items in a language (e.g. black, white, red, yellow, purple). While we might naively think that they are straightforwardly comparable, a wealth of research demonstrates that they are not: domains such as colour are structured (Berlin and Kay 1969; Davies and Corbett 1995). We start from the assumption that the classifiers we wish to investigate will exhibit similar structuring, and we therefore turn first to established methods for uncovering it. Given the varied situation described above, which we had chosen deliberately for our project, a free-listing task was a useful first step. We used a free-listing task (i) to establish the current semantic domains represented by each classifier; (ii) to give an indication of the frequency of overlap between a possessed noun and the inventory of classifiers; and (iii) to create a list of possessed nouns that could be used for further experimentation.

The free-list was one of three experiments that participants took part in.[3] The other two experiments were a video vignette task and a card sorting task. In the video vignette task participants watched short video clips of people interacting with different objects to investigate whether the manipulation of different interactional contexts prompts different possessive classifiers. In the card sorting task participants grouped together images based upon different instructions. The task investigated whether the possessive classifier system affects general cognition. What is important to note is that the whole set of experiments lasted on average between two and three hours. This offered much more opportunity for building rapport and engaging with participants' knowledge of the classifier systems than if we were running a single experiment. Spreading out the different experiments across several days with each participant would have been ideal, as this would have allowed repeated contact and in turn help build a better relationship between the participant and researcher. Unfortunately, time constraints did not allow this more leisurely timetable.

We first discuss the planning and implementation of free-listing (section 3). We then look at data processing as this too can affect replicability and comparison of data across similar experiments (section 4). In section 5 we discuss initial descriptive results gained from the free-listing experiment about the current state of affairs of the possessive classifiers in our sample languages.

## 3 Planning and implementation

The planning and implementation of free-lists appears at the outset to be relatively simple. We ask participants to list entities that are associated with a particular semantic or cultural domain. For example, we may want to understand how a particular linguistic community defines certain linguacultural domains such as the botanical or zoological concepts of 'tree' or 'bird'. Items that are elicited first within a participant's list and/or appear frequently across all participants' lists are in some sense more prototypical or salient (Smith et al. 1995; Weller and Romney 1988). On the other hand, items mentioned low down on a list and/or given by just one participant may be more peripheral to a target concept. Thus, an investigator can build up a picture of the structure of superordinate categories such as 'tree' or 'bird' – which entities can form the core and which are peripheral. However, though superficially simple, there are many decisions the investigator must make when implementing a free-list task, which can impact how the results can be used.

We discuss the different stages of decision-making that an investigator must go through in planning a free-list task (section 3.1) and how this affects the consistency of the data collected. We then discuss the challenges that we faced during the implementation of our free-list experiment (section

[3]These three experiments are the first step in part of a wider research project investigating possessive classifiers from a psycholinguistic perspective: https://www.smg.surrey.ac.uk/projects/optimal-categorisation.

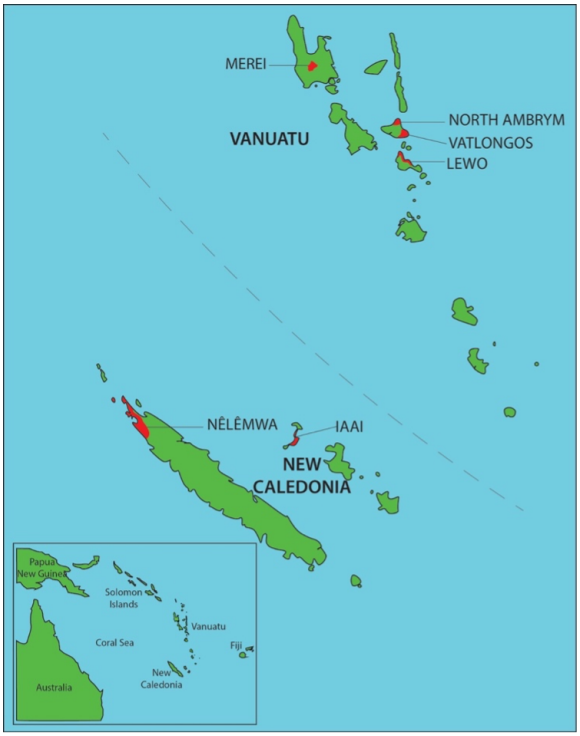


**Figure 1:** Geographic locations of the six sample languages in Vanuatu and New Caledonia.
Map created by Michael Franjieh.

3.2). Most of these challenges are specific to our task, though some are more general. We also highlight unexpected issues that can arise when running experiments across a range of languages. We argue that conducting a consistent experimental approach allows results to be valid (i.e., to enable a reliable analysis), and replicable (i.e., to allow the experimental method to be conducted on additional speakers and/or languages for further comparison). At the same time more spontaneous discussions should be allowed but should not affect the validity of the results. This balancing of an objective and replicable experimental approach with the need to engage with the participants is particularly important in endangered and indigenous language communities.

To briefly summarise the experiment, participants heard a recording of each of the classifiers found in their language. For example, speakers of Lewo heard a recording of the three classifiers in their inventory: *soma* 'GENERAL.CL.2SG' *moma* 'DRINK.CL.2SG' and *koma* 'FOOD.CL.2SG'.[4] After each classifier recording participants had to list as many nouns as possible that are associated with that classifier. Once a participant had finished, the next classifier recording was played and the process repeated until lists from all classifiers were completed. For example, after hearing the FOOD classifier, Lewo participant 17 gave the following list of nouns: *kinanena* 'food', *pleit* 'plate', *ika* 'fish', *lokovei* 'garden' *lakaboro* 'chicken', *kub lakaboro* 'chicken egg', and *mar* 'dry coconut'.

[4]The classifier's vowel in Lewo is /o/ when inflected for 2SG possessors, but /a/ for all other person number combinations. For example, the FOOD classifier is *ka-na* when suffixed with the 3SG possessor, but *ko-ma* when the possessor is 2SG. Furthermore, there are two dialectal forms of the GENERAL classifier in Lewo. Twenty participants heard the form *soma* and a further three participants heard their dialect equivalent form, *kioma*.

### 3.1 Planning the experiment

We discuss five key points in the planning of a free-list task.

#### 3.1.1 Responses: written or oral

The participant's mode of response is important. Participants can either write down their answers in a list (e.g., Davies and Corbett 1994; Özgen and Davies 1998; Tham et al. 2020) or say them out loud (e.g., Davies et al. 1994; Grandison et al. 2014; Roberson et al. 2004) and the experimenter writes them down. Both of these modes have advantages and disadvantages. What must be avoided is a mix of modes which would lead to variations in the experiment that could potentially lead to differing results across participants. For example, if participant A wrote down their list, whereas participant B gave their list orally, it might be that participant A gave more items in their list due to the written items acting as prompts for further items. If the variables are controlled, then we know that a longer list is not due to a different mode of response.

Written responses are ideal as participants can see their list as they add more items to it, which aids in avoiding the repetition of items. With oral responses, participants often repeat items, which may need to be excluded later.[5] The more the data requires editing at a later stage, the more likely mistakes will be made. However oral responses have one key advantage over written responses in that this medium does not exclude participants who are illiterate, or speak a non-codified language. Using written responses could therefore exclude potential participants, especially from the elderly, and in particular, the female population in rural indigenous language settings who are more likely to have had little access to education. The disadvantage of oral responses is that the experimenter may mishear or miswrite the responses, which can lead to confusion later when checking the meaning of items.

For our free-list task we chose to have oral responses from all participants as many of the languages have not been fully codified and some elderly participants had had no schooling and could not write.

#### 3.1.2 Participants: individuals or groups

Should you work with a single individual speaker, with multiple individual speakers, with an individual group of speakers or with multiple groups of speakers?

The first choice here is between an individual speaker or multiple individual speakers. If we were to simply test just one speaker, the list that is produced could be highly idiosyncratic. Whereas, if we test multiple individual speakers, we can smooth out the idiosyncratic responses which will just be mentioned by one or two individuals out of many respondents. Testing multiple individual speakers enables a comparable and reproducible experiment that results in reliable data that can be used for analysis and provide a platform for further enquiry. An adequately powered sample of individual speakers can provide generalizability with a 95% confidence interval. In other words, if we were to take 100 different samples and calculate a 95% confidence interval for each sample, then approximately 95 of the 100 confidence intervals would contain the theoretical true mean value, and so we can be 95% confident that our sample contains the theoretical true mean value. Power analyses can be conducted using G*power (Faul et al. 2009) to determine the minimum number of

[5] This is an analytical decision and there may be reasons not to exclude repetitions.

speakers needed to achieve this 95% confidence interval. Other researchers offer a rough guide of between 20 and 30 participants to achieve useful results (Stausberg 2011; Weller and Romney 1988). The researcher should record basic speaker metadata and can aim for a generalisable sample of the population dependent upon sociolinguistic factors such as sex, age and education.

Working with a group of participants can be a quick way to build consensus about a particular semantic domain, but particular group dynamics can influence the list. For example, certain participants can be overbearing or contribute more to the list, whereas other participants may remain quiet (Chelliah and de Reuse 2011: 182–183; Quinlan 2017). If a researcher works with multiple groups of speakers then the influence of one speaker may be mitigated. Of course, a group of participants can be equated to multiple single participants if each individual writes their responses down without a group discussion. This is a useful way of quickly working with a large number of participants.

In our free-list task we chose to work with multiple individual participants. This was partly due to the participants having to undertake a suite of experiments in our study that could not be completed in groups. Furthermore, due to the complex kinship networks in Vanuatu and New Caledonia the dynamics in a group setting may be unbalanced leading to a lack of representativeness in the data. We aimed for 22 speakers for each of the languages in our sample. However, finding participants for two languages, Iaai and Nêlêmwa, proved more difficult, due to our short time in the language communities, and for these languages we worked with considerably fewer speakers. Participants were chosen using the snowball/convenience method (Goodman 1961) via recommendations from community leaders, and partly through recommendations by other participants.

### 3.1.3 Timing of responses

The collected responses can either be temporally bounded by imposing a limited time for participants to respond, or unbounded by allowing participants as much time as they need to respond. The colour term free-list collected by Grandison et al. (2014) with the Otjiherero was an unbounded task and participants took between one and two minutes to complete the task. A temporally unbounded task is feasible when working on domains with limited numbers of possible responses – in the case of Otjiherero colour terms a maximum of ten lexical items were given. Davies and Corbett (1995) used a temporally bounded task to elicit colour terms in English, with a limit of five minutes, though in practice the majority of participants took no longer than two minutes.

For our free-list task we imposed a three-minute time limit for responses. Each language differed in the size of the classifier inventories from two to 23. For languages with smaller inventories it would have been feasible to have a temporally unbounded task, but for languages such as Iaai and Nêlêmwa with large classifier inventories an unbounded task would be unfeasible. Limiting the task to three minutes ensured that participants were able to maintain attention and energy throughout the whole experimental session. Furthermore, having a temporally bounded task ensured that participants from a language with just a few classifiers would spend the same amount of time per classifier as a participant from a language with many classifiers, though the overall task time would differ. A temporally bounded task ensures comparability and replicability across classifiers and languages.

### 3.1.4 Experiment location

The location or immediate surroundings of the experiment can play an important role in the responses given by the participants. For example, if we are interested in colour terms, sitting in front of a participant and asking them to name all the colours they know while wearing a bright red t-shirt

might influence the responses. Of course, it is impossible to limit or block out all possible visual stimuli that might confound the results. However, if possible one should try to ensure a setting that is similar for all participants.

For our experiment we were unable to ensure the same setting or location for all participants as we were working with six different language communities across two countries and multiple islands. The recording location was in some instances dictated by the participants, and in other instances by us. In many cases we recorded participants in their homes or at our hosts' homes. It was evident that some speakers would use the environment around them as a visual aid in coming up with items for the classifiers – often pointing to the types of fruit they could see when discussing the FOOD classifier, or even our computer, headphones and recording equipment when discussing the GENERAL classifier. Although this does mean that there was variability in the visual prompts that participants were likely to use, these prompts were likely to reflect the cultural salience that we were aiming to capture within the free-list data.

### 3.1.5 Creation of instructions

Should instructions be written for participants to read or should they be read out by an experimenter? Furthermore, which language should one choose – the target language or a lingua franca? The first question is similar to the one posed for the mode of response and depends on whether participants are literate in a lingua franca or the target language. The second question depends on different factors: (i) Is the experimenter fluent in the target language? (ii) Are the participants fluent in the lingua franca? (iii) Will the use of the target language induce priming effects?

For our experiment, instructions were written in English first, then translated into Bislama, the lingua franca of Vanuatu, and French, the lingua franca of New Caledonia. The translations were then checked by native speakers before being translated back into English by a speaker fluent in Bislama and a speaker fluent in French to ensure consistency in translation across the instructions.[6] In order to avoid using the possessive classifiers in the instructions, and thus avoid any priming effects, we used the appropriate lingua franca instead of the target language for the instructions – as neither Bislama nor French have possessive classifiers. To further avoid priming we also did not use examples of possessed items during the instructions. We recorded the possessive classifiers, inflected in the second person, from mother tongue speakers of the sample languages. These were played to participants and acted as cues to start the list of associated items. This ensured participants heard the correct pronunciation.

## 3.2 Implementation

The free-listing task was part of a larger suite of three experiments, the order of which was randomised for all participants. An information sheet about the project was read out to the Ni-Vanuatu[7] participants in Bislama and informed consent was recorded on a Zoom H6 audio recorder. Information sheets and consent forms in French were given to the New Caledonian participants to sign.[8]

[6]The experiment instructions and translations can be found in our Zenodo repository: https://zenodo.org/record/6378820.

[7]Ni-Vanuatu is the correct designation for a person from Vanuatu. It translates as 'of Vanuatu'.

[8]The participant information sheet (https://zenodo.org/record/6393395) and consent form (https://zenodo.org/record/6393324) are available in our Zenodo repository.

Metadata concerning sex, date of birth, education, languages spoken and brief histories of the participants were recorded. Researchers, who were fluent in either Bislama or French, read out the instructions to participants. Participants then heard one of the pre-recorded classifiers through the laptop speakers. The words analysed as possessive classifiers in each of our sample languages were pre-recorded in the target language, as there are no words in Bislama or French for the classifiers. The recordings were played as a set of three repetitions to the participants in order to avoid replaying the recording which could lead to less time for responses.

The order of the classifiers was randomised for all participants. All participant responses were recorded on a Zoom H6 audio recorder and simultaneously written down by the experimenter in a spreadsheet on a laptop. The researcher wrote down participant responses in the order they were given. Participants had three minutes to respond, signalled by a timer on the researcher's phone. At the end of the time limit the researcher asked the participant for a list of translations of unknown items.

Despite thorough planning of the experiment there were several minor issues that occurred during the implementation that required ad-hoc changes. These are discussed below.

### 3.2.1 Comprehension of instructions

How should the experimenter deal with non-comprehension of instructions? Ideally the most replicable experiment has a secondary protocol that can be read out or given to participants who do not fully understand. This ensures that all participants who did not understand the primary instructions receive identical follow up explanations. In practice, even the secondary protocol may still lead to further questioning and this may lead to ad-hoc explanations.

During implementation of our task the instructions were not necessarily read out verbatim and variations on the instructions were used in order to ensure that participants understood the task. Researchers need to be sensitive to how participants manifest their misunderstanding, e.g., through facial gestures or body language, and to help minimise participants' insecurities through rephrasing instructions and creating consensus. If participants were unclear, the instructions were rephrased. With some participants the task did not become clear until they heard the possessive classifier and the main instruction repeated – to list as many words that go with that classifier. Importantly, neither were examples of the different possessive classifiers mentioned in the instructions, nor were example members of classifier categories given, in order to avoid a priming effect.

We found that some participants thought that they had to inflect the classifier in different ways, rather than give associated possessed nouns, that is they repeated the possessive classifier with different possessor suffixes. At this point we reiterated that we wanted the items that are associated with the classifier, which denote possessions. The three-minute timer was not restarted in these instances.

### 3.2.2 Comprehension of category label

What should the researcher do if a participant is unsure of the meaning of the classifier? This is a problem we encountered, especially with the New Caledonian languages with large numbers of classifiers. Not all classifiers were known to all participants, due to attrition of the system (see Dotte, 2017, for loss of parts of the classifier system in Iaai). Should the experimenter explain the meaning of the category label and continue or should the experimenter skip that category? Ultimately, whether the responses from non-comprehended semantic categories are used in the analysis impacts

on the comparability of the experiment. If it is decided to use the responses that required further explanation, then the results will be less comparable as the participants will have been primed with an explanation. It is better to explain, but not use the resulting data, if a replicable experiment is needed. We believe that not explaining a non-comprehended target category is bad practice as it may appear as if the experimenter is withholding knowledge about a term the participant should know in their language, which may have implications for the researcher-participant dynamics and the participant's willingness to respond to subsequent prompts.

A problem we faced during the experiment with Iaai speakers was that the second person singular inflected forms of two of the classifiers turned out to be homophonous with each other, due to an original vowel length distinction no longer being recognised. Thus, *hnââm* [ʰnɔːm] 'PATIENT.CL.2SG' and *hnâm* [ʰnɔm] 'THOUGHT.CL.2SG' are now both pronounced as *hnâm* [ʰnɔm]. Recording and playing the 1SG, 2SG and 3SG inflections of these two classifiers disambiguated the homophony as there is a stem change in third person inflections: *hnân* [ʰnɔn] 'PATIENT.CL.3SG' and *hnen* [ʰnɛn] 'THOUGHT.CL.3SG'. The set of three repetitions of these two classifiers was different to the rest of the recordings for Iaai.

We aimed for comprehension rather than replicability in this instance. Furthermore, throughout the experiment across the sample languages if a participant was unsure of the target classifier, the experimenter would offer the first, second and third person inflections to aid comprehension. If a participant was unsure of the target category, the experimenter explained the rough semantic definition and recorded the participant's responses. As this may have resulted in a priming effect, these responses were not used for further cross-linguistic empirical analysis. However, we can still use these responses for a comprehensive language internal description of the classifiers. In hindsight it would have been advisable to leave explanation of the unknown classifiers until after all other classifier lists had been elicited. This would have avoided any priming effects.

### 3.2.3 Encouraging generation of more items

A choice needs to be made as to whether participants should be actively encouraged to list more items or not, and if so how to do this. A highly replicable experiment will have the same amount of intervention for each participant. Therefore, no further intervention in asking for more items is the most replicable. If the decision to ask for more items is made, then this must be done in a comparable way for all participants. Brewer (2002) discusses three techniques for encouraging more items in a free-list: (i) non-specific prompting – asking the participant for more items for the target category; (ii) reading back already listed items as prompts for more; and (iii) using each item on a participant's list as a cue for more items.

In our study we asked participants for more items if they had more time left. We also said that if they could not think of any more examples they could stop. One line of encouragement was that we plan to make a thematic dictionary with the results from the free-list experiment on what nouns occur with which classifier. Participants were told to list as many items as possible as the more items the better the dictionary will be. Sometimes a participant's list was read back to promote further responses. These methods of prompting were only done when we felt that the participant needed an extra push and therefore not every participant or every classifier category received prompting. This does make our results less objective as participants received differing levels of prompting. However, this increased our engagement with participants who struggled with the task and ensured that they felt their responses were deemed interesting and valid. We avoided asking participants to list more of a

specific sub-category. For example, if a participant gave the noun denoting 'bird', we refrained from asking for specific birds as this approach focusses the participant on one sub domain of animals to the detriment of others. Instead, participants should be simply asked if they know more 'animals'. This is consistent with Brewer's third technique, which uses each individual item given as a springboard for listing more items. When done for each elicited item, this method is highly replicable, as each item given by every participant would be a prompt, rather than just some items given by some participants. However it is time-consuming and would have taken more than our allocated three minutes for each classifier.

### 3.2.4 Follow up discussion

At the end of the experiment the experimenter can ask follow-up questions either about items that were mentioned or not, or general meanings of the target categories discussed. It is important to have the discussion at the end of the full experiment if multiple free-lists are being elicited as any discussion about potential members of a particular list may have a priming effect for subsequent lists. As any discussion will be particular to each participant, the follow up discussion elicits purely exploratory data. We believe that follow up discussions are essential for the valorisation of a speaker's knowledge of categories. By asking such questions as 'you said X was part of this category, can you think of any other similar items?' or 'can X also go in this category?' the experimenter actively engages with the participant as a language consultant, rather than just an experimental participant.

During our experiment we noted when an item on a list occurred in more than one list and would discuss with participants if there was a change in meaning dependent upon the use of the possessive classifier. For example, if the word 'pig' occurred in the FOOD classifier list and also in the GENERAL classifier list we asked if there was a difference in the conception of the referent, i.e. dead or alive? Furthermore, we often asked participants if there was an overarching semantic label that could be given to each of the classifiers.

Between eliciting the lists we tried to avoid any discussion with the participants beyond a clarification of the meaning of the elicited items. However, often participants themselves would start a discussion about the inherent meaning of the classifiers or offer further insights into a classifier's members by giving further examples. In these instances we felt that cutting short a participant's discussion would not be helpful for building rapport, not at all polite, and the resulting information was useful for a language internal description of the classifier system.

Similarly, at the beginning of the experimental session we stated that we are interested only in individual responses and requested that the participant did not ask others' opinions or that if there were any additional persons present they could observe but remain silent. Occasionally, other family members would walk in and offer their responses. It is very difficult for a researcher to stop this happening, especially when the experiment is conducted in the particiant's home. The researcher should simply note any responses that were given by any others present and exclude them later. The researcher can also reiterate that they would like just the participant's responses and invite any additional onlookers to join in the discussion at the end instead. To avoid these types of situations it may be best to conduct the experiment in a community space, such as a school room, or at the researcher's place of residence.[9]

[9]The place of conducting the experiment should be informed by knowledge of what is acceptable in each language community. The experimenters in our study were both familiar with community contexts in Vanuatu and New Caledonia.

### 3.3 Planning and implementation summary

Throughout the planning and implementation, we aimed for consistency so that the experiment would be replicable across additional participants and additional languages. However, in reality there were factors that worked against this. An important one that is often overlooked in the experimental approach is how an experimenter interacts with the participants. It is vital, especially when working with minority and endangered language communities, to build rapport and ensure that participants' voices are heard. Not only are participants unlikely to have taken part in an experimental study before, but they are also not likely to want to be treated as a strict experimental subject. By increasing engagement with participants we give up some aspects of a strict objective experimental approach (see Davidson 2020 for the notion of 'experimental' as gradable). However, by ensuring that consistent experiment protocols are in place the experiment still produces useable and replicable results. Increasing engagement with indigenous language speakers adds value to the participants' knowledge, helps build a more authentic interaction between the experimenters and the language communities, and leads to broader informative discussions around the subject matter. Furthermore, establishing rapport with participants aids in building a long-term partnership between the researcher and the community, helping the researcher return for further studies. Through our partnership with local stakeholders in education, we plan on sharing key findings from our work via the production of thematic dictionaries and teachers' guides to possessive constructions, to aid local efforts in vernacular literacy education.

## 4 Data processing

Before any analysis can be conducted the data must be processed and collated. Here we highlight a few key decisions we took that influenced the analysis. We emphasise that our approach to processing the data is one of many options available to a researcher. Whichever approach is taken, it must be systematic, that is the researcher must make any alterations to the data based on documented protocols that are replicable across data sets. Furthermore, any alterations must be retraceable, as in any alterations must be made in a stepwise fashion on separate versions of the data, so that no data is ever truly deleted and everything is fully recoverable. By being explicit in our approach we are allowing other researchers to replicate the analysis on our data, and to allow further comparisons with new data, processed the same way.[10]

### 4.1 Initial processing of lists

As the lists collected in our study were given orally there was a high chance that a participant would repeat the same noun in the same list. We decided to delete duplicates that occurred in any individual classifier list. During the elicitation of the lists, duplicates were still written down but deleted in the processing stage.

Sometimes participants said an item and then changed their minds. The item was still written down, but with an asterisk to show that it would be deleted later.[11] Similarly if an item was given

[10] Original data files and processed data will be uploaded to our Zenodo repository at the end of our project.

[11] When we say 'deleted', we do not mean an item is no longer recoverable. All deletions and alterations were not done on the original lists that participants gave, but on subsequent versions of the data. This allows us to trace back and 'recover' any items or changes later from the preceding versions.

by another speaker who was present, this was also written down and then deleted later. Sometimes the researcher gave encouragement for further responses of particular semantic sub-domains, rather than simply asking for more items. Items from encouraged sub-domains were also deleted. Any items given after the three minute time limit were still noted, but later deleted. Very rarely setting the timer was forgotten, and in this instance an extended time was given to the participant for that particular classifier category. Later the recording was checked to find the three minute cut-off point.

As the membership of a classifier category is open we decided to include borrowings from French or Bislama. If an indigenous noun and a borrowed noun denoting the same entity were given in the same list we did not consider it a duplicate unless they were adjacent in the list. If an indigenous term was followed by a French or Bislama term, this was judged to be a consecutive translation by the participant. If the indigenous term directly followed the borrowed term, then the borrowed term was thought to be a place-holder while the participant was thinking of the indigenous term. Recordings were checked to ensure context in these situations.

We also excluded directly possessed noun usage of classifiers. For example in Iaai, the form *xavââ-* 'clothes of' can be used either as a directly possessed noun (6a) or as a classifier (6b):

(6) a. xavââ-m
clothes-2SG
'your clothes'

b. xavââ-m ûxaaû
CLOTHES.CL-2SG clothes
'your clothes'

When used as a directly possessed noun, the form may be followed by a preposition, e.g *xavââm hnâân he* 'your clothes for travelling (clothes.2SG for travel)' or *xavââm hnâân ûne* 'your clothes for school (clothes.2SG for school)'. In contrast, a classifier usage must be followed by the possessed noun itself, as in (6b) above. The directly possessed noun usages were removed during our data processing stage.

What is important is that during the experiment the researcher should transcribe everything and only later discard particular entries under certain conditions.

### 4.2 Standardisation

Standardisation is important for any automated processing of the data, especially in creating scripts that can count instances of nouns. We created a spreadsheet for each classifier list, with columns for participant ID and item order. Additional columns were created for the different versions of the data, e.g. noun-v1, gloss-v1. The first step in standardization was to ensure that differences between entries were genuine differences. For this purpose, two new columns were added for new versions of the data, noun-v2 and gloss-v2, which were standardised spellings of the original two columns. Any changes were highlighted.

### 4.3 Translations

Translations of the lists were needed to allow comparison across languages. Here there is ample scope for problems, which we forestalled by applying a series of rules to our translation of the data. Three new columns were added to each spreadsheet – noun-v3, gloss-v3 and notes-v3. Three different rules were then applied.

The first rule 'most direct translation' ensured that all words in the noun column were glossed and that different nouns had different glosses. For example there are several wild yam species, and

in the Vatlongos lists we had *ramano* and *balbat*, both glossed as 'wild yam sp.' We added the suffix 1 and 2 to differentiate the glosses. Another example is from Lewo, where some speakers had given the translation for *kub lakaporo* as 'egg', however the most direct translation would be 'chicken egg', as *kub* means 'egg' and *lakaporo* means 'chicken'.

The second rule 'most basic translation' ensured that if different translations were given for the same item by different participants, the most basic translation covering all meanings was chosen.[12] For example in North Ambrym, *bwelaye* was translated as 'cup, plate' so the more basic translation 'vessel' was chosen to ensure comparability.

The third rule 'if no basic translation use multiple translations' was implemented when no basic translation was sufficient. For example in Iaai, the noun *mena* means both 'bird' and 'animal' so was translated as 'bird/animal' in all instances. All changes were highlighted and when a rule was applied this was noted.

Creating and applying these basic standardization procedures means that we are indeed comparing like with like. Furthermore, if there is some unexpected interest in, say, yam species, we can trace this back through the steps we have taken.

### 4.4 Modification and correspondence

Modified examples were the next issue to be addressed. Imagine we had asked for colour terms and a respondent gave the basic colour term 'red' and also a non-basic term, i.e. a modifier and colour term, e.g. 'dark red'. We would need both to be able to treat this as 'red' but also as being different from 'red'. For similar examples in our data, three new columns were created – noun-v4, gloss-v4 and notes-v4. Two rules were created to handle the incidence of modifiers.

The first rule 'look within lists and across lists' was applied when an item also occurred with a modifier such as a property concept or quantifier. If the same participant had listed the noun without a modifier in the same list, the modifier was kept. If the noun and modifier occurred in a participant's list without the unmodified noun, then the modifier was deleted, though this was cross checked against other participants' lists and across all classifier lists within the language. For example, participant 13 in Lewo gave *sur tai* 'one thing' as their first item in the list for the food classifier. The fourth item on their list was *sur* 'thing'. Thus the modifier *tai* 'one' was retained. An example of modifier deletion is also found in the same classifier list in Lewo, this time from participant 15 who gave *mra laki lala* 'fruit; fruit tree PL'[13]. Across all of participant 15's classifier lists they did not give *mra laki* without the modifier and the two other participants who gave *mra laki* 'fruit' used the non-modified form, i.e. without the plural modifier. The modifier *lala* was then removed from participant 15's item.

The second rule 'same noun or different' looked at whether two different forms offered were different instantiations of the same noun. For example, North Ambrym *pus* and *puskat*, from Bislama, denote 'cat'. As *pus* is an abbreviation of *puskat* we changed all instantiations of *pus* to *puskat*. Another example involved the nouns *trak* and *amkumku* from North Ambrym; these both mean 'truck' but as the former is clearly a borrowing then these are left as two different nouns.

[12] Though translations were not sought as part of the standard experimental protocol, it was essential for the researcher when working with a new language community to ask for translations with the first few participants until the most often repeated nouns became recognised.

[13] The translation of *mra laki* is made up of *mra* 'fruit of' and *laki* 'tree', giving a generic meaning of 'fruit'. *Lala* is the plural modifier.

### 4.5 Data processing summary

It is important that methods for processing the data are acknowledged and are made in a stepwise fashion. This is important for two reasons. Firstly, from a workflow perspective, if a researcher decides to change how the data are processed, this should be relatively simple as all decisions regarding the processing have been made and noted in the spreadsheets. Secondly, from the perspective of replicability, if the goal of an experimental approach is replicable results, then data processing must be explicit as part of the research methodology as this has implications for comparing results from independent experiments.

## 5 Overview of descriptive results

The free-list task offers up a bounty of data, and we present some of the results since they illustrate just how useful the task can be. We found data that were interesting both from a descriptive perspective and from an analytical perspective. If we had aimed for a purely exploratory free-list task we would have had many interesting findings, giving a helpful perspective on grammatical and semantic description of the classifier systems of the different languages, as currently used. But by pursuing a more replicable free-list task, we are also able to compare the results from participants within a language and across languages from an analytical perspective. The descriptive findings were also borne out by a more replicable approach to the experiment; thus, there is value in aiming for a strict experimental protocol. In this paper we shall only focus on general descriptive trends to demonstrate the power of the free-listing technique.

The free-list task gave us an up-to-date understanding of both the inventory size of each classifier system and the related semantic domains (c.f. Table 1 for a brief overview). Just by implementing the free-list task, interesting discussions with speakers around the use of the classifiers were instigated. We were also informed about possessive classifiers that had not been documented in previous grammatical descriptions of the languages. Thus one participant from the Iaai language community reported a classifier *tue-* for possessions that are smoked, such as cigarettes and cannabis. Similarly, one Nêlêmwa speaker reported a classifier, *thaxilo-*, for unripe fruits, such as green mangoes or green papayas. Both of these classifiers are interesting from a typological perspective as there are no Oceanic languages, except Nêlêmwa or Iaai, to our knowledge, that have specific classifiers for these types of possessions. This is information that we might have expected to be available in published sources but was not, hence it was rewarding to find out something new so early in the research. Furthermore, since the unripe fruit classifier was reported to us during the recording of the target classifiers, we were able to include it in our experiment for Nêlêmwa. A further six participants recognised this classifier and were able to give items that are associated to it. Unfortunately, the smokable classifier in Iaai was reported to us after the recording of target classifiers and so we were unable to include it as part of the experimental protocol. This classifier will need to be corroborated by further elicitation and experimentation.

**Table 1:** Descriptive statistics

| Language | Classifiers | Participants | Total unique nouns |
|---|---|---|---|
| Merei | 2 | 21 | 176 |
| Lewo | 3 | 23 | 249 |
| Vatlongos | 4 | 24 | 323 |
| North Ambrym | 5 | 23 | 296 |
| Nêlêmwa | 20 | 11 | 299 |
| Iaai | 23 | 16 | 447 |

One of the important findings from the free-list experiment was discovering the different semantic ranges of items that can occur with the different classifiers and that had previously evaded description. For Iaai, we included six classifiers that had previously been claimed to be unique by Ozanne-Riviere (1976), that is classifiers that only occur with one individual item. Results from our free-listing experiment indicate that all six supposedly unique classifiers are able to occur with more than one noun. But for these six classifiers, many nouns were only given by a maximum of one participant, suggesting that there is a good deal of idiosyncracy in the use of the classifiers. Dotte (2017: 345) gives examples of the ROAD and the SUGARCANE classifiers being only able to occur with one noun each – *gethen* 'path' and *aakû* 'sugarcane', respectively. Our research has found that the ROAD classifier also occurs with *garanyai* 'ascending path' and *e* 'road'. Furthermore, the SUGARCANE classifier also occurs with *waasu* 'chewing gum', *ûvëëu* 'hibiscus bark', and *öiö* 'Pueruarua sp.'. It must be noted that relatively few speakers gave these as examples of members of these classifiers, suggesting that these uses signify either expert or idiosyncratic knowledge (c.f. Stausberg 2011).

An interesting finding came from Merei, which only has two classifiers. For all the other languages there is some extent of semantic heterogeneity in at least one of the classifiers – that is the possessed nouns that occur with a particular classifier are members of different semantic groups. For example, as mentioned before, the FOOD classifier in North Ambrym also occurs with items used for cooking or preparing food. This is not the case in Merei, which has collapsed the separate food and drink classifiers found in other languages into one CONSUMABLE classifier that covers drinks and food. Merei only allows items of food or drink in the CONSUMABLE classifier and does not show any signs of semantic extension.[14]

During the experimental design phase we decided to test several incipient classifiers in Nêlêmwa. Incipient classifiers are directly possessed nouns that function as possessive classifiers only if the possessed noun that follows is also modified by an attributive modifier. For example, in (7a) the noun *wany* 'boat' occurs in a non-possessive construction. In (7b) the directly possessed form *wajany* 'my boat' occurs. In (7c) the same directly possessed form functions as an incipient possessive classifier as it is followed by the modified noun phrase *wany hnap* 'sail-boat'. Note that the modification of the directly possessed form is ungrammatical and both the noun and modifier needs to be present for the directly possessed noun *waja-* to function as an incipient classifier.

[14]Only one item, a borrowing, ultimately from English via the lingua franca Bislama, *tin* 'tin, can' could be construed as non-edible. However, tin is often used as a metonymic extension for its contents, such as *tin fis* 'tinned fish' or *tin kakae* 'tinned food'.

(7) a. waneda wany na bwa on?
how.many boat LOC on beach
'How many boats are on the beach?' (Bril 2002: 384)

b. iva waja-ny
where boat-1SG
'Where is my boat?' (Bril 2002: 259)

c. waja-ny wany hnap (*waja-ny hnap)
boat-1SG boat sail
'my sail-boat' (Bril 2012: 72)

We included the incipient classifier *waja-* for boats along with two additional incipient classifiers: *mwa-* for houses and *kee-* for baskets. For the most part, participants used these as either directly possessed nouns, or as incipient classifiers. However, several participants used these forms as actual classifiers and allowed non-modified possessed nouns to follow the classifier. For example, participants gave items such as *tigi* 'dinghy' or *karava* 'canoe' for the *waja-* classifier. One participant used the *waja-* form as a classifier not just for boats, but for all types of transport and gave *loto* 'car' as an example. This highlights that these incipient classifiers are in flux and that there is great inter-speaker variation in their usage. By using a free-listing experiment we were able to access a wide sample of speakers and discover the variation in usage. We can now use this finding as a basis for further investigation. These novel uses will need further elicitation to fully understand how they function.

Nêlêmwa also shows an interesting development of the *shâ-iyu-* classifier which was described by Bril (2002) as being used to designate fish that have been caught that are intended to be sold.[15] The most often listed possessed item *nok* 'fish', given by seven participants, conforms to Bril's description of the classifier. However, many participants use it for different items that can be bought or sold, and not exclusively for fish. The second to sixth frequently listed items, all given by four participants, were *lai* 'rice', *loto* 'car', *miit* 'meat', *shuka* 'sugar' and *shâlaga* 'crab'. Here it is clearly no longer being used for catch from the sea to be sold, except for *shâlaga* 'crab'. Nêlêmwa has no general classifier, which is found in all the other sample languages in the study. It might be that *shâ-iyu-* is developing the semantics of a general classifier. Nêlêmwa does have a general prepositional construction that has similar usage to GENERAL classifiers in other languages. We included the general preposition in the free-list and there is one major distinction in usage. The general preposition allows the possession of nouns denoting abstract entities such as *hnahnamiwo* 'thought' and *faram* 'forgiveness', whereas only nouns denoting concrete entities occurred with the *shâ-iyu-* classifier.

Due to the scope of this article we have given only a brief descriptive overview of our results to highlight just some of the more interesting findings from the free-list experiment. These findings can now be used as a basis for further analysis to answer our main research questions and to inform subsequent experimental tasks to further understand how the classifier systems in Oceanic languages function.

[15]The *shâ-iyu* classifier compromises *shâ-* 'contents of' and *iyu* 'exchange, trade'. There are three other comparable classifiers in Nêlêmwa all with the bound element *shâ-* and are all related to how fish are caught: *shâ-doo-* is the classifier for fish caught by spear; *shâ-pwee-* is the classifier for fish caught by line; and *shâ-pwia-* is the classifier for fish caught by net.

## 6 Further analyses

The data from a free-listing experiment can be analysed in a myriad of ways. First, free-listing gives us an up-to-date understanding of the members of a particular category, in our case the possessive classifiers of several Oceanic languages as we have shown in section 5. We can use the data to investigate the most salient entities within each classifier's domain and compare similar classifiers across languages by using different types of salience analyses such as Smith's Salience (Smith et al. 1995) or the Cognitive Salience Index (Sutrop 2001). Salience is useful for comparing similar classifiers across languages as it helps uncover the differences and similarities between languages. See section 1 for an overview of these techniques.

Second, we can use the data from the free-listing task as an indication of how many nouns listed by participants occurred in more than one classifier list, giving an indication of the amount of overlap between a noun and a classifier. We expect languages with a classifier system that functions like that of North Ambrym to have less overlap than languages whose classifier systems function like Lewo, which have a freer association between noun and classifier (c.f., examples 1-3). Category membership is at the heart of all research on categories, and not just nominal categorisation systems. See the useful discussion on the overlapping membership between the categories of fruit and vegetable in free-list data in Weller and Romney (1988) and for a potential analysis using Multi Dimensional Scaling.

Third, we can use free-listing data as a starting point for designing further experiments to understand how the classifiers are used. We plan on creating a list of nouns generated from the free-list results that will be used in a noun categorisation experiment. In this experiment participants hear a noun and then have to say which classifiers they would use with it. Are more highly salient category members more likely to occur with just one classifier? And are more highly salient category members more likely to be answered with quicker reaction times than less salient category members? A pilot study by Franjieh (2018) indicates this is so, at least for North Ambrym – but this is a language that shows less overlap. How do more typical classifier systems function?

## 7 Conclusion

Free-listing is a relatively straightforward technique that results in a wealth of data and offers opportunities for discussing and valorising different speakers' understanding of linguistic categories. The data that resulted from our experiment shows that a simple free-listing task can uncover not just new classifiers, but new uses of classifiers, and give an up-to-date picture of linguistic categorisation systems. With the exception of psycholinguistic studies, free-listing has been overlooked as a technique for gathering data on categories in linguistics. It is especially apt as a technique for the field of semantics. We note that free-listing does not provide negative evidence and is important to use this experiment as a basis for further experimentation or elicitation to collect negative evidence on the restrictions of nouns occurring with different classifiers.

Designing an experiment that is replicable across participants and languages ensures that the results are valid for quantitative analysis. However, such experiments must be implemented with the understanding that speakers of minority and endangered languages in rural settings will require clear guidance throughout the process and must be treated with respect and consideration to enable partnership. Including aspects of Participatory Research in our experimental design helps to improve a simple list task by increasing engagement with participants through the addition of invaluable dis-

cussions around the lists themselves. The resulting discussions not only valorise the participants' knowledge and help build rapport between speaker and researcher quickly, but also identify interesting avenues for further research.

## References


Benton, Richard A. 1968. Numeral and attributive classifiers in Trukese. *Oceanic Linguistics* 7(2). 104–146. (doi:10.2307/3622896)

Berlin, Brent and Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley and Los Angeles: University of California Press.

Brewer, Devon D. 2002. Supplementary interviewing techniques to maximize output in free listing tasks. *Field Methods* 14(1). 108–118. (doi:10.1177/1525822X02014001007)

Bril, Isabelle. 2002. *Le nêlêmwa (Nouvelle-Calédonie): Analyse syntaxique et sémantique*. Paris: Peeters.

Bril, Isabelle. 2012. Ownership, part–whole, and other possessive–associative relations in Nêlêmwa (New Caledonia). In *Possession and ownership*, ed. by Aikhenvald, Alexandra Y. and Dixon, R. M. W., 65-89. Oxford, New York: Oxford University Press. (doi: 10.1093/acprof: oso/9780199660223.003.0002)

Chambers, Robert. 1994. Participatory rural appraisal (PRA): Analysis of experience. *World Development* 22(9). 1253–1268. (doi:10.1016/0305-750X(94)90003-5)

Chelliah, Shobhana L., and Willem J. de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.

Chung, Ying Shing Anthony. 2005. *A descriptive grammar of Merei (Vanuatu)*. Canberra: Pacific Linguistics, The Australian National University.

Davidson, Kathryn. 2020. Is experimental a gradable predicate? In *NELS 50: Proceedings of the Fiftieth Annual Meeting of the North East Linguistic Society*, ed. by Mariam Asatryan, Yixiao Song, and Ayana Whitmal, 125-144. Amherst, MA: GLSA.

Davies, Ian and Greville Corbett. 1994. The basic color terms of Russian. *Linguistics* 32(1). (doi:10.1515/ling.1994.32.1.65)

Davies, Ian, Greville Corbett, Harry McGurk, and David Jerrett. 1994. A developmental study of the acquisition of colour terms in Setswana. *Journal of Child Language* 21(3). 693–712. (doi:10.1017/S030500090000951X)

Davies, Ian R. L., and Greville G. Corbett. 1995. A practical field method for identifying probable basic colour terms. *Languages of the World* 9(1). 25–36.

Dotte, Anne-Laure. 2017. Dynamism and change in the possessive classifier system of Iaai. *Oceanic Linguistics* 56(2). 339–363. (doi:10.1353/ol.2017.0017)

Early, Robert. 1994. A grammar of Lewo, Vanuatu. Doctoral dissertation. The Australian National University.

Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41(4). 1149–1160. (doi:10.3758/BRM.41.4.1149)

Franjieh, Michael. 2016. Indirect possessive hosts in North Ambrym: Evidence for gender. *Oceanic Linguistics* 55(1). 87–115. (doi:10.1353/ol.2016.0011)

Franjieh, Michael. 2018. North Ambrym possessive classifiers from the perspective of canonical gender. In *Non-canonical gender systems*, ed. by Sebastian Fedden, Jenny Audring, and

Greville G. Corbett, 37-67. Oxford, New York: Oxford University Press.

Gatewood, John B. 1983. Loose talk: Linguistic competence and recognition ability. *American Anthropologist* 85(2). 378–387. (https://doi.org/10.1525/aa.1983.85.2.02a00100)

Gaventa, John, and Andrea Cornwall. 2006. Challenging the boundaries of the possible: Participation, knowledge and power. *IDS Bulletin* 37(6). 122–128. (https://doi.org/10.1111/j.1759-5436.2006.tb00329.x)

Given, Lisa. 2008. Rapport. In *The SAGE encyclopedia of qualitative research methods*, ed. by Lisa M. Given, vol. 1–10. Thousand Oaks, CA: SAGE Publications, Inc. (doi:10.4135/9781412963909.n366)

Goodman, Leo A. 1961. Snowball sampling: The annals of mathematical statistics. *Institute of Mathematical Statistics* 32(1). 148–170. (doi:10.1214/aoms/1177705148)

Grandison, Alexandra, Ian R. L. Davies, and Paul T. Sowden. 2014. The evolution of GRUE: Evidence for a new colour term in the language of the Himba. In *Colour studies: A broad spectrum*, ed. by Wendy Anderson, Carole P. Biggam, Carole Hough and Christian Kay, 53–66. Amsterdam: John Benjamins Publishing Company.

Grinevald, Colette. 2007. Linguistic fieldwork among speakers of endangered languages. In *The vanishing languages of the Pacific Rim*, ed. by Osahito Miyaoka, Osamu Sakiyama, and Michael E. Krauss, 35–76. Oxford, New York: Oxford University Press.

Henley, Nancy M. 1969. A Psychological Study of the Semantics of Animal Terms. *Journal of Verbal Learning and Verbal Behaviour* 8. 176–184.

Lichtenberk, Frantisek. 1983. Relational Classifiers. *Lingua* 60(2–3). 147–176.

Lichtenberk, Frantisek. 2009. Oceanic possessive classifiers. *Oceanic Linguistics* 48(2). 379–402.

Lynch, John. 1973. Verbal aspects of possession in Melanesian languages. *Oceanic Linguistics* 12. 69–112.

Lynch, John, Malcolm Ross, and Terry Crowley (eds.). 2002. *The Oceanic languages*. Richmond, Surrey: Curzon.

Ozanne-Rivierre, Françoise. 1976. *Le iaai: langue mélanésienne d'Ouvéa, Nouvelle-Calédonie: phonologie, morphologie, esquisse syntaxique*. Paris: Societe d'etudes linguistiques et anthropologiques de France.

Özgen, Emre, and Ian R. L. Davies. 1998. Turkish color terms: Tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics* 36(5). (doi:10.1515/ling.1998.36.5.919)

Quinlan, Marsha B. 2017. The Freelisting Method. In *Handbook of research methods in health social sciences*, ed. by Pranee Liamputtong, 1-16. Singapore: Springer. (doi:10.1007/978-981-10-2779-6_12-2)

Rice, Keren. 2006. Ethical issues in linguistic fieldwork: An overview. *Journal of Academic Ethics* 4(1–4). 123–155. (doi:10.1007/s10805-006-9016-2)

Roberson, Debi, Jules Davidoff, Ian R. L. Davies, and Laura R. Shapiro. 2004. The development of color categories in two languages: a longitudinal study. *Journal of Experimental Psychology* General 133(4). 554–571. (doi:10.1037/0096-3445.133.4.554)

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8. 382-439. (doi:10.1.1.149.3392)

Sandoval, Jennifer A. 2017. Researcher–participant relationships. In *The SAGE encyclopedia of communication research methods*, ed. by Mike Allen, 1466–1468. Thousand Oaks: SAGE Publications, Inc. (doi:10.4135/9781483381411)

Smith, J. Jerome, Louanna Furbee, Kelly Maynard, Sarah Quick, and Larry Ross. 1995. Salience counts: A domain analysis of English color terms. *Journal of Linguistic Anthropology* 5(2). 203–216. (https://doi.org/10.1525/jlin.1995.5.2.203)

Stausberg, Michael. 2011. Free-listing. In *The Routledge handbook of research methods in the study of religion*, ed. by Michael Stausberg and Steven Engler, 245–255. London, New York: Routledge.

Sutrop, Urmas. 2001. List task and a cognitive salience index. *Field Methods* 13(3). 263–276. (doi:10.1177/1525822X0101300303)

Tham, Diana Su Yun, Paul T. Sowden, Alexandra Grandison, Anna Franklin, Anna Kai Win Lee, Michelle Ng, Juhyun Park, Weiguo Pang, and Jingwen Zhao. 2020. A systematic investigation of conceptual color associations. *Journal of Experimental Psychology*. General 149(7). 1311–1332. (doi:10.1037/xge0000703)

van der Riet, Mary and Merridy Boettiger. 2009. Shifting research dynamics: Addressing power and maximising participation through participatory research techniques in participatory research. *South African Journal of Psychology* 39(1). 1–18. (doi:10.1177/008124630903900101)

van Vlaenderen, Hilde and David Neves. 2004. Participatory action research and local knowledge in community contexts. In *Critical psychology*, ed. by Derek Hook, Nhlanhla Mkhize, Peace Kiguwa, and Anthony Collins, 445–464. Cape Town: UCT Press.

Vogl, Christian R. and Brigitte Vogl-Lukasser, and Rajindra K. Puri. 2004. Tools and methods for data collection in ethnobotanical studies of homegardens. *Field Methods* 16(3). 285–306. (doi:10.1177/1525822X04266844)

Wallerstein, Nina. 1999. Power between evaluator and community: Research relationships within New Mexico's healthier communities. *Social Science and Medicine* 49(1). 39–53. (doi:10.1016/S0277-9536(99)00073-8)

Weller, Susan C. and Antone Kimball Romney. 1988. *Systematic data collection*. London: SAGE Publications, Inc.