



Universiteit Utrecht

The Interpretation of Mass Collectives by Dutch Adults

An analysis of native Dutch-speaking adults' interpretations of mass collectives and the influence of contextual information.

Hannah Aalbers

Student number: 4114728

Research Master Linguistics

Utrecht University

Supervisor: Dr. Luisa Meroni

Second reader: Prof. Dr. Peter Coopmans

June, 2021

Acknowledgements

Throughout the process of writing this thesis, I have received a very valuable amount of support and assistance. First of all, I would like to thank my supervisor, Dr. Luisa Meroni, who has supported me through the whole thesis process. Even when unexpected difficulties emerged as a result of the COVID-19 pandemic, she remained available through video calls to support me and help me adjust the research project as needed because of COVID-19 measures. Luisa's expertise and feedback pushed me to continuously improve myself, and video meetings with her often gave me a burst of renewed motivation and self-confidence. A special thanks to Prof. Dr. Peter Coopmans for being the second reader.

Next, I am grateful to all participants that filled out the questionnaire. As I originally aimed for around 50 participants, I was overwhelmed by the enthusiasm of those around me to spread the questionnaire among their friends and family. Because of you, I reached a surprising amount of 100 questionnaire replies within one day.

I would also like to thank the friends that proofread chapters of my thesis and provided feedback, and the friends that always supported me through thesis-related worries, stress, and fears. In no particular order, thank you my dear friends: Marjolein Talsma, Leonie Barbas-Weil, Maaïke de Jager, Godelieve Voorhoeve, Rémy Jodrey, Annemarie Boonstra, and Tessa Groen. A special shout-out to Chiara Staal for being such a great support in figuring out everything related to statistics. On top of this, my parents and sister provided much needed space for me to relax and unwind with the joys of food and game nights. Dear friends and family, I could not have done this without you!

Finally, I would also like to mention Utrecht University. I could not have finished my thesis without access to a study spot in the library and am grateful that the university worked so hard to offer support to students during the COVID-19 pandemic.

Table of Contents

Acknowledgements.....	2
Table of Contents.....	3
Abstract.....	4
1. Introduction.....	5
1.1 Mass and Count.....	5
1.2 Mass Collectives: Introduction to the Present Study	7
2. Theoretical Background.....	11
2.1 Mass Collectives	11
2.2 Barner and Snedeker	12
2.3 Huang and Meroni.....	15
2.4 Mass Collectives in Dutch.....	18
3. Methodology	21
3.1 Individual-Oriented Context Items.....	24
3.2 Substance-Oriented Context Items.....	25
4. Results.....	27
4.1 Data Analysis	28
5. Discussion.....	31
5.1 Data Interpretation.....	32
5.2 Discussion of Individual Test Items.....	33
5.3 Discussion of Individual Participants.....	36
5.4 Discussion of the Experimental Design	37
6. Concluding Remarks and Future Research.....	39
References.....	41
Appendix 1: Original Questionnaire in Dutch.....	43
Appendix 2: Translated Questionnaire in English.....	51
Appendix 3: Data Recoded	59

Abstract

The aim of this study is to add to the existing body of research on mass collectives (collectives hereafter). These collectives, such as the English word *furniture*, are argued to have mass syntax, while semantically allowing a mass reading, an individual reading, or both. Recent studies show that collectives are individual-denoting in a wide range of typologically distinct languages such as English (Barner & Snedeker, 2005), Mandarin Chinese (Lin & Schaeffer, 2018), Hebrew (Hacohen, 2008), and Dutch (Van Witteloostuijn & Schaeffer, 2018). However, these conclusions are questioned by further research based on methodological issues. Huang and Meroni (to appear) suggest that the individual-denoting reading is the most accessible reading of collectives, but not the only possible one. Their research on collectives in Mandarin Chinese shows that manipulating the context and morpho-syntactic information allows participants to give both the individual-denoting and the substance-denoting reading. In the present study, using an online version of the Truth Value Judgment Task, I extend this line of research to a typologically different language, namely Dutch. I investigate if a substance-oriented context can elicit the substance-denoting reading of collectives in native Dutch speaking adults in the same way it does for Mandarin speaking adults and children (Huang & Meroni, to appear). Participants were asked to judge both collectives and count nouns in both substance- and individual-oriented contexts. Results indicate that the substance-oriented context did not elicit a substance-denoting reading of collectives; the difference between the interpretation of collectives in both contexts was not significant. Several potential explanations for these results are discussed as it could be possible that the online setting of the experiment and/or the collectives used for test items made the substance-denoting reading less accessible. Another explanation could be that the mass-count distinction is stronger or different in Mandarin Chinese because of the typological differences between Dutch and Mandarin Chinese. Further research on a language typologically similar to Dutch could shed more light on the results of this experiment.

1. Introduction

Most people are at least vaguely familiar with the concept of mass and count nouns and everyone applies the concept of mass and count every day, whether they are aware of it or not. The difference between mass and count nouns is an important backbone of this thesis, which will ultimately focus on the topic of mass collectives and their syntactic and semantic properties. The background of the current thesis is shaped by previous research by Barner and Snedeker (2005) on the syntax and semantics of mass collectives, its replications in other languages (Lin & Schaeffer, 2018; Inagaki & Barner, 2009; Van Witteloostuijn & Schaeffer, 2018; among others), and the subsequent critical analysis by Huang and Meroni (to appear). Before discussing mass collectives and previous research, I will discuss the differences between mass and count nouns as this difference is the key in understanding the research on mass collectives.

1.1 Mass and Count

The distinction between mass nouns and count nouns in English has clear morpho-syntactic criteria, which have been thoroughly described by Gillon (1999). Gillon discussed seven criteria that distinguish mass nouns from count nouns. The first criterion is about the presence or absence of, or contrast between, singular and plural form; mass nouns do not have a plural form, while count nouns do. For example, the mass noun *advice* cannot be pluralized (**advices*) whereas the count noun *suggestion* has the plural form *suggestions*. The second criterion is based on whether the noun can be modified by cardinal numerals: for count nouns this is grammatical (*three suggestions*), but for mass nouns it is not (**three advice*). The third criterion distinguishes mass nouns from count nouns by using quasi-numerals as modifiers. Again, this is grammatical for count nouns (*several suggestions*) but not for mass nouns (**several advice*). The fourth criterion distinguishes mass nouns from count nouns by

determining whether the noun can be modified by an indefinite article. This is ungrammatical for mass nouns as it is ungrammatical to say that ‘*someone gave you *an advice*’, while it is grammatical to receive a suggestion from someone: ‘*someone gave you a suggestion*’. The fifth criterion utilizes the *one*-antecedent as a test to distinguish between mass and count nouns. Again, this is grammatical for count nouns but not for mass nouns, as can be seen in (1).

(1) Mass: Marjolein gave Hannah advice and Luisa gave her *one too.

Count: Marjolein gave Hannah a suggestion and Luisa gave her one too.

The sixth criterion states that count nouns can be modified by *many/few*, while mass nouns cannot (*many/few suggestions*; **many/few advice*). The seventh and last criterion looks at both noun classes being modified by *much/less*; this is the only criterion that is grammatical for mass nouns but ungrammatical for count nouns (*much/less advice*; **much/less suggestion(s)*). An overview of all criteria with examples can be found in Table 1.

Table 1

The Seven Morpho-Syntactic Criteria for Distinguishing Mass and Count Nouns from Gillon (1999) with Examples.

Morpho-syntactic criteria	Mass noun example	Count noun example
Singular / plural contrast	Advice - *advices	Suggestion - suggestions
Modified by cardinal numerals	*Three advice	Three suggestions
Modified by quasi-cardinal numerals	*Several advice	Several suggestions
Modified by indefinite article	*An advice	A suggestion
<i>One</i> antecedent	Marjolein gave Hannah advice and Luisa gave her *one too.	Marjolein gave Hannah a suggestion and Luisa gave her one too.
Modified by many/few	*Many/few advice	Many/few suggestions
Modified by much/less	Much/less advice	*Much/less suggestion(s)

1.2 Mass Collectives: Introduction to the Present Study

With this information on mass and count nouns as a foundation, I will now discuss mass collectives. The definition of mass collection used in this thesis states that collectives are “high-level categories in natural kind and artifact taxonomies, containing perceptually diverse members” (Wisniewski, Imai, & Casey, 1996, p. 270). Following from this definition, there is a difference between mass collectives such as *furniture* and count collectives such as *animal*. The difference between mass and count has been explained above and applies to this distinction between mass collectives and count collectives as well: *furniture* can be modified by ‘much’ and ‘less’, whereas *animal* cannot be preceded by these modifiers (*much furniture*, **much animal*). On top of that, count collectives have both singular and plural forms, whereas mass collectives do not. The difference between mass collectives and count collectives only applies to number marking languages such as English and not to classifier languages such as Mandarin

Chinese, in which the count-mass status of nouns is not marked by a grammatical marker (Huang & Meroni, to appear).

According to Gillon (1996), the difference between mass collectives and count collectives is that mass collectives (e.g. *furniture* and *silverware*) clearly denote ‘minimal parts’ or ‘atoms’. This distinction also highlights the difference between typical mass nouns such as *water* and mass collectives such as *furniture*. Mass nouns and mass collectives are syntactically similar as both have mass syntax, but collectives are semantically different as they allow both mass (substance) and count (individual) readings. Mass nouns such as *water* or *wine* do not have a linguistically specified minimal part, whereas mass collectives always have these; the collective *furniture* can be separated into pieces of furniture, such as chairs and tables (Nicolas, 2008). This shows how mass collectives have different semantics from typical mass nouns, even though both mass nouns and mass collectives differ from count nouns because of their mass syntax.

Previous research, which will be more thoroughly discussed in Chapter 2, has focused on the possible readings that people can assign to mass collectives (collectives hereafter). Barner and Snedeker (2005) conducted multiple experiments to test the interpretation of collectives compared to other noun classes. The results show that English-speaking adults and children quantify collectives by cardinality. This means that, in English, collectives are assigned an individual-denoting interpretation, and are thus semantically interpreted as count nouns. However, Huang and Meroni (to appear) have suggested that these results may simply mean that the individual-denoting interpretation is the most accessible interpretation, but not the only interpretation people can access. They suggest that providing specific contexts could make participants able to also access the substance-denoting interpretation. The experiments they have conducted with Mandarin Chinese speaking adults and children confirm this hypothesis: participants can access the substance-denoting reading of collectives if the relevant

contextual information is provided. In other words, contrary to previous research, Huang and Meroni claim that collectives allow both the individual-denoting and the substance-denoting reading, and the context in which the collectives are presented is crucial. In the present experiment I would like to extend such research to a language typologically different from Mandarin Chinese to investigate whether these results hold true in a language that has different means to mark mass and count. Due to the COVID-19 pandemic, the research unfortunately had to be altered in two ways; first it became impossible to test children in schools, which led to the second change; the decision to use an online questionnaire to test adults. Because of these changes to the circumstances, the research question was as follows: *Can a substance-oriented context elicit the substance-denoting interpretation of mass collectives in native Dutch speaking adults?* As all previous studies tested both adults and children, the results from the present study can be compared to the previous results from adult participants and add valuable insights to the existing body of research on the properties of collectives.

The prediction is that the independent variable ‘context’ will have an effect on whether the participants assign the substance-denoting reading or the individual-denoting reading to collectives. Thus, the expectation is that a substance-oriented context will more often result in a substance-denoting interpretation of collectives when compared to the results in an individual-oriented context. Participants will also judge count nouns in both contexts; this is done to compare the effect of the context on the interpretation of collectives to the effect that context has on the interpretation of count nouns. As count nouns have count syntax and count semantics, it is expected that participants assign the individual-denoting reading, even when the count noun is presented in a substance-oriented context.

In Chapter 2, I discuss previous literature on collectives and the properties of collectives in Dutch. Chapter 3 consists of a discussion of the methodology of the current experiment, followed by Chapter 4 in which the results are reported. In the discussion in Chapter 5, I review

the experiment and discuss the findings. Finally, Chapter 6 consists of the conclusion that will complete this thesis.

2. Theoretical Background

This chapter introduces previous studies on the topic of collectives, mainly focusing on the study by Barner and Snedeker (2005) and the study by Huang and Meroni (to appear). The study by Barner and Snedeker had a great impact on studies of collectives and has been replicated in different languages (Lin & Schaeffer, 2018; Inagaki & Barner, 2009; Van Witteloostuijn & Schaeffer, 2018). After discussing these key studies, I will give an account of collectives in Dutch.

2.1 Mass Collectives

Collectives are nouns that are syntactically singular but refer to a collection of items. *Furniture* is the most common example of a collective and previous research has shown that collectives appear to be involved in a syntax-semantics mismatch (Huang & Meroni, to appear). As discussed in Chapter 1, collectives such as *furniture* behave syntactically like mass nouns, following the criteria as defined by Gillon (1999): *furniture* has no plural form (**furnitures*) and it cannot be modified by cardinal numerals, quasi-cardinal numerals, indefinite articles, or quantifiers such as *many/few* (**three furniture; *several furniture; *an furniture; *many/few furniture*), while it can in fact be modified by *much/less* (*much/less furniture*). Moreover, collectives cannot undergo *one*-anaphora (*Mary bought furniture and John bought *one too*).

Semantically, however, it has been argued that collectives can be either mass (substance-denoting) or count (individual-denoting) or allow both readings despite their mass syntax. This syntax-semantics mismatch was reaffirmed when Barner and Snedeker (2005) found that English-speaking adults and children quantify collectives by cardinality. This means that, in English, collectives denote individuals and thus have count semantics, despite their mass syntax. The research by Barner and Snedeker has since been replicated in a wide array of languages, in which all results showed that the mass collectives predominantly have count

semantics and thus an individual-denoting reading (Bale & Barner, 2009; Inagaki & Barner, 2009; Lin & Schaeffer, 2018; MacDonald & Carrol, 2018; Van Witteloostuijn & Schaeffer, 2018).

Based on the research methods used in all previous research, Huang and Meroni (to appear) suggest another interpretation. The broadly provided evidence towards the count semantics of collectives could show the preferred reading of collectives, meaning that when the context is manipulated, another reading could become available. They proceeded to analyze the experimental design that was used by Barner and Snedeker (2005) and the experiments following up on them, as well as explore other possible factors that could have influenced the interpretation of collectives. Huang and Meroni (to appear) performed three experiments in Mandarin Chinese and their research and results are the basis for the current research project.

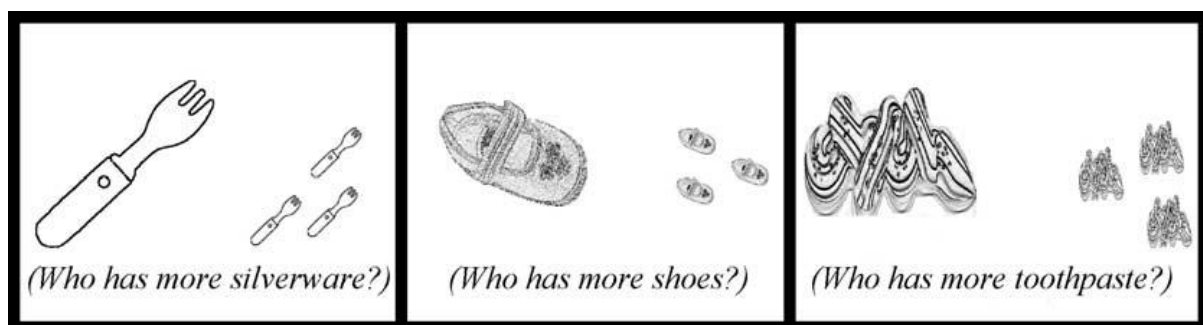
2.2 Barner and Snedeker

Historically, it was believed that collectives “capture knowledge about an unindividuated group of objects” (Wisniewski, Imai, & Casey, 1996, p. 292). This stance was supported by, amongst others, Quine (1960) and Link (1998); based on their analyses, collectives should not be quantified over individuals but instead should be interpreted based on their mass syntax. Barner and Snedeker (2005) further explored the semantics of the mass-count distinction in both young children and adults. Their experiments show that some mass nouns can actually refer to individuals, which partially refutes the previous positions mentioned above (Quine, 1960; Link, 1998; Wisniewski, Imai, & Casey, 1996). Barner and Snedeker (2005) conducted three experiments to investigate the interpretation of collectives (e.g. *furniture, silverware*) and compare it to the interpretation of count object nouns (e.g. *cups, shoes*) and mass nouns (e.g. *sand, toothpaste*) in English speaking adults and 4-year-old children.

Barner and Snedeker (2005) used a question-answering Question Judgment Task (QJT) and tested children (age 4;0–4;6, M=4;3) and adults (Harvard University undergrad students). Participants of the first experiment were shown two characters and asked who had more of the target noun (i.e. ‘*who has more X?*’, where X can be a mass noun, a collective, or a count noun). Adult participants would judge pictures whereas children would see the actual scenes played out by the experimenters. Figure 1 shows images of selected stimuli from Experiment 1 by Barner and Snedeker. If participants interpret collectives such as *silverware* as denoting individuals, they will say the character with more small items has more cutlery. In contrast, if participants interpret collectives such as *silverware* as denoting substance, they will say the character with one big item has more cutlery as the volume is bigger. In all experimental items, the overall volume of the large objects was larger than the overall volume of the number of small items of the same kind.

Figure 1

Images of selected stimuli from Experiment 1: object-mass: silverware; count: shoes; substance-mass: toothpaste (source: Barner & Snedeker, 2005, p. 50).



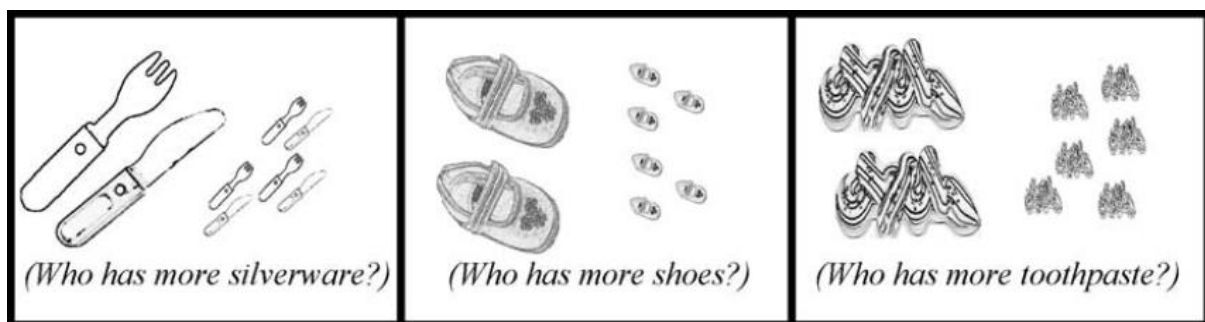
As mentioned before, the participants were asked to compare amounts of items as shown in Figure 1. If collectives such as *silverware* do not quantify over individuals, the participants should choose the larger volume on the left over the larger amount on the right. However, if collectives denote individuals, participants should choose based on the number of pieces of silverware and not on the overall volume of the silverware pieces. The third possibility

would be that adults interpret collectives as quantifying over individuals but that children do not. This would imply that children start acquisition with a semantics-to-syntax mapping that will develop into an individual-denoting interpretation later in development.

The results of the first experiment showed that the quantity judgments of adult participants are significantly more based on the number individuals for count nouns (100%) and collectives (97%) than for mass nouns (0%). These results suggest that collectives are interpreted as individual-denoting. Barner and Snedeker discussed several objections to their first experiment, for which they controlled with their second and third experiment. One of the objections to their first experiment was about the items/images used; if *silverware* is depicted by forks only, it is possible that participants replace the experimenters' question with their own paraphrase (e.g. '*who has more silverware?*' gets glossed as '*who has more forks?*'). This is potentially problematic as the word *fork* is a count noun. Figure 2 shows the adjusted images that were used in experiments 2 and 3.

Figure 2

Images of selected stimuli from Experiment 2 and 3: object-mass: *silverware*; count: *shoes*; substance-mass: *toothpaste* (source: Barner & Snedeker, 2005, p. 53).



Despite controlling for the factors that could have biased their first experiment, the results did not change in the second and third experiment. The results of all three experiments show that both 4-year-old children and adults interpreted the collectives as quantifying via cardinality (denoting individuals): the character with multiple small objects was judged to have

more than the character with one or two big objects as judged by 97.9% of the adults and 91.7% of the children (Barner & Snedeker, 2005). Quantifying over cardinality was also used when the participants were asked to interpret object nouns such as *shoes* (93.8% in adults, 97.9% in children; Barner & Snedeker, 2005). However, the results in the condition where participants were asked to interpret substance nouns (e.g. *toothpaste*) were different: the quantity judgement in this condition was made based on volume rather than cardinality (judgement by cardinality was 0% for adults; 39.6% for children). This means that the character with one big pile of toothpaste was judged to have *more* than the character with multiple small piles of toothpaste. These results show that participants interpreted collectives similarly to count nouns, and significantly different from their quantity judgement in the mass noun condition.

The results from Barner and Snedeker (2005) have widely been taken as evidence for the argument that collectives have count semantics in English, regardless of their mass syntax, and that collectives are thus interpreted as denoting individuals. Similar results were found in subsequent studies that used the same methodology and experimental design in a wide array of typologically different languages, such as Mandarin (Lin & Schaeffer, 2018), Japanese (Inagaki & Barner, 2009), Dutch (Van Witteloostuijn & Schaeffer, 2018), and others.

2.3 Huang and Meroni

After analyzing the study by Barner and Snedeker (2005) and the studies that followed, Huang and Meroni (to appear, p. 2) hypothesize that “the widely attested individual-denoting (count) reading may be a preferred reading in neutral contexts, and not the only reading that mass/bare collectives can have”. Huang and Meroni’s motivation for this hypothesis is twofold and based on Barner and Snedeker’s (2005) use of the question-answering QJT, the method that was adopted by all following studies as well. This method consists of presenting the participants with a situation – whether it is acted out or shown in images – followed by a

question phrased as “Who/what has more of item [X]?”. The participant then points to or names one of the characters. The first concern with this method is that the question-answering QJT demands that participants choose one character over the other in answering the question “Who has more ___?”. When presented with a question like this, the participants are forced to choose one reading or the other, while no context was provided to support the substance-denoting reading. This means that the results can be taken to indicate a *strong preference* for one reading, but it does not rule out the possibility that the other reading is possible as well, albeit less favorable. In this situation, the less favored reading may still be part of the participants grammar, but it does not have a chance to emerge because the context to activate this reading has not been provided.

The second concern is with the contexts used by Barner and Snedeker (2005) and all studies that followed from this. While the information about the volume of the items (e.g. furniture) is provided to the participants, the contexts are neutral as they do not provide an explanation about why the volume matters. Making the importance of volume salient is relevant as humans are inclined to focus on shape above other characteristics. This is the shape bias: the language acquisition strategy that categorizes novel objects based on shape rather than based on other characteristics (Diesendruck, Markson, & Bloom, 2003; Potrzeba, Fein, & Naigles, 2015). While the shape bias becomes weaker from age 5 onwards, adults still utilize the same strategy (Landau, Smith, & Jones, 1992). Because of the shape bias active in the interpretation of nominal expressions, participants may base their quantity judgment on shapes, which leads to a judgement based on cardinality (Landau, Smith, & Jones, 1988; Huang & Meroni, to appear). To override this bias, Huang and Meroni decided to provide salient contexts that highlight non-cardinal dimensions as measurement (e.g. volume). By doing so, they are using context to make the substance-denoting reading relevant to investigate whether the

substance-denoting reading can be accessed in specific situations, even though the individual-denoting reading may be dominant in neutral contexts.

To test the hypothesis that both the substance-denoting and the individual-denoting reading are possible, Huang and Meroni performed three experiments in Mandarin Chinese. They not only used the Quantity Judgment Task as used by Barner and Snedeker (2005), but also adapted a Truth Value Judgment Task. By doing so, participants were not only asked “Who has more X?”, but were also asked to judge statements such as “A has more furniture than B”. Including these statements, which participants had to judge as ‘true’ or ‘false’, gave the participants more opportunity to assign the substance-denoting reading instead of just the individual-denoting reading which is more prominent in neutral contexts.

Another crucial aspect of Huang and Meroni’s approach is their consideration of both morpho-syntax, such as the absence or presence of a classifier, and contextual information. The morpho-syntactic aspects may reveal more information about the functioning of collectives in classifier languages such as Mandarin Chinese. Examples (2) and (3), taken from Huang and Meroni (to appear), show a sentence with and without a classifier. In this example, when the Mandarin word *jiaju* (furniture) is combined with the individual classifier *ge*, only individual denoting readings are allowed.

(2) *Qingwa yaoguai chi le gengduo jiaju*

Frog Monster eat Asp more furniture

‘Frog Monster ate more furniture.’

Huang & Meroni, to appear, p. 14

(3) *Qingwa yaoguai chi le gengduo ge jiaju*

Frog Monster eat Asp more CL_{ge} furniture

‘Frog Monster ate more pieces of furniture.’

Huang & Meroni, to appear, p. 14

The results found by Huang and Meroni show that both contextual information and morpho-syntax are relevant to the interpretation of Mandarin Chinese collectives. According to Huang and Meroni (to appear), these results suggest that the existence of ‘ground stuff’ (elements that have been put through the grinding machine) does not necessarily have a direct relation to the assignment of the substance-denoting interpretation. The authors argue that the substance-denoting reading is instead a basic lexical meaning of bare collectives, rather than an interpretation that has been triggered by the context of the grinding machine. Bare collectives only exist in classifier languages; in number marking languages, the collectives cannot be used without the syntactic markers for mass/count status.

These results are used to generalize the findings in this research project to the following conclusion: in both adult and child grammar in Mandarin Chinese, both the presence or absence of an individual classifier and the non-linguistic contextual information provided by the stories have an important role in the interpretation of Mandarin collectives. The results found by Huang and Meroni (to appear) confirm that the individual-denoting interpretation of collectives is indeed the preferred interpretation in neutral contexts, but when the context is manipulated, the substance-denoting interpretation can be triggered. The results of their second experiment show that both adults and children only assigned the individual-denoting reading when the individual classifier *ge* was used, regardless of the context in which the sentence was presented. This confirms the hypothesis that collectives do not *just* have the individual-denoting reading available; at least in Mandarin, both interpretations are available depending on the context and morpho-syntax.

2.4 Mass Collectives in Dutch

In Dutch, the distinction between mass and count nouns is determined the same way as in English (as shown in Chapter 1), because both Dutch and English are number marking

languages. Summarizing, this means that the distinction between mass nouns and count nouns in Dutch has clear morpho-syntactic criteria. These criteria were discussed in Chapter 1 and can be found in Table 1. Following from this, the distinction between count collectives and mass collectives applies to Dutch as well. An example of a Dutch collective is *bestek* (cutlery). This collective can for example be modified by *veel* (much) and *weinig* (little), but not by *een* (a(n)), as shown in (4). An example of a Dutch count collective is *kudde* (herd), which cannot be modified by *veel* (much) or *weinig* (little), but can be modified by *een* (a(n)), as can be seen in (5). These examples show that the Dutch mass and count collectives follow the same morpho-syntactic criteria as proposed by Gillon (1999) for English.

(4) A. *ik heb veel bestek.*

I have much cutlery

‘I have a lot of cutlery.’

B. **ik heb een bestek.*

*I have a cutlery

*‘I have a cutlery.’

(5) A. **er staat veel kudde in de wei.*

*there stands much herd in the meadow

*‘there is much herd in the meadow.’

B. *er staat een kudde in de wei.*

there stands a herd in the meadow

‘there is a herd in the meadow.’

It is important to mention that, while there are many similarities between Dutch and English with regard to collectives, nouns that are collectives in English, are not necessarily

collectives in Dutch as well. For example, the famous English collective *furniture*, would translate to *meubilair* in Dutch. However, the word *meubel* (piece of furniture) is also very common and to complicate matters further, both the plural form with *-en* and the plural form with *-s* are grammatical (*meubelen* and *meubels*). While *meubilair* is a collective, *meubels* and *meubelen* are count nouns. If one were to use images of furniture to test Dutch interpretations of collectives, one should be cautious that participants could apply the unintended *meubel* to the image, interpreting it as a regular count noun.

For the purpose of this study, it is assumed that both the distinction between mass and count, and collectives in Dutch function similar as in English. However, this does not mean that words that are a collective in English, are automatically a collective in Dutch as well. The motivation for the collectives used in the present study will be discussed in Chapter 3, in which I discuss the methodology of the experiment.

3. Methodology

This study aims to replicate the study by Huang and Meroni (to appear) on Mandarin Chinese and thus planned on using the same methodology: the Truth Value Judgement Task (TVJT; Crain & Thornton, 1998) to test children and a control group of adults. Originally, the experiment was planned to be a replica of the experimental design by Huang and Meroni (to appear), meaning that the goal was to test both native Dutch children and adults. Unfortunately, the COVID-19 pandemic made it impossible to test children at primary schools and testing adults in a face-to-face setting became impossible as well. I will explain how the test materials were adapted for an online questionnaire aimed at native Dutch adults, and what the current design of the study is.

In the present study, native Dutch speaking adults have been tested with an online picture version of the Truth Value Judgement Task (TVJT). Due to an error in the program used for the questionnaire, the question asking participants for their specific age was removed, something I only realized after the questionnaire went online. Because of this, there are no data about the participant's ages; we only know that they were all adults, meaning that they were all at least 18 years old. Participants were asked whether Dutch is (one of) their native language(s); if the participant answered that Dutch was *not* one of their native languages, the questionnaire would end. This ensured that only people with Dutch as a native language participated in the questionnaire. Participants were not asked for other demographic information such as gender, because the only requirement for participation was for to be an adult with Dutch as a native language. Participants were found by posting the link to the questionnaire on Facebook and asking people to spread the link to friends, acquaintances, and family.

The online TVJT was conducted through an online questionnaire made with Qualtrics software, Version January 2021 of Qualtrics (Qualtrics, Utah, USA). The participants were

tested on their interpretation of (1) collectives and (2) count nouns, hereafter called collective stimuli and count stimuli. The stimuli can be seen in Table 2. As this questionnaire was originally designed to test children, I deliberately chose words that I assumed all children would know. The count stimuli all have regular plural forms with *-en* as endings. As for the collective stimuli, all of them syntactically behave like mass nouns, and are commonly used words in Dutch. As previously mentioned, there is a potential for confusion around the Dutch word for furniture (*meubilair*). I made sure that the collectives selected for this experiment did not have the potential for such confusions that could negatively impact the experiment. All participants saw the test items in the same order; the order of the questions was not randomized. This decision was made because I wanted to ensure that participants would see a mix of items, rather than for example seeing three collective items in a row. A deliberate mixed order of the test items, with filler questions distributed evenly, was therefore considered the best option.

Table 2

Overview of Collective Stimuli and Count Stimuli Used in the Experiment

Collective Stimuli	Count Stimuli
<i>Bestek</i> (cutlery)	<i>Schoenen</i> (shoes)
<i>Beleg</i> (sandwich toppings)	<i>Bloemen</i> (flowers)
<i>Kleding</i> (clothing)	<i>Wielen</i> (wheels)
<i>Post</i> (mail)	<i>Kaarsen</i> (candles)

Test items were manipulated on the independent variable ‘context’ to test the effects of contextual information on noun interpretation. This means that half of the items in the experiment were presented in an individual-oriented context, and half of the items were presented in a substance-oriented context, which followed the experimental design of Huang and Meroni (to appear). As collective stimuli and count stimuli were both tested in the two different contexts, this created the following four conditions:

- Collective stimuli in individual-oriented context
- Collective stimuli in substance-oriented context
- Count stimuli in individual-oriented context
- Count stimuli in substance-oriented context

For all questions, the participants listened to a voice recording and looked at an image. The voice recording created the context, by telling the participants a very short story. While listening to the story, the participants could already see the image. After listening to the recording and looking at the image, the participants were then asked to judge a true/false statement. The questionnaire consisted of 12 questions: two for each condition and four filler questions. The filler questions targeted principle A and B of the Binding Theory (Büring, 2005) and included a voice recording and an image as well, so as to not diverge from the other questions. Table 3 shows which question in the questionnaire had which function. The original material (in Dutch) can be found in Appendix 1; an English translation is added in Appendix 2. The voice recordings could not be included for obvious reasons. Because of this, transcriptions of the voice recordings were added to the Appendices.

Table 3

Distribution of Questions amongst Test Conditions

Condition	Question Number
Collective stimuli in individual-oriented context	Questions 6 and 9
Collective stimuli in substance-oriented context	Questions 1 and 10
Count stimuli in individual-oriented context	Questions 3 and 12
Count stimuli in substance-oriented context	Questions 4 and 7
Filler questions	Questions 2, 5, 8, and 11

3.1 Individual-Oriented Context Items

The individual-oriented context was shaped by stories about two fairies (the red fairy and the blue fairy) participating in a magic contest. For each question, the question was which fairy conjured the most of an item. One of the individual context items is the word *schoenen* (shoes), a countable noun. An image is shown with the red fairy having conjured four small shoes and the blue fairy having conjured two big shoes, where the overall surface size of the two big shoes is visibly larger than the overall surface size of the four small shoes as can be seen in Figure 3. The participants were then presented with the statement “*De rode fee heeft meer schoenen gemaakt*” (“The red fairy made more shoes”). As ‘shoes’ is a count noun, it is expected that participants will judge this statement to be true: four shoes is more than two shoes. If participants judge amount based on overall volume, which is the substance-denoting reading, they will judge the blue fairy to have ‘more shoe(s)’ instead.

Figure 3

The Individual-Oriented Context as Used in the Experiment; Showing the Countable Test Item Schoenen (Shoes).

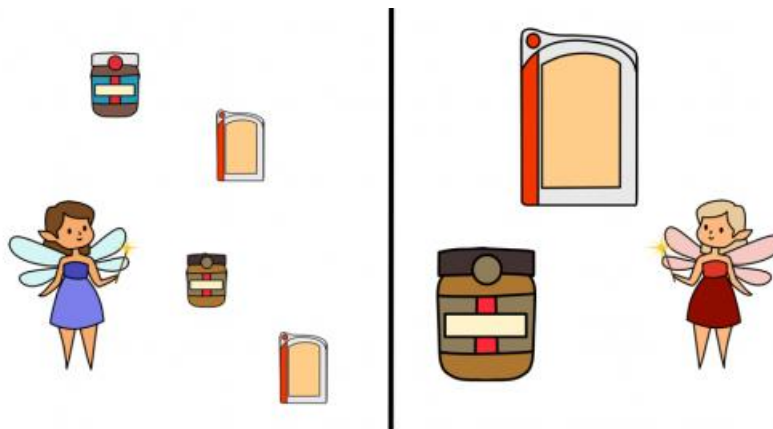


The individual-oriented context was executed exactly the same way with collective stimuli, as can be seen in Figure 4 with the collective *beleg* (sandwich toppings). As ‘sandwich toppings’ is a collective in Dutch, but it is here presented in an individual-oriented context, it

is expected that the substance-denoting reading is not being triggered. Because of this, it is expected that participants will give the individual-denoting reading.

Figure 4

The Individual-Oriented Context as Used in the Experiment; Showing the Collective Test Item Beleg (Sandwich Toppings).



3.2 Substance-Oriented Context Items

The substance-oriented context was shaped by stories about two characters (the fox and the turtle). The story is that the fox and the turtle cannot eat solid food because they have no teeth, so they use a ‘grinding machine’ to grind items into an edible substance. On the images shown, the participants saw how many items each character had and what size they were, as well as a doughy ball of substance after the items have been put through the ‘grinding machine’. Again, the participants were presented with a statement for them to judge as true or false. For the question targeting the count noun *bloemen* (flowers), this resulted in the image as seen in Figure 5, and the participants were asked to judge the following statement as true or false: “*Vos heeft meer bloemen gegeten*” (“Fox has eaten more flowers”). Again, the context was the same for collective test items, as can be seen in Figure 6. The collective test item *bestek* (cutlery) is presented in the substance-oriented context. The expectation is that this context activates the substance-denoting interpretation to become salient; participants could thus interpret two big

items of cutlery as being ‘more’ than the four small items of cutlery because the overall mass of the two cutlery items is bigger.

Figure 5

The Substance-Oriented Context as Used in the Experiment; Showing the Countable Test Item Bloemen (Flowers).

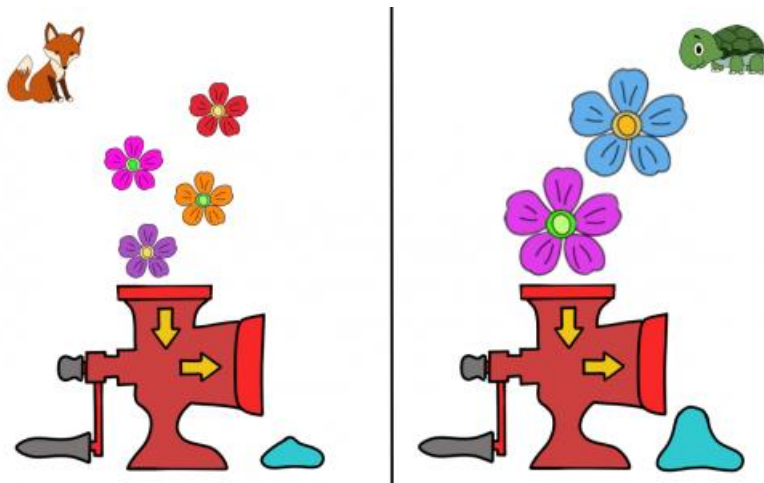
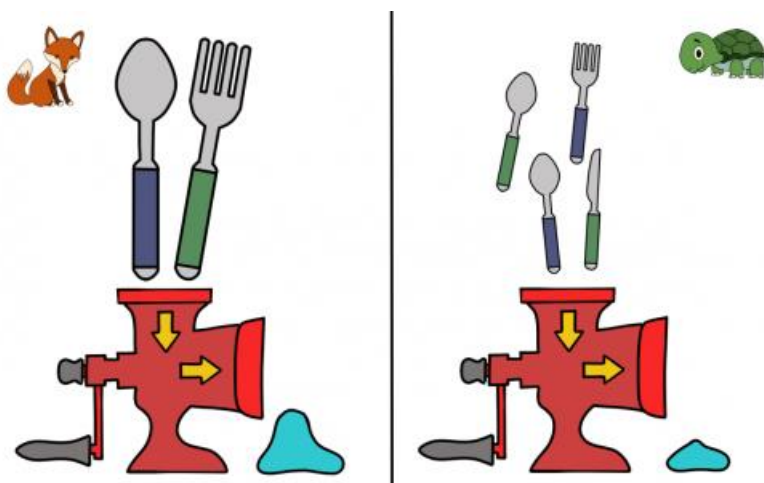


Figure 6

The Substance-Oriented Context as Used in the Experiment; Showing the Collective Test Item Bestek (Cutlery).



4. Results

In this chapter I will discuss how I analyzed the data and elaborate on the results. After extracting the data from Qualtrics, I removed the entries that were marked as spam by Qualtrics' Fraud Detection (one entry), a function of Qualtrics that marks questionnaire entries that is described as follows on their website:

Some of the biggest threats to data quality are bots and cheaters. Often bots will complete surveys en masse, or a person will take surveys on behalf of someone else multiple times. Thankfully, Expert Review tracks these patterns, so you can ensure your surveys are collecting only the highest quality of data (*Fraud Detection*, Qualtrics 2021).

Other entries that were removed were the preview entries that were done to test the questionnaire (three entries), and the unfinished entries (26 entries). After doing so, 80 of the initial 110 questionnaire entries remained. For the next step, I analyzed the answers participants gave to the four filler questions (Q2, Q5, Q8, and Q11) that targeted binding theory judgments. Considering that all participants were adults and native speakers of Dutch, participants that made more than one mistake in the filler questions were excluded from data analysis. This resulted in the exclusion of five more participants, meaning that the data analysis includes 75 participants. For all further analysis, only the eight target questions were used; the four filler questions had no additional value.

The next step was to recode the data; the Microsoft Office Excel file I extracted from Qualtrics contained the true/false answers per participant. These true/false answers did not yet offer a reliable image: depending on the phrasing of the statement the participant judged, for half of the questions the 'true' answer referred to the substance denoting interpretation, but for

the other half of the questions, the ‘true’ answer referred to the individual denoting interpretation. To control for this, I used Microsoft Excel to change all answers that translate to the substance response into ones (1), and all answers that translate to the individual response into zeroes (0). This resulted in the table in Appendix 3.

4.1 Data Analysis

It was decided to use a non-parametric test to analyze the data, more specifically the Fisher’s Exact test (Fisher, 1954). The data were organized per condition to perform Fisher’s Exact tests to calculate the p-value to compare the results between conditions. Each of the 75 participants answered two questions for each of the four conditions, resulting in 150 data entries per condition. Table 4 shows an overview of the data for both the collective stimuli and the count stimuli tested in the substance-oriented context, divided by how often participants gave the individual response and how often they selected the substance response. Table 5 shows the same for both the collective and count stimuli as tested in the individual-oriented context.

Table 4

Data from the Stimuli Presented in the Substance-Oriented Context

Context: Substance-Oriented		
	Individual Response	Substance Response
Collective Stimuli	93	57
Count Stimuli	122	28

Table 5

Data from the Stimuli Presented in the Individual-Oriented Context

Context: Individual-Oriented		
	Individual Response	Substance Response
Collective Stimuli	109	41
Count Stimuli	148	2

The Fisher's Exact test was performed four times: twice to compare the stimuli presented in the same context (A and B below) and twice to compare the same stimuli presented in different contexts (C and D below). This means that the Fisher's Exact test was applied to the following four sets, of which the results will be discussed below:

- A. Collective stimuli in the substance-oriented context versus count stimuli in the substance-oriented context (Table 4);
- B. Collective stimuli in the individual-oriented context versus count stimuli in the individual-oriented context (Table 5);
- C. Collective stimuli in the substance-oriented context versus collective stimuli in the individual-oriented context;
- D. Count stimuli in the substance-oriented context versus count stimuli in the individual-oriented context.

First, the Fisher's Exact test was performed on the responses given to all stimuli in the substance-oriented context, to measure the significance between responses to the collective stimuli and the count stimuli. The result is significant ($p < 0.001$). The second Fisher's Exact test was done in the individual-oriented context between collective and count stimuli, which is also significant ($p < 0.001$). This means that in both context conditions (individual-oriented and substance-oriented) there is a significant difference in response rates between the collective and the count stimuli.

The third test was executed on the collective stimuli to compare the interpretations between the substance-oriented and the individual-oriented context. This result was not significant ($p = 0.06$). This means that, for the collective stimuli, there is no significant difference between how collective nouns are interpreted in a substance-oriented context and

how they are interpreted in an individual-oriented context. Lastly, the Fisher's Exact test was used to calculate the p-value of the count stimuli, comparing the substance-oriented context with the individual-oriented context. This result was significant ($p < 0.001$). This means that, for the count stimuli, there is a significant difference in interpretation between the two contexts.

5. Discussion

In this thesis I aimed to answer the following research question: *Can a substance-oriented context elicit the substance-denoting interpretation of mass collectives in native Dutch speaking adults?* Based on previous research as discussed in Chapter 2, I predicted that the independent variable ‘context’ would have an effect on whether the participants gave a substance response or an individual response to collectives. In the experiment, this meant that participants were presented with collectives in two different contexts: a substance-oriented context and an individual-oriented context. The participants also answered questions with count nouns in the same two contexts. The expectation was that a substance-oriented context would more often result in a substance response of collectives when compared to the results of collectives presented in an individual-oriented context.

The motivation behind these two separate contexts lies in previous research to the interpretation of collectives. Previous research by Barner and Snedeker (2005) and subsequent studies in a wide array of languages all reached the same conclusion, collectives are indeed interpreted as denoting individuals, and thus have count semantics (Bale & Barner, 2009; Inagaki & Barner, 2009; Lin & Schaeffer, 2018; MacDonald & Carrol, 2018; Van Witteloostuijn & Schaeffer, 2018). Huang and Meroni (to appear) proposed a different view, namely that the individual-denoting reading could be a preferred reading in neutral contexts, but not the *only* reading that collectives can have. The results of the experiments by Huang and Meroni (to appear) confirmed their hypothesis that in Mandarin Chinese, both the count and the mass reading of collectives are available, depending on specific contexts and the presence of an individual classifier.

5.1 Data Interpretation

In this thesis, I aimed to replicate the study by Huang and Meroni (to appear) for Dutch adults, as because of the COVID-19 pandemic it became impossible to test children. The results, reported in Chapter 4, show that in both contexts there is a significant difference between how the count stimuli and the collective stimuli are interpreted. There is also a significant difference between how participants interpreted count nouns in both contexts, meaning that the context had an effect on the way the participant read the count noun. In both contexts, count nouns were overwhelmingly interpreted as individual (122 out of 150 in the substance-oriented context, 148 out of 150 in the individual-oriented context). The difference between these responses to the count stimuli in both contexts was substantial enough for a significant difference, even though in both cases the preference for the individual reading was overwhelmingly stronger.

Lastly, there is no significant impact of the context on the interpretation of collectives; the difference between the results of the collectives in the substance-oriented context and the collectives in the individual-oriented context is not significant. This is also shown in the raw data: in the substance-oriented context, 57 out of 150 answers are a substance reading of the collectives. Most answers (93 out of 150) point towards the individual-denoting reading of collectives, despite the collectives being presented in a substance-oriented context. Collectives presented in an individual-oriented context elicited 41 (out of 150) substance-oriented interpretations and 109 (out of 15) individual-oriented interpretations. The differences between contexts are thus, as shown by the Fisher's Exact test, not significant. This means that the hypothesis of this experiment – that a substance-oriented context would more often result in a substance response of collectives than an individual-oriented context – is not supported by the results. To analyze these findings further, I will first take a closer look at individual patterns in

the data set, both on item-level and on participant-level. Next, I will discuss aspects of the experimental design that could potentially have impacted the results.

5.2 Discussion of Individual Test Items

In Chapter 4, the data were discussed per condition. Each experimental condition consisted of two items in the questionnaire. After analyzing and interpreting the results, the question emerged whether the results can be further explained by individual outliers in the dataset; perhaps there are certain items that elicited surprising results, or big differences between two items that belonged to the same test condition.

When taking a closer look at the test items, the items that target count nouns in an individual-oriented context had consistent results: both items in this category had a 99% individual response rate, with a 1% substance response rate. Perhaps more interesting are the count nouns presented in a substance-oriented context. The count noun *wielen* (wheels) elicited a 75% individual response. The other count noun presented in the substance-oriented context, *bloemen* (flowers), elicited a 79% individual response. These percentages show that a majority of the participants still selected the individual-denoting reading, but there was a significantly bigger amount of substance responses as shown by the Fisher's Exact test, which fits the expectations as these nouns were presented in a substance-oriented context. However, overall it seems that the count nouns were consistently judged; there are no surprising differences between two nouns in the same contexts and there is no reason to worry as to whether one of the nouns could have confused the participants. The four count nouns used are all commonly used and clearly countable.

The collectives that were presented in the substance-oriented context show a very similar outcome. The first collective, *bestek* (cutlery), had a 61% individual response rate. The second collective, *post* (mail), had a 59% individual response rate. This means that 61% and

59% of the answers respectively pointed towards an individual (count) reading of the collective, despite the collective being presented in a context that was expected to elicit a substance (mass) reading. I will discuss possible explanations for these results in section 5.2.1.

The collective nouns presented in the individual-oriented context show a different pattern. The collective *beleg* (sandwich toppings) has a 47% individual response rate, meaning that this was the only item that elicited slightly more substance responses – even though this collective was presented in an individual-oriented context. The second collective presented in an individual-oriented context was *kleding* (clothing) and elicited a 91% individual reading. The surprising difference between these two items may make the results of this experimental condition (mass collective nouns presented in an individual-oriented context) less reliable. The results to the question targeting the collective *kleding* (clothing) fit within the expectations: a collective presented in an individual-oriented context does not activate the less preferred substance-denoting reading, so participants gave the individual response. However, the results from the question targeting *beleg* (sandwich toppings) are surprising and do not fit the expectations based on the research by Huang and Meroni (to appear). Perhaps this has to do with the semantic and/or syntactic properties of the noun *beleg*. For this experiment, we considered *beleg* to be a collective because it seems to behave that way. However, the unexpected results could mean that *beleg* in fact behaves more like a mass noun such as *pasta*, which refers to all different shapes and sizes of pasta. Similarly, perhaps *beleg* refers to all different kinds of things one can put on their sandwich, without syntactically and semantically behaving like a collective. If *beleg* is indeed a mass noun, that could explain why it was interpreted in a substance-denoting manner 53% of the time, despite being presented in an individual-oriented context. Another possibility is that perhaps these difficulties can be explained by the fact that the individual items making up the collective *beleg*, such as *kaas* (cheese) and *jam* (jam), are mass nouns themselves. In contrast, the individual items that make

up the collective *clothing* (kleding) are count nouns such as *t-shirt* (t-shirt), *schoen* (shoe), and *trui* (sweater). This example demonstrates that it can be difficult to determine which words are collectives and which are not. A further analysis of the syntactic and semantic properties of the word *beleg* and Dutch collectives in general could provide the information needed to reliably determine whether a word is a collective or another kind of mass noun.

5.2.1 Collectives in the Substance-Oriented Context: What Happened?

As discussed in the previous section, the two test items of collectives presented in a substance-oriented context gave unexpected results. Based on previous research, it was expected that collectives presented in a substance-oriented context would elicit the substance-denoting reading. However, the results showed that, for the two questions targeting this condition, 61% and 59% of the participants gave an individual-denoting answer, despite the substance-oriented context. There are several possible explanations for this, all of which demand further research.

The first possible explanation is that the substance-oriented context was not designed well enough to elicit the substance response. It is possible that the created context would work better when the experimenter tells the participant the context-shaping story to shape the context; an online questionnaire means that the experimenter has less control over the experiment. This means there is no way to know whether the participants paid attention to the context, which is key in this research method. A second possible explanation is that the difference between mass and count nouns in Dutch is simply less strong than in Mandarin Chinese. This could explain why the results found by Huang and Meroni (to appear) were not replicated in this study. In Mandarin Chinese, information about the noun is encoded in a separate classifier, whereas Dutch does not have these classifiers as Dutch is not a classifier language. Perhaps having a separate classifier makes the encoded information more evident,

giving it enough weight to activate the substance-denoting reading in Mandarin Chinese. Huang and Meroni (to appear) concluded in test sentences *without* classifiers, the contextual information affected the interpretation. They also concluded that sentences *with* classifiers were not significantly impacted by the contextual information and concluded that “these patterns suggest that it is classifiers that determine the countability of superordinates, and bare superordinates on their own are underspecified in countability” (Huang & Meroni, to appear, p.24). Because Dutch does not have classifiers, it is possible that the information provided by the context was not enough to affect the interpretation. Further research on languages typologically more similar to Dutch, or a repeated experiment in Dutch, could show whether these results can be repeated, confirming the results found in previous research by, amongst others, Barner and Snedeker (2005). Testing another classifier language could confirm whether the results found by Huang and Meroni (to appear) can indeed be ascribed to the fact that Mandarin Chinese encodes information on the noun in a separate classifier. Perhaps this could shine more light on the role of the classifier, and further explain the difference in results found by Huang and Meroni and the results found in this experiment.

5.3 Discussion of Individual Participants

Having looked at the individual test items, the next step is to analyze the participants at an individual level. The first thing that stands out when looking at the data (as seen in Appendix 3), is that eleven participants gave the individual response 100% of the time. This means that these eleven participants interpreted every noun – both countable and collective – as individual-denoting, regardless of whether it was presented in substance-oriented context or in an individual-oriented context. As discussed before, it is possible that this is because the substance-oriented context was not explicit enough to activate the substance-denoting reading. There are no participants that gave the substance response to 100% of the questions.

Both count stimuli in the individual-oriented context had a 99% individual response rate, meaning that only one participant gave a substance response. However, for these items it was a different participant that gave the substance response; it is not one participant that consistently interpreted these items as substance-denoting.

While these individual differences sound interesting, it is difficult to determine what they mean. Perhaps for future research, a question can be included that elicits more information about the participant – perhaps adult participants can explain what they think they did at the end of the experiment. If participants are asked for their reasoning, perhaps this could give an insight into individual differences. However, as the statistical test shows, it can be assumed that the data can be generalized despite the differences on an individual level.

5.4 Discussion of the Experimental Design

There are a couple things that need to be considered with regards to the experimental design used in this experiment. Because of the COVID-19 pandemic, this experiment was executed through an online questionnaire as face-to-face meetings were prohibited. This means that, differently from previous research, I had less control over the experiment: rather than listening to the experimenter telling the story that shapes the context, the participants had to listen to voice recordings. This means there was no possibility to control whether the participants paid attention to the context, which is crucial in this research method.

Another aspect which was discussed above is the decision for the test items, especially the collective nouns. For example, Barner and Snedeker (2005) used the collective *meubilair* (furniture) in their experiment, but I decided against using this collective for the simple reason that I strongly doubt whether children still regularly use that word, or whether it has become old-fashioned in Dutch. Another example is that Van Witteloostuijn and Schaeffer (2018) state that the plural of *deeg* (dough), a mass noun, cannot be *degen* (doughs). However, the Dutch

dictionary Van Dale (2020) states that *degen* is in fact the plural of *deeg*. This further shows that there is very little research available on which Dutch words are collectives and which are mass nouns, and as discussed above, my decision to use the word *beleg* (sandwich toppings) potentially had unintended effects as the results on the *beleg*-item were unexpected – however, it is uncertain whether this is because of the noun itself, or because of other factors.

6. Concluding Remarks and Future Research

For this thesis, I investigated the interpretation of Dutch collectives by native Dutch adults by using the online Truth Value Judgment Task. This research was originally planned to replicate the research by Huang and Meroni (to appear), but the experiment had to be adjusted because of the COVID-19 pandemic. This experiment aimed to answer the research question: *Can a substance-oriented context can elicit the substance-denoting interpretation of mass collectives in native Dutch speaking adults?* The results indicate that the substance-oriented context did not elicit a substance-denoting reading, also referred to as a substance response.

This conclusion based on the present study does not support the hypothesis and opposes the results that Huang and Meroni (to appear) found in their experiments on Mandarin Chinese collectives. Several potential explanations for these surprising results have been discussed in the Discussion in Chapter 5. Future research could provide more information about the interpretation of collectives in Dutch and whether or not the substance-denoting reading can be triggered in Dutch by manipulating context. It is possible that the online setting of the experiment had an impact, or that the collectives that were chosen as test items were perhaps not ideal. Another interesting possibility is that the substance-denoting interpretation cannot be as activated in Dutch as in Mandarin Chinese because of typological differences between these languages: Mandarin Chinese is a classifier language, Dutch is a number marking language. Future research on a language typologically similar to Dutch, could shed more light on what is happening. While the data in this study do make it seem unlikely for the substance-denoting reading to be as available in Dutch as it seemed in Mandarin Chinese, it would be interesting to see more research in languages typologically similar to Dutch to expand on what we currently know about collectives and mass-count issues.

If the results found in the present study can be confirmed, this would support the position taken by Barner and Snedeker (2005) and all subsequent studies as discussed in

Chapter 2. Thus, the present study adds new data to the existing body of data on the interpretation of collectives in a wide array of languages and the ever-evolving insights in the mass-count debate.

References

- Bale, A., & Barner, D. (2018). Quantity judgment and the mass-count distinction across languages: Advances, problems, and future directions for research. *Glossa: A Journal of General Linguistics*, 3(1).
- Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition*, 97(1).
- Crain, S. & Thornton, R. 1998. *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: The MIT Press.
- Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science*, 14(2), 164-168.
- Fisher, R.A. (1954). *Statistical methods for research workers*. Oliver and Boyd. ISBN 0-05 002170-2.
- Fraud Detection*. (2021, April 12). Qualtrics. Retrieved May 5, 2021, from <https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/>
- Gillon, B.S. (1996). *The lexical semantics of English count and mass nouns*. Paper presented at the workshop on the breadth and depth of semantic lexicons, Santa Cruz.
- Gillon, B. S. (1999). The lexical semantics of English count and mass nouns. In E. Viegas (Ed.), *Breadth and Depth of Semantic Lexicons* (pp. 19–37). Springer Dordrecht.
- Hacohen, A. (2008). Acquiring the mass/count distinction in Hebrew: How does it compare with English. *Online Proceedings Supplement of BUCLD-32*.
- Huang, A. & Meroni, L. (to appear). Grammatical and contextual factors in the interpretation and acquisition of superordinate collectives in Mandarin Chinese. *Linguistics*.
- Inagaki, S. & Barner, D. (2009). Countability in absence of count syntax: Evidence from Japanese Quantity Judgements. In S. Inagaki et al. (eds.), *Studies in Language Sciences* 8.
- Landau, B., Smith, L.B., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development* 3(3). 299-321.
- Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, 31(6), 807-825.
- Lin, J., & Schaeffer, J. (2018). Nouns are both mass and count: Evidence from unclassified nouns in adult and child Mandarin Chinese. *Glossa: A Journal of General Linguistics*,

3(1).

- Link, G. (1998). *Algebraic semantics in language and philosophy*. Stanford, CA: Center for the Study of Language and Information.
- MacDonald, D., & Carroll, S. E. (2018). Second-language processing of English mass-count nouns by native-speakers of Korean. *Glossa: A Journal of General Linguistics*, 3(1).
- Nicolas, D. (2008). Mass nouns and plural logic. *Linguistics and Philosophy*, 31(2), 211-244.
- Potrzeba, E. R., Fein, D., & Naigles, L. (2015). Investigating the shape bias in typically developing children and children with autism spectrum disorders. *Frontiers in Psychology*, 6, 446.
- Quine, W. V. O. 1960. *Word and Object*. Cambridge: MIT Press.
- Van Dale Uitgevers. (2020). *deeg*. Van Dale. Retrieved May 20, 2021, from <https://www.vandale.nl/gratis-woordenboek/nederlands/betekenis/deeg>.
- Van Witteloostuijn, M., & Schaeffer, J. (2018). The mass-count distinction in Dutch speaking children with specific language impairment. *Glossa: A Journal of General Linguistics*, 3(1).
- Wisniewski, E. J., Imai, M., & Casey, L. (1996). On the equivalence of superordinate concepts. *Cognition*, 60(3).

Appendix 1: Original Questionnaire in Dutch

Note: participants listened to voice recordings, which cannot be included in this file. In this Appendix, the voice recordings have been replaced by transcriptions of what the participants listened to.

Welkom bij de online vragenlijst voor de masterscriptie van Hannah Aalbers

FETC-GW referentienummer: 4114728-02-07-2020

Ik wil u vragen deel te nemen aan wetenschappelijk onderzoek voor mijn masterscriptie. Het is van belang dat u deze informatie goed leest voordat u akkoord gaat.

Het onderzoek is deel van mijn afstuderen bij de onderzoeksmaster Linguistics aan de Universiteit Utrecht. Het onderzoek is goedgekeurd door de Facultaire Ethische ToetsingsCommissie van de Faculteit Geesteswetenschappen (FETC-GW), Universiteit Utrecht. De vragenlijst is volledig anoniem en het invullen van de vragenlijst duurt ongeveer 10 à 15 minuten.

Wanneer u een klacht wilt indienen over de procedure omtrent dit onderzoek, dan kunt u contact opnemen met de secretaris van de FETC-GW, e-mail: fetc-gw@uu.nl, of met de functionaris voor gegevensbescherming van de Universiteit Utrecht, e-mail: privacy@uu.nl

Het doel van het onderzoek is te bepalen hoe bepaalde zelfstandig naamwoorden worden geïnterpreteerd. Tijdens het onderzoek zult u verschillende afbeeldingen te zien krijgen. Vervolgens wordt van u gevraagd om een uitspraak over de afbeeldingen te beoordelen als 'waar' of 'niet waar'. Er zijn geen goede of foute antwoorden; alle informatie is waardevol.

U heeft zelf geen voordeel van deelname aan dit onderzoek. Voor de toekomst kan het onderzoek wel nuttige gegevens opleveren, bijvoorbeeld over de verwerking van taal. Deelname aan dit onderzoek is ook niet nadelig voor u, behalve dat het u tijd kost.

Deelname is vrijwillig. Als u na het lezen van deze informatie besluit niet mee te doen, kunt u het scherm sluiten. Ook tijdens het invullen van de vragenlijst kunt u op ieder gewenst moment besluiten uw deelname te stoppen door het venster te sluiten; wanneer u dat doet verdwijnen de antwoorden die u al heeft ingevuld.

Zoals eerder genoemd, is deze vragenlijst volledig anoniem; er wordt alleen van u gevraagd uw leeftijd in te vullen. De data zullen op geen enkele manier herleidbaar zijn naar u als persoon. Wij zijn verplicht de onderzoeksgegevens minstens 10 jaar te bewaren op een beveiligde server van de Universiteit Utrecht. Daarvoor geeft u toestemming als u meedoet aan dit onderzoek. Als u dit niet wilt, kunt u niet deelnemen aan het onderzoek. De anonieme gegevens kunnen met andere onderzoekers gedeeld worden.

Mocht u voor, tijdens, of na het onderzoek nog vragen hebben, of na het invullen meer willen weten over het doel van het onderzoek, dan kunt u contact opnemen met Hannah Aalbers (e-mail: h.h.aalbers@students.uu.nl).

Toestemming

Ja, ik geef toestemming om mijn antwoorden te gebruiken voor wetenschappelijk onderzoek.

Welke uitspraak is op u van toepassing?

Moedertaal: de taal/talen die u tijdens het opgroeien hebt verworven

- Nederlands is mijn enige moedertaal
- Ik heb Nederlands als moedertaal, maar daarnaast heb ik nog een taal/talen als moedertaal, namelijk... _____
- Nederlands is niet mijn moedertaal

Welke uitspraak is op u van toepassing?

- Ik spreek alleen Nederlands
- Naast het Nederlands, ben ik ook (bijna) vloeiend in de volgende taal/talen:

Uitleg en instructies

Deze vragenlijst bestaat uit 12 vragen en iedere vraag bestaat uit een klein verhaaltje waar u naar luistert met één afbeelding erbij. Het is van belang dat u het verhaaltje luistert en de bijbehorende afbeelding bekijkt. Vervolgens wordt van u gevraagd een uitspraak als ‘waar’ of ‘niet waar’ te beoordelen.

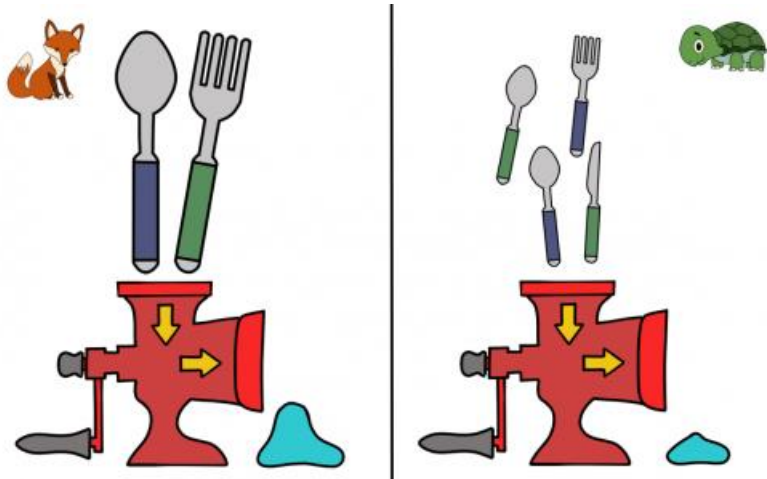
Het is voor u belangrijk om te weten dat er echt geen goede of foute antwoorden zijn; het gaat om uw persoonlijke intuïties over taal. Vul dus vooral in wat als eerste bij u opkomt!

Vraag 1

Voice recording transcription: Deze vraag gaat over Vos en Schildpad. Vos en Schildpad houden heel erg van eten en eten het liefst alles wat ze tegenkomen! Dit is wat lastig omdat Vos en Schildpad helaas geen tanden meer hebben... Hiervoor gebruiken zij een machine die hun eten pureert zodat ze kunnen eten.

Op een dag vinden Vos en Schildpad bestek. Vos vond een grote lepel en een grote vork en gebruikte zijn machine om deze in een grote berg puree te veranderen. Na het eten zat Vos vol. Schildpad vond twee kleine lepels, een kleine vork en een klein mes en gebruikte zijn machine om deze in een kleine berg puree te veranderen. Na het eten had Schildpad nog honger.

Je ziet deze scene in het plaatje hieronder

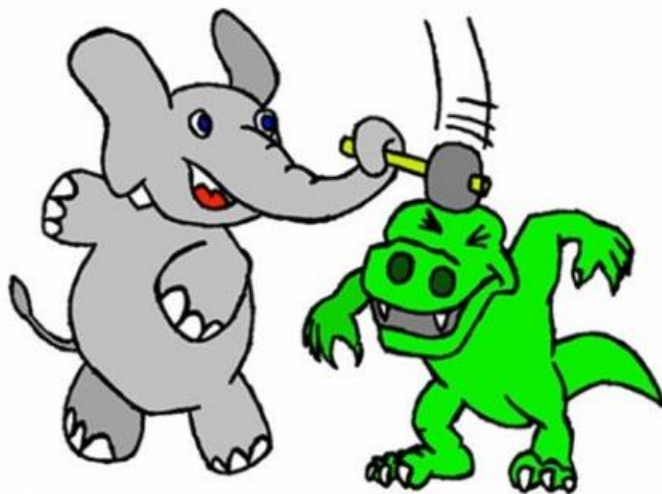


Vos heeft meer bestek gegeten

- Waar
- Niet waar

Vraag 2

Voice recording transcription: Deze vraag gaat over Olifant en Dino. Olifant en Dino waren samen aan het spelen met de hamer. Deze scène zie je in het plaatje hieronder.



De olifant slaat hem

- Waar
- Niet waar

Vraag 3

Voice recording transcription: Deze vraag gaat over de blauwe fee en de rode fee. Op een dag hielden zij een wedstrijd om te kijken wie er het beste kon toveren. Wie het meeste kan

toveren, is de winnaar! De blauwe fee toverde twee grote schoenen en de rode fee toverde vier kleine schoenen.

Deze scene zie je in het plaatje hieronder

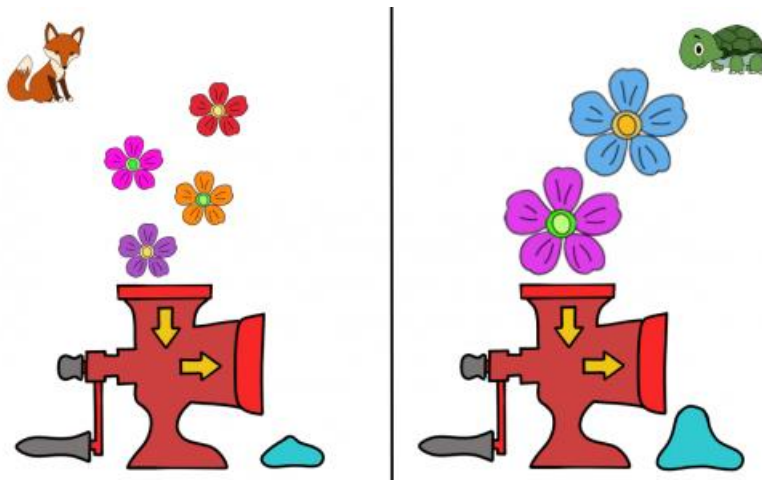


De rode fee heeft meer schoenen gemaakt

- Waar
- Niet waar

Vraag 4

Voice recording transcription: Vos en Schildpad hadden weer eens onwijs honger. Vos vond vier kleine bloemen en gebruikte zijn machine om deze in een kleine berg puree te veranderen. Na het eten had Vos nog honger. Schildpad vond twee grote bloemen en at een grote berg puree. Na het eten zat schildpad vol.

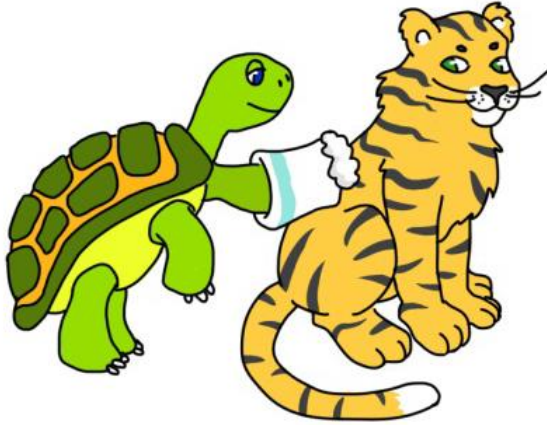


Vos heeft meer bloemen gegeten

- Waar
- Niet waar

Vraag 5

Voice recording transcription: Deze vraag gaat over Schildpad en Tijger. Na een lange dag gingen Schildpad en Tijger douchen voordat ze naar bed gingen. Deze scene zie je in het plaatje hieronder.

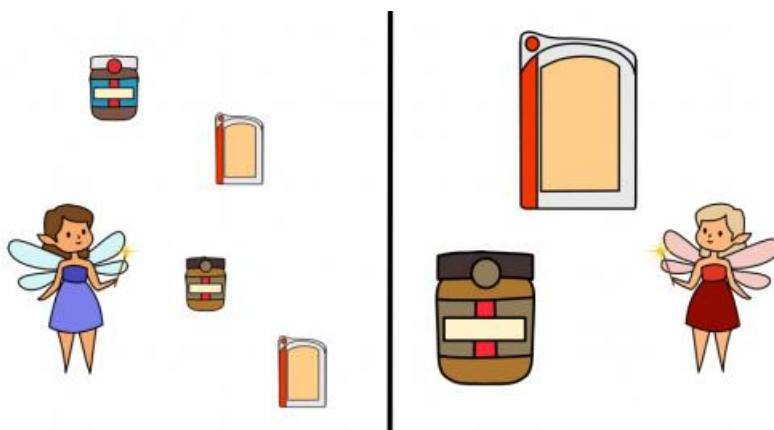


Schildpad wast zich

- Waar
- Niet waar

Vraag 6

Voice recording transcription: De rode en de blauwe fee besloten een tweede toverwedstrijd te houden met dezelfde regels: wie het meeste tovert, wint. De blauwe fee toverde een kleine pot pindakaas, een kleine pot chocopasta en twee kleine pakjes kaas. De rode fee toverde een groot pak kaas en een grote pot pindakaas. Deze scene zie je in de afbeelding hieronder.

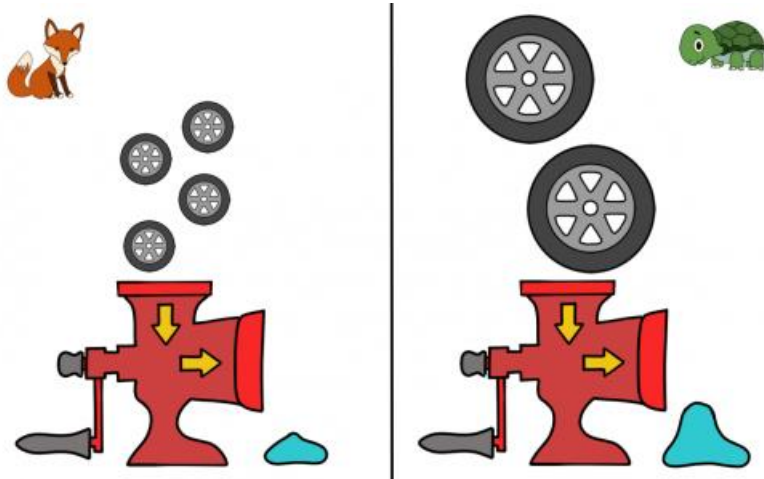


De rode fee heeft meer beleg

- Waar
- Niet waar

Vraag 7

Voice recording transcription: Het was geen verrassing dat Vos en Schildpad alwéér honger hadden. Dit keer hadden ze wielen gevonden. Vos pureerde vier kleine wielen en had na het eten nog steeds honger. Schildpad stopte twee grote wielen in zijn pureermachine en zat na het eten vol. Deze scene zie je in het plaatje hieronder.

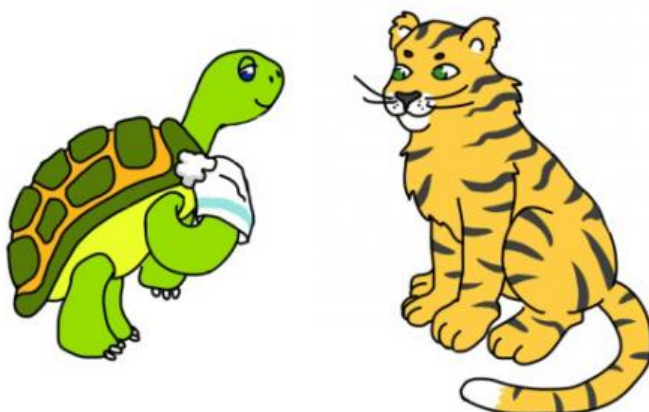


Schildpad heeft meer wielen gegeten

- Waar
- Niet waar

Vraag 8

Voice recording transcription: Deze vraag gaat weer over Schildpad en Tijger. Ze willen natuurlijk goed schoon worden van het douchen en zepen zich goed in. Deze scene zie je in het plaatje hieronder.



Schildpad wast zich

- Waar
- Niet waar

Vraag 9

Voice recording transcription: Voor de derde ronde van de toverwedstrijd, toveren de feeën kleding. De blauwe fee tovert een groot t-shirt en een grote jurk. De rode fee tovert twee kleine t-shirts, een klein jasje en een kleine broek. Je ziet deze scene in het plaatje hieronder.

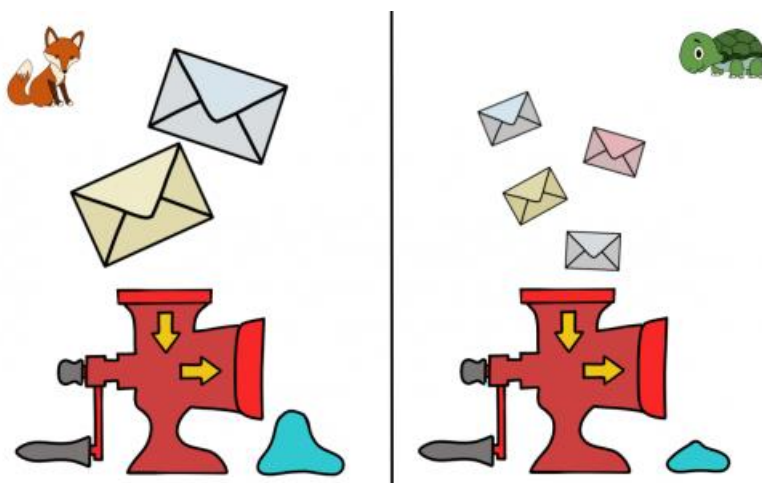


De blauwe fee heeft meer kleding

- Waar
- Niet waar

Vraag 10

Voice recording transcription: Voor hun volgende maaltijd hebben Vos en Schildpad post gevonden. Vos pureert twee grote enveloppen en eet een grote berg puree. Na het eten zit vos vol. Schildpad pureert vier kleine enveloppen en eet een kleine berg puree. Na het eten heeft schildpad nog honger. Deze scene zie je in de afbeelding hieronder.

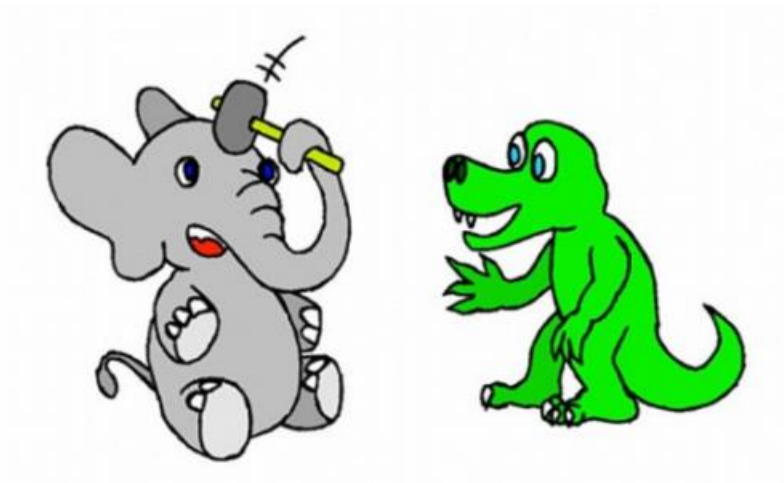


Schildpad heeft meer post gegeten

- Waar
- Niet waar

Vraag 11

Voice recording transcription: Voor deze vraag gaan we terug naar de olifant en de dino die aan het spelen zijn met de hamer. Je ziet de scene in het plaatje hieronder.

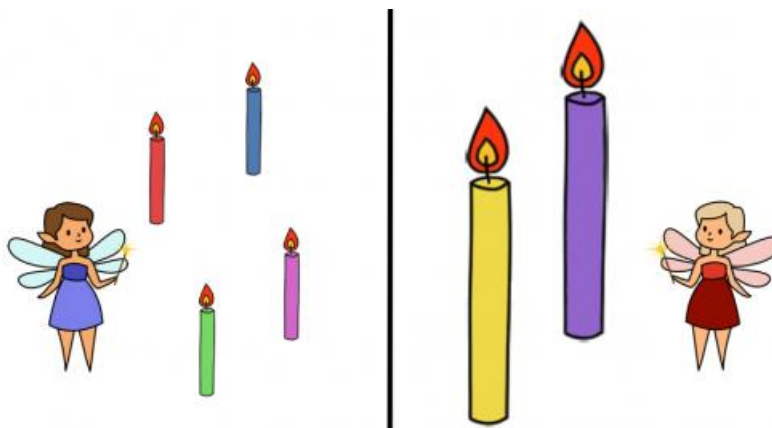


De olifant slaat hem

- Waar
- Niet waar

Vraag 12

Voice recording transcription: Voor de laatste toverwedstrijd, toveren de feeën kaarsen. De blauwe fee tovert 4 kleine kaarsen en de rode fee tovert twee grote kaarsen. Je ziet deze scene in de afbeelding hieronder.



De rode fee heeft meer kaarsen

- Waar
- Niet waar

Appendix 2: Translated Questionnaire in English

Note: This Appendix is an English translation of the original Dutch materials used.

Welcome to the online questionnaire for the master's thesis of Hannah Aalbers

FETC-GW (Ethical committee) reference number: 4114728-02-07-2020

I would like to ask you to participate in scientific research for my master's thesis. It is important that you carefully read this information before you agree to participation.

This research is part of my graduation of the research master Linguistics at Utrecht University. This investigation has been approved by the Linguistics Chamber of the Faculty Ethics Assessment Committee of the Faculty of Humanities, Utrecht University (FECT-GW). The questionnaire is completely anonymous and lasts 10 to 15 minutes.

If you want to form a complaint about the procedure of this research project, you can contact the secretary of the ethical committee, e-mail: fetc-gw@uu.nl, or with the data protection officer of Utrecht University, e-mail: privacy@uu.nl.

The purpose of this research is to determine how certain nouns are interpreted. During the questionnaire you will see different images. Next, you will be asked to judge a statement about the images as 'true' or 'not true'. There are no good or wrong answers; all information is valuable.

You do not profit from participating in this research. This research can provide useful data for future research, for example about language processing. Participation in this research does not have any inconvenience, other than taking some time.

Participation is voluntary. If you decide not to participate after reading this information, you can close the screen. You can also decide to stop participating at any point during the questionnaire; if you do this, the answers you have already given will disappear.

As mentioned before, participation is completely voluntarily. You are only asked to enter your age. It will not be possible to trace the data back to you as an individual. The coded data will be stored on protected servers of the University of Utrecht for 10 years. You agree to this by participating. If you do not agree with the use of data that is explained above, you can not participate in this questionnaire. The anonymous data can be shared with other researchers.

If you have any questions before, during, or after the research, or want to learn more about the purpose of this study after finishing the questionnaire, you can contact Hannah Aalbers (e-mail: h.h.aalbers@students.uu.nl).

Permission

- Yes, I authorize the use of my answers for scientific research.

Which statement applies to you?

Native language: the language/languages you acquired growing up

- Dutch is my only native language
- Dutch is my native language, but I have another language/languages as native language, namely... _____

- o Dutch is not my native language

Which statement applies to you?

- o I only speak Dutch
 - o Besides Dutch, I am also (near) fluent in the following language(s):
-

Explanation and instructions

This questionnaire consists of 12 questions and each question consists of a small story that you will listen to and one image. It is of importance that you listen to the story and look to the accompanying image. Next, you will be asked to judge a statement as ‘true’ or ‘not true’.

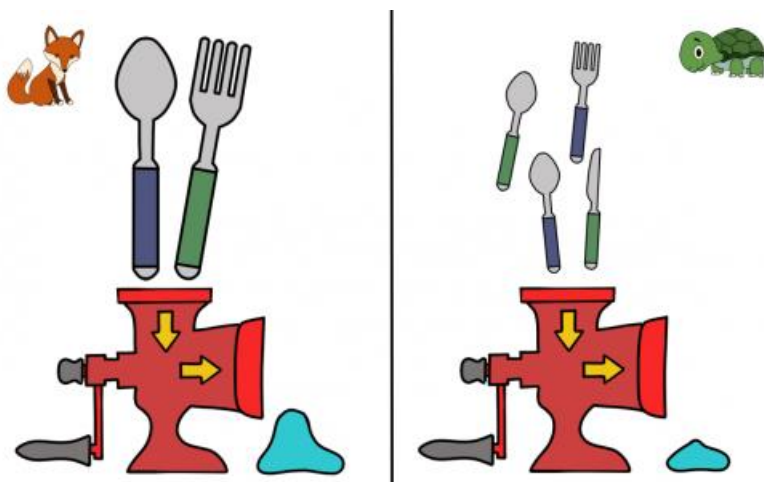
It is important for you to know that there are not good or wrong answers; this is about your personal intuitions about language. You are encouraged to write down the answer that first comes to mind!

Question 1

Voice recording transcription: This question is about Fox and Turtle. Fox and Turtle love food very much and would prefer to eat everything they see! This is a bit difficult because Fox and Turtle sadly don’t have teeth anymore... Because of this, they use a machine that grinds their food into puree so they can eat it.

One day, Fox and Turtle find cutlery. Fox found a big spoon and a big fork and used his machine to grind them into a big pile of puree. After eating, Fox was full. Turtle found two small spoons, a small fork, and a small knife and used his machine to grind them into a small pile of puree. After eating, Turtle was still hungry.

You see this scene in the image below.

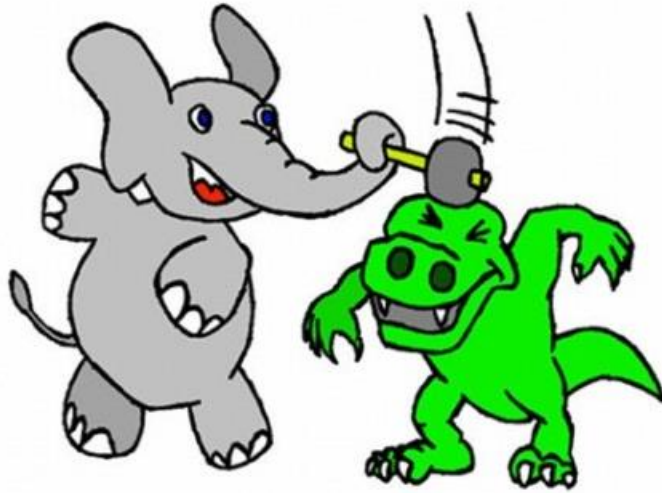


Fox ate more cutlery

- True
- Not true

Question 2

Voice recording transcription: This question is about Elephant and Dino. Elephant and Dino were playing with the hammer together. You see this scene in the image below.



The elephant hits him

- True
- Not true

Question 3

Voice recording transcription: This question is about the blue fairy and the red fairy. One day, they held a competition to see who was the best at magic. Who can conjure the most items, is the winner! The blue fairy conjured two big shoes and the red fairy conjured four small shoes.

You can see this scene in the image below.

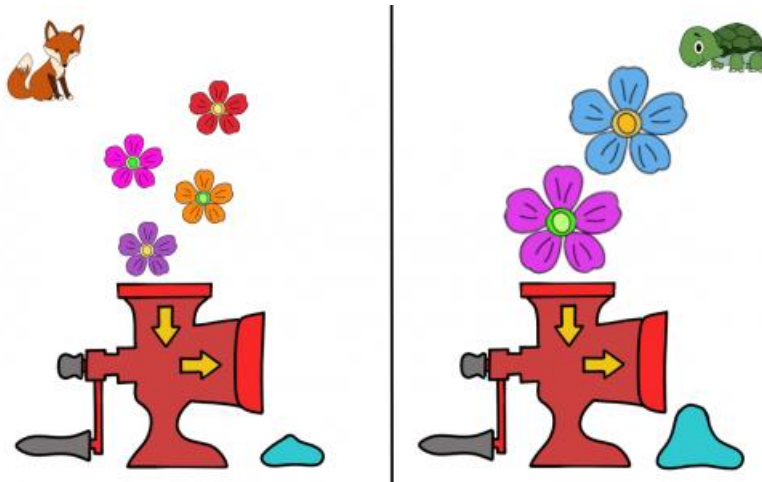


The red fairy made more shoes

- True
- Not true

Question 4

Voice recording transcription: Fox and Turtle were very hungry again. Fox found four small flowers and used his machine to turn them into a small amount of puree. After eating, Fox was still hungry. Turtle found two big flowers and ate a big mountain of puree. After eating, Turtle was full.

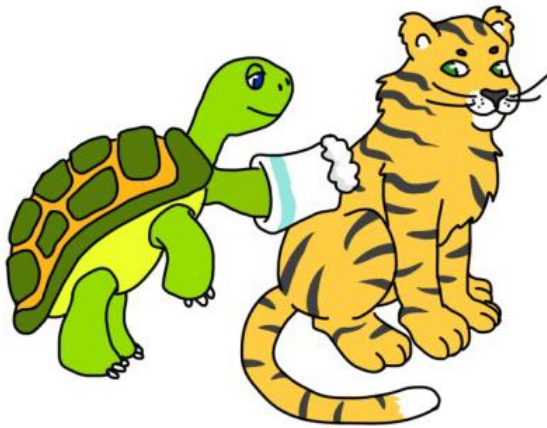


Fox ate more flowers

- True
- Not true

Question 5

Voice recording transcription: This question is about Turtle and Tiger. After a long day, Turtle and Tiger took a shower before going to bed. This scene can be seen in the image below.

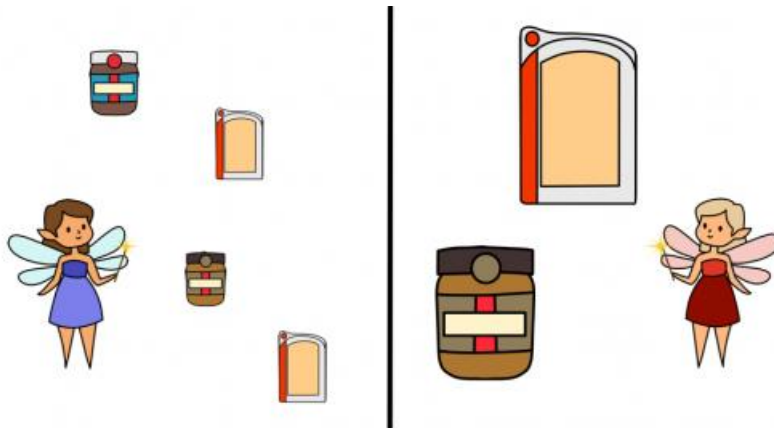


Turtle washes himself

- True
- Not true

Question 6

Voice recording transcription: The red and the blue fairy decided to have a second magic competition with the same rules: who conjures the most, wins. The blue fairy conjured a small jar of peanut butter, a small jar of chocolate spread, and two small packages of cheese. The red fairy conjured a big pack of cheese and a big jar of peanut butter. You see this scene in the image below.

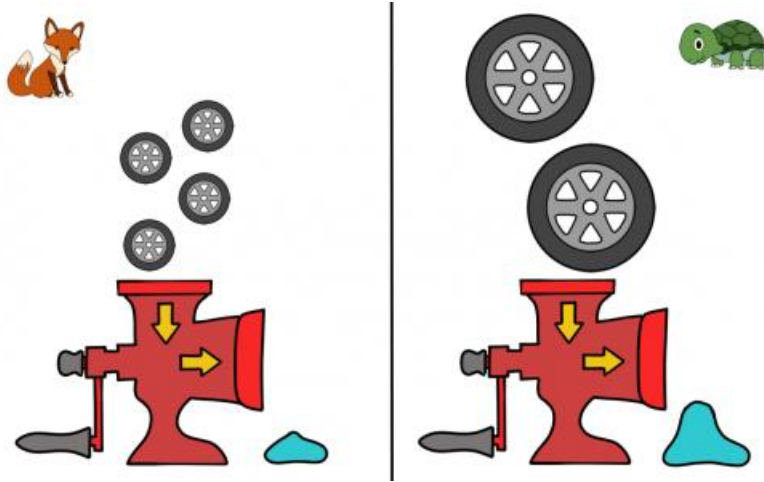


The red fairy has more sandwich toppings

- True
- Not true

Question 7

Voice recording transcription: It was not a surprise that Fox and Turtle were hungry *again*. This time they found wheels. Fox ground four small wheels and was still hungry after eating. Turtle put two big wheels in his grinding machine and was full after eating. You see this scene in the image below.

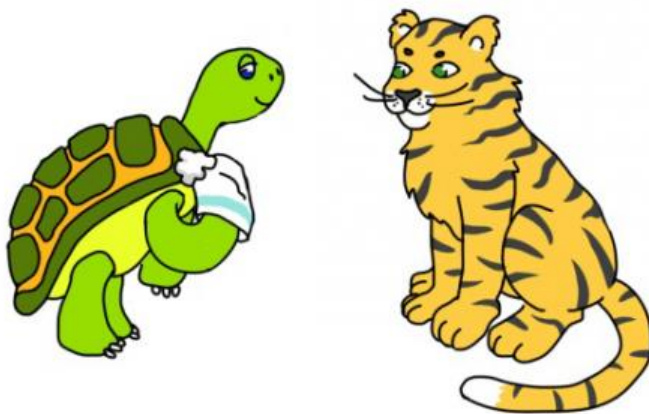


Turtle ate more wheels

- True
- Not true

Question 8

Voice recording transcription: This question is about Turtle and Tiger again. They want to get really clean in the shower so they're putting on soap. You see this scene in the image below.



Turtle is washing himself

- True
- Not true

Question 9

Voice recording transcription: For the third round of their magic competition, the fairies conjure clothing. The blue fairy conjures a big t-shirt and a big dress. The red fairy conjures two small t-shirts, a small jacket, and small pants. You see this scene in the image below.

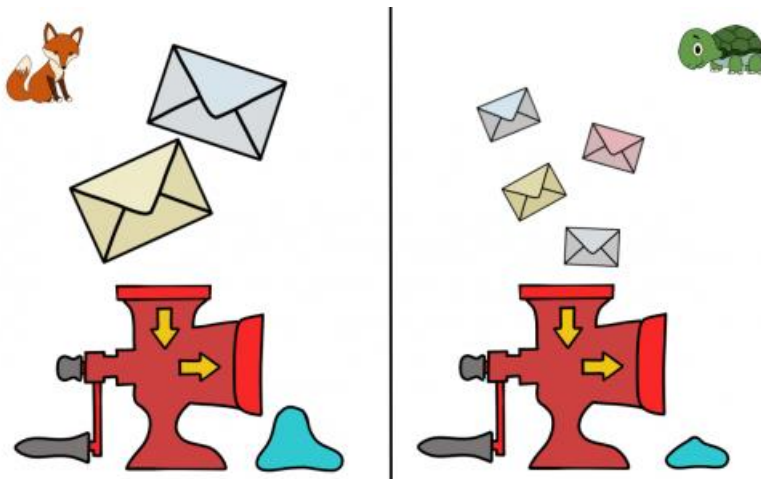


The blue fairy has more clothing

- True
- Not true

Question 10

Voice recording transcription: For their next meal, Fox and Turtle found mail. Fox grinds two big envelopes and eats a big mountain of puree. After eating, Fox is full. Turtle grinds up four small envelopes and eats a small mountain of puree. After eating, Turtle is still hungry. You see this scene in the image below.

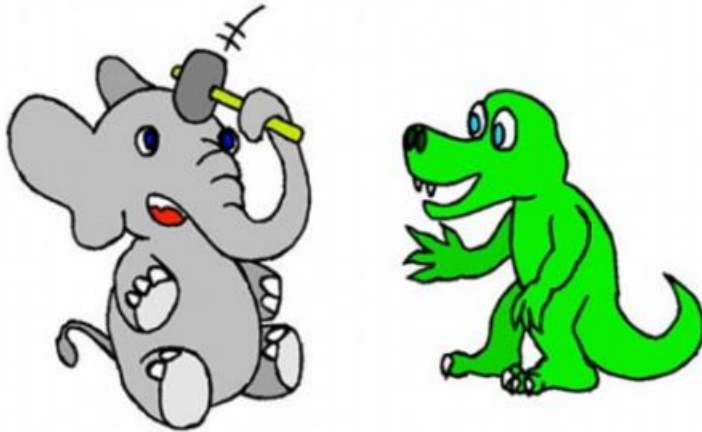


Turtle ate more mail

- True
- Not true

Question 11

Voice recording transcription: For this question we return to the elephant and the dino that were playing with the hammer. You see the scene in the image below.

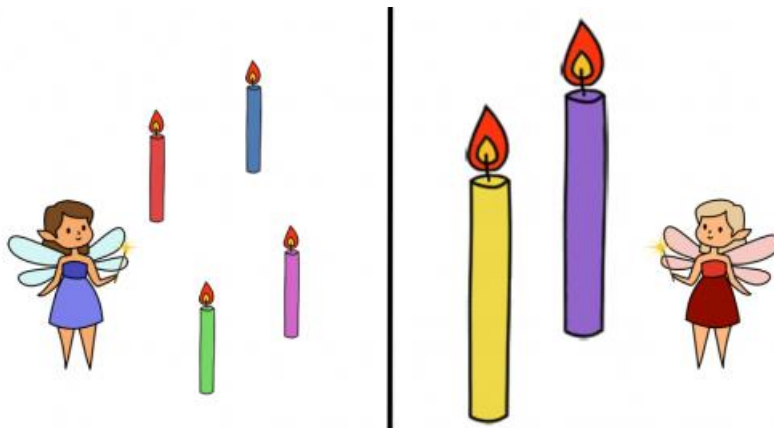


The elephant hits him

- True
- Not true

Question 12

Voice recording transcription: For the last magic competition, the fairies conjure candles. The blue fairy conjures four small candles and the red fairy conjures two big candles. You see this scene in the image below.



The red fairy has more candles

- True
- Not true

Appendix 3: Data Recoded

1 = the participant gave the substance response;

0 = the participant gave the individual response.

PARTICIPANT	Q1	Q3	Q4	Q6	Q7	Q9	Q10	Q12
1	0	0	0	1	0	0	0	0
2	0	0	1	1	0	1	1	0
3	1	0	1	0	1	0	1	0
4	1	0	0	0	0	0	1	0
5	1	0	1	1	1	0	1	0
6	0	0	0	1	1	0	0	0
7	0	0	1	0	0	0	0	0
8	1	0	0	1	0	0	0	0
9	1	0	1	1	1	1	1	0
10	1	0	0	1	0	0	1	0
11	1	0	0	0	1	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0
15	0	0	0	1	0	0	0	0
16	0	0	0	0	0	0	0	0
17	1	0	0	1	0	0	1	0
18	0	0	1	0	0	0	1	0
19	0	1	0	0	1	0	1	0
20	0	0	0	1	0	0	1	0
21	0	0	0	1	0	0	0	0
22	0	0	0	0	1	0	0	0
23	1	0	0	1	1	0	0	0
24	0	0	0	0	1	0	1	0
25	0	0	0	1	0	0	0	0
26	1	0	0	0	0	1	0	0
27	0	0	0	0	0	0	0	0
28	1	0	0	0	0	0	0	0
29	0	0	0	1	0	0	0	0
30	0	0	0	1	0	0	1	0
31	0	0	0	1	0	0	0	0
32	0	0	0	1	0	0	0	0
33	0	0	0	1	0	0	1	0
34	1	0	1	0	1	0	1	0
35	1	0	0	0	0	0	1	0
36	0	0	0	0	0	0	1	0
37	0	0	0	1	0	0	0	0
38	0	0	0	0	0	0	1	0
39	0	0	0	0	0	1	0	0
40	1	0	0	0	0	0	0	0

41	0	0	0	0	0	0	0	0
42	1	0	0	1	0	0	1	0
43	0	0	0	1	0	0	0	0
44	1	0	1	0	1	0	1	0
45	1	0	0	1	0	0	0	0
46	0	0	0	1	0	0	0	0
47	1	0	0	1	0	0	0	0
48	0	0	0	0	1	0	0	0
49	0	0	0	0	0	0	1	0
50	0	0	0	1	1	0	1	0
51	0	0	0	0	0	0	0	0
52	0	0	1	1	0	0	0	0
53	0	0	0	0	0	0	0	0
54	0	0	0	1	0	0	0	0
55	1	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	0	0
58	0	0	0	1	0	0	0	0
59	1	0	0	0	0	0	0	0
60	1	0	1	1	0	0	1	0
61	0	0	0	0	0	1	1	0
62	1	0	0	0	0	0	0	0
63	1	0	0	0	1	0	0	0
64	1	0	0	0	0	0	1	0
65	1	0	0	0	0	1	0	1
66	1	0	0	0	0	0	0	0
67	0	0	0	1	0	0	1	0
68	0	0	0	1	0	0	0	0
69	0	0	0	0	0	0	1	0
70	0	0	1	0	0	0	1	0
71	1	0	1	1	1	0	1	0
72	0	0	0	0	0	0	0	0
73	1	0	0	0	0	1	1	0
74	0	0	0	1	1	0	0	0
75	0	0	0	1	0	0	0	0