Grammatical gender and observational word learning: useful addition or useless information?

Bachelor's Project Thesis

Floris Uithof, f.h.uithof@student.rug.nl,
Supervisors: dr J.K. Spenader

Abstract: This research investigated whether including grammatical gender information influences word learning for both children and adults. Previous research with adults showed that word learning improved when the target picture belonged to a consistent semantic category, even if that picture hadn't been shown before (Dautriche and Chemla, 2014). This raised the question: would including extra linguistic context in the form of grammatical gender information also improve learning? We extended an observational word-learning task by Woodard, Gleitman, and Trueswell (2016) to a 2x2x2 between-subjects design, by having 3 instead of 2 pictures in each trial and by adding a new between-subject factor: grammatical gender. Participants had to match one of three unknown animals with an unknown name during 2 presentations, with some filler trials in between. We looked into the effects of age (preschool children versus adults), the article that was used in the sound files (definite article in Dutch ('de' or 'het') versus indefinite article ('een')) and the target picture on the second presentation (the picture matched with the name on the first presentation versus one of the other two pictures from the first presentation) on performance.
We found that for both children and adults, participants did not perform better when grammatical gender was added. However, similar to the results of Woodard et al. (2016), participants performed better when asked to identify the same picture in the second presentation. Furthermore, the adult participants performed significantly better than the child participants.

1 Introduction

Language acquisition is one of the most astonishing feats humans are capable of. Being able to verbally communicate with other human beings opens up a world of possibilities and allows for much more complex communication than is possible for other species. The process of language learning starts in infancy with the ability to recognize one's own name, something that happens at just 4 months of age ((Depaolis, Vihman, and Keren-Portnoy, 2014)) and grows until adulthood, when most adults know the finer details of a language. Language acquisition can be divided into sub-processes, like word learning and the acquisition of syntax and semantics. All these processes are necessary to master a language. In this paper, we look into acquiring the meaning of words: how does one learn what a certain word refers to?

1.1 Word learning: fast mapping

Word learning is one of the most important features of language: communication is nearly impossible without knowledge of words and their meanings. Words are the building blocks of a language, and the process of acquisition has caused some debate in the academic world. According to Bloom (2000), individuals have an average vocabulary size of about 60,000 to 80,000 words. This is an incredible amount, especially considering the fact that it is difficult to retrieve facts, but words can be retrieved instantaneously. A vocabulary of 60,000 words exhibits that one would need to learn about 10 words a day from age 1 to 18 to reach this amount (Bloom, 2000). How are these words learned? And what factors might influence word learning?

First, we will consider how words are learned.

A concept that is often named regarding word learning is fast mapping (Bloom, 2000), which is called fast mapping because people can map the name of an object to an object really quickly: just a few encounters are needed. Evidence for fast mapping is given in the work of Bloom and Markson (Bloom, 2000), whose experiments show that fast mapping is used in learning new meanings for words.
In their experiment, Bloom and Markson had a group of 3-year-olds, a group of 4-year-olds and a group of adults learn new words. These new words were introduced to the children in speech and accompanied by either a linguistic fact or a visual fact. The participants were divided into three groups that were tested directly, after a week or after a month of learning the new words. The researchers found that when a new word was accompanied by a linguistic fact, all 3 groups recalled that new word with an accuracy significantly above chance, even after a month. When the word was accompanied by a visual fact, however, they saw that after a month performance dropped to chance for the children, but the adults remained above chance.

The results of Bloom (2000) show that fast-mapping is part of at least object learning (based on the performance with a linguistic fact, which was above chance), but not for any memorization task (based on the performance with a visual fact, which was at chance). In our experiment we work with mapping a word onto an object, and even though the 2 groups of children and the adults in this experiment performed at or above chance, the use of a visual aid could mean that adults would perform better.

1.2 Word learning: two theories

Does fast-mapping influence word learning? One theory in word learning leans heavily on the concept of fast mapping. This theory is called 'Propose-but-verify', which according to Trueswell, Medina, Hafri, and Gleitman (2013) is like hypothesis testing using fast mapping: one possible meaning is connected to a word and evaluated. When this meaning is correct, it will remain connected to the word, but when it is incorrect another meaning is connected to the word until one has found the correct meaning. In this paradigm, a word only has one possible meaning at the same time.
However, 'Propose-but-verify' is not the only explanation research offers for word learning. Another theory is called 'Accumulative learning', and is distinct from 'Propose-but-verify'. According to 'Accumulative learning', people who do not know the meaning of a word yet (the object or concept it is referring to) keep a set of possible referents in their mind, pruning referents that are not plausible anymore (after one learns what that referent actually is called, for example) until only one is left. Evidence for this theory comes from research by Dautriche and Chemla (2014), who showed that the meaning of made-up words could be learned based on under-informative trials. In their experiment, participants were shown a made-up word with 4 different pictures. This happened 12 times, for 12 different words. By altering the picture shown with the words in consecutive sets of trials, the participants would give more correct answers over time because they would reason which picture belonged to the word based on previous trials. After seeing the word with different sets of pictures, participants would start pruning the pictures that they did not see every time, until only picture was left.

Many objects can be the referent of a word, for example all objects around the object a word is actually referring to. These other objects are context to the actual referent of a word, and could influence the way people actually learn words.

1.3 Context

Context is an intriguing factor in word learning. Words are rarely encountered in isolated form: they are often located in a sentence, which is surrounded by other sentences in a text. Because words are often surrounded by some form of context, could this context be used in the learning of new words? Evidence for this claim has be found in several studies (Dautriche and Chemla, 2014; Borovsky, Kutas, and Elman, 2010; Borovsky, Elman, and Kutas, 2012).
Dautriche and Chemla (2014) showed that when new, made-up words were presented in a semantically consistent context, they were remembered

better than when the context was random. They performed a second experiment, and compared it with the experiment that is described above. In the first experiment, as described above, the pictures that were shown in each trial of the first block were random pictures. In the second experiment, the 12 trials of the first block would have pictures from just one of three categories: animals, clothes and a mixed set of both. These pictures would provide the participants with context for the meaning of the word. The results showed that the participants performed significantly better when the words were first shown in a semantic context than when they were not. Dautriche and Chemla (2014) argued that this was caused by the context that the participants had stored together with the word, which was proven by the fact that participants would choose a picture from the semantic category the word was first shown in more than from the other category, even if that picture came from the mixed category (It shared the same semantic category, but had never been shown with that word). These results show that context does have an influence in word learning.

More evidence comes from Borovsky et al. (2010, 2012), who found that unknown words are learned better when placed in a highly constrained context compared to a poorly constraining context. This is based on the effects that they found in a certain Event-Related Potential (ERP): the N400. This ERP is linked to word learning and it behaved similarly for known words as for unknown words, but only when these unknown words first appeared in a sentence with a highly constrained context. This shows that the unknown words are actually learned better when appearing in a highly constraining context than in a poorly constraining context.

Word learning is a complex matter, with different paradigms trying to describe it. Dautriche and Chemla (2014) shows that adults can use certain contextual features to their advantage in the process, and Borovsky et al. (2010, 2012) shows that linguistic context in the form of sentences can be useful for word learning as well. But what about other linguistic features? In the next paragraph, we will go over one of these features: grammatical gender.

1.4 Grammatical gender

Languages contain linguistic features, an example of which is grammatical gender. But what exactly is the grammatical gender?
In his classic work, Charles Hockett gives a definition of grammatical gender: "Genders are classes of nouns reflected in the behavior of associated words" (Hockett, 1958, as cited in Bloom, 2000). The prevalence of grammatical gender differs per language group. In the Indo-European language family, having a grammatical gender system is common, while in the Uralic language family it does not occur at all(Corbett, 1991). According to Corbett in Haspelmath, Dryer, Gil, and Comrie (2005), 44% of the world's languages have a grammatical gender system, while 56% do not have one.
There is a lot of variation in which grammatical gender can occur. Several common systems are:

- A system with 2 genders: masculine and feminine. This is common in Romance Languages like French. (Boloh and Ibernon, 2013)
- A system with 2 genders: masculine-feminine and neuter. A system like this can be found in Dutch. (Unsworth, 2007).
- A system with 3 genders: masculine, feminine and neuter. A good example of this system can be found in German. (Lemhfer, Schriefers, and Hanique, 2010)
- A system with 3 genders and subgenders: this is a construction found in Slavic languages, where besides the masculine, feminine and neuter gender there is a subgender: animate/inanimate. (Ibrahim, 1973)
- A system with more than 3 genders: the amount of genders can be large, as can be observed in a language like Arapesh: a language from Papua New-Guinea which has 13 genders. (Haspelmath et al., 2005)

When do children start using grammatical gender? Depaolis et al. (2014) showed that word learning in children starts around 12 months: this is when children can recognize words from running speech (single words can be recognized earlier, but in on-line language processing the onset is later).
Research on the acquisition of grammatical gender

has been done by van Heugten and Christophe (2015). They designed an experiment in which French toddlers (on average just 18 months old) were tested on their knowledge of grammatical gender. The researchers measured the time the children looked at the movie used to capture their attention while listening to somebody pronouncing an article-noun combination. They found that the children focused longer on the movie when the article-noun combination was correct (a masculine article with a masculine noun) than when they were incorrect (a masculine article with a feminine noun). According to van Heugten and Christophe (2015), this shows that young children already have sensitivity for gender cues and that they have a preference towards correct article-noun combinations, which is quite impressive.

Research has also been done regarding the use of grammatical gender in both children and adults. For example, in Lew-Williams and Fernald (2007), the researchers performed an experiment to see whether grammatical gender was informative in both Spanish-speaking children and adults. In the experiment, two pictures of objects were shown. These objects could either be from the same grammatical category (for example, both masculine) or from different categories (masculine-feminine). A sentence with one of the objects in it was played and the reaction time (the time a child took to focus on the correct object) was measured. The results showed that both a group of children and a group of adults were significantly faster to focus on the correct object when the two objects did not have the same grammatical gender than when the objects did. This left the researchers with the idea that people can use grammatical gender as an informative feature.

1.5 Research question

In our current study, we are going to combine the two subjects of language acquisition discussed above: word learning and grammatical gender. Previous research found that grammatical gender can be an informative feature, and that words can be preceded by items with an article with grammatical gender. Based on these findings, our research question is as follows:

Does grammatical gender in articles help in word learning?

1.6 Hypothesis

We based our hypothesis on research we mentioned above. Here we found that context can be of aid in word learning and that grammatical gender is an informative linguistic feature. This research shows, together with that of Dautriche and Chemla (2014), that context does influence performance, and although this is between different levels of context, it shows that context does have an effect. Because of this, our hypothesis is that the grammatical gender does have influence on the word learning performance of our participants.

1.7 Experiment

We have designed an experiment based on an earlier experiment performed by Woodard et al. (2016). In this experiment, participants performed a so-called observational word learning task, in which words are learned by showing pictures with the corresponding names in a sound file. By creating sound files containing articles with and without grammatical gender, we tried to find any differences in word learning performance.

2 Method

2.1 Participants

71 participants took part in the experiment, split into 2 groups. The first group of participants were Dutch children from a local daycare centre. This group consisted of 34 participants, of whom 18 were male. The average age of the children was 2;11 (between 2;0 and 3;11, SD = 7.2 months).
We chose to have children of this age as participants because this is commonplace in developmental linguistic research: according to Bloom (2000), you can pick participants of any age, but children between the ages of 2 and 5 are relatively easy to test: they are often enthusiastic and when they are at (pre)school, a larger group can be tested at once. However, because we used articles in our test, we needed to be sure that children could use articles correctly. If they could not do this, it would hard

to get extra contextual information out of the articles: they would not be "attached" to words firmly. According to Maratsos (1976), children around the age of 36 months know how to use articles properly, so they can use any information embedded in them correctly. This is why we tried to test children who were around that age or older. The parents of the children signed an informed consent form due to their young age. They spent an average of 15 minutes performing the experiment and they 12 were rewarded with a sticker.
The second group were adult participants. This group consisted of 37 Dutch participants, of whom 18 were male. The participants were mostly students and staff of the University of Groningen. The average age of these participants was 28;1 years (between 18 and 70, SD = 12;9 years). All participants signed an online informed consent form and performed the experiment on their own computers at home. They spent somewhere between 5 and 10 minutes performing the experiment and did not receive a reward for their participation.

2.2 Design

The experiment was based on the observational word learning experiment that was conducted by Woodard et al. (2016), although we made some notable changes. The most important change can be found in the design: we created a 2 X 2 X 2 design (Condition X Gender X Age), in which Condition stands for the Same and Switch conditions described in Woodard et al. (2016) . Gender is whether the words are preceded by an article with or without grammatical gender. Age refers to the two age groups we have: preschool children and adults. The original experiment only looked at the influence of the Same and Switch case in children, which makes this experiment more comprehensive.

The observational word learning task consisted of 6 separate blocks, in which 6 new words were learned. These words can be found in Table 2.1. They were generated using an on-line fake word generator (Feldarkrealms, Unknown Year) and were selected because they could be pronounced as Dutch words (there were no fake Dutch word generators available, so these words were generated using a fake English word generator).

Table 2.1: New words that were learned during the experiment

Words to be learned:
Trusp
Zwoel
Lysel
Korin
Skart
Nebak

Each block consisted of 5 different trials. Just like in Woodard et al. (2016), the trials consisted of pictures of animals, with a button that can be pressed to play a sound file. In Figure 2.1, an overview of a block is shown. A difference with the design of Woodard et al. (2016) is that instead of two pictures of animals, three pictures were shown. This was done because in Woodard et al. (2016), the children already had a high accuracy scores. As is described in Trueswell et al. (2013), children performed at chance level when the previous item was incorrect (like in our switch case). By increasing the number of pictures, we lowered the chance level and therefore hoped to lower the performance of the children. We expected that this would make the effects better recognizable, because the children do not perform at near-ceiling level.

Each block followed the same pattern: the first trial consisted of a set of three familiar animals. By pressing the sound button, a sentence would be played, according to the following structure:

(1) "Kijk, Daar is [ARTICLE] [ANIMAL]! Zie jij ook [ARTICLE] [ANIMAL]? Wijs ernaar."

(2) "Kijk, Daar is het nijlpaard! Zie jij ook het nijlpaard? Wijs ernaar."

(3) "Kijk, Daar is een nijlpaard! Zie jij ook een nijlpaard? Wijs ernaar."

The second trial was similar, except for the fact that there were 3 unknown animals on the screen. After trials 3 and 4, filler trial with known animals, the last trial of the block was reached. This trial was similar to trial 2: There were three unknown animals lined up, one of which had previously been

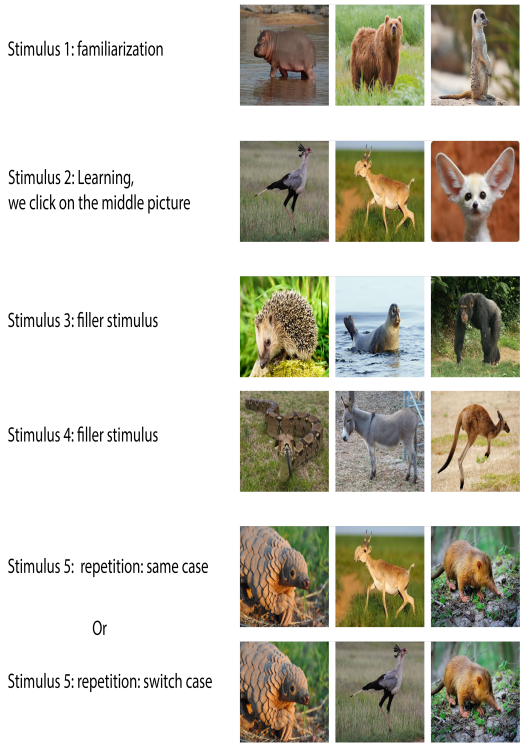

Figure 2.1: Overview of an entire trial, with 5 sets of trials. One of the last 2 is used, depending on the version

seen in trial 2. Depending on which version was used, it was either the animal that had been clicked on (the 'Same' case) or one of the other two, randomly chosen (the 'Switch' case). The goal was to click on the animal that was also shown in the second trial. The other trials (3 sets per block, 18 in total) were used to see how well participants knew animals.

2.3 Experiment materials and stimuli

The experiment was created using PHP, HTML, and Javascript and its hosted on the University server. The images used in the experiment were retrieved from the internet. The pictures had a height of 320 pixels and together spanned the whole screen.

The sound files could be listened to by pressing the sound button. When this button was pressed, the sound file attached to the trial played. Every version had its own sound file, and these could be paused, unpaused and replayed by pressing the button.

A fellow student recorded the sound files for us in order to avoid confusion between the voice in the experiment and the xperimenter's voices. All sentences were built using the sentence structure from example 1. There were two lists of sentences: one containing articles with grammatical gender (Example 2) and one with articles without grammatical gender (Example 3). The recordings were subsequently split and slightly edited (silences were cut off, volume was increased) using the Audacity program (Mazzoni and Dannenberg, 1999).

To correct for biases in our experiment, we created several different versions of our experiment, based on those used in the experiment of Woodard et al. (2016). First, to correct for possible connections between the made-up words and the pictures of the unknown animals (for example the kiki-bouba effect, where certain words are associated with certain objects based on pronunciation of the words and shape of the objects (Gómez Milán, Iborra, de Córdoba, Juárez-Ramos, Rodríguez Artacho, and Rubio, 2011)) we made 4 additional lists in which the made-up words are combined with different sets of animal pictures. To correct for possible preferences in the position, we created 4 more lists in which the animals were shown in a

different order on the screen, which will result in different clicking behaviour (the correct animal could be moved from the center position to the right, in which a different action is required). We also made 4 extra lists in which only the positions in trial 5 of each block are changed. Furthermore, we corrected for the order in which the blocks are presented: we created 16 extra versions, one for each of the before mentioned versions, in which the trials were reversed (the blocks were shown in reverse order, with all trials reversed as well, except for trial 2 and 5, for obvious reasons). This gave us a total of 32 lists, 8 for each version.

We also made sure our experiment was approved by CETO: the Research Ethics Review Committee of the University of Groningen. They found that we follow international standards for the protection of research participants. This meant that our experiment should have no adversary effects on our participants.

2.4 Procedure

For the child participants, the experiment was performed in a quiet space with minimal distractions to maximize concentration. The children were seated in front of the laptop with a researcher on either side of them. After a short talk with the child to make them feel at ease, the child was then instructed that they would be testing a small game. In this game, there were going to be 3 pictures of animals and a voice asking them to point at one of them. When this question was asked, the children were asked to explicitly point at the animal. It was also explained that some animals might be unknown to them, and that in that case they could point to the animal they thought was connected to the name. To show what would happen during the experiment, one test trial was performed. After this test trial, an intermediate screen was shown and the children were asked if they had any more questions about the 'game' they were about to play. The children were also asked if the conditions of the experiment were good: if the sound files were loud enough and whether or not the pictures were clearly visible. Any problems would be resolved and the experiment was started afterwards.
The researcher operated the computer during the trials. For each trial, the researcher pressed the sound button to play the sound file. After the recording was played, the child would point at the picture they chose and the researcher would click on it, saving the information. When a child would hesitate to answer the question, the researcher would ask if they wanted to hear the sound file again. Once a child had chosen, the researcher would click on the picture and the experiment would continue with the subsequent trial. No feedback was given to prevent any influence on the subsequent trials.
Because it can be hard for children to focus on a repetitive task, children were asked some related questions in between blocks, for example: "what is your favorite animal?". This was only done between blocks, so the questions do not influence the performance of the children on the actual experiment. At the end of the experiment, the child was rewarded with a sticker. No comment was made about the child's performance on the experiment. The total duration of the experiment was about 15 minutes.

For the adult participants, the general process of the experiment was different: they were allowed to perform the experiment online, on their own computers. The experiment was run on the secured server of the University of Groningen, and could be accessed as a web page. In the experiment, the participant first had to fill in some information, like age, version number and informed consent number; the last two given to them by the researchers. After this, the participants had to actively agree with our informed consent form before they could proceed to the explanation of the experiment. Once the experiment was explained, the participants got a practice trial, and subsequently the actual experiment started. When the experiment was done, the participants were thanked for their participation on the last screen and the experiment was done.

2.5 Data processing

We saved a comprehensive data set for every participant: the subject number (anonymous), whether or not Dutch is the native language of the participant, the gender, month and year of birth, the age of the participant, the version of the experiment, the number of informed consent, the trial number,

the answers that were given and the time stamp. For every participant, we calculated 2 accuracies. The first was an accuracy for the trials with known animals (18 trials). We would remove the data for participants who scored lower than 80% for this phase. This bar was set in Woodard et al. (2016) and used to keep out data from people who do not have ample knowledge of animals. The second was the accuracy for the unknown animals (6 trials) as well, which we used to test our hypothesis on.

2.6 Statistical analysis

After processing the data, we needed to decide on the statistical tests we needed to perform. We had one variable that was of interest, namely the accuracy on the trials, and 3 independent variables, represented in the experiment by 8 groups. Besides the independent variables, there are 2 factors that can influence the performance: the subject and the trial. We also want to take these factors into account. With both intentional and random effects, we need to use a general linear mixed effects model (Verzani, 2014). In order to perform this test, the data had to meet several requirements, with a normal distribution as the most important one. If these requirements would not be met, we would have to use a non-parametric test. We used these tests to determine whether grammatical gender had a significant influence on word learning.

3 Results

3.1 Children

First, we looked at the answers the children gave in the trials with known animals. In order to keep the results of the actual trials in the data, the children needed to answer at least 80% of these trials correctly. We can conclude that 5 out of 34 participants did not reach 80% accuracy: these participants did not want to finish the experiment. We removed their data and processed the data of the 29 children that did reach an 80% accuracy. In Figure 3.1, we have visualized their results.

The children performed this part of the task skillfully, averaging between 90.7% and 96% accuracy, depending on the version, and 93.9% in total. This was expected, because the animals in

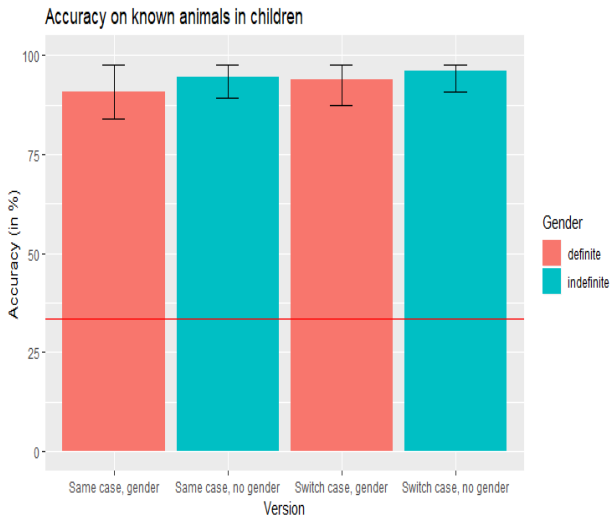

Figure 3.1: Average accuracy for the known animal trials in children

the experiment were chosen to be familiar even for young children. We can conclude that there were no significant differences in animal knowledge between children assigned to either the 'Same' or 'Switch' condition or between children assigned to the conditions with or without grammatical gender.

We then looked at the answers the participants gave in the last trial of each block: the trial in which the target animal re-appears. In Figure 3.2, The accuracy of these trials is shown.

There was quite a large difference between the performance in versions that were in the 'Same' condition compared to performance in versions that were in the 'Switch' condition: the mean performance in the 'Same' condition seemed to be higher (average = 48.8%, SD = 28.1%) than that of the 'Switch' condition (average = 22.2%, SD = 13.6%), although it is hard to say anything conclusive due to the large standard deviations. The difference between the use of a definite and an indefinite article seemed to be negligible: the average performance for the definite and indefinite article in the experiment were 38.1% and 32.2% respectively, with standard deviations of 28.1% and 23.1%. The standard deviations were large, so the chance that the difference in performance would be significant was small. It is clear that the performances for the unknown animals were lower than the performance for the

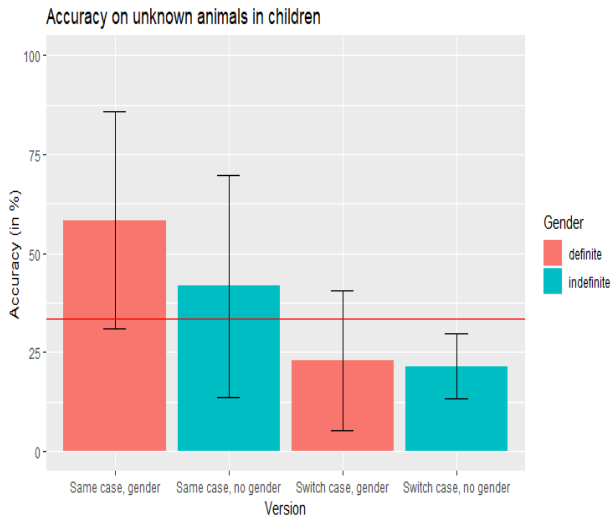

Figure 3.2: Average accuracy for the unknown animal trials in children

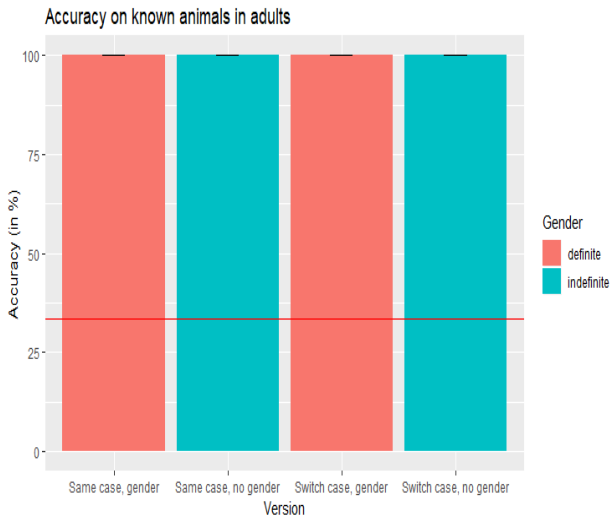

Figure 3.3: Average accuracy for the known animal trials in adults

trials with known animals. The trials with known animals were easier because there was no learning involved and the animals were known, which explains their higher performance.

The figures 3.1 and 3.2 show a large difference in performance for 'Same' and 'Switch', but not for definite and indefinite articles. To see whether these differences were significant we made a linear mixed effects model. This was possible as our data was normally distributed ($p = 0.05384$ for the Shapiro-Wilkinson test). We used the correctness of a trial (whether the answer was correct or not) as the response variable and the 'Same'/'Switch' case and the kind of article that is used as predictor variables. We also added the age of the child as a random effect. The summary of our model can be found in table 3.1. It shows that the only condition that made a significant effect is the difference between the 'Same' and 'Switch' case (p-value = 0.00028), while the effect of the article's gender was not significant, with a p-value of 0.18908. The random effect, the age of the participants, has a variance of 0.3951 with a standard deviation of 0.6285.

3.2 Adults

Secondly, we looked at the results of the adults. In Figure 3.3, we visualized the results of adult participants for the known animals.

The adults all got a 100 per cent accuracy for this part of the experiment. This was expected because the animals used in the experiment are expected to be known by adults. We did not have to exclude any data based on the 80% accuracy mark we set. We then looked at the answers the adult participants gave in the last trial of each block: the trial with the 'Same' or 'Switch' case and the definite or indefinite article. In Figure 3.4, their accuracy is shown.

Figure 3.4 shows that the performances for the 'Same' cases seem to be higher than those for the 'Switch' cases: the average accuracy for the 'Same' cases was 99.1% with a standard deviation of 3.9%, which makes it a near perfect score. the average accuracy for the 'Switch' cases was 87.7%, with a standard deviation of 14.5%. The 'Switch' case has a large standard deviation, which made it hard to say whether this difference was significant.
The average performance when a definite article was used was 96.1%, with a standard deviation of 7.3%. The average performance when an indefinite article was used was 90.8%, with a standard deviation of 14.8%. Again, because of the large standard deviation for the indefinite article, it was hard to determine the significance of this difference.

Table 3.1: Experimental results children: Model = Correct ∼ Same/Switch + Definite/Indefinite + (1|Age)

Predictor	Estimate	Standard error	z value	Pr(>|z|)
Intercept	5.1385	1.1250	4.568	$4.93 * 10^{-6}$ ***
Same/Switch(*Switch*)	-2.6748	1.0782	-2.481	0.0131 *
Definite/Indefinite(*Indefinite*)	-0.5985	0.6570	-0.911	0.3623

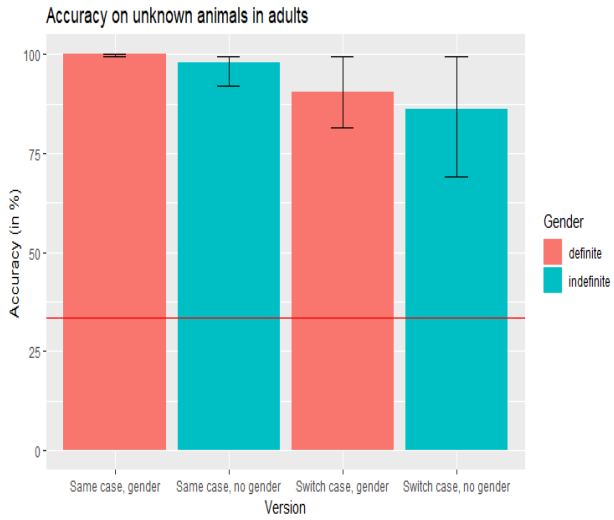

Figure 3.4: Average accuracy for the unknown animal trials in adults

Figure 3.4 shows that there did not seem to be a significant difference between the 'Same' and 'Switch' cases or the use of a definite compared an indefinite article, but we had to examine if these impressions were actually correct. Again, we built a general linear mixed effects model, which was possible because the data was normally distributed (p = 0.5767 for the Shapiro-Wilkinson test). The model had the correctness of the trial as the response variable. The 'Same' and 'Switch' case and the article that was used are the predictor variables, while the random effect was now the trial (from which block the data point came). The results can be found in table 3.2. We can see that there was one significant effect in the model: the group in which the participant was placed ('Same' and 'Switch') did have a significant effect on the performance of the adult participant: a p-value of 0.0125 resulted from of the linear mixed effects model. Participants in the 'Same' cases performed significantly better than those in the 'Switch' cases. Again, the effect of grammatical gender was not significant: this effect produced a p value of 0.3331. The random effect, the trial from which the data was collected, has a variance of 0.08849 and a standard deviation of 0.2975.

3.3 General

We compared the performances of children and adults as well. First, we performed a set of t-tests to compare the accuracy of the groups that we have to the level of chance, which is 33.33% (1 out of 3, because there were 3 pictures). All adult groups had a p-value lower than 0.05, which means they were significantly different from chance. For the children, only 'Switch, no gender' is below 0.05, meaning that only this value is significantly different from chance. There is no value for adults 'Same, no gender'. This is because there is no variation in this data set, which made doing a t-test impossible. Since all accuracies for this group were 100%, we can safely assume that this group performed above chance. The same goes for the other adult groups: their accuracies were all higher than 85%. The value for 'Switch, no gender' in children was under chance, so this value was significantly lower than chance.

It is interesting to see what the effect of age was: did adults perform better than children? We combined all data into one data frame, and added an extra variable to show whether the participant is an adult or not. After running the linear mixed effects model, of which the summary is in table 3.3, we found that the effect of being an adult was certainly significant: $< 2.2 * 10^{-16}$. Because the estimate of this value is positive when it is changed to adult, we can conclude that adults performed significantly

Table 3.2: Experimental results Adults: Model = Correct ∼ Same/Switch + Definite/Indefinite + (1|Trial)

Predictor	Estimate	Standard error	z-value	Pr(>|z|)
Intercept	5.0245	1.0865	4.625	$3.75 * 10^{-6}$ ***
Same/Switch(*Switch*)	-2.6182	1.0477	-2.499	0.0125 *
Definite/Indefinite(*Indefinite*)	-0.5971	0.6169	-0.968	0.3331

Table 3.3: Complete experimental results: Model = Correct ∼ Same/Switch + Definite/Indefinite + Age + (1 | subjectNumber)

Predictor	Estimate	Standard error	z-value	Pr(>|z|)
Intercept	0.3812	0.2960	1.288	0.1978
Same/Switch(*Switch*)	-1.5680	0.3459	-4.533	$5.8 * 10^{-6}$ ***
Definite/Indefinite(*Indefinite*)	-0.5478	0.3193	-1.716	0.0862 .
Age(*Adult*)	3.7315	0.4035	9.248	$<2 * 10^{-16}$ ***

better than children. We again find a significant effect for the 'Same'/'Switch' case ($p = 5.8 * 10^{-6}$) and no significant effect for using a definite or an indefinite article ($p = 0.0862$), although the p-value is getting close. The random effect, subject number, has a variance of 0.1734 with a standard deviation of 0.4164.

3.4 Summary

We only found one significant effect in our data: both children and adults perform significantly better in the 'Same' case than in the 'Switch' case. Our variable of interest, the use of grammatical gender in the articles, did not show to be significant. Furthermore, we can conclude that the adults performed significantly better than the children.

4 Discussion

The research question of this paper was: *Does grammatical gender in articles have an influence in word learning?* We know from previous research that word learning can be facilitated by using the conceptual context in which the word that is learned appeared (Dautriche and Chemla, 2014; Borovsky et al., 2010, 2012) and the fact that grammatical gender is also an informative feature (Lew-Williams and Fernald, 2007; van Heugten and Christophe, 2015). Grammatical gender could be used as a linguistic context and as such facilitate word learning. Our hypothesis was that grammatical gender does help with the word learning of the participants (Dautriche and Chemla, 2014).
To answer our research question, we have modified an experiment performed by Woodard et al. (2016), by giving the participants three instead of two options. The idea was to make the experiment more difficult in order to see a possible effect of grammatical gender.

Based on our experiment, however, we can say that our experiment did not provide us with any evidence to support this hypothesis: there is no significant effect on word learning performance when the article has a grammatical gender compared to when it has none.
A reason for the fact that Borovsky et al. (2010, 2012) found something else could be that our context was not constrained enough: in Dutch, there are only two genders, which means that by using an article with gender we can split the pool of all possible words in at most two subgroups. This is not very constrained, and apparently was not enough to cause a real effect.

When we compare our results to those of the experiment performed by Woodard et al. (2016), we can see that our participants performed better in the 'Same' case than in the 'Switch' case.

However, the difference is that in our research, the performance for both 'Same' cases is not significantly higher than chance for the children. This could be influenced by the small number of participants, but it is interesting to examine this further. From the current results, at least in children, we cannot conclude that our research supports the 'Propose-but-verify' stance that the original paper takes.
The results of the adults are contrasting to the children's results: these are almost at ceiling level (all higher than 85%, one category even reached 100%). These results seem to be more in line with those of Dautriche and Chemla (2014), albeit better, where context did play a role. This could give us more evidence towards the 'Accumulative learning' paradigm, but that is not certain. There is of course the difference that the research (Woodard et al., 2016) did was with children, so follow-up experiments are required, but it is certainly an interesting finding.

4.1 Improvements

In this section, we discuss 5 points of the methodology that could have influenced the results of the experiments. We also give an approach for improvement.

1. The children, especially the younger ones, were occasionally hesitant to participate in the experiment: this could have caused them to perform worse than their abilities. We tried to make the children feel as comfortable as possible, by first joining then in the classroom and by doing the experiment in the presence of a teacher, but this could still be of influence. For future research, we could try to either be present more often in the classroom or to perform the experiment in the classroom instead.

2. We created one practice trial for the experiment, in order to get the participants accustomed to the task. However, a few children were struggling with the task, mostly in the first few trials. It would have been better if we added more practice trials (for example 4) to get the children more accustomed to task.

3. The number of child participants was too low: Due to the large variation in task performance no clear results came out of this experiment. By testing more people, we hope to bring down the variation and get clearer results.

4. The children were tested on a laptop. Because controlling a laptop mousepad is arduous for young children, the children pointed at the picture while we clicked on it. In order to make the task more manageable for children, it would be a good idea to let children perform the experiment on a tablet: they could click on the pictures themselves and because it feels more like an actual game, this could make them feel more comfortable.

5. The performance of the adult participants is at ceiling level: they sometimes reach 100% accuracy. This makes it look like the experiment was too simple for adults. To correct for this, the experiment could be made more difficult by adding more pictures of animals, for example by using the experimental structure used in Dautriche and Chemla (2014): with more filler items in between targets, the task will become more complex and make the performance decrease.

If these points of improvement are solved, one could repeat the experiment and see if different results or significant effects emerge.

4.2 Future research

Besides the alterations mentioned in the improvements section, there are other way in which our research could be expanded in the future.
Our results show a partial replication of the results in Woodard et al. (2016): our participants performed significantly better in the 'Same' case than in the 'Switch' case. However, in our experiment the performance for both 'Same' cases is not significantly higher than chance, and that one of the 'Switch' cases was significantly lower than chance. This is somewhat surprising, it is not in line with the results connected to 'Propose-but-verify', and could be a ground to examine the debate between the 'Propose-but-verify' and 'Accumulative learning' paradigms.

This experiment focused on the influence of grammatical gender on word learning in Dutch.

However, there are more gender systems besides the one used in Dutch, which might have a different influence on word learning. In future research, we could examine word learning and the influence of other grammatical systems on the performance. We should do this in order to allow the generalization of the conclusions drawn from this experiment. It seems valid that systems with more genders will have more constraining power. More constraining power has an effect on learning made-up words (Borovsky et al., 2010, 2012). One example of a follow-up experiment is performing the same word-learning experiment, but with a language like Arapesh: with 13 genders, the context should be made more constraining and create a larger difference between using grammatical gender or not(Haspelmath et al., 2005).

Research can be extended towards different grammatical manifestations as well. In Dutch, the grammatical gender can be determined from the definite article. However, in Swedish, the gender marking for the definite article is represented as a suffix added to the noun (Josefsson, Platzack, and Håkansson, 2013). An example of a follow-up experiment is comparing current research results with results obtained from an equal experiment with Swedish children and adults. According to Schriefers, Jescheniak, and Hantsch (2002), grammatical gender is stored at the lemma level: every noun has their grammatical gender stored as a property. Because grammatical gender is stored in the noun, its manifestation should not matter and the effect should not be any different in Swedish compared to Dutch.

5 Conclusion

Using an observational word learning task, we tried to answer the research question: does the use of grammatical gender in articles help in word learning? We found that for both children and adults, the addition of grammatical gender to articles did not help them learn words better than when grammatical gender was not added. We also found that the results for the child participants did not completely replicate the results found in research by Woodard et al. (2016), which causes our research not to fully support their 'Propose-but-verify' stance. We can conclude that using grammatical gender in articles does not have a significant effect in word learning.

6 Acknowledgements

I would like to thank Daniëlle Metz for helping us recording the sentences in Dutch and Liese Schmidt for helping us with the running a pilot test and with our experimental design. I would also like to thank Jennifer Spenader for the help and support she provided as our supervisor, Sieta van Pelt for inviting us to her daycare centre, all our participants (children and their parents and the adult participants) who helped us out and Jaap Uithof for proofreading my paper.

References

P. Bloom. *How Children Learn the Meanings of Words.* Boston, MA: The MIT press., 2000.

Y. Boloh and L. Ibernon. Natural gender, phonological cues and the default grammatical gender in french children. *First language*, 33:449–468, 2013.

A. Borovsky, M. Kutas, and J. Elman. Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, 116: 289–296, 2010.

A. Borovsky, J. Elman, and M. Kutas. Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8:278–302, 2012.

G. G. Corbett. *Gender.* Cambridge, UK: Cambridge University Press., 1991.

I. Dautriche and E. Chemla. Cross-situational word learning in the right situations. *Journal for experimental psychology. Learning, memory and cognition.*, 40:892–903, 2014.

R.A. Depaolis, M. M. Vihman, and T. Keren-Portnoy. When do infants begin recognizing familiar words in sentences? *Journal of Child Language*, 41:226–239, 2014.

Feldarkrealms. Fake word generator. https://www.feldarkrealms.com/words/.W-rDP5NKg2w, Unknown Year. Accessed: 2018-10-02.

E. Gómez Milán, O. Iborra, M.J. de Córdoba, V. Juárez-Ramos, M.A. Rodríguez Artacho, and J.L. Rubio. The kiki-bouba effect: A case of personification and ideaesthesia. *Journal of Consciousness Studies*, 20:84–102, 2011.

M. Haspelmath, M.S. Dryer, D. Gil, and B. Comrie. *The World Atlas of Language Structures.* Oxford, UK: Oxford University Press., 2005.

M.H. Ibrahim. *Grammatical Gender : Its Origin and Development.* Den Haag, NL: Mouton & Co., 1973.

G. Josefsson, C. Platzack, and G. Håkansson. *The Acquisition of Swedish Grammar.* Amsterdam, NL/Philadelphia, PA: John Benjamins Publishing Company, 2013.

K. Lemhfer, . Schriefers, and I. Hanique. Native language effects in learning second-language grammatical gender: A training study. *Acta Psychologica*, 135:150–158, 2010.

C. Lew-Williams and A. Fernald. Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18:193–198, 2007.

M.P. Maratsos. *The use of definite and indefinite reference in young children: An experimental study of semantic acquisition.* Cambridge: Cambridge University Press., 1976.

D. Mazzoni and R. Dannenberg. Audacity program, 1999.

H. Schriefers, J.D. Jescheniak, and A. Hantsch. Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28:941–950, 2002.

J. C. Trueswell, T. M. Medina, A. Hafri, and L.R. Gleitman. Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66:126–156, 2013.

S. Unsworth. Age and input in the acquisition of grammatical gender in dutch. *Second Language Research*, 24:365–295, 2007.

M. van Heugten and A. Christophe. Infants acquisition of grammatical gender dependencies. *Infancy*, 20:675–683, 2015.

J. Verzani. *Using R for introductory statistics.* Boca Raton, FL: CRC Press., 2014.

K. Woodard, L. R. Gleitman, and J. C. Trueswell. Two- and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development*, 12: 252–261, 2016.