

# Using experimental design for descriptive work: reflections on why it is helpful and how to do it\*

Suzi Lima  
University of Toronto

**Abstract:** In this paper I discuss the use of small-to-medium scale production and comprehension studies to further the description of languages. This discussion is based on studies performed in Yudja, a Brazilian Indigenous language. The discussion centers around the considerations one should make while designing a study. Furthermore, I review how production and comprehension studies with a small-to-medium group of speakers - aligned with other techniques (such as analysis of naturalistic data and one-on-one context-based elicitation) - can contribute to the description of language use across different generations of speakers and to the advancement of typological and theoretical studies on a given topic. Finally, I also discuss the importance of investigating the same topic using more than one method.

**Keywords:** experimental design, semantics, Brazil

## 1 Introduction

This paper presents a discussion of the role of small-to-medium scale studies in the process of describing languages. I will discuss methodological aspects of studies carried out on Yudja (Tupi), a Brazilian Indigenous language whose population numbers around 880 people (Siasi/Sesai 2014),<sup>1</sup> most of whom live in the Xingu Indigenous Territory.

The literature estimates that there are 150 to 180 Indigenous languages spoken in Brazil (Moore, Galucio and Gabas Jr 2008). Moore, Galucio and Gabas Jr (2008) observe that many of these languages have incipient (32%) or little to no significant description (23%) available. In this context, efforts for language description and analysis are both imperative and urgent. These descriptions are valuable materials for future generations of speakers and are essential for language maintenance and revitalization initiatives (e.g., the creation of resources for speakers, including pedagogical materials). Furthermore, linguistic description of underrepresented languages is necessary for the advancement of typological and theoretical studies.

There are several strategies for language description and analysis, ranging from the use of naturalistic data (recordings of spontaneous speech) —which provides only positive evidence—to different types of production and comprehension tasks (e.g., storyboards, context-based elicitation, judgment tasks, etc.) which may be used to gather positive and negative evidence. Among these many possibilities, we can consider the use of small-to-medium scale studies. In this paper, *small-to-medium scale* studies will be used to describe tasks of short duration that involve the use of experimental design (controlled design of conditions) with a small group of speakers consulted individually. For example, within the context of Brazil, given the size of the Indigenous peoples'

---

\* I am very thankful to the Yudja people especially to Tawaiku and Chadawa Juruna who generously contributed their time and knowledge to the studies described in this paper. I also would like to thank two anonymous reviewers and the editors, Jérémy Pasquereau and Lisa Matthewson for their very valuable comments.  
<sup>1</sup><https://pib.socioambiental.org/en/Povo:Yudjá/Juruna>

populations and speech communities (which are often smaller than the population size, see Moore 2006; Moore, Galucio and Gabas Jr. 2008 for details), we may find that a small-to-medium scale study may consist of 20 members of the speech community (but, of course, this number depends on the context of each language community).

This paper discusses a workflow of the process of designing and running a study on understudied languages based on studies conducted among the Yudja (Tupi, Brazil). In Section 2, I present an outline of different steps necessary for designing a study (Section 2.1). Then, I discuss the implementation of these steps in the investigation of different phenomena in Yudja (Section 2.2). In Section 3 I discuss the importance of investigating the same topic using more than one method.

## 2 Steps in the process of designing studies

To illustrate the steps involved in designing a study, I will refer to studies on the interpretation of nouns (Lima 2014, 2018) and container phrases in Yudja (Lima 2016).

### 2.1 Workflow

#### 2.1.1 Preliminary steps before designing the study

Studies do not exist in a vacuum. The designing of an experimental study is preceded by the description of a given construction in the language in which the study is going to be carried out. The work of semantic description involves consultations of previous work written in/about the language, spontaneous data (narratives), and context-based elicitation (see Matthewson 2004). Based on the initial description of a construction, the researcher might identify questions they want to explore further. I will illustrate the role of description prior to designing experimental studies by discussing two examples.

One of my projects (which will be described in more detail in Section 2.2) involved exploring the interpretation of container phrases in constructions with substance-denoting nouns. In initial elicitation sessions studying this topic, I worked one-on-one with language consultants describing constructions with container phrases. In early sessions, I observed that container phrases (1) – that were not obligatory in constructions with numerals and substance denoting nouns (see Lima 2014 for details) – had the same structure as of locative phrases (2) in the language:

- (1) **Karaha he** una txabiũ awĩla wĩ  
bottle in 1SG three honey bring  
'I brought three bottles of honey'

- (2) **Aka he** na apĩ apayũ  
house in 1SG dog tie  
'I tied the dog in the house'

(Lima 2014: 159 – examples 23 and 24)

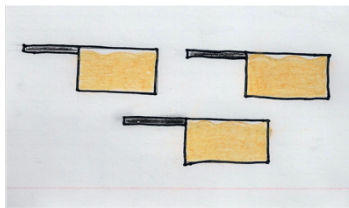
This observation was central to the designing of one of my first studies that explored the interpretation of these constructions. More specifically, given the similarities between container phrases and locative phrases, I wanted to know whether container phrases in constructions with numerals and substance denoting nouns would allow a locative interpretation. In one of my first studies on

this topic,<sup>2</sup> I investigated whether container phrases would allow an interpretation where they are not denoting the counting or measuring unit (4a) but only indicate where portions of a substance are located (4b). The study confirmed that a locative interpretation is available for container phrases, across the different age groups of participants (see Lima 2014, 2016 for a detailed discussion).

- (3) Awīla ‘honey’  
 txabiū awīla wā’ē he  
 three honey pan in

(4)

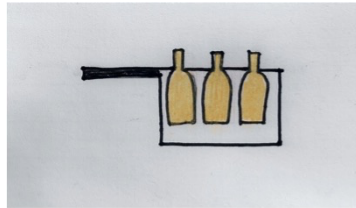
a Condition 1



*Counting unit:* pans

*Location of concrete portions:* pans.

b Condition 2



*Counting unit:* bottles.

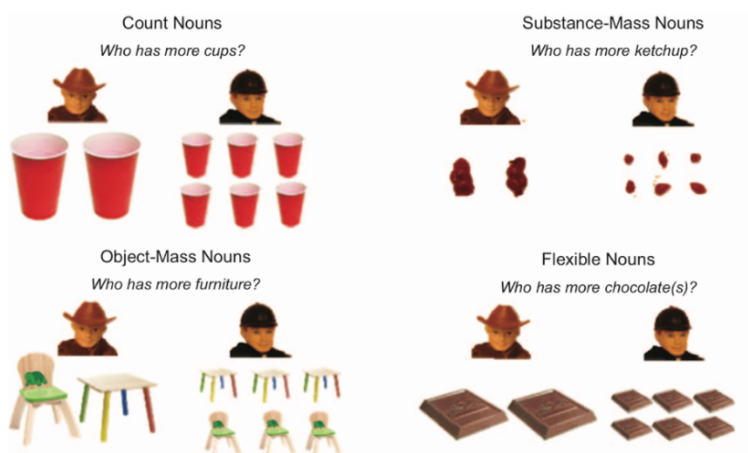
*Location of concrete portions:* pan.

(Lima 2014: 164 – example 27; Lima 2016)

An additional example to illustrate the importance of language description prior to running the studies in under-described languages comes from quantity judgment tasks (Lima 2014, building on Barner and Snedeker 2005). These tasks are often used to investigate the interpretation of count and mass nouns across languages (see Bale and Barner 2018 for an expanded discussion on this).

Participants are presented with two images or different object displays, as illustrated in Figure 1. One of the images displays one or more large portions of a substance, or one or more large-sized individuals (in Figure 1, this corresponds to the images below the man with a brown hat). The second image displays a higher cardinality of portions of substances/individuals (in Figure 1, this corresponds to the images below the man with a black hat). Participants are then asked to answer, ‘who has more *x*?’ where *x* can be a count (such as *cup* in English) or a mass noun (such as *ketchup*, a substance mass noun in English and *furniture*, an object-mass noun in English).

<sup>2</sup>20 adults and 26 children (8, 4-to-6-year-old children; 18, 7- to 12-year-old children) participated in this study. In small communities (for example, the largest Yudja community in the Xingu Indigenous Territory, which has approximately 232 members) the researcher might not be able to recruit participants for their study that have the same or very close age (for example, twenty 20-25 year-old adult speakers; twenty 6-year-old children). This issue may arise even in larger communities because not all individuals of a particular age group will be available or interested in participating in such studies. This absence of uniformity in the age group of participants may be used to the advantage of language description. It allows the researcher to have a broader understanding of how different age groups of speakers (including adults) produce and interpret a construction in a language. In this process, a researcher might be able to identify aspects of the language in variation, which is one aspect that is often not discussed in semantic language descriptions of under-studied languages.



**Figure 1:** Sample stimuli, quantity judgment tasks (Barner, Li and Snedeker 2010: 4)

In the process of designing such a study in Yudja, one central aspect was the choice of the quantifier that would appear in the question to ensure that both answers would be possible (a volume and a number answer). In other words, it was critical that the quantifier used in the questions would not create a bias for a particular answer. As such, prior to designing this study – which intended to investigate the interpretation of nouns – I worked on the description and interpretation of quantifiers in different contexts. The quantifier chosen for the quantity judgment task was *bitu* ‘more,’ which allows both volume and cardinality readings, as illustrated in (5) and (6).

(5) Volume interpretation:

Context: dividing flour with Pedro.

Bitu na asa upadjihu yahã a hae  
 More 1SG flour half NMLZ want ASP  
 ‘I want more than half of the flour.’

(Lima 2018, example 7)

(6) Number interpretation:

Scenario (spontaneous speech): a child wants to know if Suzi has more hair clips to give:

Suzi, au de bitu taba pĩdikaha  
 Suzi have Q more hair clip  
 ‘Suzi, do you have more hair clips?’

(Lima 2018, example 8)

Follow-up quantity judgment tasks also included questions with a quantifier that only allows a cardinal interpretation with notional count and notional mass nouns (*itxibĩ* ‘many’) and an adjective that only allows a volume interpretation with notional count and notional mass nouns (*urahu* ‘big’) to confirm that indeed these expressions could only be associated with a cardinal and a volume interpretation, respectively (see Lima 2014, 2018 for a discussion of the results for these tasks).

In sum, designing studies was always preceded by describing a given structure over the course of several elicitation sessions. This is an essential aspect of the process: designing studies requires having an initial description of morphosyntactic/semantic aspects of a structure. Based on descriptions of different topics I investigated over the years, I was able to identify phenomena and conditions that could be further explored through new techniques, across different age groups of speakers.

### 2.1.2 Choosing a method and design

One critical challenge in the process of designing a study is evaluating the translatability of the task from context/language A to context/language B. The researcher needs to make sure that the task is culturally appropriate and understandable to all participants. The participant needs to understand what is required from them to feel comfortable with their participation and for the results to be reliable.

As such, the process of choosing a method depends on two factors. First, a method needs to be suitable for exploring a particular falsifiable hypothesis and its predictions. Second, the method must be appropriate for the particular context and age group. For example, self-paced listening tasks might be a better choice than self-paced reading tasks for online studies in communities where literacy in the object language is low or very variable for several reasons: low availability of materials to read in the target language, a significant discrepancy of reading proficiency among participants, or competing orthographies in the target language (see Wagers, Borja and Chung 2015 for a discussion on the self-paced listening method in the investigation of *wh*-dependencies in Chamorro; see also Wagers and Chung (to appear) on other methodological aspects of conducting studies with the Chamorro community in the Northern Marianas Islands).

Later in this paper, in Section 2.2, I will describe in more detail a study designed to explore the features Partee and Borschev (2012) associated with measure and non-measure interpretations of container phrases. Prior to traveling to the field, I prepared the stimuli to be used. This involved: (a) the selection of substances and containers that would be manipulated in the study; (b) the preparation of pictures and randomization of items in different lists; and (c) the translation of sentences to be used in the study.

As mentioned above, when designing a study, we need to be sensitive to cultural aspects that might make the conducting of a study possible or not. For example, in the studies exploring the interpretation of substance-denoting nouns, the selection of nouns was based on findings from previous elicitation sessions. I selected substance-denoting nouns that would be familiar to different generations of speakers (since the goal was to run the studies with children and adults). The same holds for containers.

The medium of presentation of visual stimuli is also essential. For example, in Yudja communities, many children and adults draw by hand. Drawings can be seen in multiple places, such as posted on the walls of schools. Therefore, the use of drawings in these studies is not foreign to the participants. Photographs and videos are often displayed at community events. As such, the presentation of visual stimuli in my studies involved not only drawings but also photographs and videos, as well as physical objects.

For all my studies, I keep a list of specific details about the protocol that would be useful for writing the materials and methods sections of papers and during the process of designing follow-up tasks. These details include:

1. Topic
2. Motivation/Hypothesis
3. The number of expected participants and age group
4. Materials (independent and dependent variables)
5. Methods (description of the methods)

6. Type of design (*within or between subjects*) and latin square
7. Expected answers (if a hypothesis is confirmed).

Another important document that I prepare for all my studies is an answer sheet (one per participant) where I take notes during the studies. This document includes a table with the following columns: a) the sentences to be tested; b) the conditions; c) the study the sentences belong to, d) expected answers, e) a column for the actual answers, and f) notes. In the 'notes' column, I write comments from the participants (if any) and outstanding reactions of the participant to a given sentence/stimuli, such as laughs and long pauses (see Matthewson 2004 discussion on the relevance of these comments in language description). A different document contains other relevant information (instructions to the participants and verbal contexts to be presented [if any]).

### **2.1.3 In the field: checking the feasibility of the study**

Once in the community, I work with one language consultant to verify whether the sentences included in the study were grammatically correct and appropriate. In this process, we also discuss the presentation of the instructions in Yudja. A conversation we have during the preparation of the instructions is about how to present the task without suggesting a particular answer. This is an important element when conducting this type of study since the instructions should not create a bias toward a particular result. In situations where the researcher is not a native language speaker, careful work on the instructions is needed, jointly with language consultants.

### **2.1.4 Running studies**

After checking the feasibility of the study and working on the instructions, I would pilot the study with another experienced language consultant/language teacher to make sure the task was working. That is, to make sure the pictures and sentences were clear.

After that, we would proceed to the testing phase with the actual participants of the study. In the studies I conducted, participants entered the room one by one. The room used for the studies was one of the local school classrooms, when no classes were taking place in the building. The school and the area surrounding it is often used for activities unrelated to classes (community meetings, projection of movies, etc). The sessions are kept short (about 20 minutes long). A language consultant/local teacher is always in the room to communicate the content of the consent form in Yudja (to participants and their parents [when children are the participants]) and provide the study's instructions. For studies with children, their parents are always welcome to be in the room.

Working with a research assistant (language consultant) during the studies requires a clear discussion of the procedure (as it would in any other lab-type experience). This process can lead to relevant discussions about language acquisition, description, and variation. In many circumstances, these initial conversations were later expanded in different contexts (for example, many consultants find it interesting to discuss grammatical patterns and language acquisition in other languages). These topics also partially informed the Linguistics workshops for high school students and Yudja teachers conducted in the communities.

### 2.1.5 The role of training and pre-test tasks

It is crucial to ensure that the participants clearly understand the task employed in a study and that they are comfortable with it. Pre-tests introduce the participants to the main task without including the target sentences/conditions being manipulated. An example is presented in Section (2.2). I will first introduce the theoretical motivation of the study and then aspects of its design.

## 2.2 The use of studies in the process of investigating meaning: exploring the interpretation of container phrases

In the previous section, I presented an overview of the workflow involved in designing studies. In this section, I discuss in more detail the structure of one of the studies on container phrases, focusing on the designing aspects, to illustrate some of the suggestions made in Section 2.1. First, I will lay out the theoretical motivation of the study.

Much literature has argued that constructions with numerals and container phrases may allow measure and non-measure readings (Selkirk 1977; Landman 2004; Rothstein 2009, 2011, 2012; Partee and Borschev 2012; Lima 2016; Khrizman et al. 2015, among many others). This contrast is presented in (7a) and (7b). The same container phrase ('two glasses of water') may have different interpretations in each sentence:

- (7) a. Mary, bring two glasses of water for our guests!  
(non-measure interpretation most salient)
- b. Add two glasses of water to the soup!  
(measure interpretation most salient)  
(Rothstein 2012: 4 – examples 15)

In (7a), the amount of water is not being measured; one refers to the container containing water and counts the cups. In (7b), there is an expectation that the person who hears this sentence will fill out two full glasses of water, as the quantity of water is being measured for a recipe; that is, in (7b) *glass* is a measuring unit. Partee and Borschev (2012) further explore the distinction between measure and non-measure interpretations and observe distinct features associated with each of them (Table 1).

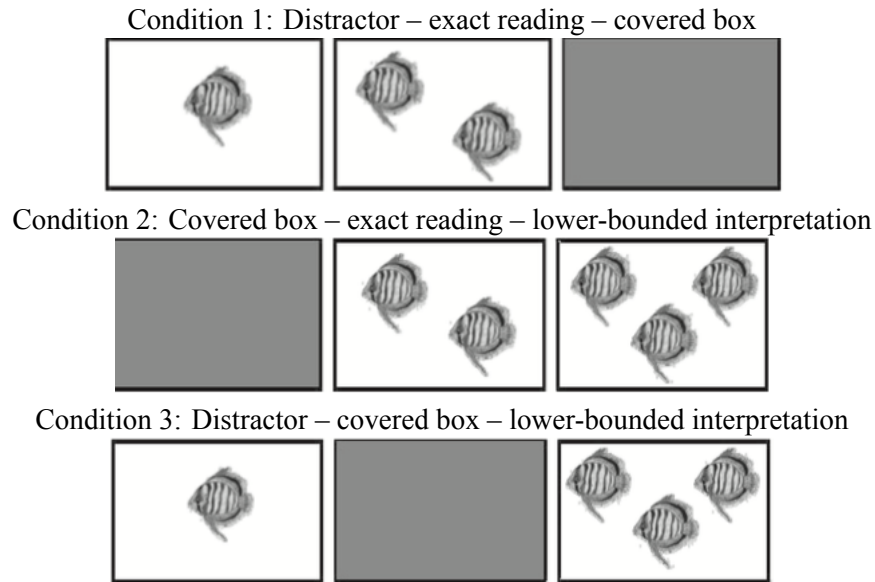
**Table 1:** Grammatical properties of interpretations of container nouns (Partee and Borschev 2012 *apud* Lima 2016: 11)

	Non-measure	Measure
Does it allow fractional numbers?	No	Yes
Does it require the container to be full?	No	Yes
Can it refer to containers of different sizes filled with the same substance?	Yes	No

Based on Partee and Borschev (2012), Lima (2016) described a study that explored whether the non-measure and measure readings across languages would have similar requirements. The background literature and results of these tasks in Yudja, English, and Kawaiwete (Tupi, Brazil) are reported in detail in Lima (2016). In this paper, I focus on the methodological aspects of these studies.

### 2.2.1 Covered box task

Covered box tasks (see Huang, Spelke and Snedeker 2013) involve asking the participant to say where an object is located. Participants are presented with three boxes (real objects or pictures); two of these are open (and, therefore the participants can see their contents), and one of these is a covered box. The participants are then asked to give the box where a particular object is located to the researcher. Huang, Spelke and Snedeker (2013) used a covered box task to study how children interpret numerals; they investigate whether 2-to-3-year-old English-speaking children interpret numbers as having an exact interpretation (exactly two) or a lower-bounded interpretation (at least two). In the task, participants were instructed by the researcher, “Give me the box with two fish” while being presented with the items in Figure 2 (one on each trial):



**Figure 2:** Sample stimuli, covered box task

(Huang, Spelke and Snedeker 2013: 8, Figure 1)

In this task, the critical condition is the third one. If participants choose the covered box rather than the third box (that has three fish) this could indicate that numerals have an exact semantics. An interesting aspect of the task is that it leaves open the possibility for the participant to provide an answer describing an object that is not in the scene. That is, this method can be set up as a comprehension task that includes a production component.

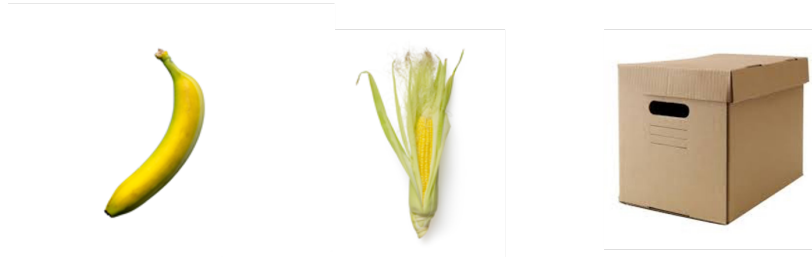
As mentioned above, a covered box task was performed in Yudja to explore the interpretation of container phrases. I will now comment on designing this task and conducting it.

**Training and Pre-test tasks** This study was divided into three parts: a training task, a pre-test task and the main task that included the target items being tested. First, I will discuss the use of the training task and the pre-test task.

The training task was designed to familiarize the participant with the method employed in the study. As illustrated in Figure 3, in this task, participants had to answer questions unrelated to the main study (interpretation of container phrases), such as *a'i de ataũ* ‘where is the potato?’ (Figure



3). At this phase, participants received feedback if they provided the wrong answer or if they had questions about the structure of the task.



**Figure 3:** Sample stimuli, familiarization task 1 [training task] - Question: *a'i de ataũ* 'where is the potato?'

In Figure 3, speakers were expected to point to the box and say that the potato was in there. The training task included both items where the object that was asked about was visible and questions where the object that was asked about was in the covered box. This step was essential to get the participant comfortable with the task. The training task is also one way for the researcher to validate the task/design, that is, to guarantee that the results in the main task are not due to extraneous factors.

After the training phase, we had a pre-test task. In the pre-test task, participants were presented with the containers and substances that were part of the main study. This task was important to guarantee that the visual stimuli used in the main task were clear and that the label associated with each container was appropriate (Figure 4).



**Figure 4:** Sample stimuli, familiarization task 2 [pre-test task] – *A'i de awatxi'i saku he?* 'Where is the bag of rice?'

Figure (4) illustrates an item from the pre-test task where the expected answer is visible (left picture). This task included questions where the object that was referred to was in the covered box and questions where the object that was referred to was visible. Crucially, since children participated in this study, it was imperative to verify whether they were familiar with the terms used.

The main task followed the pre-test task (in both the main task and pre-task, feedback was not provided). The conditions manipulated in the main task were based on representative features of measure and non-measure constructions, according to Partee and Borschev (2012) (Figure 5).

Condition 1: Full and identical containers (compatible with individuation and measure interpretations):



Condition 2: Full containers of different sizes (incompatible with measure interpretations, if containers need to be of the same size in order to indicate a particular quantity of a substance):



Condition 3: Containers of the same size, but with different amounts (incompatible with measure interpretations, if containers need to have the same size/amount of a particular substance):

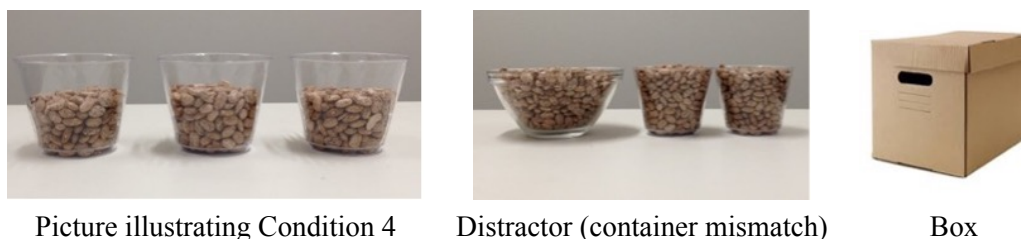


Condition 4: Containers of the same size, but not full (incompatible with measure interpretations, if containers need to be full):



**Figure 5:** Conditions of the main task (covered box task) (Lima 2016)

The main task was intended to explore whether the size of the containers or the amount of a substance within a container would impact how speakers answered the questions. Consider Figure 6, which illustrates the options offered to the participants to answer the question *A'i de txabiü puju kaneku he?* ‘Where are the three cups of beans?’



**Figure 6:** Sample item, main task (target sentence: *A'i de txabiü puju kaneku he?* ‘Where are the three cups of beans?’)

The main study included, in each question, a picture of a covered box, a picture that corresponded to one of the conditions presented in Figure 5, and a picture that represented the distractor item (mismatch in the number of containers, type of container mismatch or substance mismatch). For example, in Figure 6, if a Yudja speaker had chosen the box, that could indicate a preference for a measure-oriented reading (full containers). The participants were encouraged to describe the content of the covered box, if they chose it. As such, while this is a forced-choice task (participants are presented with options to choose from), they had the option to present a better alternative if they wished to do so.

The results of this task show that speakers did not select the distractor (therefore supporting the method’s validity; participants were not confused by the task or its items). Furthermore, the participants did not select the covered box in the main task trials, suggesting that, in the absence of a context, sentences including numerals and container phrases can be used to describe containers of different sizes and with different amounts of a substance. Knowing this, the next step was to evaluate these container phrases in contexts that biased measure and non-measure interpretations, as will be discussed in Section 2.2.2.

To conclude this section, it is important to highlight that children welcomed this task (covered box) with great enthusiasm because it was structured as a game. Creating engaging tasks, whenever possible, is ideal (see Louie 2015 on the positive effects of the use of humorous stories on participant engagement when using storyboards and Silva and Anderbois 2016 on the positive consequences of using a game [*Mastermind*] on the elicitation of evidentials).

## 2.2.2 Truth value judgment tasks

An often-used task in semantic studies is the truth value judgment task (see cf. Crain and Thornton 1998; Gordon 1998). In these tasks, a context is introduced verbally and/or visually to the participant and then the participant must evaluate if a sentence adequately described the context.

I discuss in this section the process of designing a truth-value judgment task that explores the interpretation of container phrases in Yudja.<sup>3</sup>

<sup>3</sup>Participants of this task were also participants in the covered box task, during the same session.

In this task, the goal was to explore the measure and non-measure interpretation of container phrases, following Partee and Borschev's criteria (2012). From the covered box task results, we learned that locative-container phrases, out of context, can refer to sets of containers that could vary in size and fullness (amount of the substance within the containers). In this task, the goal was to explore the effect of context and verb choice in sentences that included container phrases. For example, we were expecting that verbs such as *drink* in 'Paula drank two cups of water' would be more likely to be associated with an interpretation where the container is being used to measure portions of a substance as opposed to verbs such as *bring* in 'The woman brought two bowls of rice' where we may be just referring to the containers and not be using the bowl as a measuring unit.

The procedure for designing and running the study follows the workflow described in the Section 2.1. Similar to the previous task, this study was designed before the fieldwork trip. Once in the field and after working on the instructions and piloting the task with language consultants, the task was conducted with Yudja children and adults (20 adults and 26 children<sup>4</sup>). This task was structured in three steps:

Step 1 (Instructions): participants learned that they would be presented with a short story followed by short videos that presented different version of the continuation of the story. After each video, they would then hear a question, to which they would answer yes or no.

Step 2 (presentation of verbal context): participants were then verbally presented with a context. Pictures of the women who would later appear in the videos were presented along with the verbal context. An example of the type of verbal context that was provided to them is presented below:

“Drink” context

This is Maria [picture: woman with long hair] and this is Paula [picture: woman with short hair]. Paula needs to drink two cups of water every day. Maria will help her to make sure she drinks two cups of water.

Participants were presented with either a measure context (including the verbs *awi* 'drink' or *itu* 'pour') or a non-measure context (including the verb *dju wi* 'bring'; one group was presented with 'bring two bowls of rice' and another group was presented with 'bring two bottles of water'). A *between* subjects design for verbs was chosen to avoid the judgments of a verb impacting participants' choices when presented the same task using a different verb. That is, participants exposed to the 'drink water' context were not exposed to the 'pour beans' context.

Step 3 (conditions and target sentences): participants were then presented with a sequence of five videos (four videos: target conditions; one video: control). After each video, they were presented with the target sentence. As in the covered box task, this study manipulated Partee and Borschev's characterization of measure and non-measure readings, therefore manipulating the size of the containers and the amount of a substance inside them (see Lima 2016 for a detailed description). A *within* subjects design was chosen for the presentation of the videos that represented these different conditions (Figure 5). That is, participants were presented with all the conditions being manipulated. The task included a control video/item (container or number mismatch) to ensure the answers were

---

<sup>4</sup>26 Yudja children (7, 5-to-6-year-olds, 6, 7-to-8-year-olds, 6, 9-year-olds, and 7, 10-to-13-year-olds.)

reliable. If participants answered ‘yes’ to a control question whose answer was expected to be ‘no’, this would suggest that something was not working due to methodological (method, instructions, etc.) or extraneous factors (e.g., a participant who is distracted).

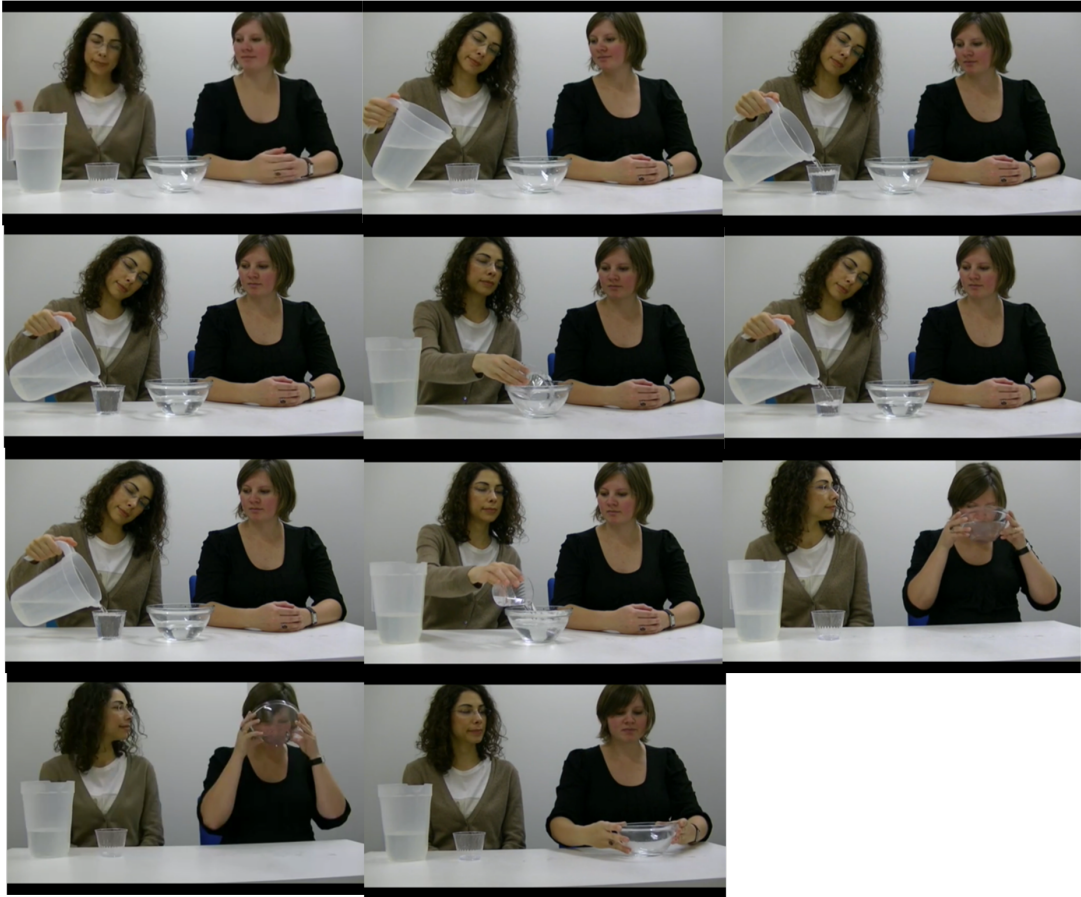


Figure 7: Screenshot of condition 1, ‘Maria drink two cups of water’

Critical question:

- (8) Awi de Paula yauda kaneku he iya be  
 drink Q Paula two cup in water DAT  
 ‘Did Paula drink two cups of water?’

‘Yes’ answers were taken as indicating that the structure presented was compatible with the context provided (true in the given context). As expected, the results show that all conditions (containers of different sizes/full or not full) were taken to be compatible with sentences that include a verb that did not bias a measure interpretation (*bring bowls of rice/bring bottles of water*). In the measure conditions, Lima (2016: 37) reports that “(...) half of the [adult] participants accepted the description provided in conditions where containers were not full nor identical. This could be due to an interference of the other possible interpretations of container phrases in this language, in particular the locative interpretation, which seems to be the most basic interpretation of those phrases given

that they have the same syntax as do locatives.” The same study was also carried out in English<sup>5</sup> and in Kawaiwete<sup>6</sup>(Tupi, Brazil). See Lima (2016) for details.

In sum, the covered box task (described in 2.2.1) was an important step to undertake before the truth-value judgment task. It showed that, out of context, container phrases could be associated with different types of scenarios, represented by the pictures in Figure 5. The truth value judgment task was used to show how these constructions are interpreted in measure and non-measure contexts. Most importantly, these tasks – performed in genetically unrelated languages– provided evidence to argue that the features described by Partee and Borchev impact the interpretation of container phrases. Furthermore, these studies allowed a better understanding of the developmental path of container phrases. The results turned out to be well aligned with proposals which predict that the measure interpretation is more complex and derived from the non-measure interpretation (see footnote 5).

### 3 Investigating the same topic using more than one method

In this section I will discuss the relevance of investigating the same topic using more than one method based on studies on object-denoting nouns in Yudja. First, I will present a brief background overview of the topics investigated in this study.

Lima, Li and Snedeker (2017) explored the possible interpretations of object-denoting nouns in constructions that involved numerals; more precisely, what counted as an atom for such nouns. The definition of what can be considered an atom for counting has been a much-debated question in the count/mass literature. According to some formal theories, for example, Chierchia (2010), semantic atomicity (what can be considered as an atom available for counting in language) reflects natural atomicity. One possible definition of natural atomicity is presented as follows:

---

<sup>5</sup>In English, this task was performed with thirty-three English-speaking children (3- to 6-year-olds) and 37 English-speaking adults. The results for children and adults show that in the course of language development, the learning of the requirements for measurement occurs after five years of age. While six-year-old children accepted only Condition 1 (described in Figure 5) in measure scenarios, younger children accepted the other conditions in the same context. These results are compatible with Partee and Borschev’s analysis. Partee and Borschev argue that the measure interpretation of container phrases results from a series of lexical shifts. Under this analysis, the use of container phrases to refer to a container (non-measure, individuation reading) is the most basic interpretation of such constructions. Lima (2016)’s results support this analysis by providing language development evidence for this claim. Lima (2016) also discusses other studies that show that children younger than six do not master the interpretation of measure units. See Gal’perin and Georgiev 1969; Levin and Wilkening 1989 for details.

<sup>6</sup>As discussed in the introduction, in Yudja, container phrases are syntactically locative phrases and may be interpreted as such (see Lima 2014, Lima 2016 for details). In Kawaiwete, container phrases may be encoded as locatives (i) and non-locative constructions (ii). Only non-locative constructions pattern similarly to English. See Lima (2016) for details.

- (i) Maria mukūi yrerusinga pype ‘ya werut.  
Maria two bottle in water bring
  
- (ii) Maria mukūi yrerusinga ‘ya werut  
Maria two bottle water bring

Natural atomicity:

$$\forall x \forall k \forall k' [x \in \pi_1(N_k) \wedge x \in \pi_1(*N_{k'}) \rightarrow \pi_1(N_{k'})]$$

‘If N is naturally atomic, then for any two contexts k and k’, if x is an atom of N<sub>k</sub>, and x is in the denotation of N<sub>k'</sub>, x is also an atom in N<sub>k'</sub>.’ (Rothstein 2010: 373)

A similar definition is provided by Chierchia (2010: 121), according to whom count nouns are characterized by stable atomicity (“the stable atoms of a property are the atoms of that property in every precisification of the original context. By construction, everything which is a P-atom relative to a context c, will remain such in every precisification.”). Under Chierchia’s proposal, the stability of atoms for count nouns will differentiate them from mass nouns. As mass nouns do not denote stable atoms, to interact with numerals, such nouns require an intervening container/measure phrase (such as *bottles of* in ‘three bottles of water’) which will define the counting unit in a particular context.

While it is indeed the case that natural atoms are very likely to be semantic atoms—across languages, objects are more likely to be count nouns—much literature has shown that at least three empirical facts suggest that semantic atomicity is not tightly coupled to natural atomicity: 1) nouns such as *fence*, and *wall* can be atomized in different ways depending on the context (Rothstein 2010); 2) some object mass nouns (such as *silverware* and *furniture* in languages like English) refer to object kinds that have natural atomicity (cf. Chierchia 2010, Schwarzschild 2011, among many others); 3) some substance denoting nouns such as *water* and *flour* can, in some languages, be directly combined with numerals, at least at a superficial level<sup>7</sup> (see Davis and Matthewson 1999; Gillon 2010, 2012; Mathieu 2012; Wiltschko 2012; Deal 2017; Lima 2014; Lima and Rothstein 2020, among many others). Lima, Li and Snedeker (2017) explore this topic (natural atomicity vs. semanticity atomicity) using two different methods in order to verify that the results were stable across different tasks. I describe both tasks and their results below.

In the first task, participants were told by a native speaker of Yudja that a person would describe some pictures to them in Yudja. The participants were told that the person providing the sentences (a non-native Yudja speaker) was learning how to speak their language and that she could sometimes make mistakes. Participants were then encouraged to provide a better description if the description provided by the speaker was not the best way of describing a given picture. That is, the task involved correcting the speech of a person who was not a member of the community; the process of correcting this person was set up in a positive way (“you are correcting them to help them improve their language skills”) precisely to avoid a situation where participants would feel uncomfortable to correct the researcher. Before taking part in this study, the same groups of speakers participated in another task that involved correcting a non-native speaker of Yudja (same type of task, on a different topic [position of quantifiers]). In this other task, the participants (children and adults) were not reticent in correcting the speaker when needed. As such, going into the first task, we had a second measure indicating that the method *per se* (correcting a non-native speaker of Yudja) would not be an issue.

Three conditions were manipulated in this study (whole object, pieces, groups), as illustrated in Figure 8.

The alternative descriptions provided by the speakers as a response to the task were essential in the process of advancing the description of the language, as they provided evidence on how different

<sup>7</sup>In some analyses, nouns are not directly combined with numerals in constructions with substance denoting mass nouns. A null classifier is postulated in such constructions according to some proposals. See Chierchia 2015 and Deal 2017 for details.



**Figure 8:** Sample stimuli (left picture: ‘whole object’ condition; middle picture: ‘piece’ condition; right picture: ‘group’ condition)

speakers of different age groups would describe the stimuli provided. For example, in the ‘piece’ condition, some of the alternative sentences provided by adults included nominalizations that described that the object was broken without the inclusion of a (pseudo)partitive (as in 9) and in much more rare occasions they included a word that can be used to refer to pieces (10).<sup>8</sup>

- (9) Context (visual stimulus): a banana broken in three pieces.

Pakua txabiü lakīri yahā  
 banana three break NMLZ  
 Lit.: a banana broken in three [parts]

- (10) Alternative description: two adults, in two trials (one trial each), used a pseudopartitive word:

Pakua akuata txabiu a’i  
 banana piece three here  
 Lit.: three long pieces of a banana.

(Lima, Li and Snedeker 2017)

The second task explored the same research question, but now using a different task (a felicity judgment task). As previously mentioned, the goal was to verify whether the data elicited in the first task (production, language teaching task) was reliable or an effect of the task used. 46 Yudja speakers participated in this task. 19 children were under 10 years old (5, 3-to-5-year-olds; 5, 6-to-7-year-olds; and 9, 8-to-9-year-olds; M=6.9 years old, stdev=1.9), 19 were 10 to 16 years old (9, 10-to-11-year-olds; 6, 12-to-13-year-olds and 4, 13-to-16-year-olds; M=12 years old, stdev=1.9), and eight were 18 years old or older.

The task involved asking questions such as “Are there two X?” where X was a noun (paca [animal], canoe [artifact], shirt [artifact], or banana [fruit]). The task included four different types of visual stimuli:

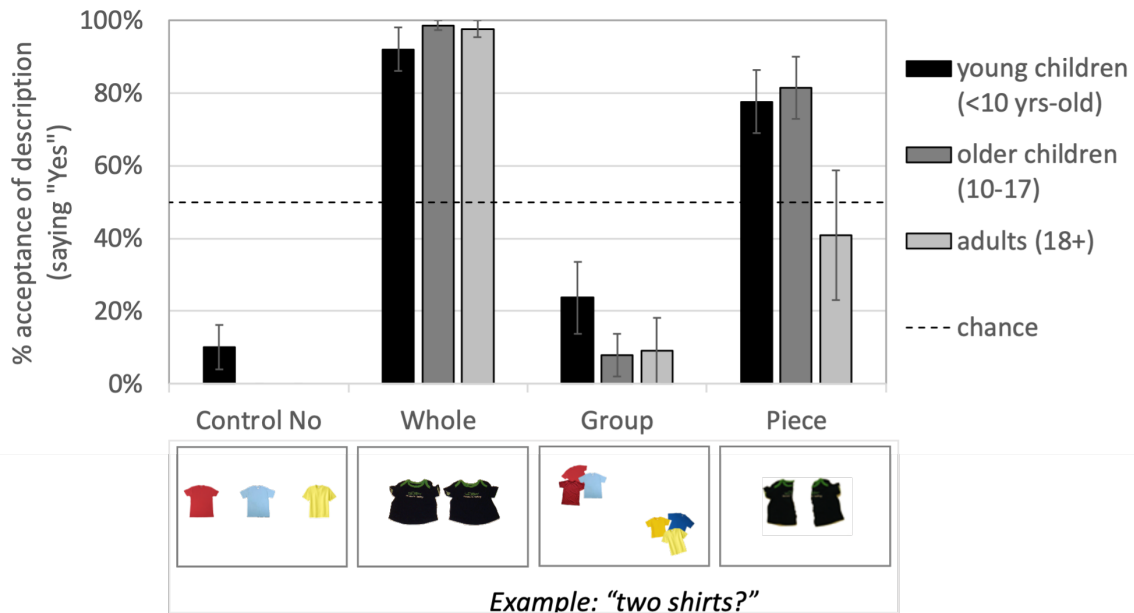
1. A picture that presented two whole objects (e.g. two canoes).
2. A picture that presented two pieces of a given object (e.g., a canoe broken in two).
3. A picture of two groups of a given object (e.g., two groups of three canoes).

<sup>8</sup>See Lima (2014) for a discussion on words for describing pieces in Yudja.



- A control item where there was a number mismatch (e.g., three shirts, while hearing the question “are there two shirts?”).

In this task, the control questions were an important measure to evaluate whether the results were reliable (a ‘yes’ answer to a control No question would indicate that either the participant was distracted, or the method was not working). The results have shown the same pattern observed in the first study for two of the three conditions (pieces and whole objects), suggesting that the judgments for the ‘piece’ condition, the critical condition of the study, were reliable.



**Figure 9:** Felicity Judgment Results. Adult and children’s acceptance of the numeral noun phrase for a given picture. Pictures were spatially and temporally discrete individuals or groupings that matched in number and kind of object mentioned in the queried phrase (e.g., *yauḍa pīza*; ‘two canoe(s)’). The ‘individual’ pictures depicted individual whole objects (two canoes), the ‘piece’ pictures depicted an individual cut into pieces (two pieces of a canoe), and the “group” pictures depicted groups of whole objects (two groups of three canoes) (Lima, Li and Snedeker 2017).

In sum, in both studies children were more likely to accept the use of bare nouns to refer to pieces. The first task was communicatively relevant (the participants were helping someone to improve their Yudja language skills) and flexible. By providing speakers the opportunity to describe the items as they saw fit, the task also allowed the researchers to expand their understanding of the language by analyzing the new items provided by the speakers and also gave the participants the opportunity to refute the options provided. The second task (forced-choice task) allowed us to verify whether the results of the first task were reliable by means of employing a different methodology.

#### 4 Final remarks

In this paper, I presented a workflow for experimental design in linguistic fieldwork. I provided some examples of how experimental design may be used to further the description and analysis of a language. I also argued that combining different sources of data (spontaneous speech, context-based elicitation, experimental studies) can bolster the description of a phenomenon in a language. In these final remarks, I will emphasize some of the lessons learned in designing studies and conducting semantic fieldwork in South America.

**Experimental design** Throughout most of this paper, I discussed experimental design. In my experience, the process of discussing the design of the studies with local Indigenous teachers (who would not become participants in the study) was also an additional opportunity to better understand the language. I have also learned from local teachers that stimuli used in the studies were helpful for them in preparing language classes. In working closely with speakers in study design, we can also achieve another critical goal: to contribute to building local capacity (that is, having local Indigenous researchers invested in language documentation). As discussed in this paper, in designing a study, a researcher must make sure that the method is adequate not only for the question being asked but also for the community of speakers where it is going to be used. As mentioned in Section 2, one must also consider familiarization tasks (Section 2.2.1) to ensure the participants are comfortable with the task.

In order to control the conditions that might impact results, knowing how the phenomenon manifests across languages is not enough; descriptive work based on spontaneous data (e.g., recording of traditional narratives), semi-spontaneous data (e.g., data from storyboards) as well as data elicited in elicitation sessions (e.g., context-based elicitation, grammaticality/truth value/felicity judgment tasks with a small group of speakers or in one-on-one sessions) can serve as the base for small-to-medium scale studies in a community. That is, while our theoretical/typological knowledge on a topic is instrumental and informative, a clear understanding of a phenomenon in a given language precedes experimental-type studies with a higher number of speakers.

**Language variation** In her list of lessons learned in the field, Rice (2001: 230) indicates that one should not “think that language is a monolithic entity within a community. There is variation within language, and this must be part of any analysis.” Small-to-medium scale studies as the ones described in this paper can be used to explore language variation in a given language. The case studies presented (on the interpretation of container phrases and object-denoting nouns) involved studies with Yudja speakers of different age groups. While these studies can contribute to the understanding of synchronic variation, it can also shed light on language development (if a study is performed with children), and language change (if one analyzes, for example, adults of different generations). This can be achieved by a) using the same stimuli with different age groups in a community and b) employing a quantitative and qualitative analysis of the data (cf. the analysis of quantity judgment tasks with bilingual Yudja speakers and the impact of late exposure to Brazilian Portuguese when these tasks were performed in Brazilian Portuguese (Lima 2014, 2018)).

**Holding methods constant across languages** It is undeniable that the advancement of typological research and theoretical models depends on the description of under-described languages. Examples of the relevance of these contributions can be observed in a variety of domains. In the count/mass distinction domain, Chierchia (2010) presented an updated version of his 1998 nominal parameter typology that included a new type of language described by Wilhelm (2008). Another example of the impact of the description of underrepresented languages in the theoretical literature is the

emergence of descriptions of languages that lacked D-quantification (for example, Mohawk [North America, Baker 1995], Strait Salish [North America, Jelinek 1995], Asurini do Trocara [Guarani/South America, Vieira 1995], among others). These descriptions impacted Barwise and Cooper (1981: 179)'s Determiner Universal.<sup>9</sup>

Cross-linguistic investigations of a phenomenon can benefit from the use of the same method across languages. This does not mean that the same task should be used in different languages without any modifications: while critical aspects of the task should be kept constant, some adaptations will be necessary in order to make the task appropriate across languages (aspects of a context, lexical choices). Using the same method consistently across languages (see Section 2.2.2) makes it possible to formulate more rigorous cross-linguistic generalizations since the data for each language has been elicited under similar conditions (see Lima and Rothstein 2020 for an illustration in the domain of the count/mass distinction). Methods can only be kept constant if a clear description of the tasks and materials is provided in publications; as such, researchers working on individual projects of language description should keep in mind the value of including a 'materials and methods' section in their publications, regardless of the number of participants (see Tonhauser and Matthewson 2015 for a discussion on this issue).

## References

- Bale, Alan, and David Barner. 2018. Quantity judgment and the mass-count distinction across languages: Advances, problems, and future directions for research. *Glossa: a journal of general linguistics* 3(63): 1-23.
- Baker, Mark C. 1995. On the absence of certain quantifiers in Mohawk. In *Quantification in natural languages*, ed. by Bach Emmon, Eloise Jelinek, Angelica Kratzer, and Barbara H. Partee, 21-58. Dordrecht: Springer.
- Barner, David, and Jesse Snedeker. 2005. Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97 (08): 41-66.
- Barner, David, Peggy Li, and Jesse Snedeker. 2010. Words as windows to thought: The case of object representation. *Current Directions in Psychological Science* 19(3): 195-200.
- Barner, David, and Jesse Snedeker. 2005. Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97 (1) (08): 41-66.
- Barwise, Jon, and Robin Cooper. 1981. Generalized Quantifiers and Natural Language. *Linguistics Philosophy* 4: 159-219.
- Crain, Stephen and Rosalind Thornton 1998. Truth Value Judgments. In *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*, ed. by Stephen Crain and Rosalind Thornton. 209-213. Cambridge: MIT Press.
- Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174(1): 99-149.
- Chierchia, Gennaro. 2015. How universal is the mass/count distinction? In *Three grammars of counting*. *Chinese syntax: A cross-linguistic perspective*, ed. by Yen-hui Audrey Li, Andrew Simpson, Wei-Tien Dylan Tsai, 147-177. Oxford: Oxford University Press.
- Davis, Henry, and Lisa Matthewson. 1999. On the functional determination of lexical categories. *Revue québécoise de linguistique* 27(2): 29-69.

---

<sup>9</sup>Every natural language contains basic expressions (called determiners) whose semantic function is to assign to common noun denotations (i.e., sets) a quantifier that lives on A (Barwise and Cooper 1981: 179).

- Deal, Amy Rose. 2017. Countability distinctions and semantic variation. *Natural Language Semantics* 25(2): 125-171.
- Gal'perin, Piotr, and Lachezar S. Georgiev. 1969. The formation of elementary mathematical notions. *Soviet studies in the psychology of learning and teaching mathematics* 1: 189-216.
- Gillon, Carrie. 2010. The mass/count distinction in Innu-aimun: Implications for the meaning of plurality. In *WSCLA 15: The fifteenth workshop on structure and constituency in languages of the Americas*, ed. Beth Rogers and Anita Szakay, 12-29. Vancouver: University of British Columbia.
- Gillon, Carrie. 2012. Evidence for mass and count in Inuttut. *Linguistic Variation* 12(2): 211-246.
- Gordon, Peter. 1998. The truth-value judgment task. In *Methods for assessing children's syntax*, ed. by Dana McDaniel, Cecile McKee, Helen Smith Cairns. 211-231. Cambridge, Mass: MIT Press.
- Huang, Yi Ting, Elizabeth Spelke, and Jesse Snedeker. 2013. What exactly do numbers mean?. *Language Learning and Development* 9(2): 105-129.
- Jelinek, Eloise. 1995. Quantification in straits Salish. In *Quantification in natural languages*, ed. by Emmon Bach, Eloise Jelinek, Angelika Kratzer, and Barbara Partee, 487-540. Dordrecht: Kluwer Academic Publishers.
- Khrizman, Keren, Fred Landman, Suzi Lima, Susan Rothstein, and Brigitta R. Schvarcz. 2015. Portion readings are count readings, not measure readings. In *Proceedings of the 20th Amsterdam Colloquium*, ed. by Thomas Brochhagen, Floris Roelofsen and Nadine Theiler, 197-206. Amsterdam: ILLC, University of Amsterdam.
- Landman, Fred. 2004. *Indefinites and the type of sets*. Oxford: Blackwell.
- Levin, Iris, and Friedrich Wilkening. 1989. Measuring time via counting: The development of children's conceptions of time as a quantifiable dimension. In *Advances in psychology*, ed. Iris Levin and Dan Zakay (Volume 59), 119-144. Amsterdam, New York: North-Holland.
- Lima, Suzi. 2014. The grammar of individuation and counting. Doctoral Dissertation, University of Massachusetts Amherst.
- Lima, Suzi. 2016. Container constructions in Yudja: locatives, individuation and measure. In *Baltic International Yearbook of Cognition, Logic and Communication (Number: Cognitive, Semantic and Crosslinguistic Approaches)*, ed. by Susan Rothstein and Jurģis Šķilters, 1-40. Manhattan: New Prairie Press.
- Lima, Suzi, Peggy Li and Jesse Snedeker. 2017. Acquiring the denotation of object denoting nouns. In *Language acquisition at the interfaces: Proceedings of GALA 2015*, ed. Jiyoung Choi, Hamida Demirdache, Oana Lungu, Laurence Voeltzel, 176-195. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lima, Suzi. 2018. Quantity judgment studies in Yudja (Tupi): Acquisition and interpretation of nouns. *Glossa: a journal of general linguistics*, 3(1).
- Lima, Suzi, and Susan Rothstein. 2020. A typology of the mass/count distinction in Brazil and its relevance for mass/count theories. *Linguistic Variation*, 20(2): 174-218.
- Louie, Meagan. 2015. The problem with no-nonsense elicitation plans (for semantic fieldwork). In *Methodologies in semantic fieldwork*, ed. by Ryan Bochnak and Lisa Matthewson, 47-71. Oxford: Oxford University Press.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International journal of American linguistics* 70(4): 369-415.
- Mathieu, Eric. 2012. On the mass/count distinction in Ojibwe. In *Count and mass across languages*,

- ed. by Diane Massam, 172-98. Oxford: Oxford University Press.
- Moore, Denny. 2006. Brazil: Language situation. In *Encyclopedia of Language and Linguistics*, ed. by Keith Brown, 117-28. Elsevier Science.
- Moore, Denny, Ana Vilacy Galucio and Nilson Gabas Jr. 2008. Desafio de documentar e preservar línguas. *Scientific American (Brasil): Amazônia, A Floresta e o Futuro* (3), 36-43.
- Partee, Barbara H., and Vladimir Borshev. 2012. Sortal, relational, and functional interpretations of nouns and Russian container constructions. *Journal of Semantics* 29(4): 445-486.
- Rice, Keren. 2001. Learning as one goes. In *Linguistic fieldwork*, ed. by Paul Newman and Martha Ratliff, 230-249. Cambridge: Cambridge University Press.
- Rothstein, Susan. 2009. Measuring and counting in Modern Hebrew. *Brill's Annual of Afroasiatic Languages and Linguistics* 1: 106-45.
- Rothstein, Susan. 2010. Counting and the mass/count distinction. *Journal of semantics*, 27(3), 343-397.
- Rothstein, Susan. 2011. Counting, measuring and the semantics of classifiers. *The Baltic International Yearbook of Cognition, Logic and Communication* 6: 1-42.
- Rothstein, Susan. 2012. Numericals: Counting, measuring and classifying. In *Proceedings of Sinn und Bedeutung* 16(2), ed. by Ana Aguilar Guevara, Anna Chernilovskaya, and Rick Nouwen, 527-542. Cambridge: MIT Working Papers in Linguistics.
- Selkirk, Lisa. 1977. Some remarks on noun phrase structure. In *Formal syntax*, ed. by Peter W. Culicover, Thomas Wasow and Adrian Akmajian, 285-316. New York: Academic Press.
- Schwarzschild, Roger. 2011. Stubborn distributivity, multiparticipant nouns and the count/mass distinction. In *Proceedings of NELS* 39 (2), ed. by Brian Smith Suzi Lima, Kevin Mullin, 661-678. Amherst: Graduate Linguistics Students Association.
- Silva, Wilson and Scott AnderBois. 2016. Fieldwork game play: Masterminding evidentiality in Desano. *Language Documentation and Conservation*. 10: 58-76.
- Stolze Lima, Tania. 2018. Yudjá/Juruna. In *Povos Indígenas no Brasil*. <https://pib.socioambiental.org/en/Povo:Yudj>
- Tonhauser, Judith, and Lisa Matthewson. 2015. Empirical evidence in research on meaning. Ms., The Ohio State University and University of British Columbia.
- Vieira, Marcia Damaso. 1995. The Expression of Quantificational Notions in Asurini Do Trocará: Evidence Against the Universality of Determiner Quantification. In *Quantification in natural languages*, ed. by Emmon Bach, Eloise Jelinek, Angelika Kratzer, and Barbara Partee, 701-720. Dordrecht: Kluwer Academic Publishers.
- Wagers, Matthew, Manuel F. Borja, and Sandra Chung. 2015. The real-time comprehension of WH-dependencies in a WH-agreement language. *Language* 19(9): 109-144.
- Wagners, Matthew and Chung, Sandra. To appear. Language processing experiments in the field. In *Oxford Handbook of Experimental Syntax*, ed. by Jon Sprouse. Oxford: Oxford University Press.
- Wilhelm, Andrea. 2008. Bare nouns and number in Dëne Sų́iné. *Natural Language Semantics* 16(1): 39-68.
- Wiltschko, Martina. 2012. Decomposing the mass/count distinction: Evidence from languages that lack it. In *Count and mass across languages*, ed. by Diane Massam, 146-171. Oxford: Oxford University Press.