



**Cognitive pressures in the emergence of
agreement systems?**

B171213

Word count: 7956

MSc in Linguistics

The University of Edinburgh

August 2021

Abstract

Despite being more complex compared to other types of languages (Lupyan & Dale, 2010, Trudgill, 2011), languages with noun classification and agreement systems are widespread across the world and present in languages genetically and geographically non-related. Such systems are interesting because they seem to provide an advantage in terms of language processing, facilitating word recognition (Tsang, Chambers & Mozuraitis, 2017; Dye et al., 2016). On the other hand, noun classification and agreement systems appear to be more complex in terms of language acquisition: in fact, they require learning several morphological encodings. However, recent research has shown that this morphological redundancy might facilitate noun class acquisition. For example, Demuth and Weschler (2012) investigate language acquisition in Sesotho, finding that children remember noun classes more accurately when the agreement is provided.

Here we set out to investigate how agreement affects noun class acquisition through an artificial language learning paradigm. We ran a behavioural experiment with three experimental conditions: reduplicative, arbitrary and no agreement, which differ for type and number of suffixes. We predicted that the agreement conditions would yield better learning results of the noun class system compared to no agreement. However, we found that only reduplicative agreement provides a clear advantage in language acquisition.

Acknowledgements

I would like to thank my supervisors Professor Kenny Smith and Dr Jennifer Culbertson for the support, the insightful advice and for providing the stimuli used in this experiment. Coding of the experiment, recruitment of participants, running of the experiment and analysis was done by the author. Examples from various languages, unless otherwise stated, are from the author's own knowledge of those languages.

Table of Content

1. Introduction	1
2. Method	10
2.1 Participants	10
2.2 Materials	10
2.3 Procedure	11
2.4 Session 1	12
2.5 Session 2	12
2.6 Coding	15
2.7 Results	16
2.8 Noun Class Learning	16
2.9 Noun Class Learning in Agreement Test	18
2.10 Agreement Learning	21
3. Discussion	23
4. Conclusion	25
5. References	26

1. Introduction

There are over 7000 languages spoken across the world. These languages vary greatly in terms of their structural (morphology and syntax) and non-structural (vocabulary) properties (Reali et al., 2018). For example, some languages have a rich inflectional morphological system: in Italian, verbs take a different inflection for each of the eight tenses and the six persons. On the contrary, other languages have very few inflections: for instance, English verbs are only inflected for the past tense, the gerund and the third person singular of the present tense. It has long been debated whether some of these features are more complex than others. For most of the 20th century, there was a general consensus that all languages were of equal complexity. However, more recent works on language diversity and creoles have led to a revaluation of this position (Joseph & Newmeyer, 2012; Spike, 2017). Indeed, these works have shown that some languages have a degree and specificity of morphological encoding which is astounding. For example, Karok, a language spoken in California, has morphological suffixes for forms of containment, such as pa:h-kirih “throw into fire”, pa:h-kuriah “throw into water”, pa:h-rupriah “throw in through a solid” - the affixes are unrelated to the lexemes for water, fire, and solid (Lupyan & Dale, 2012). The debate on linguistic complexity is still open and there is some disagreement on what constitutes complexity but, generally, scholars tend to agree that the following features can be regarded as complex: opacity, irregularity, multiple morphological categories and syntagmatic redundancy. On the other hand, transparency, regularity, few morphological categories and little or no syntagmatic redundancy are regarded as simple features. (Lupyan & Dale, 2010; Trudgill, 2011). Syntagmatic redundancy refers to the existence of two or more forms to express the same meaning, as in Italian “la ragazza alta” (the.F girl tall.F “the tall girl”) in which the gender of the noun “ragazza” is expressed three times by the suffix -a: on the article, on the noun and on the adjective.

The existence of redundant morphological forms seems to be in contrast with a cultural evolutionary account of the origins of language. According to this view linguistic structure results from a trade-off between two pressures: simplicity and expressivity (Culbertson & Kirby, 2016; Kirby et al., 2015). The first one arises during language acquisition, while the second one arises during communication. The pressure for simplicity is imposed on language by learners’ prior cognitive biases: language learners have a domain general cognitive bias for simplicity, which causes languages to become easier to learn and, therefore, more compressible. A compressible language is a language which allows for grammatical

representations which are more concise than a listing of all the possible utterances of a language. Kirby et al. (2015) show that if language is subjected to a pressure for simplicity only, the end point of a transmission chain is a type of language which they call degenerate. Degenerate languages are maximally compressive, meaning that they use one form (signal) to express all the meanings. Obviously, such languages do not exist, as they would be extremely ambiguous and inefficient. As illustrated by Kirby et al. (2015), if the pressure for simplicity is counterbalanced by a pressure for expressivity, the resulting language is a language which is both compressible and apt to solve the communicative task. A key characteristic of such languages is compositionality, in which the meaning of a complex utterance is derivable in a predictable way from the meanings of its subparts.

If we have a cognitive bias for morphologically simpler structures, why are structurally redundant structures, so widespread? To answer this question, we look at a particular example of morphological redundancy, that is grammatical agreement. Grammatical agreement seems an area of linguistic complexity particularly interesting to investigate: in spite of its structural complexity, it is widespread and present in languages geographically and genetically non-related such as Cantonese, Italian and Sesotho (a Bantu language). An example of the type of redundancy expressed by agreement is the Italian ‘il gatto’ (the M. cat M, ‘the cat), in which the article ‘il’ specifies gender and number (masculine singular which is already expressed in the noun). Nominal agreement in gender is a type of noun classification device, a broader category found in almost all languages. We will now look at three of the most common types of noun classification systems and their corresponding agreement (or lack of agreement): grammatical gender, noun class and noun classifier systems (Aikhenvald, 2021; Seifart, 2010; Senft, 2007).

We will first look at grammatical gender and noun class systems, highlighting their differences and similarities. Grammatical gender systems usually have two (like Italian) or three (like German) classes. The classes are identified based on a simple semantic distinction, usually masculine vs feminine (examples of this are the Romance languages), animate vs non-animate (as in Siouan, a language spoken in North America) or human vs nonhuman (as in some Dravidian languages) (Aikhenvald, 2021). According to Aikhenvald (2021), the label noun class is typically used for languages which have several classes, grouped according to more complex semantic criteria. For example, some Bantu languages have up to 20 noun classes. Dyirbal, an Australian language with four noun classes, exemplifies the semantic complexity behind noun groupings: class I contains male humans and non-animate entities,

class II is for female humans, drinkable liquids, fire and fighting, class III contains non-flesh food and class IV is a residue class which contains all remaining nouns. In spite of this difference, noun class and grammatical gender behave in a very similar way: they are both characterised by an agreement system. Formally, this is realized through morphological encodings (usually affixes, less often clitics) on one or more elements which refer to the noun, such as adjectives, demonstratives, articles, verbs, and sometimes prepositions. It can also be marked on the noun itself, which is called overt gender. Examples of a language with grammatical gender and a language with noun classes are given in (1) and (2) respectively:

(1) Italian

Masculine	Feminine
Lo scoglio piccolo	La tazza piccola
the m rock small m	the f mug small f
‘the small rock’	‘the small mug’

Italian has a two-class gender system: nouns are divided between masculine and feminine based on semantic as well as phonological criteria. Agreement is expressed through a suffix on the article, the noun and the adjective.

(2) Swahili

M-toto m-dogo a-lianguka CLASS1-child CLASS1-little CLASS1-fell_down ‘The little child fell down’
Ki-kapu ki-dogo ki-lianguka CLASS7-basket CLASS7-little CLASS7-fell_down ‘The little basket fell down’
Vi-kapu vi-dogo vi-lianguka CLASS8-basket CLASS8-little CLASS8-fell_down ‘The little baskets fell down’

(Senft, 2007)

Swahili has 11 noun classes. Nouns are grouped based on more complex semantic criteria compared to Italian: for example, class 1 is for human nouns, class 7 is chiefly for artifacts and tools in the singular, class 8 is for artifacts and tools in the plural. Agreement is

expressed through a prefix on the noun, the adjective and the verb. Beyond agreement, noun class and grammatical gender systems share another similarity: they are only loosely based on a core semantic feature (as mentioned above masculine vs feminine, animate vs non-animate, or human vs nonhuman) and often include exceptions in their classes. For example, in German the man ‘der Mann’ is masculine, the woman ‘die Frau’ is feminine, the boy ‘der Jungen’ is masculine but the girl ‘das Maedchen’ is neuter. Italian behaves similarly: neuter objects like the table ‘il tavolo’ and the knife ‘il coltello’ are masculine, while others such as the chair ‘la sedia’ and the mug ‘la tazza’ are feminine. Finally, Bantu languages are also semantically unreliable: class 1, containing human nouns, is the most reliable, the other classes include a mix of semantic categories such as animals, plants, inanimate entities, and abstract nouns. From these examples, it is clear that noun classes in these language types are not based on semantics alone, but also on morpho-phonological features: for instance, the suffixes -o and -a in Italian are very reliable indicators of masculine and feminine class respectively.

The third type of nominal classification device are noun classifier systems. These systems are different compared to the other two classification devices outlined above (grammatical gender and noun classes). To classify their nouns, such systems use specific classifying morphemes, so-called classifiers. The classifier is usually placed in proximity of the noun (often just before the noun). Unlike noun classes and grammatical gender, noun classifiers do not trigger agreement, i.e., a particular classifier does not co-occur with morphological markings on other elements, such as adjectives and articles. An example of a noun classifier language is given in (3):

(3) Cantonese

sam[1] go[3] jan[4] Three cl people ‘Three people’	sam[1] zek[3] gau[2] three cl dogs ‘Three dogs’	sam[1] zi[1] bat[1] three cl pens ‘Three pens’
---	---	---

(Culbertson, Jarvinen, Haggarty & Smith, 2019)

Cantonese has a numeral classifier system consisting of many distinct forms, used with nouns depending on semantic features. For example, the classifier go[3] is used with nouns referring to humans, zek[3] is used for animals and zi[1] is used to refer to nouns denoting long, narrow objects (Tsang & Chambers, 2011). Finally, unlike the first two classification devices,

noun classifiers are not semantically arbitrary: their classification is based on the meaning of the noun (Senft, 2007). To sum up, the three types of classification devices outlined above differ along two dimensions: their semantic reliability and their use of grammatical agreement.

We will now look in more detail at grammatical agreement which is our primary research interest. Noun classifier systems lack agreement, thus, they have only one formal cue to noun class: the classifier which accompanies the noun. On the contrary, grammatical gender and noun class systems, which have agreement, show multiple cues to noun class: the morphological encodings which mark agreement on the elements of the noun phrase (e.g. the article, the adjective, etc...) or beyond the noun phrase (e.g. on the verb). The form different agreement morphemes in a noun class system take can be reduplicative, i.e., expressed with the same phoneme or morpheme across nouns and the elements which refer to the noun. An example of this can be found in (1): Italian marks masculine with the morpheme -o on all elements of the noun phrase, and feminine with -a. Intuitively, this is an efficient manner of agreement, as the same meaning corresponds to the same form, i.e., the class or gender. Alternatively, agreement can also be arbitrary, i.e., each different agreement target has a different morpheme or marker. An example of this is German ‘der grosse Mann’ (the.M big.M man, ‘the big man’) in which the indicating of masculine agreement is completely different in the article and the adjective. Intuitively, this is a less efficient form of agreement because there two different forms to express the same meaning. To sum up, noun class and grammatical gender languages differ from classifier languages in their semantic reliability and their use of grammatical agreement. Further, the patterns of agreement amongst noun class and grammatical gender languages may differ between being reduplicative or arbitrary.

Despite these semantic and formal differences, all noun classification systems show morphological redundancy by expressing the same meaning several times and often with several forms. For example, Italian repeats the grammatical gender of the noun, which, even if arbitrary (i.e. not based on biological gender or any other reliable semantic feature) is already expressed on the noun itself. Similarly, in the example (2) a. from Swahili, there are three morphological markers (m-, m- and a-) to express the same meaning, that is the noun class of the noun, ‘child’. Even though this morphological redundancy might appear arbitrary, the cultural evolutionary account of language evolution tells us that language is a system which is shaped by selectional pressures from human learning and processing mechanisms (Christiansen & Chater, 2008). Therefore, within this framework, grammatical

agreement should result from general learning and processing biases from each generation of language users and learners. Studies on language processing do reveal that agreement serves several beneficial purposes during communication (Dye et al., 2016; Tsang, Chambers & Mozuraitis, 2017). There are fewer studies which looked directly at the role of agreement in language acquisition (Demuth & Weschler, 2012), but if we combine their results with those from studies on noun class acquisition, we find some evidence that formal features might play a facilitative role in noun class acquisition (Culbertson et al., 2017; Culbertson et al., 2019). In the following paragraphs, we will review some studies on the effects of agreement during communication and in language acquisition.

There are several studies which have investigated the effects of agreement in language processing. The general hypothesis here is that agreement ‘assists comprehension processes by linking temporally separated elements in discourse, establishing local and global coherence’ (Dye et al., 2016). This is evident in a language such as Latin in which attributive adjectives need not appear in a fixed position relative to their noun. This is because in Latin nouns and adjectives are declined for gender, number and case in a way which makes clear which adjectives and nouns belong together. This is an example from Dye et al. (2016):

ultim-a Cumae-i ven-it iam carmin-is aet-as

last-NOM-FEM Cumai-GEN-NEU came now song-GEN-NEU last-NOM-FEM

“The last age of the Cumaean song has now arrived.”

As Dye et al. (2016) note, the function of agreement seems to vary from one language to the other. For example, in German agreement seems to reduce entropy, that is the uncertainty over which noun will appear in a given context. According to Dye et al. (2016), a language like English which has no gender has a higher entropy compared to German, as illustrated by the following examples:

(4a) Yesterday I visited the doctor.

(4b) Gestern besuchte ich den Arzt.

yesterday visited I the.M doctor

“Yesterday I visited the doctor”.

In German, the set of possible nouns after the article is greatly reduced compared to English: in fact, only nouns belonging to the masculine class are allowed in the given context. Another

example comes from Tsang, Chambers & Mozuraitis (2017), who assess the role of Cantonese noun classifiers. One of the difficulties of language processing in Cantonese is differentiating compounds from isolated words. This task is made more complicated by the stress-timed nature of Cantonese. However, this task is facilitated by classifiers which help to identify nouns and which noun is the head in a compound. It therefore seems that, during communication, grammatical agreement serves various functions: tracking the dependency of non-adjacent elements (Latin), reducing entropy (German) and signalling the presence of a noun or a noun compound (Cantonese).

In terms of language acquisition, agreement and noun class systems present various challenges to learners, as they exploit various cognitive skills. Firstly, they require the tracking of so called non-adjacent dependencies, which is a skill that develops only after around 14-18 months (Culbertson et al., 2016; Gomez, 2002; Lany & Gomez, 2008). Secondly, noun classes are based on a mixture of semantic and morphophonological cues (Aikhenvald, 2000). For example, French has a prototypical mixed gender system, in which natural gender is a highly reliable cue and there are 34 distinct suffixes which cue one class or the other one (Culbertson et al., 2017; Surridge, 1986). This means that to acquire noun classes and agreement, learners must track both deterministic and probabilistic cues across domains, including exceptions that need to be memorized (Culbertson et al. 2019; Arnon & Ramscar, 2012; Siegelman & Arnon, 2015; Frigo & McDonald, 1998).

In spite of these difficulties, studies show that children learn noun classes and their corresponding agreement patterns by the age of 3 or earlier (Mills 1985; Levy 1983; Demuth, 2000; Demuth & Ellis 2008; Gagliardi & Lidz 2014). Interestingly, in a study on the acquisition of Sesotho noun classes, Demuth & Weschler (2012) shows that children by the age of 2;1 performed near or above 90% accuracy in their use of noun class agreement. Interestingly, the study also points out that the vast majority of mistakes were made for nouns where null prefixes are permitted in the language. This suggests that overt markers of agreement—reduplicative agreement in this case—may facilitate noun class acquisition. Further, some recent studies suggest that multiword units facilitate the acquisition of certain grammatical relations such as agreement. Particularly, Siegelman and Arnon (2015) showed that, when adult learners were asked to learn an artificial language with an agreement system, those who were exposed to unsegmented speech – and, thus, encouraged to treat article and noun as a multiword unit – acquired the agreement relations better. This also indicates a better learning of noun class, given that agreement is used as a marker for noun class in the

artificial language used in these experiments. The same result has also been found with children (Havron & Arnon, *in press*).

The studies presented so far on language acquisition suggest that the presence of grammatical agreement might strengthen the correct acquisition of noun classes both in adults and children. Finally, some studies show that there is a preference for formal cues over semantic cues in noun class acquisition (Culbertson et al., 2017; Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981). Karmiloff-Smith (1981) found that French-speaking children as old as 9, privilege phonological cues over semantic cues in determining the noun class of novel nouns. A similar result is presented by Frigo and McDonald (1998). In an artificial language learning experiment, participants were taught pairs of greetings and names. Different greetings were used depending on the person's social status. Additionally, 60 % of the names in each class were phonologically marked, while 40 % of the names were unmarked. The result is that phonologically marked names were learnt far better than phonologically unmarked ones. Gagliardi and Lidz (2014) report similar findings for Tsez, a Northeast Caucasian language. Importantly, they found that even when phonological cues were statistically less reliable than competing semantic cues, they were still preferred to determine class membership. Culbertson et al. (2017) use an artificial language learning experiment to investigate what might drive this preference. They find that there is an interaction of saliency and early-availability of cues: particularly, if semantic and phonological cues have the same saliency, earlier available cues are preferred to determine noun class. It is, indeed, the case that in natural language learning contexts, learners have access to phonological properties of nouns before semantic features of their referents, which explains why the former are preferred (Carroll, 1999; Culbertson et al., 2017; Culbertson & Wilson, 2013; Demuth, 2000; Polinsky & Jackson, 1999). The finding that formal cues are prioritised during learning suggests that, due to its grammatical nature, morphological agreement might be equally prioritised and particularly helpful in noun class acquisition.

While a number of studies have looked separately either at the role of agreement in communication (German and Latin: Dye et al., 2016; Cantonese: Tsang, Chambers & Mozuraitis, 2017) or at the acquisition of noun classes and non-adjacent dependencies (Culbertson et al., 2017; Gomez, 2002; Siegelman & Arnon, 2015), only one study has addressed the potentially role of agreement in noun class acquisition (Demuth & Weschler, 2012). As mentioned above, the findings of Demuth & Weschler (2012) suggest that reduplicative agreement, as in Sesotho, might facilitate noun class acquisition. This study sets

out to investigate whether this is in fact the case, and whether any facilitative effect depends on the type of agreement: reduplicative or arbitrary. Using an artificial language learning experiment, we compared the effects of three experimental conditions which vary in their type of morphological agreement on the acquisition of a language with three noun classes. Each noun class contains ten noun and is based on a core semantic criterion (60% of the total nouns), but also contains exceptions to the semantic rule (40% of the nouns), following Frigo and McDonald (1998): class 1 contains six humans, class 2 six animals and class 3 six plants. Further, each class contains four objects to reduce its semantic reliability. The first one is reduplicative agreement: participants are taught a language with an agreement system in which same form correspond to same meaning (similar to (2) b., (2) c., (1) a. and (1) b.). The second condition is arbitrary agreement. It is similar to (2) a.: participants are taught a language with an agreement system in which two different forms are used to express the same meaning. Finally, in the third condition, participants are exposed to a language which is more similar to a noun classifier system: there is one morphological encoding which accompanies the noun, but no corresponding agreement.

The three conditions are predicted to have different learning advantages and costs. Reduplicative agreement provides multiple cues to noun class—predicted to facilitate noun class learning—and requires learners to acquire only one agreement morpheme. Arbitrary agreement similarly has multiple cues to noun class available, but participants need to learn multiple morphemes instead of one. Finally, in the no agreement condition, learners only have one cue to noun class, but on the other hand, there is no agreement to learn.

We therefore expect the three agreement conditions to yield different learning results. Specifically, our main prediction is that reduplicative and arbitrary agreement will lead to better noun class learning compared to no agreement. This is because the latter does not provide multiple cues to noun class, which are expected to facilitate the learning process. The second prediction is that, between the reduplicative condition and the arbitrary condition, the former will better facilitate learning. This is because, even though both conditions have the advantage of multiple cues to noun class, arbitrary agreement comes with the additional cost of learning multiple agreement forms for each class. To summarize, we predict that our experimental conditions will yield different learning results. Particularly, we expect to have the following ranking for learning results from best to worst: reduplicative agreement, arbitrary agreement and no agreement.

2. Method

2.1 Participants

Participants were recruited on the online platform Prolific. There was a total of 150 participants: 50 per each experimental condition. Participants were all adults and monolingual native English speakers, as indicated by Prolific filters. Before the start of the experiment, they were shown the consent form and asked to provide their consent to participate by pressing a button. 13 participants were excluded (3 from reduplicative agreement, 6 from arbitrary agreement and 4 from no agreement) because they failed more than the allowed number of attention checks, as illustrated in session 1 and session 2 below.

2.2 Materials

The materials for this experiment were audio recordings of nonsense words and visual stimuli. By pairing the audio recordings with the visual stimuli, we created an artificial language consisting of 3 noun classes, 30 nouns (10 per each class), 2 numerals and 6 suffixes: -ki, -pa, -mu, -to, -ne and -guh.

For each participant, the 30 nouns and the 2 numerals were drawn randomly from a list of 144 CVC monosyllabic nonsense words, which were taken over from a previous noun class learning experiment by Culbertson, Gagliardi and Smith (2017). These are some examples: *clound, snork, frag, flaid, speat, stope, plound, clount, plock, moint, crail, yort, stird, treast* and *slome*. Each word was randomly assigned to a class and matched to its appropriate suffix, depending on experimental condition.

Each noun was presented to participants in two contexts: (i) in the singular form and (ii) in the plural form accompanied by one of the numerals. Therefore, each noun appeared three times: once in the singular and twice in the plural as illustrated in table 2 below:

	Reduplicative Agreement	Arbitrary Agreement	No Agreement
Class 1	Kripki Kripki gleafki Kripki gloutki	Kripki Kripki gleafpa Kripki gloutpa	Kripki Kripki gleaf Kripki glout
Class 2	Forchmu Forchmu gleafmu Forchmu gloutmu	Forchmu Forchmu gleafto Forchmu gloutto	Forchmu Forchmu gleaf Forchmu glout

Class 3	Cladene Cladene gleafne Cladene gloutne	Cladene Cladene gleafguh Cladene gloutguh	Cladene Cladene gleaf Cladene glout
---------	---	---	---

Table 1 shows examples of the three types of agreement systems that were taught to participants.

The visual stimuli consisted of photos of humans, animals, plants and objects. The images were randomly drawn from a set of 156 images and randomly allocated to the three classes. The result is a different randomisation of pairs of words and images for each participant. As mentioned in the introduction, the classes were based on a semantic principle, but the semantic cue was only partially reliable: as in Frigo and MacDonald (1998), 6 out of 10 nouns followed the semantic rule (either human, animal or plant), while 4 out of 10 were exceptions belonging to the object category.

Finally, the experiment was coded by the author with in jsPsych (de Leeuw, 2015).

2.3 Procedure

The experiment consisted of two sessions. In session 1, participants were exposed to the words of the artificial language, whereas, in session 2, they were trained and then tested on the full stimuli set (audio recordings and images). The experiment was split into two sessions to facilitate the learning process. Acquiring thirty new nouns and their noun class and corresponding agreement system is a demanding task in terms of memory. However, some studies have shown that previous exposure to unsegmented speech facilitates mastery of certain grammatical relationships between words. For instance, Siegelman and Arnon (2015) investigated the acquisition of noun class in an artificial language paradigm. They found that learners who were exposed to unsegmented article-noun units first and segmented speech later showed better learning results of the noun class system of their artificial language compared to the participants that were exposed to segmented speech first and unsegmented speech afterwards. Session 1 was based on this result which suggests that indeed exposure to multiword units facilitates noun class acquisition.

Participants were randomly assigned to one of the three conditions on session 1. Since each randomisation of words and images pairs differed from one another, on session 1 participants' id were recorded and used to present the same randomisation to the same participant on session 2.

2.4 Session 1

Participants were exposed to the words of the artificial language: the audio recordings were concatenated and played consecutively with a short pause (750 ms) between each phrase consisting either of singular nouns or of noun and numeral pairs. Participants listened once to the complete set of audio recordings, that is singular and plural phrases (see Materials) accompanied by the appropriate suffix. The stimuli were presented in a random order.

Every five phrases (approximately 20/30 seconds) participants performed an attention task: they heard the audio recording of an English number and were asked to select the button corresponding to the number they just heard out of three options. If they made the wrong selection, the last audio segment was played again, until they made the correct selection. The accuracy of responses in this task was used as an exclusion criterion: we decided to exclude the data from participants who selected the wrong number more than 15% of the times, as they did not show a high level of attention.

2.5 Session 2

Session 2 consisted of three phases: exposure, noun testing and agreement testing. In the exposure phase, participants saw and listened to the complete set of stimuli: the audio recordings from session 1 were paired with their corresponding image and their written form. The audio and the visual stimuli were presented simultaneously. Each stimulus was presented once, in other words participants were exposed to each noun once in the singular, once in the plural with the numeral 2 and once in the plural with the numeral 3. Figure 1 shows examples of the stimuli presented to participants during the exposure phase.



vokene



cruitne gleafguh



sterkmu gloutto

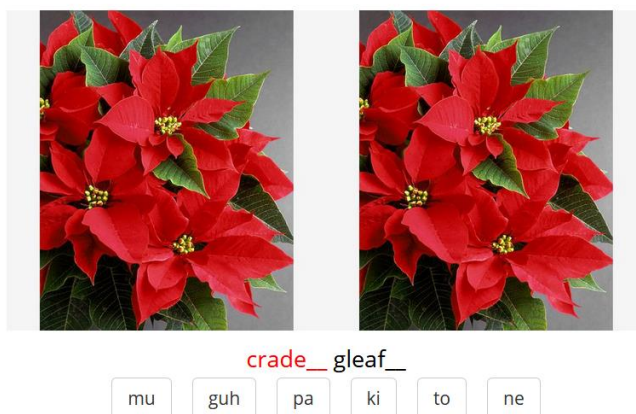
Figure 1 shows three examples of the stimuli shown to participants in the arbitrary agreement condition. The image on the top left corresponds to a singular noun, the one on the top right corresponds to a noun in the plural with the 2 numeral and the one on the bottom corresponds to a noun in the plural with the 3 numeral. The stimuli for the other two conditions were similar: the only difference was in the type and number of suffixes.

After each trial, there was an attention task: participants were asked to select, among four buttons, the one corresponding to the phrase they just heard and saw. Only one of the four buttons was correct, the other three buttons had either the wrong suffix or the wrong root, as exemplified in figure 2.



Figure 2 shows one trial from the exposure phase (on the left) and its corresponding attention check (on the right). Only one button in the right image has the correct alternative (the last one). The other ones have either the wrong root (button 1), the wrong suffix (button 3) or both (button 2).

As for session 1, this task was used as an exclusion criterion data of participants who failed more than 15% of the attention checks were excluded. After the exposure phase, participants performed two tests: the noun classification test, which assessed how accurately the noun class has been learnt, and the agreement test, which assessed the correct acquisition of the agreement system (figure 3 shows an example).



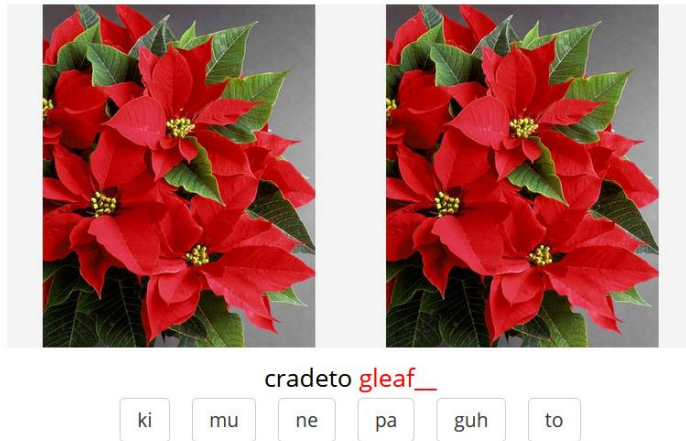


Figure 3 show two examples from the agreement testing phase: participants must select the suffix on the noun (image on the top) and then the suffix on the numeral (image on the bottom).

In the noun classification test, participants were tested on singular nouns. They were shown the same stimuli as in the exposure phase. However, the suffix of each noun was replaced by a ‘beep’ in the audio recording and by an underscore in the written form. Participants were given a list of buttons each containing one of the possible suffixes (6 for the arbitrary agreement and 3 for the reduplicative and no agreement conditions). They were asked to select the button with the correct suffix. The agreement test is similar (suffixes are replaced by a ‘beep’ in the audio recording and by an underscore in the written form), but, instead of one suffix, participants must provide two suffixes: one for the noun and one for the numeral, with the exception of the no agreement condition which has only one suffix on the noun.

2.6 Coding

Participants’ responses were subsequently coded in the R software. In the noun classification test, each trial was given a binary outcome score - ‘1’ if the target form and the response matched, ‘0’ if they did not. In other words, a point was given if participants selected the correct suffix for each noun. Likewise, each trial in the agreement test was also given a binary outcome score. However, since this type of testing has two points of interests – noun class acquisition as well as the agreement acquisition – two coding schemes are used. For the noun class acquisition, ‘1’ was given if the target form and the response matched, ‘0’ if they

did not, as above. Instead, for the agreement acquisition, ‘1’ was given if the agreement was correctly acquired (i.e., the suffix on the noun and the suffix on the numeral matched), ‘0’ was given if the agreement was not correctly acquired. In this case, participants were given 1, even if the noun suffix (i.e., the suffix indicating the noun class) was wrong, as the point of interest was the correct acquisition of the agreement system.

2.7 Results

In the introduction, we made the following two predictions: i) the no agreement condition will yield worse learning results than the other two conditions, and ii) participants in the reduplicative agreement condition will perform better than those in the arbitrary agreement condition. These patterns of results are expected as morphological agreement should facilitate noun class acquisition in reduplicative agreement and arbitrary agreement compared to no agreement, but having two morphological encodings should make acquisition harder in arbitrary agreement compared to reduplicative agreement. Further, we predict that inanimate trials will be particularly interesting to answer our hypotheses. This is because, while in animate trials participants have two cues to noun class (a semantic and a morphological one), in inanimate trials they can only rely on morphology. Therefore, we predict, that participants in reduplicative agreement will show better learning results for inanimate trials compared to arbitrary and no agreement. To assess our hypotheses, we used two statistical models. With the first one, we evaluated how well participants learnt the noun classes across the three conditions, in other words, how well participants remembered the morphological encodings the nouns. With the second model, we looked at how well participants learnt the agreement system, in other words which suffixes matched.

2.8 Noun Class Learning

Firstly, we looked at the correct selection of noun class suffix in the noun testing phase. Across all three conditions, participants had an accuracy rate above 54%, which means that they performed well above chance in learning the noun classes. Figure 4 shows participants’ accuracy rate across all conditions, including animate and inanimate trials.

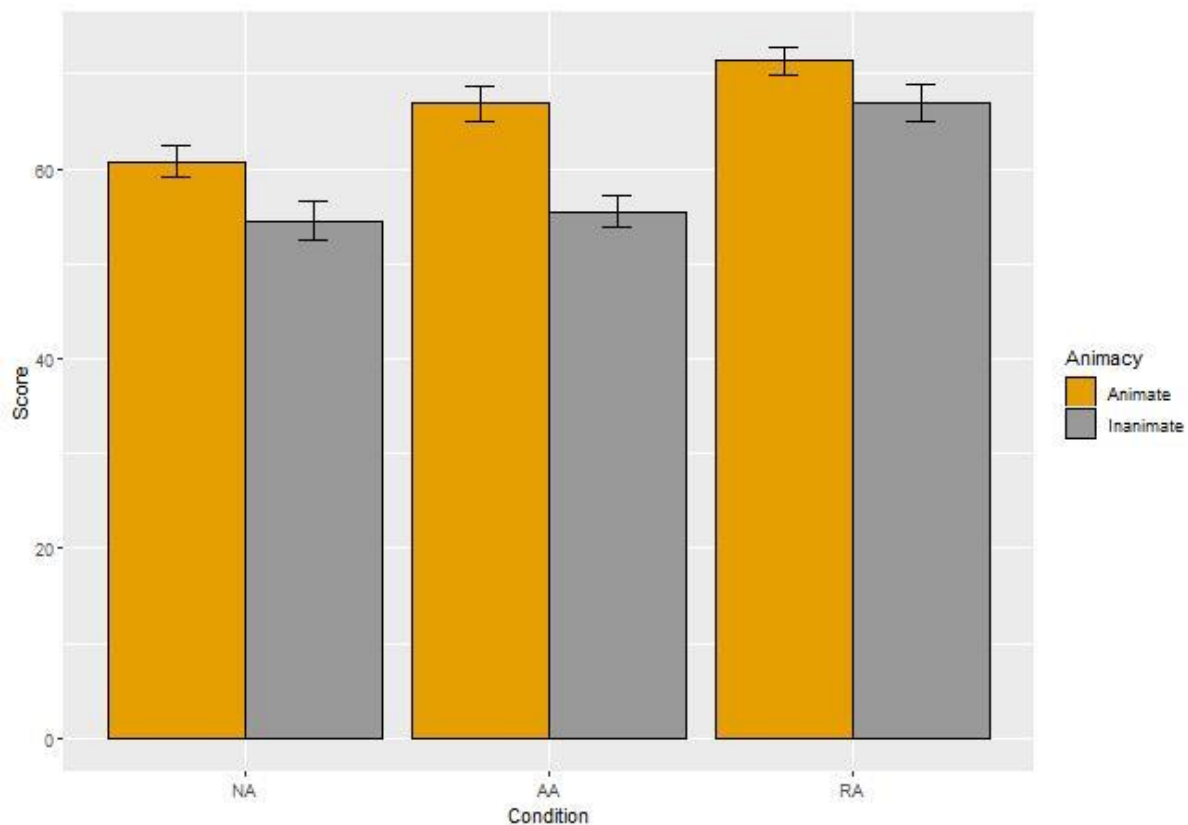


Figure 4: This bar graph shows a comparison of participants' accuracy across the three morphological conditions in the singular noun testing phase (AA: arbitrary agreement, NA: no agreement, RA: reduplicative agreement). The yellow bars show the proportion of correct responses for animate trials, whereas the blue bars indicate correct responses for inanimate trial. Error bars show a 95% confidence interval for each condition.

As shown in figure 4, participants show better learning outcomes for animate trials, where they have a reliable semantic cue, compared to inanimate ones. The difference in performance is the smallest in RA suggesting that indeed reduplicative agreement facilitates noun class acquisition compared to AA and NA. Contrary to our expectations, the discrepancy between animate and inanimate trials seems largest in the arbitrary agreement condition: this suggests that participants found learning two morphological encodings quite challenging and not so beneficial in terms of noun class acquisition.

The results of the noun class test were then analysed using a logistic mixed effects regression model with the lme4 package (Bates et al., 2015) in R (R Core Team, 2018). Condition, animacy and their interaction were set as fixed effects with participant and item as random

intercepts and animacy as random slope. No agreement is set as the reference level; inanimate trials were set as the reference level for animacy using simple treatment coding.

Noun Class Model: $\text{Score} \sim \text{Condition} * \text{Animacy} + (\text{Animacy} | \text{Participant}) + (1 | \text{Item})$

The results of this model are summarized in Table 2. There is a significant positive interaction between reduplicative agreement and no agreement, indicating that participants in RA performed significantly better than those in NA. This is in line with our main hypothesis that reduplicative agreement would be a particularly helpful type of agreement to acquire noun classes. No significant relationship is shown between arbitrary and no agreement, which does not confirm our main hypothesis that any kind of agreement is better than no agreement. Instead, animacy is significant ($b=0.3734$, $SE = 0.1814$, $z\text{-value} = 2.059$, $p = 0.0395$), which is in line with our hypothesis that animate trials should yield better result because they provide two cues (semantic and morphological). Finally, the interaction of condition and animacy are also non-significant. This is not in contrast with our two main hypotheses, but does not confirm our expectation that the discrepancy between animate and inanimate trials should be significantly smaller in the two agreement conditions.

	Estimate	Std. Error	z value	P value
(Intercept)	0.1805	0.1461	1.235	0.2169
Arbitrary Agreement	0.1375	0.1509	0.911	0.3623
Reduplicative Agreement	0.6316	0.1508	4.189	<.001***
Animacy	0.3734	0.1814	2.059	0.0395 *
Arbitrary Agreement: Animate	0.1049	0.1851	0.567	0.5710
Reduplicative Agreement: Animate	-0.1272	0.1839	-0.692	0.4891

Table 2: Generalized linear mixed model fit by maximum likelihood with results in logits (log likelihood = -2720.1).

2.9 Noun Class Learning in Agreement Test

We also looked at the correct use of noun class markers on nouns in the agreement testing phase. This reveals whether participants performed differently when having to provide the marker on numerals as well as the one on nouns. Figure 5 reveals that, compared to figure 4, participants in the reduplicative agreement condition performed better for animate trials, whereas participants in the no agreement condition performed worse for both animate and inanimate trials. Further, the discrepancy between animate and inanimate trials is higher in the reduplicative agreement and no agreement condition, but slightly lower in the arbitrary agreement condition.

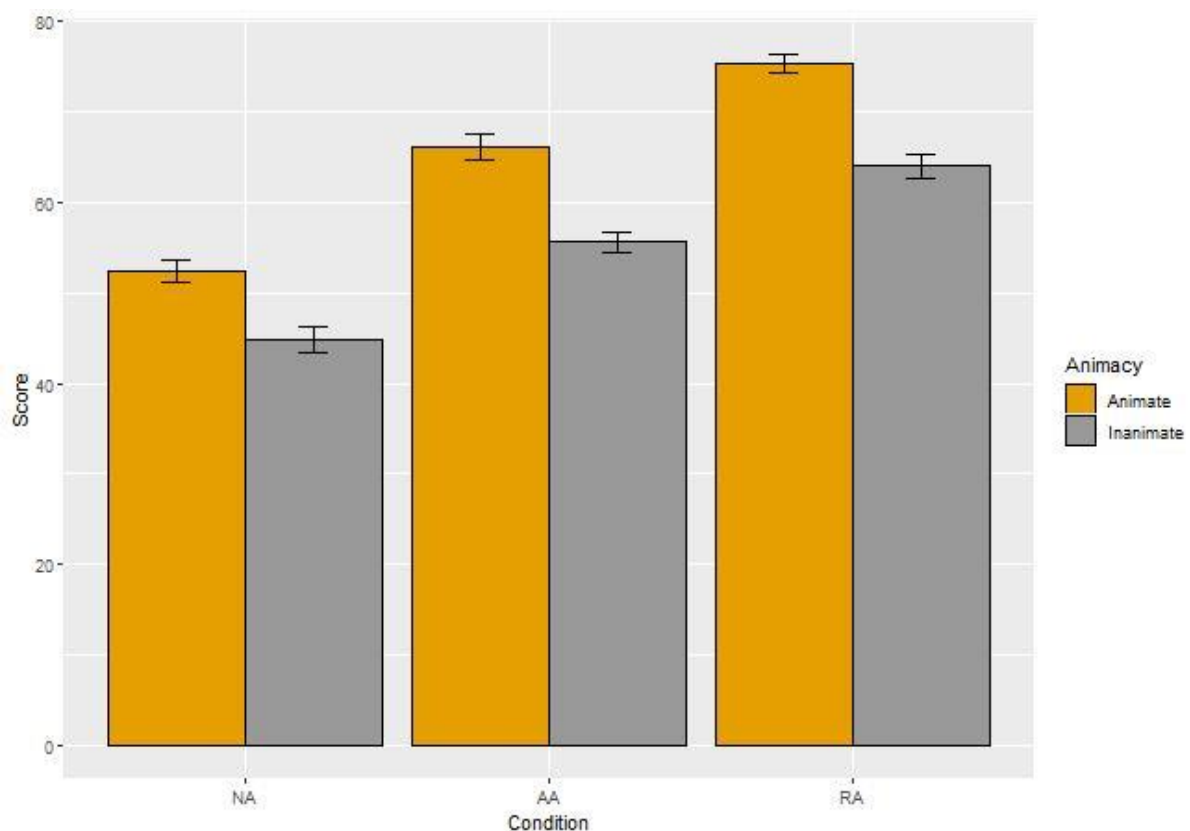


Figure 5: This bar graph shows a comparison of participants' accuracy across the three morphological conditions in the agreement testing phase (AA: arbitrary agreement, NA: no agreement, RA: reduplicative agreement). The yellow bars show the proportion of correct responses for animate trials, whereas the blue bars indicate correct responses for inanimate trial. Error bars show a 95% confidence interval for each condition.

The same Noun Class Model (as was used in the noun class test above) was applied to results of the agreement test. As shown in table 3, there is a significant positive relationship between no agreement and arbitrary agreement, and no agreement and reduplicative agreement. This is in line with our main hypothesis that morphological agreement facilitates

noun class learning. Finally, there is a positive significant interaction between condition and animacy both for reduplicative agreement and animate trials, and arbitrary agreement and animate trials. This indicates that the discrepancy between animate and inanimate trials in reduplicative agreement and arbitrary agreement is significantly bigger than that in no agreement. This is not in contrast with our main hypotheses, but also does not confirm our prediction that the discrepancy in accuracy for animate and inanimate trials should be smaller in the reduplicative agreement and/or the arbitrary agreement compared to the no agreement one.

	Estimate	Std. Error	z value	P value
(Intercept)	-0.5455	0.2464	-2.214	0.02684 *
Arbitrary Agreement	0.5650	0.487298	-2.375	0.0175 *
Reduplicative Agreement	1.1359	0.2049	5.543	<.001***
Animacy	0.3972	0.2886	1.376	0.16874
Arbitrary Agreement: Animate	0.3842	0.1927	1.994	0.04619 *
Reduplicative Agreement: Animate	0.4610	0.2019	2.283	0.02241 *

Table 3: Generalized linear mixed model fit by maximum likelihood with results in logits (log likelihood = -4731.7).

Given the different performance in the two tests (noun classification only and noun classification in the agreement phase), we ran a combined analysis of the two tests to assess whether they produce different patterns of behaviour.

Combined Analysis Model: Score ~ Condition * Animacy*Test+ (Animacy*Test| Participant

As shown in table 4, there is a significant positive effect of test, suggesting that participants performed significantly better in noun classification in agreement test compared to noun classification only ($b = 0.47508$, $SE = 0.11315$, $z\text{-value} = 4.199$, $p < .001***$). Additionally, there is a significant negative interaction of arbitrary agreement and test, and reduplicative agreement and test. This indicates that the discrepancy between the two tests in reduplicative

agreement and in arbitrary agreement is smaller than the discrepancy between the two tests in no agreement.

	Estimate	Std. Error	z value	P value
(Intercept)	0.01632	0.18374	0.089	0.9292
Arbitrary Agreement	0.18191	0.15820	1.150	0.2502
Reduplicative Agreement	0.67553	0.16440	4.109	<.001***
Animacy	0.22199	0.21771	1.020	0.3079
Test	0.47508	0.11315	4.199	<.001***
Arbitrary Agreement: Animate	0.41905	0.17109	2.449	0.0143 *
Reduplicative Agreement: Animate	0.24193	0.17847	1.356	0.1752
Arbitrary Agreement: Test	-0.35755	0.15714	-2.275	0.0229 *
Reduplicative Agreement: Test	-0.37596	0.16382	-2.295	0.0217 *
Animacy: Test	-0.16484	0.11535	-1.429	0.1530
Arbitrary Agreement: Animate: Test	0.08894	0.16173	0.550	0.5824
Reduplicative Agreement: Animate: Test	-0.10659	0.17208	-0.619	0.5244

Table 4: Generalized linear mixed model fit by maximum likelihood with results in logits (log likelihood = -6122.8).

2.10 Agreement Learning

Finally, we looked at the acquisition of the agreement system. As can be seen in figure 6 below, participants performed much better in learning the agreement in the reduplicative agreement condition than with arbitrary agreement. There is almost no discrepancy between animate and inanimate trials.

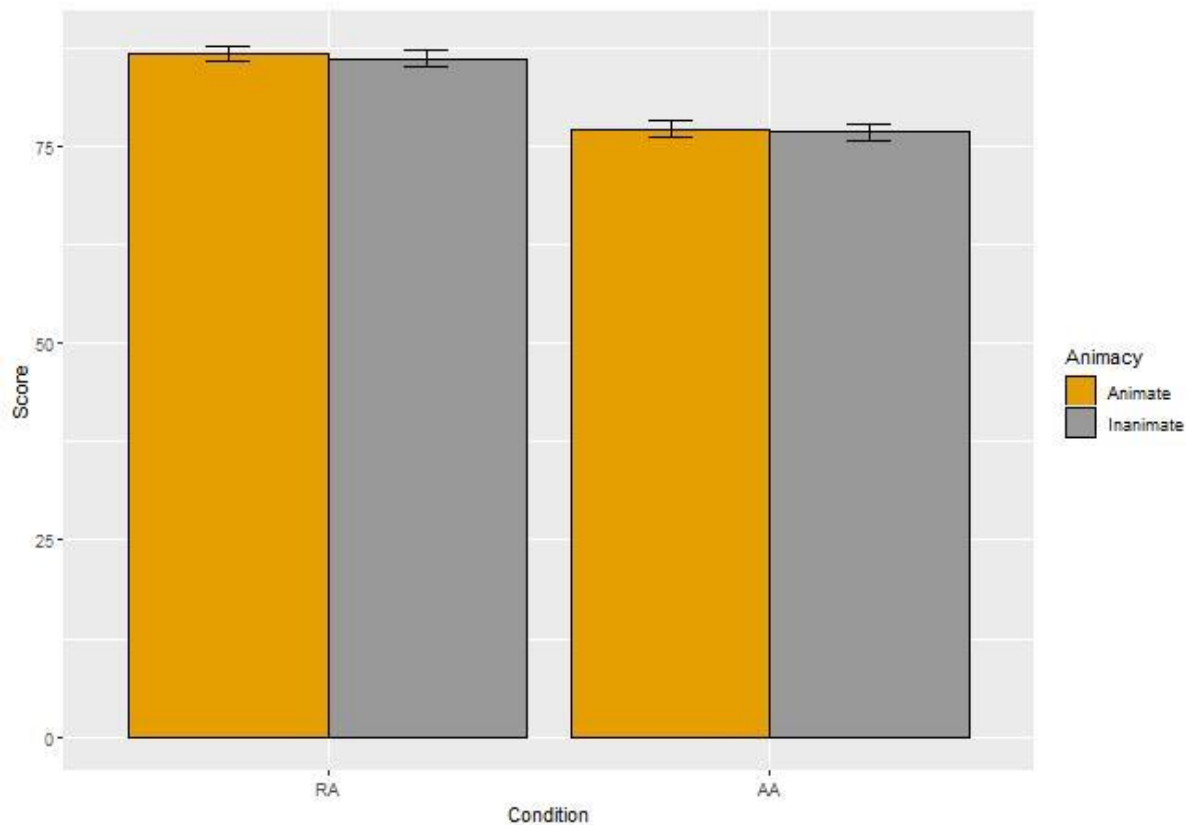


Figure 6: This bar graph shows a comparison of participants' accuracy in learning the agreement system across the three morphological conditions (AA: arbitrary agreement, NA: no agreement, RA: reduplicative agreement). The yellow bars show the proportion of correct responses for animate trials, whereas the blue bars indicate correct responses for inanimate trial. Error bars show a 95% confidence interval for each condition.

For the analysis of the acquisition of agreement system, we ran a logistic mixed effects regression model comparing reduplicative agreement to arbitrary agreement (the question of agreement was irrelevant for the no agreement condition). Condition and animacy are set as fixed effects, participants and items as random effects. Reduplicative agreement and inanimate were set as the reference levels.

Agreement Model: $\text{agreement} \sim \text{condition} * \text{animacy} + (\text{animacy} \mid \text{id}) + (1 \mid \text{Item})$

	Estimate	Std. Error	z value	P value
--	----------	------------	---------	---------

(Intercept)	3.494100	0.391415	8.927	<2e-16 ***
Arbitrary Agreement	-1.157541	0.487298	-2.375	0.0175 *
Animacy	-0.006679	0.272200	-0.025	0.9804
Arbitrary Agreement: Animate	0.116975	0.226341	0.517	0.6053

Table 5: Generalized linear mixed model fit by maximum likelihood with results in logits (log likelihood = -1681.2).

As shown in table 5, the results of this analysis confirm our expectations: there is a significant negative relationship between reduplicative and arbitrary agreement. This means that participants in the arbitrary agreement condition performed significantly worse than those in the reduplicative agreement condition.

3. Discussion

Overall, our results present some evidence in support of our two main hypotheses (i) that morphological agreement is better than no agreement in learning noun class, and (ii) that reduplicative agreement is better than arbitrary agreement. We ran the noun class model twice: (1) in noun classification only and (2) in noun classification in agreement test. We obtained different results. In (1), participants in the reduplicative agreement condition performed significantly better, confirming our hypothesis that this is a particularly efficient type of agreement. However, (1) does not confirm our hypothesis that any type of morphological agreement provides better learning results than no agreement. The results of (2) seem to be in favour of this hypothesis: in fact, in (2) both the interaction between no agreement and arbitrary agreement, and no agreement and reduplicative agreement are significant, indicating that in the agreement conditions participants learnt the morphological markers significantly better. The results of (2) present a more robust claim in favour of our hypotheses. One explanation for this could be that the agreement system boosts participants' performance more when it is recalled explicitly as in the agreement testing (i.e., participants must provide both the suffix for the noun and the suffix for numeral). This hypothesis, however, is not supported by our combined analysis of the two tests. In fact, this analysis reveals that test has a significant effect, but it seems that it is participants in no agreement

who perform worse in (2) test and not participants in arbitrary and reduplicative agreement who perform better. This result is unexpected, and it is difficult to explain why the performance of participants in no agreement should drop from (1) to (2). Given that participants in no agreement perform the same task in noun classification and agreement test, whereas participants in arbitrary and reduplicative agreement perform two slightly different tasks (i.e., in agreement test they must provide two suffixes instead of one), we would expect to find a discrepancy in performance in the last two conditions and not in no agreement. One possibility is that, given the length of the testing phases (there were 90 trials in total: 30 in the noun classification test and 60 in the agreement test), participants found the second test (agreement test) more challenging, as they might experience more tiredness. However, it is not clear why this should be the case only in no agreement and further investigation in a follow-up experiment is probably required.

It is somewhat surprising that animacy did not play a significant role in any testing phase. Model (2) found a significant interaction of condition and animacy. However, this was in contrast with our prediction that the discrepancy between animate and inanimate trials should be smaller in the agreement conditions. One explanation for this could be that to learn the noun classes participants relied on semantic criteria more than on other criteria even in the agreement conditions. To explain why in (2) there is a significant positive interaction between reduplicative agreement and animate trials, and also arbitrary agreement and animate trials, a possibility could be that instead of acquiring the underlying structure participants could rely on memory and learn each instance of noun and suffix combination individually. This would explain why the morphological agreement conditions are not more helpful in acquiring the class of those items lacking a semantic cue, because participants learnt each noun individually regardless of morphological cues. To assess whether participants relied on memory or acquired the underlying structure, the experiment could be modified in various ways. Firstly, the difficulty of learning relying on memory could be increased: for instance, more nouns could be added to each class, so that participants would have to find an alternative learning strategy to memory. Gomez (2002) suggests that if the number of items is too low, participants seem to rely on memory. Secondly, generalization trials could be introduced as in Frigo and Macdonald (1998) to determine which learning strategy was used.

Overall, the results presented here provide evidence in favour of our secondary hypothesis (that reduplicative agreement facilitates noun class learning more than arbitrary agreement). This is supported by the significant effect of reduplicative agreement in our first two analysis:

noun classification only and noun classification in agreement. Further, the agreement model analysis shows that reduplicative agreement was learnt significantly better (with an accuracy of above 86%) than the arbitrary agreement. Such a robust learning of reduplicative agreement suggests that, indeed, participants might rely on this morphological structure in noun class learning. The evidence in favour of our primary hypothesis (i.e., that morphological agreement, including the arbitrary type, facilitates noun class learning) is not so straightforward. The discrepancy between (1) and (2) does not allow to make strong conclusions about arbitrary agreement: in fact, in (1) arbitrary agreement is a non-significant effect, whereas in (2) it is a significant one. As shown by the result of the agreement model, having two forms for the same meaning makes arbitrary agreement a more difficult system to acquire. This could mean that we might have to revise our primary hypothesis: not all type of morphological agreement facilitates noun class acquisition, only reduplicative agreement does. However, given that (2) yields a significant result for arbitrary agreement, another possibility is that, to some extent, this type of agreement is helpful in noun class acquisition, indicating that the problem lies not in the hypotheses but in the experimental design. A follow-up experiment should aim to clarify the role of arbitrary agreement in noun class acquisition and the reason behind the drop in performance in noun classification in agreement test in no agreement but not in the other two conditions. Firstly, the length of the two tests could be reduced: instead of having to provide suffixes for each singular noun and each noun and numeral pair, participants could be tested only on a fraction of the language. This would make the tests less challenging, but still shed light on how well the noun classes and agreement systems have been learnt. Further, the agreement system could be made more salient: instead of two elements (noun and numeral) with the morphological marker(s) for noun class, there could be three, a noun a numeral and an adjective. This should give participants in the morphological agreement conditions an additional cue on which to rely

4. Conclusion

This study investigated the role of agreement in noun class acquisition. Our main claim was that morphological agreement would facilitate the learning of noun classes compared to the no agreement condition. Our second claim was that, within morphological agreement, some types of agreement are more efficient than others: particularly, we reduplicative agreement to show better learning results than arbitrary agreement, as the first one requires learning only one form. Our findings provide strong evidence in favour of our second claim, and mixed evidence in favour of our first claim. All our analysis revealed a significant effect of

reduplicative learning in noun class acquisition. Instead, only the noun classification in agreement test yielded a significant result for the arbitrary agreement condition, suggesting that this type of agreement is less efficient than reduplicative agreement in noun class acquisition. These results do not allow us to make any claim on the role of morphological agreement as such in noun class learning.

5. References

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. New York: Oxford University Press.

Aikhenvald, A.Y. (2021). Nominal Classification Systems. In *The International Encyclopedia of Linguistic Anthropology*, J. Stanlaw (Ed.).
<https://doi.org/10.1002/9781118786093.iela0286>

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292-305.

Bates, D. (2010). lme4: Mixed-effects modeling with R <<http://lme4.r-forge.r-project.org/book>>.

Carroll, S. E., (1999). Input and SLA: adults' sensitivity to different sorts of cues to French gender. *Language Learning* 49.37–92. doi: 10.1111/1467-9922.00070.

Christiansen, M. H., & Chater N. (2008). Language as shaped by the brain. *Behavioural Brain Sciences*, 31(5):489-508. doi: 10.1017/S0140525X08004998. PMID: 18826669.

Culbertson, J., Jarvinen, H., Haggarty, F., & Smith, K (2019). Children's Sensitivity to Phonological and Semantic Cues During Noun Class Learning: Evidence for a Phonological Bias. Advance online publication on *Project MUSE*. doi:10.1353/lan.0.0234.

Culbertson, J., Gagliardi, A., & Smith, K., (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language* 92.343–58. doi: 10.1016/j.jml.2016.08.001.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain general biases have domain specific effects. *Frontiers in Psychology*, 6.

Culbertson, J., & Wilson, C., (2013). Artificial grammar learning of shape-based noun classification. *Proceedings of the 35th annual meeting of the Cognitive Science Society*

(CogSci 2013), 2118–23. online: <https://mindmodeling.org/cogsci2013/papers/0389/paper0389.pdf>.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1-12. doi:10.3758/s13428-014-0458-y.

Demuth, K., & Weschler, S. (2012). The acquisition of Sesotho nominal agreement. *Morphology*, 22(1), 67-88.

Demuth, K. (2000). Bantu noun class systems: Loan word and acquisition evidence of semantic productivity. *Classification systems*, 270-292.

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2016). *A functional theory of gender paradigms. Morphological paradigms and functions*. Leiden: Brill.

Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, 90.58–89. doi: 10.1353/lan.2014.0013.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39(2), 218-245.

Havron, A., & Arnon, I. (*in press*). Starting Big: The Effect of Unit Size on Language Learning in Children and Adults. https://d96b8fe2-1b77-4b2a-a4a8-07afdb2e6975.filesusr.com/ugd/56a7f5_608ae3ed29164864b3363d149b4d67a1.pdf

Karmiloff-Smith, A., (1981). *A functional approach to child language: A study of determiners and reference*. New York: Cambridge University Press.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87-102.

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological science*, 19(12), 1247–1252.
<https://doi.org/10.1111/j.1467-9280.2008.02233.x>

Levy, Y., (1983). It's frogs all the way down. *Cognition*, 15.75–93. doi: 10.1016/0010 - 0277(83)90034-3.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1), e8559.

- Mills, A. E., (1985). The acquisition of german. In I. Slobin, *The crosslinguistic study of language acquisition*, 1, 141–254. Hillsdale, NJ: Lawrence Erlbaum
- Polinsky, M., & Jackson, D., (1999). Noun classes: Language change and learning. In Fox, B., Jurafsky, D., & Michaelis, L., (Eds), *Cognition and function in language* (pp. 29–50). Stanford, Ca: CSLI Publications.
- Seifart, F. (2010), Nominal Classification. *Language and Linguistics Compass*, 4: 719-736. <https://doi.org/10.1111/j.1749-818X.2010.00194.x>
- Senft, G. (2007). Nominal classification. In D. Geeraerts, & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 676-696). Oxford: Oxford University Press.
- Siegelman, N., & Aron, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, 85, 60-75.
- Surridge, M. (1986). Genre grammatical et dérivation lexicale en français. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique*, 31(3), 267-283. doi:10.1017/S0008413100011749
- Trudgill, P. (2011). *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford, UK: Oxford University Press.
- Tsang, C., & Chambers, C. G. (2011). Appearances aren't everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1065.