RESEARCH REPORT

OPTIMAL CATEGORISATION: THE NATURE OF NOMINAL CLASSIFICATION SYSTEMS

Alexandra GRANDISON ⓘ ✉

School of Psychology – University of Surrey

Michael FRANJIEH ⓘ ✉

Surrey Morphology Group – University of Surrey

Lily GREENE ⓘ ✉

School of Psychology – University of Surrey

Greville G. CORBETT ⓘ ✉

Surrey Morphology Group – University of Surrey

ABSTRACT

The debate as to whether language influences cognition has been long standing but has yielded conflicting findings across domains such as colour and kinship categories. Fewer studies have investigated systems such as nominal classification (gender, classifiers) across different languages to examine the effects of linguistic categorisation on cognition. Effective categorisation needs to be informative to maximise communicative efficiency but also simple to minimise cognitive load. It therefore seems plausible to suggest that different systems of nominal classification have implications for the way speakers conceptualise relevant entities. A suite of seven experiments was designed to test this; here we focus on our card sorting experiment, which contains two sub-tasks — a free sort and a structured sort. Participants were 119 adults across six Oceanic languages from Vanuatu and New Caledonia, with classifier inventories ranging from two to 23. The results of the card sorting experiment reveal that classifiers appear to provide structure for cognition in tasks where they are explicit and salient. The free sort task

did not incite categorisation through classifiers, arguably as it required subjective judgement, rather than explicit instruction. This was evident from our quantitative and qualitative analyses. Furthermore, the languages employing more extreme categorisation systems displayed smaller variation in comparison to more moderate systems. Thus, systems that are more informative or more rigid appear to be more efficient. The study implies that the influence of language on cognition may vary across languages, and that not all nominal classification systems employ this optimal trade-off between simplicity and informativeness. These novel data provide a new perspective on the origin and nature of nominal classification.

RÉSUMÉ

Le débat concernant l'influence de la langue sur la cognition existe depuis longtemps et a produit des résultats contradictoires dans des domaines tels que les catégories de couleur et de parenté. Rares, cependant, sont les travaux qui ont étudié la variation à travers les langues des systèmes de classification nominale (genre, classificateurs) afin d'examiner les effets de la catégorisation linguistique sur la cognition. Une catégorisation efficace doit être informative pour maximiser l'efficacité de la communication, mais aussi simple pour minimiser la charge cognitive. Il semble donc plausible que différents systèmes de classification nominale puissent influer sur la façon dont les locuteurs conceptualisent les entités pertinentes. Nous avons conçu une série de sept expériences pour tester cette hypothèse. Nous nous concentrons ici sur une expérience de tri de cartes, qui contient deux sous-tâche — un tri libre et un tri structuré. Les 119 participants sont des adultes, locuteurs de six langues océaniques du Vanuatu et de la Nouvelle-Calédonie. Ces langues ont des inventaires de classificateurs allant de deux à 23. Les résultats de l'expérience de tri de cartes révèlent que les classificateurs semblent fournir une structure pour la cognition dans les tâches quand ils sont explicites et saillants. Nos analyses quantitatives et qualitatives montrent que la tâche de tri libre n'a pas incité à la catégorisation au moyen de classificateurs, sans doute parce qu'elle nécessitait un jugement subjectif, plutôt qu'une instruction explicite. En outre, nous observons moins de variation dans les groupes produits par les participants dans les langues présentant un système de classification

plus complexe. Ainsi, les systèmes plus informatifs ou plus rigides semblent plus utiles pour la cátegorisation. Cette étude indique que l'influence de la langue sur la cognition peut varier d'une langue à l'autre, et que les systèmes de classification nominale n'utilisent pas tous un compromis optimal entre simplicité et informativité. Ces nouveaux résultats offrent une perspective originale sur l'origine et la nature de la classification nominale.

INTRODUCTION

Linguistic communication is a fundamental aspect of human experience, and one which shows great variation, given that there are over 7,000 different languages spoken worldwide (EBERHARD; SIMONS; FENNIG, 2021). It has been widely debated whether the language spoken influences speakers' perception and cognition; the claim that it does is known as the Sapir-Whorf hypothesis (CARROLL, 1956). Some researchers argue that structural differences in language influence the way a native speaker thinks about the world (e.g. BROWN; LENNEBERG, 1954; LUPYAN et al., 2020) whilst others claim that language does not shape cognition (e.g. SCOVEL, 1991; SPEED et al., 2020). Slobin (1996) reinterprets the Sapir-Whorf hypothesis as 'thinking for speaking', a special type of thought process involved in speech production, and argues that the variation in obligatory grammatical elements across languages enables speakers to focus on different aspects of experience. Categorisation is a process central to both communication and thought, and therefore provides an appropriate arena to explore this much debated relationship between language and cognition. Hawkins (2004) suggests that systems of categorisation should be informative to maximise communicative effectiveness, and simple to minimise cognitive effort, yet there is much variation in the types and numbers of categories utilised across different languages of the world. For example, a system with many different categories is highly descriptive and so highly informative but the result of having many descriptive categories is that this relative complexity leads to high cognitive load.

Research investigating the Sapir-Whorf hypothesis has focused on a number of different domains (GUMPERZ; LEVINSON, 1996) and a significant amount of work has been conducted on linguistic colour categories (BERLIN; KAY, 1969). For example, it has been found that differences in colour naming systems across languages reflect a near optimal trade-off between simplicity and informativeness (JAMESON; D'ANDRADE, 1997). Specifically, in colour naming systems that only have two terms those terms occur at maximally distant points within colour space, so as to be optimally descriptive of the space. An example of this is the Dani language of Western New Guinea, which uses the terms *mola* for bright, warm colours and *mili* for dark cold colours. Systems with additional terms tend to locate these referents at points that are located furthest from the existing referents within colour space. Such investigations on colour naming systems have investigated British English speakers (FRANKLIN et al., 2008; GILBERT et al., 2006) and made comparisons with monolingual and bilingual populations in Europe (DELGADO, 2004; WINAWER et al., 2007), East Asia (THAM et al., 2020), Africa (PILLING; DAVIES, 2004; ROBERSON et al., 2006) and Oceania (ROBERSON; DAVIES; DAVIDOFF, 2000). The findings from such studies investigating the influence of language on colour categories is conflicting. It can be concluded that there are universal boundaries in colour naming that may govern

prelinguistic colour cognition (e.g. CLIFFORD et al., 2009; MAULE; FRANKLIN, 2019), but also that differences across languages in colour naming create variations in colour cognition (KAY; REGIER, 2006), which appears to begin during colour term acquisition (FRANKLIN et al., 2008). See Regier and Kay (2009) for a review. However, Bi (2017) argued that speakers' conceptualisation may be altered by personal experience and culture-related knowledge, and so the differences found in cognition may be due to culture, not language. Gibson et al. (2017) highlight this, suggesting that American English speakers (with 11 colour terms) were more efficient at naming and distinguishing colours than the Tsimane people of the Amazon, who have only three colour terms. The results were attributed to industrialisation, and colour terms being developed for artificial colours. On the other hand, the results could also be due to a lack of task familiarity for the Tsimane speakers compared to the American English speakers. Thus, it is important to consider the interplay between task familiarity, cultural, environmental, and linguistic differences when making comparisons between speakers of different languages.

Further research exploring the Sapir-Whorf hypothesis has focused on kinship categories. Researchers have argued that languages differ in their categorisation of kinship terms, but similarly to colour categories, the range of variation appears to be limited. Kemp and Regier (2012) state that kin classification systems exhibit a near optimal trade-off between simplicity and informativeness, to minimise cognitive load and maximise efficient communication. However, these systems also show differences across cultures, as they often relate to and are shaped by the cultural norms of marriage and living. Differences in kinship categories across languages may also be shaped by communicative needs, and as these communicative needs vary across cultures, the semantic systems should vary also (KEMP; XU; REGIER, 2018). As semantic category systems are explained by the same general principles across disparate domains such as colour and kinship categories, it is arguable that categorisation has a domain-general basis. Therefore, it is important to investigate linguistic relativity in other domains of categorisation, such as grammatical gender and classifier systems.

It can be argued that aspects of grammar are a better means by which relativity can be investigated compared to colour research, as grammar is not affected by sensory input. Thus, language does not need to breach psychophysical barriers (SAMUEL; COLE; EACOTT, 2019). Vigliocco et al. (2005) argue that language may have less effect on discrete domains, such as objects and sex, than on continuous domains such as colour, time, and space. However, few studies have set out to investigate specific language systems across different languages, to examine how speakers of those languages perform in classification tasks (SERA et al., 2002). A broad variety of nominal classification systems exist across languages, ranging over gender and classifier systems (see, for instance, SENFT, 2000). Grammatical gender is extremely widespread across the world, present in some 44% of the

languages in Corbett's 256-language sample (2005). There is considerable variation; all gender systems have 257 semantic core, most often including female versus male, but a sizable minority have animate versus inanimate. The origin of such systems is naturally of great interest, provoking debate as to whether the origin of gender lies in meaning (GRIMM, 1831) or form (BRUGMANN, 1889).

Until recently, typologists tended to oppose gender systems to classifier systems. This was in part the result of historical accident, since the languages that were best known represented only a part of the typological space. Now that "in-between" systems have been documented, it is time to see these systems as part of a large typological space, in which we can identify criteria which exist in a wide range of combinations (FEDDEN; CORBETT, 2017, 2018; CORBETT; FEDDEN; FINKEL, 2017; CORBETT; FEDDEN, 2018). In order to provide an entry to the literature, however, it is useful to describe traditional classifier systems (as in GRINEVALD, 2000). In the clearest instances, classifiers are separate words used to classify nouns according to the shape, size and use of the referent. Similar to gender, classifiers have underlying conceptual meanings, but also exhibit large variability in the number of categories they contain, the size of these categories and in the consistency of their use (SPEED et al., 2016).

Grammatical gender has been shown to alter object concepts for German and Spanish speakers, as differences in similarity judgements between people and objects were attributed to grammatical gender (BORODITSKY; SCHMIDT; PHILLIPS, 2003). However, other researchers (BENDER; BELLER; KLAUER, 2011; LANDOR, 2014) found no evidence of a positive correlation between grammatical gender and the conceptualisation of objects, suggesting that gender does not alter conceptual representations. In a similar vein, Speed et al. (2016) argued that classifier systems reflect conceptual structure rather than affect it, supporting the view that language may reflect thought rather than constrain it. Specifically, speakers of Dutch (a language without classifiers) and speakers of Mandarin (a language with classifiers) were asked to assess the similarity of a target object to four comparison objects. One of the four comparison objects shared a classifier with the target object, while the others did not. Despite the fact that Dutch does not have classifiers, both Dutch and Mandarin speakers assessed the target object as more similar to the object with a shared classifier compared to the other distractor objects. Furthermore, Dutch and Mandarin speakers performed identically, suggesting that even speakers of a language without classifiers are sensitive to the object similarities that underlie classifier systems. This finding provides support for the idea that classifier systems may be closely linked to how humans conceptualise objects. In subsequent work Speed et al. (2020) did not find any difference between speakers of classifier vs. non-classifier languages (Mandarin and Dutch respectively) when looking at numeral classifiers specifically. This demonstrates that speakers of a non-classifier language are also affected by the types of conceptual similarities that

underpin classifier systems. Such systems may therefore reflect this conceptual structure, rather than shaping it. This indicates that the influence of language is perhaps more restricted than was previously suggested. Furthermore, Malt and Wolff (2010) found no evidence that classifier categories are used as foundations for categorisation, compared to thematic or taxonomic relations, when classifiers are not explicitly stated. Therefore, classifiers may not function to structure and influence cognition. However, previous research separated stimuli according to language. For example, object names were given in Mandarin for Mandarin speakers and in English for English speakers (SPEED et al., 2016). Thus, the differing behaviour of speakers across languages may be attributed to the grammar or lexis of the speakers' native language, rather than due to any underlying differences in cognition (PINKER, 1997). Beller et al. (2015) aimed to disentangle cultural and linguistic effects on cognition by asking participants to assign a male or female voice to a range of nouns from different semantic categories. It was found that gender systems did have an impact on voice assignment, but these grammatical effects were small in comparison with effects brought about by culturally conveyed associations. This suggests that cultural factors are a stronger influence than linguistic factors within and across semantic domains. Thus, the evidence on the relationship between grammatical gender and cognition is very mixed to date. This has been recently exemplified by Samuel et al. (2019) in a systematic review of 43 articles involving almost 6000 participants. The overall conclusion was that support for linguistic relativity within the domain of grammatical gender was strongly task and context dependent. Indeed, as Bender et al. (2016) summarise, it appears that cognitive effects of gender are more likely to occur when stimuli are presented linguistically than when they are presented as pictures when measures explicitly ask for gender-relevant assignments, and also when languages have two genders (feminine and masculine) rather than three (feminine, masculine, neuter class). However, Bender et al. (2016) did find gender congruency effects in German, a language with three genders, when taking into account different semantic categories of nouns. Despite this, previous studies have utilised a range of different methodologies and focused on different noun categories, making it difficult to draw concrete conclusions about the interplay between language and thought within this domain (for an overview see CUBELLI, PAOLIERI, LOTTO & JOB, 2011). This highlights a need for further experimental work exploring grammatical gender, including classifier systems, to enable direct comparison across tasks within the same context. This will enable the mechanisms underlying any relative effects of language on cognition to be more clearly revealed.

1. THE CURRENT STUDY

Oceanic languages contrast inalienable and alienable possession using two different possessive constructions – direct and indirect. Possession here is understood broadly, including 'possession' of body parts and kin, through to legal possession of items, and temporary association. Inalienable possession is typically encoded by the direct construction, in which the possessor indexing is marked directly on the possessed noun (1a). Alienable possession, on the other hand, uses the indirect possession construction; it is typically marked using one of a set of possessive classifiers where the possessor indexing attaches to the possessive classifier rather than the possessed noun (1b), hence the term 'indirect construction'.

Iaai, New Caledonia (DOTTE, 2017, p. 341)

1a. *hinyö-k*   b. *haalee-k*   *kuli*
   mother-1SG   CL.PET-1SG   dog
   'my mother'   'my dog (pet)'
   inalienable possession   alienable possession
   direct construction   indirect construction

Our study on optimal categorisation focuses solely on the indirect construction with possessive classifiers which, in the Oceanic languages, range from two (e.g. Merei, Vanuatu) to more than thirty classifiers found in the Micronesian sub-group within Oceanic. The inventory size of the possessive classifiers in our sample languages ranges from two to 23 (Table 1). An important point about these Oceanic possessive classifiers is that they have been termed 'relational' (LICHTENBERK, 1983). Relational indicates that a possessed noun can co-occur with different classifiers depending on the relation of the possessor to the referent of the noun, that is, how the possessor intends to use the possessed noun referent, as in (2).

Lewo, Vanuatu (EARLY, 1994, p. 216)

2a. *ma-na*   *wi*   b. *sa-na*   *wi*
   CL.DRINK-3SG   water   CL.GENERAL-3SG   water
   'her water (for drinking)'   'her water (for washing)'

The classifier systems in the Oceanic languages map the principles of simplicity and informativeness onto the speaker-hearer distinction. Simple systems with fewer classifiers favour the speaker, whereas informative systems with more classifiers favour the hearer. There is an added principle of the degree of semantic transparency held by the members of each classifier. We therefore see a trade-off in the Oceanic classifier systems between the principles of simplicity, informativeness and semantic transparency.

Figure 1. Map showing the location of the six languages included in the current study.

We have chosen carefully a sample of six Oceanic languages spoken in Vanuatu and New Caledonia (Figure 1). Our choice is based on several criteria. First, the sample languages are closely related so that we can use the differences between the languages as a proxy for development through time. Second, the sample languages vary in their inventory size of classifiers so that we can investigate how optimal a simple system with two classifiers is compared to an informative system with 23 classifiers. Third, the classifiers themselves vary in their semantic transparency in terms of the membership of nouns. For example, the drink classifier in Iaai and Nêlêmwa only allows nouns whose referents are intended to be drunk, while the corresponding classifier in Lewo, Vatlongos and North Ambrym not only includes nouns whose referents are able to be drunk, such as water or green coconut, but also nouns whose referents refer to houses and mats. Finally, the sample languages represent different stages in the grammaticalisation from noun to classifier to marker of grammatical gender. Due to the varying inventory size of the classifier systems in the sample languages not all classifiers are shared across all languages. Even the three common classifiers - general, food and drink - reconstructed for Proto Oceanic are only shared across four languages in our sample, as shown in Table 1.

Language	Number of Classifiers	Common Classifier Categories		
		General	Food	Drink
Merei	2	✓	[1]	
Lewo	3	✓	✓	✓
Vatlongos	4	✓	✓	✓
North Ambrym	5	✓	✓	✓
Nêlêmwa	20	✓[2]	[3]	✓
Iaai	23	✓	✓	✓

Table 1. The size of classifier inventories and the common classifier categories.

The first stage of grammaticalisation is the emergence of classifiers from other parts of speech. There is evidence of this first stage in Nêlêmwa where some directly possessed nouns can be used as incipient classifiers, as in (3). Bril (2002) reports that the directly possessed noun *waja-ny* 'my boat', given in (3a) can be used as an incipient classifier for boats but only if the possessed noun is modified, as in (3b) with the modifier *hnap* 'sail'. These appositional phrases are precursors to classifiers.

Nêlêmwa, New Caledonia (BRIL, 2002, pp. 259, 369)
3a. *iva*   *waja-ny*   b. *waja-ny*   *wany*   *hnap*
    where   boat-1SG   boat-1SG   boat   sail
    'where is my boat?   'my sail boat'

However, work with speakers of Nêlêmwa revealed that there is considerable variation in how speakers use these incipient classifiers. Some speakers are able to use the *waja-* as a classifier, without the possessed noun being modified. Other speakers use *waja-* as a classifier with extended semantics, which includes all transport items such as cars and bicycles.

Classifiers themselves can become more rigid in terms of their noun-classifier collocation and thus function more similarly to a grammatical gender system in terms of noun assignment. For example, the noun *we* 'water' in North Ambrym can only ever occur with the drink classifier. Compare (4) with the comparable Lewo example (2).

North Ambrym, Vanuatu (FRANJIEH, 2016, p. 95)
4a. **ma**-*n*   *we*   b. *****mwena**-*n*   *we*
    CL.DRINK-3SG   water   CL.GENERAL-3SG   water
    'his/her water (for any purpose)'   [not possible]

---

1   Merei has collapsed the two originally distinct edible and drinkable classifiers into one 'consumable' classifier.

2   Nêlêmwa does not have a possessive classifier that is used for general possessions (i.e. possessions not included in the other classifiers' semantic domains). Instead it employs either direct possessive constructions or a prepositional construction. We include the prepositional construction in our count of the inventory size of possessive classifiers as it functions semantically in a way similar to the general classifiers found in other Oceanic languages.

3   Nêlêmwa does not have a dedicated classifier covering all edible possessions. Instead it has developed independent possessive classifiers for fruit and vegetables, unripe fruit, starchy food, meat, chewable food, and sugarcane.

Turning now to a particularly relevant investigation, Franjieh (2016, 2018) investigated the nominal classification system of possessive classifiers in North Ambrym, an Oceanic language spoken in Vanuatu (see Figure 1). Participants were asked to list all nouns occurring with each classifier, in a free list task. Participants also completed a noun categorisation task, and were presented with a set of nouns and asked to provide the associated possessive classifier. Finally, participants were asked to describe video vignettes depicting different contextual interactions. Franjieh (2016) argued that the possessive classifiers in North Ambrym are markers of gender, due to the more rigid collocation between a noun and classifier, compared to more typical relational classifier systems found in other Oceanic languages. The study found that for North Ambrym, nouns that are more central to a classifier's semantic categories and are typical possessions, only occur with one classifier, whereas nouns that are less typical possessions and less central to the classifier's semantic categories can be assigned to different classifiers (FRANJIEH, 2018).

We aim to explore the findings of Franjieh (2016, 2018) further, through investigation of the semantic domains of the possessive classifier systems in six related Oceanic languages, in order to shed light on the cognitive functions and efficiency of varying classifier systems. Such an investigation has not been carried out before in a set of languages showing such significant linguistic variation yet with such high cultural similarity and close geographic proximity (Figure 1).

Having selected the languages for our study, we investigated the semantic domains of the possessive classifiers in these languages using a card sorting experiment, which enables us to investigate how participants categorise objects and what governs perceived similarity. Participants sorted images in a free card sort task, where they were asked to group the images in any way they chose, followed by a structured card sort task, where they were asked to group images based on the classifiers in their language. All participants were presented with the same stimuli, to enable comparison of sorting and the relative influence of classifiers across the different languages (MALT; WOLFF, 2010). Comparison between the groupings in the free sort task and the structured sort task provides an insight into the relative influence of linguistic and cultural factors on participants' categorisation. Data were analysed using a mixed methods approach.

It was predicted that classifiers would provide a structure for cognition for the languages tested in this study. A series of additional specific hypotheses were made to enable investigation of this overarching research question:

Hypothesis 1: Speakers of languages with larger classifier inventories (Iaai and Nêlêmwa) will produce a larger number of categories in the free sort and structured sort tasks, resulting in more categories across the two tasks, compared to speakers of languages with smaller classifier inventories (Merei, Lewo, Vatlongos, and North Ambrym);

Hypothesis 2: Similar numbers of categories will be made in the free sort and the structured sort tasks for all languages;

Hypothesis 3: Languages with larger classifier inventories (Iaai and Nêlêmwa) will exhibit greater variation in number of categories across participants in comparison to languages with smaller classifier inventories (Merei, Lewo, Vatlongos, and North Ambrym);

Hypothesis 4: North Ambrym, with a more rigid gender-like system of noun-classifier assignment, will exhibit less variation in comparison to languages with more typical classifier-like systems whose systems allow variable noun assignment (Merei, Lewo, Vatlongos, Nêlêmwa, and Iaai).

2. METHOD

Participants. The participants were 119 adults from villages in Vanuatu and New Caledonia. Data from one further participant was excluded since they collaborated on their answers. A G*Power analysis was conducted for a large effect size (d = .40) and an alpha of .05. Results showed that this number of participants was an appropriate sample size to achieve .80 power (FAUL et al., 2009). Of our sample, 54.5% were male, and 45.5% were female. The age range was 18-77 years (M = 43.25, SD = 15.29). Additional demographic information for each language is presented in Table 2. The education level of participants ranged from 0-19 years of education, and the average length of education for participants was 8.5 years. Participants were recruited through convenience and snowball sampling methods via initial contacts in each of the language communities. In compensation for their time, participants were offered a small monetary amount or gift in accordance with cultural norms of each country. Prior to conducting this study, ethical approval was obtained from the University of Surrey ethics committee.

Language	No. participants	Gender		Age		Years of education
		Male	Female	M	SD	M
Merei	21	15	6	32.57	11.67	6.30
Lewo	23	13	10	45.78	12.11	7.90
Vatlongos	24	9	15	36.71	13.35	6.30
North Ambrym	23	12	11	44.04	12.60	7.50
Nêlêmwa	12	7	5	53.58	19.13	11.00
Iaai	16	9	7	54.25	13.10	12.06

Table 2. Demographic information for each language, including overall number of participants, number of female and male participants, mean age and mean education level of participants.

Experimental Setup. The card sorting experiment was conducted alongside a free listing experiment and a video vignette experiment, which will be reported elsewhere. The order of these three experiments and the stimuli used within each was randomised across participants. However, the order of the three tasks within the card sorting experiment was fixed so that all participants did a naming task, followed by a free sort task and then a structured sort task. Testing took place across 59 days between July and October 2019;

during this time between one and three participants were tested each day. Participants were typically tested seated at a table inside or on the veranda of their house, a shelter or church. Participants usually completed the card sort task on mats on the floor, or on a large table. Before the recording of any data, participants were provided with information on the study, and gave informed consent. For the Vanuatu languages, the information sheet and consent form were read aloud to participants in Bislama (the lingua franca of Vanuatu) and oral consent was provided. Informed consent was recorded with an audio recorder, since some participants are not literate. For the New Caledonia participants, the information sheet and consent form were written in French (the lingua franca of New Caledonia), and presented to them to read and sign. These details were also summarised orally by the experimenter. The protocols were translated into Bislama and French and checked by one native speaker for Bislama and one for French. These were back-translated by English native speakers, one fluent in Bislama, and one fluent in French, to validate the translation.

Materials and Design. Participants were presented with 60 cards detailing a standardised set of images. These images depicted entities and different interactional uses of an entity. For example, a live pig and a pig being roasted (see Figure 2 for all images). The choice of images aimed to represent the total inventory of semantic categories covered by the classifiers in all six languages, apart from Iaai and Nêlêmwa. In Iaai 20 out of 23 classifiers were covered, with 'idea/thought', 'noise/sound' and 'mana/strength' not being covered, as these are abstract entities that were not possible to depict with the type of images used in this study. In Nêlêmwa there are four classifiers that are associated with different ways of catching and using fish – 'catch-speargun', 'catch-line', 'catch-net' and 'catch-sell'. The pictures of fish in our set of images were not targeted to these different specific classifiers, though they may occur with at least one. Thus 17 of Nêlêmwa's 20 classifiers were targeted. Images were sized 10cm x 10cm with a resolution of 1181 x 1181 pixels.
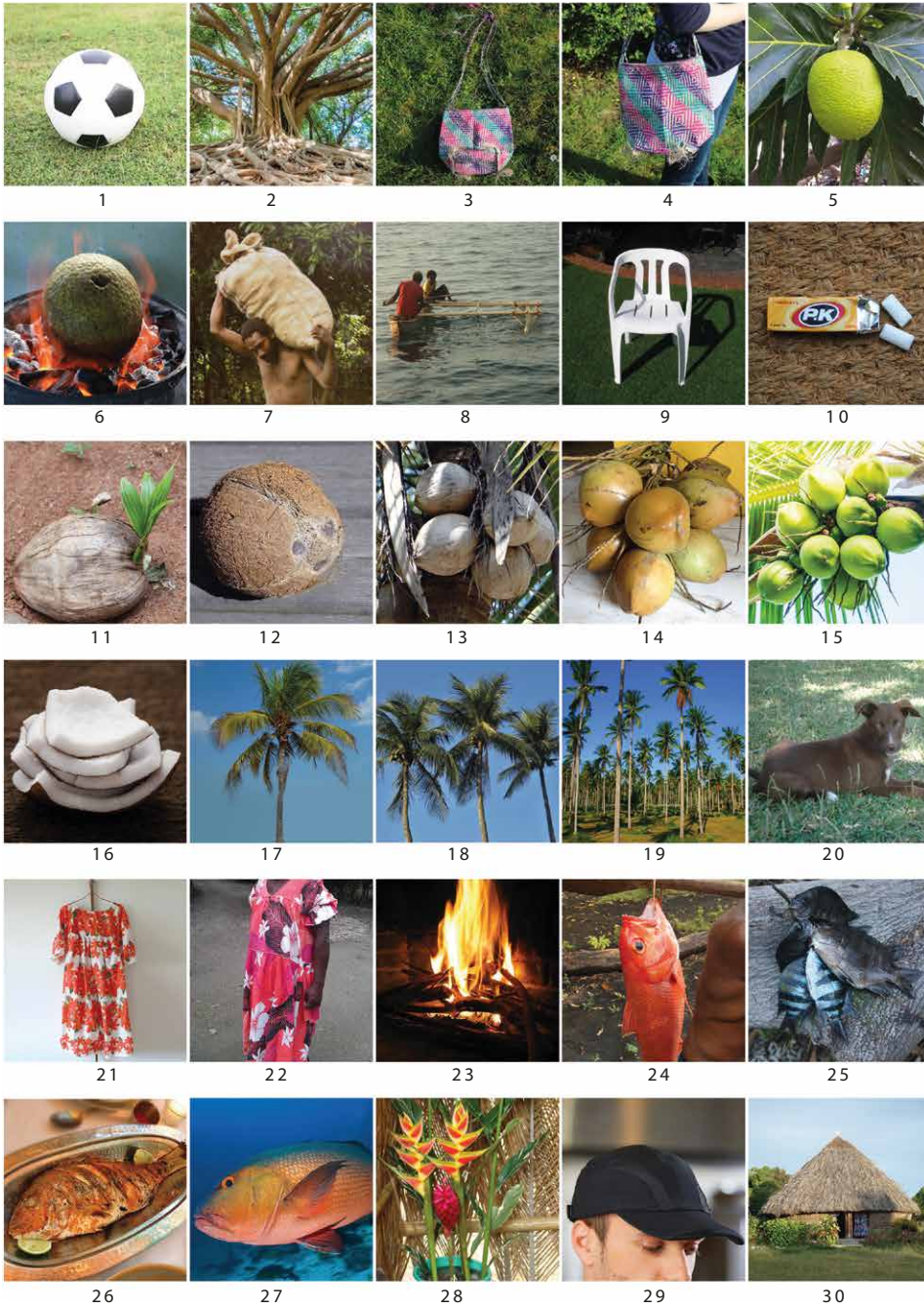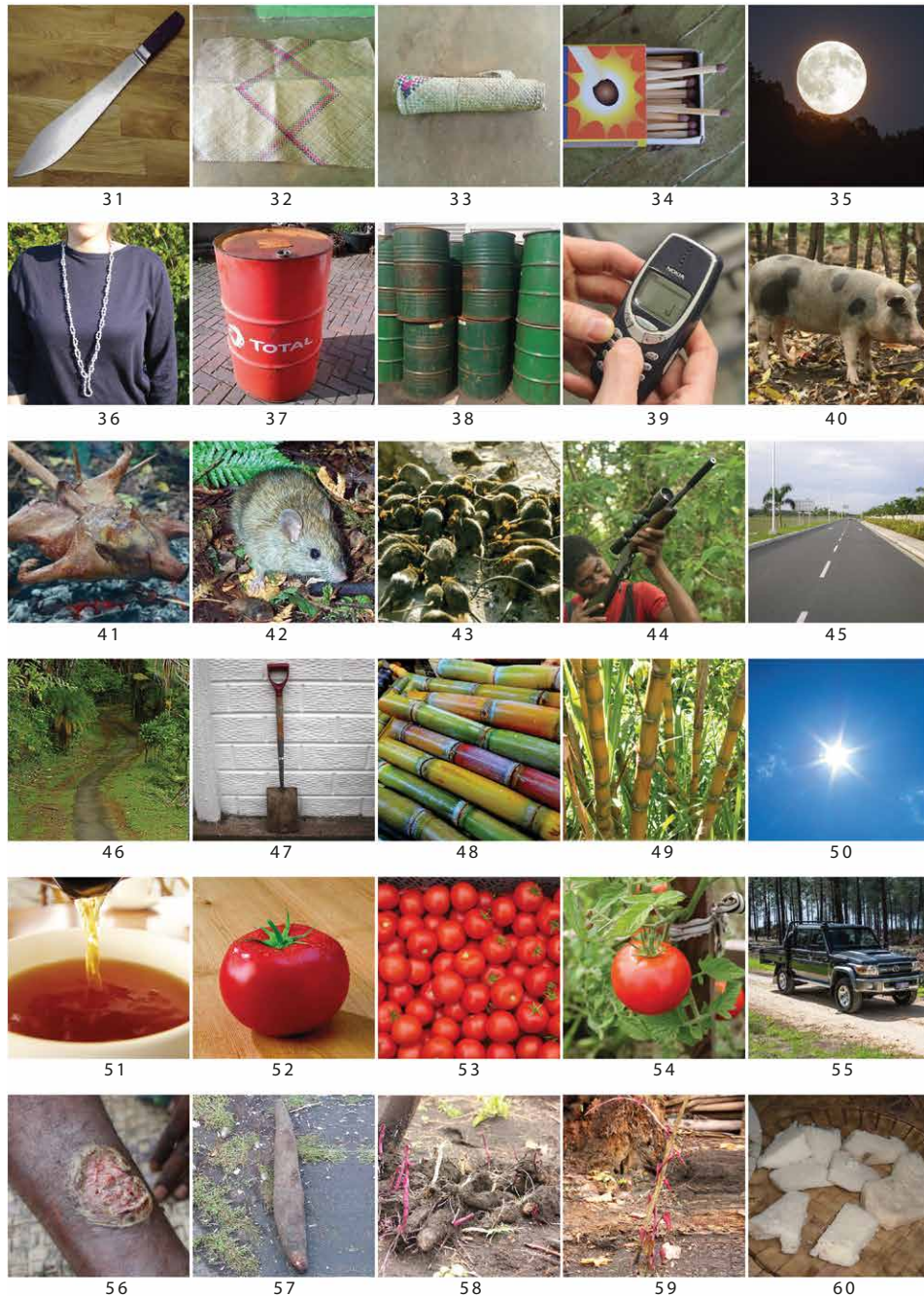
Figure 2. All images used in the pile sort experiment. The picture IDs correspond to the IDs in Table 3.

Procedure. Participants first completed a naming task. They were presented with each image in turn and asked to name the entity to ensure understanding. The images were then laid out in front of the participants, who were asked to free sort the cards into groups. The verbatim instructions for this task are given below:

There are sixty pictures in front of you. I would like you to sort these pictures into groups such that pictures in the same group are related to each other. There are many different ways you can group the pictures together. There is no right or wrong way to do it. You can group the pictures together any way you like according to your logic, your way of viewing things. You can then divide the pictures into different groups. You can make as many groups as you want to. Maybe there will be groups with lots of pictures, others with just one or two pictures, that's not a problem. Try not to spend too much time on it, It should be spontaneous. Then you will be asked to say a word or a little sentence in the language to explain each group.

After the task, participants were asked to give a summary of each pile, explaining why these cards were grouped together. The verbatim instructions are below:

Now, please tell me one word in your language that describes each of the groups that you have made. If you can't think of a single word, you can tell me several words or a sentence. In order to give a title to the group. You can start with any of the groups you have made.

The cards were then shuffled again and laid out in front of the participants who were then asked to sort the cards into groups relating to the classifiers of their mother-tongue language, in a structured card sort. This ensured that the different classifiers were used. Participants were asked to sort images into groups according to which word they would use to mean 'mine', similarly to the colour naming task in Grandison, Davies, and Sowden (2014). See below for the verbatim instructions for the structured sort task:

In your language <INSERT LANGUAGE NAME> there are many different ways of saying 'this belongs to me' depending on the function of the object. I want you to put the different pictures into groups so that the pictures in each group all go with a particular way of saying 'this belongs to me'. Each group of pictures that you make will go with a different way of saying 'this belongs to me'. It's as if for each image, you have to say to yourself 'this is mine and the name of the object' and you put it with the other images for which you would say the same for the word which means 'this is mine'.

3. RESULTS

Analysis overview and data screening. A mixed methods approach was used to analyse the data. Analysis of Variance (ANOVA) was conducted to investigate the numbers of categories made across languages for each of the tasks (free sort and structured sort). Pairwise comparisons were used to further investigate significant main effects and interactions. Statistical analyses were conducted using R (R CORE TEAM, 2020). A qualitative approach was employed to explore the labels assigned to the various piles produced by participants within the free sort. Both the labels and the cards sorted into the corresponding piles were analysed through a qualitative lens, taking cues from Thematic Analysis (BRAUN; CLARKE,

2006) in order to parse out any themes and/or narrative structure present in the labelling. Thematic Analysis involves becoming familiar with the similarities and differences within the data, generating 'codes', and defining the framework of overarching themes and sub-themes (BRAUN; CLARKE, 2020). An advantage of Thematic Analysis is that it is a flexible approach to organising data, which allows for nuanced thinking and consideration of unforeseen insights (TERRY et al., 2017). However, the loose structure of Thematic Analysis means that the data could become overwhelming, potentially leading to superficial observations and a lack of coherence (NOWELL et al., 2017). As such, to maintain the depth, richness, and quality of the observations, it was necessary to keep a rigorous track of the emerging themes from the initial analyses until the final stages of refinement (MCGRATH; PALMGREN; LILJEDAHL, 2019).

Naming task. Percentage accuracies were calculated for all languages for the images used in the card sort experiment (see Table 3). English translations of terms were assessed, and conceptually similar terms collapsed. For example, 'tea' and 'liquid tea' were collapsed as conceptually the same, as well as 'yam', 'yam for a ceremony' and 'one yam'. However, 'coconut' and 'fruit' were not collapsed, nor were 'clothes' and 'woman', for example. Images with ≤ 75% accuracy in one language were excluded from analysis across all languages, which led to our omitting five images: tea, necklace, dress worn, hat, sun. The figure of 75% was data driven and arose as a logical cut off to maximise the number of pictures included in the analysis whilst implementing stringent conceptual consistency.

ID	Picture Name	Language					
		Merei	Lewo	Vatlongos	North Ambrym	Nêlêmwa	Iaai
1	ball	100%	100%	100%	100%	100%	100%
2	banyan tree	100%	100%	100%	100%	77%	100%
3	basket not worn	100%	100%	100%	100%	100%	100%
4	basket worn	100%	100%	100%	100%	100%	100%
5	breadfruit on tree	100%	100%	100%	100%	100%	95%
6	breadfruit roasting	95%	96%	96%	96%	100%	95%
7	burden/heavy load	76%	87%	96%	100%	100%	100%
8	canoe	91%	100%	100%	100%	100%	88%
9	chair	100%	100%	100%	100%	100%	100%
10	chewing gum	81%	87%	100%	91%	100%	100%
11	coconut - germinating	100%	100%	100%	100%	100%	100%
12	coconut (dry) off tree	100%	100%	100%	100%	100%	100%
13	coconut (dry) on tree	100%	100%	100%	100%	100%	100%
14	coconut (green) off tree	100%	100%	100%	100%	100%	100%
15	coconut (green) on tree	100%	100%	96%	100%	100%	100%
16	coconut flesh	95%	100%	100%	91%	100%	100%
17	coconut palm	100%	100%	96%	100%	100%	100%
18	coconut palms	100%	100%	96%	100%	100%	100%
19	coconut plantation	95%	100%	96%	100%	100%	100%
20	dog	100%	100%	100%	100%	100%	100%
21	dress hanging	100%	100%	100%	100%	100%	100%
22	*dress worn*	*48%*	*44%*	*50%*	*52%*	*77%*	*88%*
23	fire	100%	100%	100%	100%	100%	100%
24	fish caught	100%	100%	100%	100%	100%	100%
25	fish pl caught	100%	100%	100%	100%	77%	100%
26	fish roasted	100%	100%	100%	100%	100%	100%
27	fish swimming	100%	100%	100%	100%	100%	100%
28	flower	100%	100%	100%	96%	100%	100%

29	*hat*	*62%*	*65%*	*71%*	*48%*	*77%*	*94%*
30	house	100%	100%	100%	100%	100%	100%
31	knife	100%	100%	100%	100%	100%	100%
32	mat (flat)	100%	100%	100%	100%	100%	100%
33	mat (rolled)	100%	100%	96%	100%	100%	100%
34	matches	76%	97%	96%	87%	100%	100%
35	moon	91%	100%	100%	91%	100%	100%
36	*necklace*	*67%*	*57%*	*71%*	*48%*	*100%*	*94%*
37	oil drum	100%	96%	100%	91%	100%	100%
38	oil drums	100%	100%	96%	91%	100%	100%
39	phone	95%	91%	100%	96%	100%	100%
40	pig alive	100%	100%	100%	100%	100%	100%
41	pig roasted	100%	96%	100%	100%	77%	100%
42	rat	100%	100%	100%	100%	100%	100%
43	rat swarm	95%	100%	100%	96%	100%	100%
44	rifle	95%	91%	96%	96%	100%	100%
45	road	95%	96%	83%	100%	100%	100%
46	road (jungle path)	81%	96%	100%	91%	100%	100%
47	spade	100%	100%	100%	100%	100%	100%
48	sugarcane cut	100%	100%	100%	83%	100%	100%
49	sugarcane growing	100%	96%	96%	91%	100%	100%
50	*sun*	*71%*	*83%*	*67%*	*61%*	*77%*	*94%*
51	*tea*	*5%*	*13%*	*25%*	*22%*	*77%*	*77%*
52	tomato	100%	100%	100%	96%	100%	100%
53	tomatoes	100%	100%	100%	100%	77%	100%
54	tomatoes on plant	100%	100%	96%	100%	100%	100%
55	truck	100%	100%	100%	100%	100%	100%
56	wound/sore	95%	100%	96%	96%	100%	100%
57	yam dug up	91%	100%	100%	96%	100%	100%
58	yam tubers	95%	100%	96%	96%	100%	94%
59	yam vine	100%	100%	96%	96%	100%	94%
60	yams cooked	81%	78%	96%	96%	77%	88%

Table 3. Percentage accuracies for all 60 images presented in the naming task within the card sort experiment, across all languages. Omitted items are italicised.

Free sort. Overall, descriptive statistics showed that the largest number of categories for the free sort were made by Merei participants and the fewest categories were made by Iaai participants (see Table 4). ANOVAs were calculated using the *rstatix* package (KASSAMBARA, 2020). There was a significant difference in the number of categories made for the free sort task between languages, $F_{(5, 113)} = 2.91$, $p = .017$, $\eta_p^2 = .11$. Pairwise comparisons comparing individual languages indicated that Iaai made significantly fewer categories compared to Merei ($p = .003$), Lewo ($p = .003$), and Vatlongos ($p = .004$), with all other comparisons being non-significant (smallest $p = .052$). This showed that fewer categories were produced by speakers of languages with larger classifier inventories (Iaai) compared to speakers with smaller classifier inventories (Merei and Lewo), contradicting Hypothesis 1.

	Free sort		Structured sort	
Language	*M*	*SD*	*M*	*SD*
Merei (2)	21.48	11.31	14.86	14.02
Lewo (3)	21.30	9.28	5.22	5.02
Vatlongos (4)	21.13	10.23	10.67	10.04
North Ambrym (5)	18.09	7.86	7.13	5.79
Nêlêmwa (17)	15.33	9.76	12.08	5.87
Iaai (19)	12.00	8.16	12.13	4.01

Table 4. Means and standard deviations for the numbers of categories made for the free sort and structured sort tasks across all languages. Number of classifiers included in analysis after exclusion of pictures with ≤75% accuracy in naming task shown in brackets.

Structured sort. Descriptive statistics for the structured sort task showed that Merei participants made the greatest number of categories, followed by Iaai and Nêlêmwa participants. Lewo participants made the fewest (see Figure 3). This partially supports Hypothesis 1. Further, through comparison of the common classifiers across Lewo, Vatlongos, North Ambrym and Iaai, it was found that more images were placed into the general classifier (M = 24.67, SD = 11.18) compared to the food (M = 17.67, SD = 9.45), and the drink classifiers (M = 4.66, SD = 3.03).
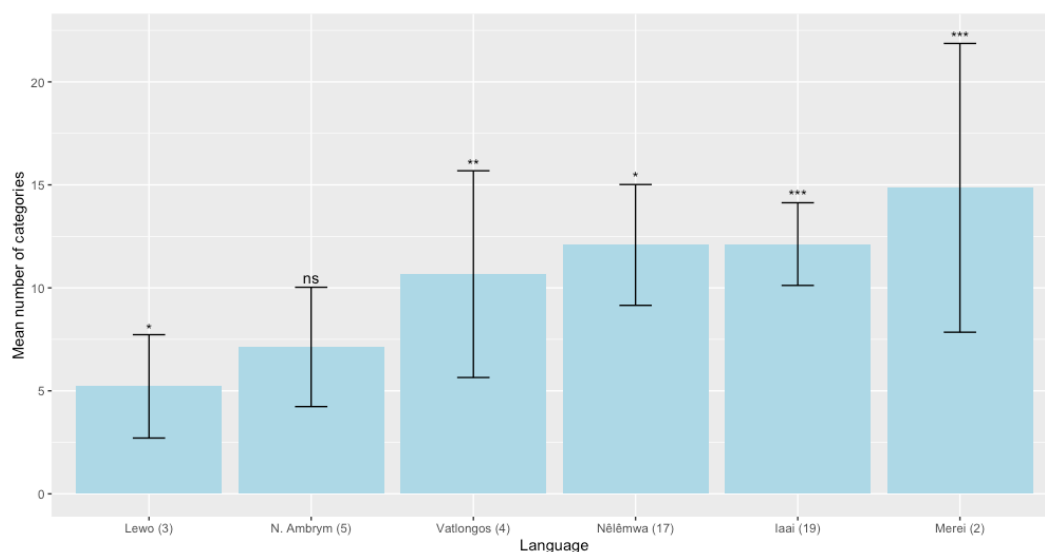

Figure 3. Bar chart showing the mean number of categories made by each language for the structured sort task, compared to the number of classifiers of the language. Error bars represent the standard deviation. Note - significant differences between number of categories and classifier inventories are indicated by *** < .001, ** < .01, * < .05, ns = not significant.

Overall, there was a significant difference in the number of categories made across languages in the structured sort task, $F_{(5, 113)}$ = 3.73, p = .004, η_p^2 = 14. Pairwise comparisons indicated that significantly fewer categories were made in Lewo compared to Merei (p < .001), Vatlongos (p = .03), Nêlêmwa (p = .025) and Iaai (p = .014). Furthermore, significantly fewer categories were made in North Ambrym compared to Merei (p = .003). Additional t-tests revealed significant differences between the number of categories made and the number of classifiers tested from each language for all languages except North Ambrym (see Table 5 for inferential statistics).

Language	Number of classifiers	t	p	d^4
Merei	2	4.20	**< .001**	0.92
Lewo	3	2.12	**.046**	0.44
Vatlongos	4	3.25	**.003**	0.66
North Ambrym	5	1.76	.092	0.37
Nêlêmwa	17	-2.90	**.014**	0.84
Iaai	19	-6.85	**< .001**	1.71

Table 5. Inferential statistics from the t-tests comparing the number of classifiers with the number of categories produced in the structured sort for each language.

Additionally, a 3 (classifier: drink, food, general) x 4 (language: Lewo, Vatlongos, North Ambrym, Iaai) mixed ANOVA investigated whether the number of images placed into the common classifier categories differed across the relevant languages (Table 1). A Greenhouse-Geisser sphericity correction was applied due to the assumption of sphericity not being met as Mauchly's test was significant, W = 0.373, p < .05.

There was a significant main effect of classifier, $F_{(1.23, 100.79)}$ = 146.10, p < .001, η_p^2 = .64. Pairwise comparisons found that significantly more images were placed into the general classifier compared to the food (p < .001) and drink classifiers (p < .001). Also, significantly more images were placed into the food classifier compared to the drink classifier (p < .001). There was a significant main effect of language on the number of items placed into the common classifiers, $F_{(3, 82)}$ = 22.52, p < .001, η_p^2 = .45. Pairwise comparisons showed that Iaai placed significantly fewer images into the common classifiers compared to Lewo (p = .01) and North Ambrym (p = .016). All other comparisons were non-significant (smallest p = .09).

There was a significant interaction between classifier and language, indicating that the relative numbers of items placed into the three common classifiers differed across languages: $F_{(3.69, 100.79)}$ = 20.12, p < .001, η_p^2 = .42. Figure 4 shows that North Ambrym participants placed the most images into the food classifier and all other languages placed the most images into the general classifier. All languages placed the fewest images into the drink classifier. Finally, when comparing the number of images placed into the general classifier across languages, pairwise comparisons showed that Lewo placed significantly more images into the general classifier compared to North Ambrym (p = .005). Furthermore, Vatlongos placed more images into the general classifier compared to North Ambrym (p < .001) and Iaai (p = .001). All other comparisons were non-significant (smallest p =.059).

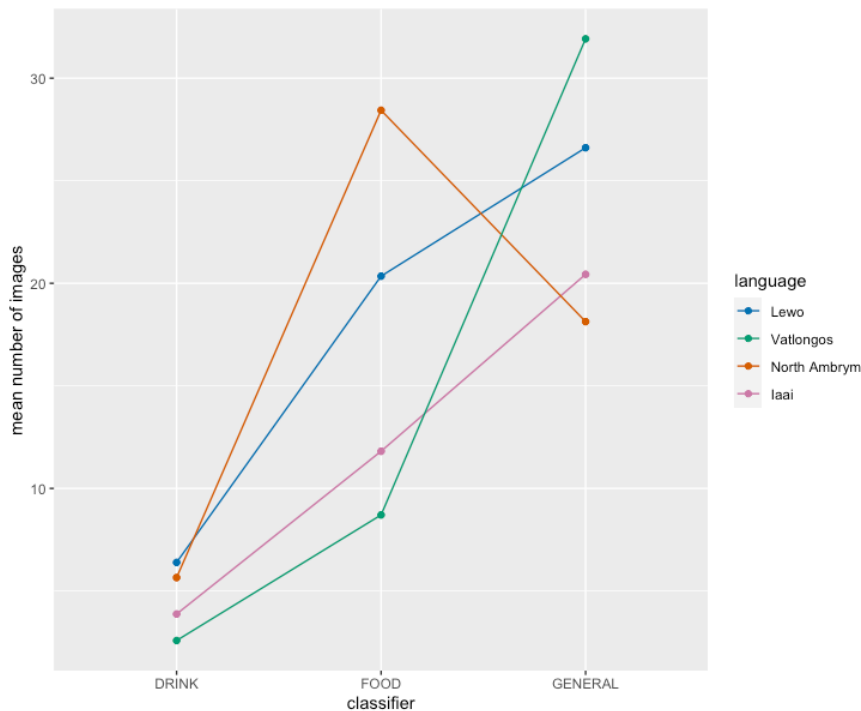4   Cohen's *d* calculated using the *lsr* package (NAVARRO, 2015).

Figure 4. The interaction effect for the mixed ANOVA conducted for the structured sort. Mean number of images placed into each classifier for the relevant languages.

Comparison of free sort and structured sort – Overall, a greater number of categories were made in the free sort (M = 18.82, SD = 9.91) compared to the structured sort (M = 10.01, SD = 8.98). Table 6 shows that all languages apart from Iaai made more categories in the free sort compared to the structured sort. This difference was significant for Lewo, Vatlongos, and North Ambrym but was non-significant for Merei, Nêlêmwa, and Iaai. This only partially supports Hypothesis 2, which states that similar numbers of categories will be made across both tasks.

Language	t-test		
	t	p	d
Merei	-1.68	.100	0.52
Lewo	-7.31	**< .001**	2.16
Vatlongos	-3.58	**< .001**	1.03
North Ambrym	-5.38	**< .001**	1.59
Nêlêmwa	-0.99	.336	0.40
Iaai	0.05	.957	0.02

Table 6. Inferential statistics from the t tests for the comparison of the number of categories made in the free sort task and the structured sort task. Note - significant results are shown in bold.

Furthermore, a 2 (task: free and structured sort) x 6 (language: Merei, Lewo, Vatlongos, North Ambrym, Nêlêmwa, Iaai) mixed ANOVA was conducted on the number of categories made. There was a significant main effect of task on the number of categories made, F(1,

113) = 60.97, $p < .001$, $\eta_p^2 = .35$, but there was no significant main effect of language on the number of categories made, $F(5, 113) = 2.01$, $p = .075$, $\eta_p^2 = .08$. A significant interaction between task and language was revealed, $F(5, 113) = 5.52$, $p < .001$, $\eta_p^2 = .20$. Figure 5 shows that all languages made fewer categories in the structured sort apart from Iaai, who made a similar number of categories in both tasks.
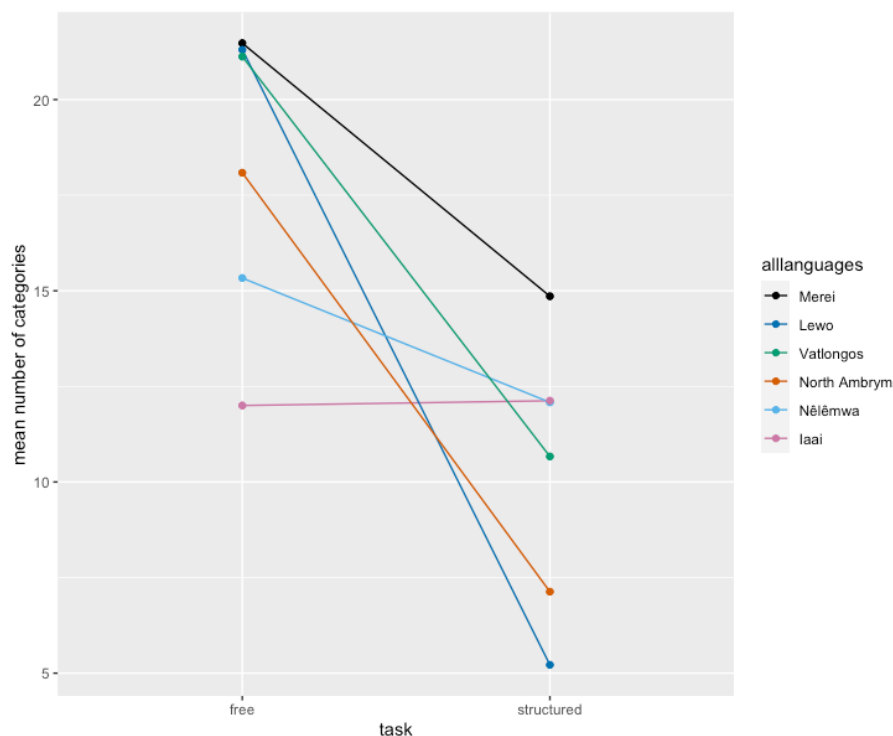

Figure 5. Graph showing the interaction effect for all languages for the mixed ANOVA, comparing the mean number of categories made across the free sort and structured sort.

Thematic analysis of category labels in the free sort. Five themes emerged from the qualitative exploration of the labels assigned to the piles produced by participants within the free sort. These five themes are presented below in order of their prevalence and significance.

The influence of classifiers. One facet of the responses that emerged is that participants did not solely create their piles explicitly in accordance with the classifiers of their language. The majority of participants did not associate their piles with a classifier and many simply labelled the piles literally, with the label stating the items contained within. For example, one speaker of Merei constructed a pile containing images 56 and 57 (each consisting of oil drums) and labelled it simply as 'drum'. This manner of labelling piles literally in referring to the items contained within was the most frequent method of labelling during the free sort. However, some participants did create groups based on classifiers, particularly speakers of Vatlongos. For example, one participant created four piles, each corresponding to a classifier: *nganak be sam xil* 'everything is yours (land

classifier); DEM COP CL.LAND.2SG', *nganak be nam* 'this is yours (general classifier); DEM COP CL.GENERAL.2SG', *am* 'yours (food classifier); CL.FOOD.2SG', and *mam* 'yours (drink classifier); CL.DRINK.2SG'. A different speaker of Vatlongos created six piles, each labelled with an associated classifier. Despite this, even the participants who exhibited more influence of classifiers on their groupings often included at least some labels that were literal descriptions of a pile's contents.

Item usage/utility. While the previous method of labelling was the most common, another frequently used method of classification described the piles in respect to the utility of the items contained within. For example, one speaker of Merei created several piles that follow the format of: 'things that x'. Their pile entitled 'things for work' included the images of the rifle, the machete, the spade, and truck, while their pile 'things for playing' included the football and chewing gum. Similarly, one speaker of North Ambrym created a pile entitled 'for playing', which included the football. However, the same participant also created a pile entitled 'for work', which was quite dissimilar to that of the Merei speaker outlined above. This pile contained items such as the road and the oil drums. Indeed, groupings based around work were common. One speaker of Vatlongos included the machete and the spade in a 'work' pile, while a Nêlêmwa speaker included many items within a similarly titled pile including the oil drums, machete, spade, and truck. Other manifestations of this 'things that x' label include 'for cooking', which was given by a speaker of North Ambrym, and 'for eating' given by a speaker of Lewo. All of these piles share the theme of labelling according to the usage or utility of the items contained within.

Sourcing of items. Another theme within the labels of piles was labelling according to where items may be sourced from. For instance, one speaker of Lewo included the labels 'something from the store', 'something from the garden', and 'something from the sea'. The first of these piles included items such as the truck and the chewing gum, items that would need to be purchased. Similarly, one speaker of Merei gave the label 'all things from the store', and also included the images of the oil drums and the plastic chair. In contrast, 'something from the garden' consisted primarily of food that would be grown, such as yams and tomatoes. 'Something from the sea' included the images of fish. Similarly, one Iaai speaker included the label 'the fruits from the earth', which contained many of the items outlined above in relation to the 'something from the garden' pile, but also several of the images relating to coconuts. This categorisation based on the manner in which something is acquired was frequent and consistent across languages.

Local or abroad. One particular distinction that several participants made was between things that could be found locally and things that needed to be sourced abroad. For example, one speaker of Nêlêmwa, labelled a pile containing the football, the phone, the chewing gum and the truck as being 'objects with no name in Nêlêmwa'. Another Nêlêmwa speaker included these items and more (the plastic chair, the rifle and the matches) as being

'things that have been introduced/imported'. This theme was also present in the other languages. One Iaai participant used this as their only basis of categorisation creating two piles that were labelled 'things that exist in Iaai' and 'foreign stuff', respectively. This distinction between 'indigenous' and 'foreign' items was often conceptualised by referring to foreign items with some variation of the label 'white people'. This was demonstrated by three participants from North Ambrym, two Lewo speakers, and a speaker of Merei, who all created labels relating to the possessions of 'white people' or objects that 'white people' make. These labels include 'things of white people' produced by one Lewo speaker (for the phone, the truck, the plastic chair, and the oil drums), and 'metal of the white man', a label given by a North Ambrym speaker to a pile including the phone, the rifle, the machete, and the truck. Other items that were labelled within this theme were matches, tea, and tomato.

Natural/unnatural and the manifestation of God. Two related concepts that arose within the labels were the distinction between the 'natural' and 'unnatural' things, and the manifestation of God. These concepts were often related and so combine to form one theme. As an example, one Iaai speaker uses the label: 'the natural resources' for a pile including all of the biological items such as the fish, coconuts, tomatoes, pigs, etc. This distinction between natural and unnatural is often made through the lens of religion, with one Merei speaker exemplifying this by including two piles entitled 'God makes life (things with life)' and 'without life'. The former grouping contained primarily biological objects, such as dog, rat, coconut, fish, and yams, while the latter group contained inanimate objects such as the mats and football. This tendency was also clear from a Iaai speaker who grouped many of the images depicting manmade objects under the labels 'things made by men' and 'the things that men built (artefacts)'. This speaker distinguishes these piles from 'the animals of God' containing all the animals, and 'the things of God', containing objects such as coconut palms and the sun and the moon. The sun and the moon, while often labelled literally, were also connected to the divine by participants. Speakers across Iaai, Merei and Lewo produced labels such as 'the creation of the sky', 'the things of God', 'power of God' and 'God creates' in relation to the sun and the moon.

Construction of a narrative. A frequent occurrence across all six languages was structuring piles around the telling of a story. In this instance, labels were used to tie the objects within the pile together through narrative. Sometimes this was done in a whimsical manner, as in the example 'the witch is going to fly' where one Iaai speaker linked the cards representing the dresses, the basket, the mats, and the moon. Another, simpler example of such a story was provided by a speaker of Merei, who placed the necklace and the chewing gum together under the label 'sister eats chewing gum'. Stories explaining practical, potentially routine actions were also found. For example, a Lewo participant produced the label 'go shoot pig' in relation to the rifle, the live pig, the cooked pig, the dog, the machete, and the jungle road. Another example of this more straightforward narrative was the label

'person plays with his toys (man puts his hat on, plays football, wears chain, plays on mobile phone)' given by a North Ambrym speaker in relation to a pile containing the football, telephone, basket, necklace and cap.

4. DISCUSSION

The current study aimed to investigate the semantic domains of classifier systems in six related languages, to reveal their impact on cognition and the relative cognitive efficiency of each system. The findings will be discussed in relation to our hypotheses and our overarching aim to investigate whether classifiers provide a structure for cognition within the languages of focus.

Hypothesis 1 predicted that speakers of languages with larger classifier inventories (Iaai and Nêlêmwa) would produce a larger number of categories in the free sort and structured sort tasks, compared to speakers of languages with smaller classifier inventories (Merei, Lewo, Vatlongos, and North Ambrym). This hypothesis was only partially supported. Indeed, speakers of Iaai and Nêlêmwa (languages with larger classifier inventories) did produce more categories than speakers of Lewo, North Ambrym and Vatlongos (languages with smaller classifier inventories), supporting Hypothesis 1. However, this was only the case for the structured sort task and not the free sort task (c.f. Table 4). Additionally, speakers of Merei with only two classifiers produced the largest number of categories in the structured sort, directly contradicting Hypothesis 1. Thus, for the structured sort, categorisation may be in part structured by classifiers. As the classifier inventory increased, the mean number of piles made increased (with the exception of Vatlongos and Merei). This is consistent with findings from other languages such as Mandarin that employ rather different classifier systems from the Oceanic type (JIANG, 2017). The pattern found in Merei may reflect the system's relative simplicity and lack of informativeness. As such a system is limited in the inventory size of classifiers, it may be less likely to influence speakers in a task of this nature. The more classifiers a system has, the more semantically transparent a system is. Thus systems with smaller classifier distinctions have potentially more opaque systems than a larger system and may be less likely to effect cognition.

Moreover, the prediction that similar numbers of categories would be made in the free sort and the structured sort tasks for all languages (Hypothesis 2) was only partially supported. Although more categories were made in the free sort compared to the structured sort for all languages, this difference was only significant for Lewo, Vatlongos and North Ambrym (c.f Table 6). This pattern suggests that classifiers are not the default system of categorisation in the context of the free sort at least. As the free sort task required

participants to sort images into groups with deliberately no guidance on how they might do this, participants may have categorised the items according to semantic information, perceptual properties, or taxonomic or thematic relations, in addition to classifier membership. Such methods of categorisation were evidenced in Mandarin speakers, who did not use their rather different classifier categories as the basis for categorisation (SAALBACH; IMAI, 2007). Additionally, the significant difference in the number of categories made across tasks for Lewo, Vatlongos, and North Ambrym may inform us about the nature of these types of systems. It is possible that these more moderate systems may be less optimal for categorisation, in comparison to Iaai, Nêlêmwa, and Merei at the extremes of classifier inventory size. In combination, these findings suggest that the linguistic relativity debate does not have a simple or concrete answer, and that the relative influence of language on cognition appears to vary across different classifier systems.

With regard to Hypothesis 3, it was found that Iaai with the largest classifier inventory displayed smaller variation on both the free sort and the structured sort, compared to Merei and Vatlongos with their smaller inventories, contradicting this hypothesis. This was shown through relatively smaller standard deviation scores for the mean number of categories produced by Iaai participants. This indicates that in Iaai categorisation was more consistent across participants, and there was lower inter-speaker variation. However, the lowest variation in the free sort was demonstrated by speakers of North Ambrym. Additionally, speakers of Lewo showed relatively low variation in the structured sort. This demonstrates that there are factors other than inventory size contributing to these patterns. Arguably, the relative cognitive salience of different classifiers may be influenced to some degree by classifier inventory, but also by informativeness and rigidity. Here we see that categorisation systems can affect conceptual structure, supporting previous evidence (BORODITSKY; SCHMIDT; PHILLIPS, 2003), but also that this may be task dependent. This offers a new perspective on the linguistic relativity debate, illustrating that influences of language on cognition may be stronger for certain linguistic systems and may vary depending on the way that language is being used.

In addition, the findings corroborate Hypothesis 4, indicating that North Ambrym, with a more rigid gender-like system of noun-classifier assignment, exhibited less variation across speakers in comparison to some languages with more typical classifier-like systems that allow variable noun assignment (Merei, Vatlongos and Nêlêmwa). Indeed, speakers of North Ambrym displayed relatively smaller standard deviation scores across both tasks, indicating smaller variation, in comparison to Merei, Lewo and Vatlongos. North Ambrym's standard deviation scores were comparable to Iaai, which has the largest classifier inventory. This suggests that more informative systems (Iaai) or more rigid gender-like systems (North Ambrym), provide more structure for categorisation. Thus, these more extreme systems may be more optimal for categorisation in comparison to more moderate

systems. This casts new light on the debate about the trade-off between simplicity and informativeness, in that cognition appears to be facilitated more optimally with either informative systems with variable assignment or more simple systems with fixed assignment. Nevertheless, evidence of less optimal systems is equally intriguing. This is arguably the case for Merei, and also for Vatlongos, which shows a more balanced trade-off between simplicity and informativeness. Although speakers of North Ambrym demonstrated relatively smaller inter-speaker variation across the free sort task and the structured sort task, there was also evidence of variation between the two tasks for this language. Indeed, speakers of North Ambrym produced significantly more categories in the free sort compared to the structured sort indicating cross-task variation. This could be attributed to the explicit instruction to use classifiers in the structured sort, compared to the more subjective judgement of the task demands in the free sort. These results again support the idea that there is variability in communicative efficiency in linguistic categories across languages (KEMP; REGIER, 2012).

Overall, there is mixed evidence for whether classifiers provide a structure for cognition and guide categorisation. Effects of classifier inventory were evident: Iaai, with the largest classifier inventory, placed fewer nouns into categories corresponding with the common classifiers (general, food, drink), compared to Lewo and North Ambrym. Further, there was a significant effect of language on the number of items placed into the general classifier (cf. Figure 4). Collectively, these findings make an important contribution to the debate that language influences cognition (BORODITSKY; SCHMIDT; PHILLIPS, 2003), providing more concrete directions into how this may occur. Rather than classifiers having a fixed and universal impact on cognition, their influence appears to be variable and task dependent.

The qualitative analysis of the piles constructed by participants during the free sort provides further insight into the extent to which classifiers influence cognition. One notable element of the free sort pile labels was the lack of explicit reference to classifiers. While there were incidents of participants directly labelling piles with a corresponding classifier, participants largely labelled their piles as simple, literal descriptions of the items within. This may suggest that when categorising objects, the default is not to use classifiers, but rather to refer to broader characteristics. This fits with the finding that even native speakers of languages that do not have nominal categorisation systems are privy to distinctions in language that classifiers reflect. This is indicated by the influence of object similarities underlying classifier systems, irrespective of the presence of classifiers within a language (SPEED et al., 2016). Indeed, it may be that classifier systems reflect conceptual structure rather than shaping it, and this seems to be reflected by many of the labels used by participants to describe their groupings within the free sort task.

In addition to literal descriptions of the items contained with piles, perhaps one of the most interesting tendencies among participants was their inclination towards constructing narratives with their piles. This sometimes enabled participants to provide a story linking seemingly disparate items within individual piles. This approach also enabled participants to explain how items were linked by practical usage and this was particularly interesting. When an individual lists dog, pig alive, pig roasted, and rifle all together, and explains that a man goes to hunt a pig with his dog, he is explaining something about his culture and his manner of thinking through the medium of storytelling. Within other themes, we observed speakers grouping items that 'the white man made' or that are 'fruits of the sea', again independent of explicit reference to classifiers, reflecting similarities in form, source, or function. Although these approaches did not make explicit reference to classifiers, they often referred to the structure of those items in the daily life of the participant. In doing that, participants frequently drew on their possessive relationships with objects. This highlights a fundamental feature of classifiers and signals the relevance of classifier systems even in the absence of explicit usage. It therefore appears that participants' categorisation within the card sort task was influenced by a range of different cultural and linguistic factors. Once again, this suggests that classifiers may reflect rather than shape cognition, in line with our quantitative analyses. This qualitative approach provides further context for our data, adding to the methodological toolkit established within this wider project.

5. CONCLUSION

We have explored further the findings of Franjieh (2016, 2018), who showed differences in categorisation between the systems of closely related languages. The current study found that classifiers may influence categorisation only when they are explicit and salient, as was the case for the structured sort task. It was found that the languages with larger classifier inventories listed the greatest number of nouns and produced the largest number of categories in the free sort task. The results suggest that classifiers do not always provide structure for categorisation, as significantly more categories were made in the free sort than in the structured sort task. Also, more extreme categorisation systems (whether more informative or more rigid) seem to be more optimal systems of categorisation, showing smaller variation across speakers in comparison to more moderate systems. Finally, our qualitative analyses reveal that participants categorise entities based on a range of properties, aligning with themes relating to the influence of classifiers in relation to items' usage, how items are sourced and where they originate from, whether they are natural or unnatural and how they can be related in terms of narrative. The study illustrates that

language's impact on cognition varies across different linguistic systems, and sheds light on the intricacies of the linguistic relativity debate.

6. ACKNOWLEDGEMENTS

REFERENCES

BELLER, SIEGHARD; BRATTEBØ, KAREN FADNES; LAVIK, KRISTINA OSLAND; REIGSTAD, RAKEL DRØNEN; BENDER, ANDREA. Culture or language: what drives effects of grammatical gender? *Cognitive Linguistics*, v. 26, n. 2, p. 331–359, 2015.

BENDER, ANDREA; BELLER, SIEGHARD; KLAUER, KARL CHRISTOPH. Grammatical Gender in German: A Case for Linguistic Relativity? *Quarterly Journal of Experimental Psychology*, v. 64, n. 9, p. 1821–1835, 2011.

BENDER, ANDREA; BELLER, SIEGHARD; KLAUER, KARL CHRISTOPH. Crossing grammar and biology for gender categorisations: Investigating the gender congruency effect in generic nouns for animates. *Journal of Cognitive Psychology*, v. 28, n. 5, p. 530–558, 2016.

BERLIN, BRENT; KAY, PAUL. *Basic Color Terms: Their Universality and Evolution*. Berkeley & Los Angeles: University of California Press, 1969.

BI, YANCHAO. Nominal classification is not positive evidence for language relativity: a commentary on Kemmerer (2016). *Language, Cognition and Neuroscience*, v. 32, n. 4, p. 428–432, 2017.

BORODITSKY, LERA; SCHMIDT, LAUREN A.; PHILLIPS, WEBB. Sex, syntax and semantics. In: *Language in mind: Advances in the study of language and thought*. Cambridge, MA, US: MIT Press, p. 61–79, 2003.

BRAUN, VIRGINIA; CLARKE, VICTORIA. Using thematic analysis in psychology. *Qualitative Research in Psychology*, v. 3, n. 2, p. 77–101, 2006.

BRAUN, VIRGINIA; CLARKE, VICTORIA. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, v. 21, n. 1, p. 37–47, 2020.

BRIL, ISABELLE. *Le nêlêmwa (Nouvelle-Calédonie): analyse syntaxique et sémantique*. Paris: Peeters, 2002.

BROWN, ROGER W.; LENNEBERG, ERIC H. A study in language and cognition. *The Journal of Abnormal and Social Psychology*, v. 49, n. 3, p. 454–462, 1954.

BRUGMANN, KARL. Das Nominalgeschlecht in den indogermanischen Sprachen. *Internationale Zeitschrift fur allgemeine Sprachwissenschaft*, v. 4, p. 100–109, 1889.

CARROLL, JOHN B. Introduction. In: CARROLL, JOHN B. (Ed.). *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. New York; London: The Technology Press of Massachusetts Institute of Technology and John Wiley & Sons, Inc., p. 1–34, 1956.

CLIFFORD, ALEXANDRA; FRANKLIN, ANNA; DAVIES, IAN R. L.; HOLMES, AMANDA. Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain and Cognition*, v. 71, n. 2, p. 165–172, 2009.

CORBETT, GREVILLE G. Number of Genders. In: DRYER, MATTHEW S.; HASPELMATH, MARTIN; GIL, DAVID; COMRIE, BERNARD (Eds.). *The World Atlas of Language Structures*. Oxford: Oxford University Press, p. 126–129, 2005.

CORBETT, GREVILLE G.; FEDDEN, SEBASTIAN. New approaches to the typology of gender. In: FEDDEN, SEBASTIAN; AUDRING, JENNY; CORBETT, GREVILLE G. (Eds.). *Non-canonical gender systems*. Oxford: Oxford University Press, p. 9–35, 2018.

CORBETT, GREVILLE G.; FEDDEN, SEBASTIAN; FINKEL, RAPHAEL. Single versus concurrent feature systems: nominal classification in Mian. *Linguistic Typology*, v. 21, n. 2, p. 209–260, 2017.

CUBELLI, ROBERTO, PAOLIERI, DANIELA, LOTTO, LORELLA, & JOB, REMO. The effect of grammatical gender on object categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, V. 37, P. 449–460, 2011.

DELGADO, ANA R. Order in Spanish colour words: Evidence against linguistic relativity. *British Journal of Psychology*, v. 95, n. 1, p. 81–90, 2004.

DOTTE, ANNE-LAURE. Dynamism and change in the possessive classifier system of Iaai. *Oceanic Linguistics*, v. 56, n. 2, p. 339–363, 2017.

EARLY, ROBERT. *A Grammar of Lewo, Vanuatu*. PhD thesis—Canberra: Australian National University, 1994.

EBERHARD, DAVID M.; SIMONS, GARY F.; FENNIG, CHARLES D. *Ethnologue: Languages of the World*. Twenty-fourth ed. Dallas, Texas: SIL International, 2021.

FAUL, FRANZ; ERDFELDER, EDGAR; BUCHNER, AXEL; LANG, ALBERT-GEORG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, v. 41, n. 4, p. 1149–1160, 2009.

FEDDEN, SEBASTIAN; CORBETT, GREVILLE G. Gender and classifiers in concurrent systems Refining the typology of nominal classification. *Glossa: a journal of general linguistics*, v. 2, n. 1: 34, p. 1–47, 2017.

FEDDEN, SEBASTIAN; CORBETT, GREVILLE G. Extreme classification. *Cognitive Linguistics*, n. 29, p. 633–675, 2018.

FRANJIEH, MICHAEL. Indirect Possessive Hosts in North Ambrym: Evidence for Gender. *Oceanic Linguistics*, v. 55, n. 1, p. 87–115, 2016.

FRANJIEH, MICHAEL. North Ambrym possessive classifiers from the perspective of canonical gender. In: FEDDEN, SEBASTIAN; AUDRING, JENNY; CORBETT, GREVILLE G (Eds.). *Non-Canonical Gender Systems*. Oxford, New York: Oxford University Press, p. 36-67, 2018.

FRANKLIN, ANNA; DRIVONIKOU, GILDA V.; BEVIS, LAURA; DAVIES, IAN R. L.; KAY, PAUL; REGIER, TERRY. Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences*, v. 105, n. 9, p. 3221–3225, 2008.

GIBSON, EDWARD; FUTRELL, RICHARD; JARA-ETTINGER, JULIAN; MAHOWALD, KYLE; BERGEN, LEON; RATNASINGAM, SIVALOGESWARAN; GIBSON, MITCHELL; PIANTADOSI, STEVEN T.; CONWAY, BEVIL R. Color

naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, v. 114, n. 40, p. 10785–10790, 2017.

GILBERT, AUBREY L.; REGIER, TERRY; KAY, PAUL; IVRY, RICHARD B. Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, v. 103, n. 2, p. 489–494, 2006.

GRANDISON, ALEXANDRA; DAVIES, IAN R. L.; SOWDEN, PAUL T. The Evolution of GRUE: Evidence for a new colour term in the language of the Himba. In: ANDERSON, W.; BIGGAM, C. P.; HOUGH, C.; KAY, C. (Eds.). *Colour Studies: A broad spectrum*. NL: Benjamins, p. 53–66, 2014.

GRIMM, JACOB. *Deutsche Grammatik [Vol. III]*. Gottingen: Dieterich, 1831.

GRINEVALD, COLETTE. A morphosyntactic typology of classifiers. In: SENFT, GUNTER (Ed.). *Systems of Nominal Classification*. Cambridge: Cambridge University Press, p. 50–92, 2000.

GUMPERZ, JOHN J.; LEVINSON, STEPHEN. *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press, 1996.

HAWKINS, JOHN A. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press, 2004.

JAMESON, KIMBERLY; D'ANDRADE, ROY G. It's not really red, green, yellow, blue: an inquiry into perceptual color space. In: HARDIN, C. L.; MAFFI, LUISA (Eds.). *Color Categories in Thought and Language*. Cambridge: Cambridge University Press, p. 295–319, 1997.

JIANG, SONG. *The semantics of Chinese classifiers and linguistic relativity*. London ; Routledge/Taylor & Francis Group, 2017.

KASSAMBARA, ALBOUKADEL. *Rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.6.0, 2020.

KAY, PAUL; REGIER, TERRY. Language, thought and color: recent developments. *Trends in Cognitive Sciences*, v. 10, n. 2, p. 51–54, 2006.

KEMP, CHARLES; REGIER, TERRY. Kinship Categories Across Languages Reflect General Communicative Principles. *Science*, v. 336, n. 6084, p. 1049–1054, 2012.

KEMP, CHARLES; XU, YANG; REGIER, TERRY. Semantic Typology and Efficient Communication. *Annual Review of Linguistics*, v. 4, n. 1, p. 109–128, 2018.

LANDOR, ROLAND VIKTOR. *Grammatical Categories and Cognition across Five Languages: The Case of Grammatical Gender and its Potential Effects on the Conceptualisation of Objects*. PhD thesis—Brisbane: Griffith University, 2014.

LICHTENBERK, FRANTISEK. Relational Classifiers. *Lingua*, v. 60, n. 2–3, p. 147–176, 1983.

LUPYAN, GARY; ABDEL RAHMAN, RASHA; BORODITSKY, LERA; CLARK, ANDY. Effects of Language on Visual Perception. *Trends in Cognitive Sciences*, v. 24, n. 11, p. 930–944, 2020.

MALT, BARBARA; WOLFF, PHILIP. *Words and the Mind: How words capture human experience*. Oxford: Oxford University Press, 2010.

MAULE, JOHN; FRANKLIN, ANNA. Color categorization in infants. *Current Opinion in Behavioral Sciences*, Visual perception. v. 30, p. 163–168, 2019.

MCGRATH, CORMAC; PALMGREN, PER J.; LILJEDAHL, MATILDA. Twelve tips for conducting qualitative research interviews. *Medical Teacher*, v. 41, n. 9, p. 1002–1006, 2019.

NAVARRO, DANIEL. *Learning statistics with R: A tutorial for psychology students and other beginners*. Version 0.5, 2015.

NOWELL, LORELLI S.; NORRIS, JILL M.; WHITE, DEBORAH E.; MOULES, NANCY J. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, v. 16, n. 1, p. 1–13, 2017.

PILLING, MICHAEL; DAVIES, IAN R. L. Linguistic relativism and colour cognition. *British Journal of Psychology (London, England: 1953)*, v. 95, n. Pt 4, p. 429–455, 2004.

PINKER, STEVEN. *How the Mind Works*. New York: W. W. Norton & Company, 1997.

R CORE TEAM. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020.

REGIER, TERRY; KAY, PAUL. Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, v. 13, n. 10, p. 439–446, 2009.

ROBERSON, DEBI; DAVIDOFF, JULES B.; DAVIES, IAN; SHAPIRO, L. Colour categories and category acquisition in Himba and English. In: PITCHFORD, NICOLA; BIGGAM, CAROLE P. (Eds.). *Progress in Colour Studies*. Amsterdam: John Benjamins Publishing Company, p. 159–172, 2006.

ROBERSON, DEBI; DAVIES, IAN; DAVIDOFF, JULES. Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, v. 129, n. 3, p. 369–398, 2000.

SAALBACH, HENRIK; IMAI, MUTSUMI. Scope of linguistic influence: does a classifier system alter object concepts? *Journal of Experimental Psychology. General*, v. 136, n. 3, p. 485–501, 2007.

SAMUEL, STEVEN; COLE, GEOFF; EACOTT, MADELINE J. Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin & Review*, v. 26, n. 6, p. 1767–1786, 2019.

SCOVEL, THOMAS. Why Languages Do not Shape Cognition: Psycho- and Neurolinguistic Evidence. *JALT*, v. 13, n. 1, p. 43–56, 1991.

SENFT, GUNTER. *Systems of Nominal Classification*. Cambridge: Cambridge University Press, 2000.

SERA, MARIA D.; ELIEFF, CHRYLE; FORBES, JAMES; BURCH, MELISSA CLARK; RODRÍGUEZ, WANDA; DUBOIS, DIANE POULIN. When language affects cognition and when it does not: an analysis of grammatical gender and classification. *Journal of Experimental Psychology. General*, v. 131, n. 3, p. 377–397, 2002.

SLOBIN, DAN I. From "Thought and Language" to "Thinking for Speaking." In: GUMPERZ, JOHN J.; LEVINSON, STEPHEN C. (Eds.). *Rethinking Linguistic Relativity*. [s.l.] Cambridge University Press, 1996. p. 70–96.

SPEED, L.; CHEN, J.; HUETTIG, F.; MAJID, A. *Do classifier categories affect or reflect object concepts?* In: A. PAPAFRAGOU, D. GRODNER, D. MIRMAN, J. TRUESWELL (Eds.). *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 2016

SPEED, LAURA J.; CHEN, JIDONG; HUETTIG, FALK; MAJID, ASIFA. Classifier categories reflect but do not affect conceptual organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2020.

TERRY, GARETH; HAYFIELD, NIKKI; CLARKE, VICTORIA; BRAUN, VIRGINIA. Thematic analysis. In: WILLIG, CARLA; STAINTON ROGERS, WENDY (Eds.). *The SAGE Handbook of Qualitative Research in Psychology*. London: SAGE Publications, p. 17–37, 2017.

THAM, DIANA SU YUN; SOWDEN, PAUL T.; GRANDISON, ALEXANDRA; FRANKLIN, ANNA; LEE, ANNA KAI WIN; NG, MICHELLE; PARK, JUHYUN; PANG, WEIGUO; ZHAO, JINGWEN. A systematic investigation of conceptual color associations. *Journal of Experimental Psychology. General*, v. 149, n. 7, p. 1311–1332, 2020.

VIGLIOCCO, GABRIELLA; VINSON, DAVID P.; PAGANELLI, FEDERICA; DWORZYNSKI, KATHARINA. Grammatical Gender Effects on cognition: Implications for Language Learning and Language Use. *Journal of Experimental Psychology: General*, v. 134, n. 4, p. 501–520, 2005.

WINAWER, JONATHAN; WITTHOFT, NATHAN; FRANK, MICHAEL C.; WU, LISA; WADE, ALEX R.; BORODITSKY, LERA. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, v. 104, n. 19, p. 7780–7785, 2007.