Word frequency and semantic violations:

Investigating the effect of word frequency on the detection of semantic anomalies

by

Marian Marchal

5709717

BA Thesis Taalwetenschap,

Universiteit Utrecht

8175 words

Supervisor:

Dr. C. Van den Hazelkamp

29 June 2018

Abstract

Although previous research suggests that high-frequency words are processed faster, Van den Hazelkamp (2018) found that they were responded to later in a semantic violation detection task. The present study examines the effect of word frequency on the detection of semantically anomalous words and the confirmation of semantically correct words. In a negative semantic judgment task, participants had to press a button when they detected a semantic anomaly, while in the positive task, they were asked to press the button when the target word was used correctly. Although the effect of word frequency was hypothesized to be task-dependent, no relation between response time and word frequency was found in either of the tasks. The findings are compared to previous research and suggestions for further studies are made.

Table of Contents

1. Introduction

When listening to speech or reading text, speakers of a language sometimes encounter semantic anomalies (e.g. *Mary drinks a tree*). These have to be detected, since a word might be misheard or used inappropriately. Although previous research has often investigated how semantic violations are processed (e.g. Hagoort, Hald, Bastiaansen & Peterson, 2004), it is unclear how they are detected. The current study sheds light on this question by investigating the effect of word frequency on the detection of semantic violations. Word frequency has often been shown to facilitate lexical processing (e.g. Schilling, Rayner & Chumbley, 1998). However, in a study investigating the detection of several types of violations, Van den Hazelkamp (2018) found that semantic violations with low-frequency words are responded to faster than anomalies with high-frequency words. She suggests that this could be due to the number of constraints low- and high-frequency words impose on their context. The present study further investigates this assumption by comparing the effect of word frequency on the detection of semantically anomalous words and the confirmation of the use of semantically correct words.

In the next section, previous research on semantic violations and the effect of word frequency on processing will be discussed. The research questions and hypotheses of the current research are presented in the third section. Section 4 describes the method used in the study and the results are reported in section 5. In the final section, the findings of the present study are discussed and conclusions are drawn.

2. Theoretical background

2.1 Semantic violations

Not all combinations of words than can be made according to the rules of syntax are semantically coherent. Consider the sentence *Mary drinks a tree.* Although it is perfectly grammatical, the sentence is semantically anomalous. The verb *drink* requires a liquid object, which a tree is obviously not. The constraints words impose on the context in which they can and cannot occur are called *selectional restrictions* (Warren & McConnell, 2007; Warren, Milburn, Patson & Dickey, 2015; Katz & Fodor, 1963) and are a part of the lexicon (Chomsky, 1965; Katz & Fodor, 1963; but see, Jackendoff, 2002; Elman, 2009). A violation of these constraints results in a semantically anomalous sentence. These constraints do not only apply to verb-argument selection, but also to the selection of modifiers. For example, the Dutch adjective *drachtig* (EN: pregnant) can only be said of animals, but not of humans. The adjective only selects nouns that are [+ ANIMAL]. Thus, both verb-argument and noun-modifier combinations can be semantically anomalous due to a violation of selectional restrictions (henceforth: SR).

Semantically anomalous sentences elicit different brain responses than normal sentences. In an EEG study investigating event-related potentials (ERP's), Kutas and Hillyard (1983) found that semantically inappropriate words elicited a large N400 component, which is a peak in negativity after 400 ms. These peaks were not present with semantically appropriate words nor with grammatical errors. This effect has often been replicated and has become known as the N400 effect (e.g. De Vincenzi et al., 2003; Hagoort et al., 2004; Paczynski & Kuperberg, 2012). However, Nieuwland and Van Berkum (2006) showed that the effect of these selectional restriction violations on the N400 can be superseded by discourse context. In their study, they found no N400 for semantically anomalous words, if the context contributed to an adjustment of the SR. Hagoort et al. (2004) reported an effect of SR violations, as in *the*

Dutch trains are sour, on the N400, but they found no difference in brain responses between these kind of semantic anomalies and sentences that are untrue according to general world knowledge. An example of the latter is *the Dutch trains are white,* although it is commonly known that Dutch trains are yellow. The authors therefore argue that the processes of detecting semantic and world knowledge violations happen at the same time and are connected. On the other hand, Paczynski and Kuperberg (2012) found that the effect of world knowledge violations on the N400 almost disappeared if the target word was semantically related. However, semantic relatedness did not weaken the N400 effect caused by violations of SR. This suggests that there is at least some difference in the processing of these two types of violations.

Studies investigating eye-movements during sentence reading also reveal differences in the processing of semantically appropriate and anomalous sentences. De Vincenzi et al. (2003) showed that reading times of words following a semantically anomalous word are longer than usual. Fodor, Ni, Shankweiler and Crain (1996) also demonstrated that regressions increased after encountering a semantically anomalous word. In addition, Rayner, Warren, Juhasz and Liversedge (2004b) found that gaze durations are longer for anomalous words than for plausible or implausible words. Contrary to the assumption of Hagoort et al. (2004) that SR and world knowledge violations are processed similarly, recent studies by Warren (Warren & McConnell, 2007; Warren et al., 2015) report differences in the eye-movements in these conditions. First fixations and single fixations are longer for words that violate SR, than for words that are highly implausible (Warren & McConnell, 2007). Also, longer reading times were found for text following words causing SR violations, but not following implausible words. Warren et al. (2015) report that total reading time of a target word was affected by SR violations, but not by whether the sentence was possible according to world knowledge. For example, target words in sentences like *Corey's hamster lifted a* *nearby backpack*, which is impossible according to general knowledge, were read faster than in sentences like *Corey's hamster entertained a nearby backpack*, which is impossible according to a violation of SR. The same effect was found for first fixation and gaze duration on the region after the target word. Warren and McConnell (2007) argue that violations of SR are probably noticed as soon as the words are semantically interpreted. Warren et al. (2015) suggest this is due to the fact that SR violations are "coarse-grained abstractions" (p. 938) which are more easily identified. This is also reflected in the longer processing time of the words following the target, since these 'abstractions' are more difficult to adjust than more specific inferences of world knowledge. Thus, evidence from eye-tracking experiments suggests that violations of SR are not only processed differently from normal sentences, but also from sentences that violate general world knowledge.

Semantic anomalies have also been found to influence reaction times, indicating higher processing costs. Marslen-Wilson, Brown and Tyler (1980) found that participants were slower in identifying a target word in speech when that word caused a semantic violation than in a normal condition. Fodor et al. (1996) used a cross-modal lexical decision task to investigate the effect of semantic anomalies. Participants had to decide whether a given word was an existing word or a non-word while listening to compressed speech containing syntactic or semantic violations. Reaction times to the lexical decision task were slower when participants heard a semantic anomaly at the same time, suggesting that semantic violations increase processing load. To examine the detection of semantic violations more directly, Fodor et al. also conducted an anomaly detection experiment. Participants listened to compressed speech and were asked to press a button when they noticed an anomaly. On average, semantic violations were identified after 1500 ms, while response times to syntactic errors were significantly shorter. The authors argue that semantic violations are likely to be detected earlier than the actual button-press, but that the final decision is delayed, because "the processor (…) anticipate[s] that the apparent problem would fade away" (p. 45). Although response times in a "baseline" condition are longer than with both types of anomalies, it is not described how this condition was assessed and accuracy in this condition was extremely low (<10%). Van den Hazelkamp (2018) investigated the detection of anomalies in reading. She asked her participants to read sentences which were presented word by word. They were instructed to press a button as soon as they noticed a syntactic or semantic violation. The semantic anomalies in this experiment were due to a violation in the selection of an adjective or a subject. Like in Fodor et al.'s study (1996), semantic violations were detected later than syntactic violations. Also, it took participants longer to identify violations of verb-selectional restrictions, compared to adjective-noun combinations. However, in this study, there was no control condition assessing response times to semantically correct sentences, so it remains unclear how quickly participants recognize the correct use of a word.

Semantically anomalous sentences are processed differently from appropriate sentences and from sentences violating general world knowledge. Several studies have shown that semantic violations increase processing cost. However, it is not yet clear how violations of SR are identified and how quickly they are detected. Also, no previous research has compared the detection of semantic anomalies to responses to semantically correct words, although this could give more insight in the way these violations are identified.

2.2 Word frequency

The number of times a speaker encounters a word affects how fast that word is processed. Many studies have shown that there is a robust effect of word frequency on reaction times, with high-frequency words being responded to faster than low-frequency words. In lexical decision tasks (LDT's), participants see a string of letters that is either an existing word (e.g. *book*) or a nonword (e.g. *flane*). They are asked to decide as quickly as possible whether the visual stimulus is a word or a nonword. In this task, word frequency has been shown to significantly affect response time, with high-frequency words being recognized faster than words with a low frequency (Hudson & Bergman, 1985; Scarborough, Cortese & Scarborough, 1997; Forster & Chambers, 1973; Balota, Cortese, Sergent-Marschall, Spieler & Yap., 2004; Keuleers, Brysbaert & New, 2010; Rubenstein, Garfield & Millikan, 1970; Balota & Chumbley, 1984). Several researchers (Hudson & Bergman, 1985; Balota & Chumbley, 1984) argue that it is important to control for word length, since low-frequency words are often longer than high-frequency words (Hudson & Bergman, 1985) and word length has also been found to affect reaction time in an LDT (Hudson & Bergman, 1985; Balota et al., 2004). However, even when word length is taken into account, high-frequency words are still responded to faster than low-frequency words (e.g. Hudson & Bergman, 1985; Keuleers, Brysbaert & New, 2010). In addition, Scarborough et al. (1977) found that the effect of word frequency was strongly reduced by repetition, suggesting that the faster recognition of high-frequency words might partly be due to recency. Similar effects of word frequency on response time have been found in word naming tasks. In this task, participants are asked to name a word as quickly as they can. As in the LDT, high-frequency words are responded to faster than low-frequency words (Forster & Chamber, 1973; McRae, Jared & Seidenberg, 1990; Schilling, Rayner & Chumbley, 1998; Balota et al., 2004; Balota & Chumbley, 1984). However, Balota and Chumbley (1984) and Schilling, Rayner and Chumbley (1998) found that the effect of word frequency is smaller than in a LDT. Although length has also been argued to play a role in word naming, this effect is less consistent than in LDT's (Balota et al., 2004). Thus, in both LDT's and word naming tasks, frequency of occurrence affects response time.

Many researchers have argued that the effect of word frequency on response time in LDT and word naming tasks indicates that word frequency facilitates faster lexical access

(Forster & Chambers, 1973; McRae, Jared & Seidenberg, 1984; Schilling, Rayner & Chumbley, 1998; Scarborough et al., 1977). However, it could be argued that lexical access does not play an important role in tasks such as lexical decision and word naming, as words can be recognized or pronounced without retrieving their meaning. Balota and Chumbley (1984) argue that frequency effects in LDT's are due to other decision processes. In a category-verification task, they found no relation between response time and word frequency in the *no*-condition. Participants were instructed to identify as quickly as possible whether a target word belonged to a certain category by pulling one of two levers. According to the authors, if word frequency affects lexical access, participants should be faster to detect a wrong categorization of high-frequency words than of low-frequency words. The small effect of word frequency in the *yes*-conditions, this might be due to priming effects of the category. The authors suggest that effects of word frequency in an LDT are due to participants' judgment of the familiarity and meaningfulness of the word. In this analysis, word frequency could be expected to have little influence on judgments of its semantic content, since frequency does not affect the speed with which the meaning of a word is retrieved. However, experiments with semantic priming reveal that word meaning is accessed in an LDT. For example, Dannenbring and Briand (1982) found that mean reaction time to a word was shorter when a previous word was semantically related than when it was non-related. This effect can only possible if word meaning is retrieved in LDT. In addition, the strong correlation between the word frequency effects in both LDT and word naming tasks indicates that word frequency must affect the shared process of lexical access in these tasks (Schilling, Rayner & Chumbley, 1998). Although there is no evidence for an effect of frequency on category verification, previous research suggests that word frequency facilitates at least partial lexical access to the meaning of a word.

Other evidence that word frequency affects lexical access comes from experiments with sentence-reading. In eye-tracking experiments, participants' eye-movement is measured while reading sentences. This task could be argued to reflect more natural language processing, since words are presented in context, instead of separately as in LDT and word naming tasks. In addition, it is necessary that word meaning is retrieved in order to comprehend the sentence. This is usually checked by comprehension questions at the end of the sentence. Several studies report that fixations, the amount of time a word is looked at when first encountered, are longer for low-frequency words compared to high-frequency words (Rayner & Duffy, 1986; Schilling, Rayner & Chumbley, 1998; Rayner, Ashby, Reichle & Pollatsek., 2004a). A similar effect has been found for gaze durations, which is the total amount of time a word is looked at (Rayner & Duffy, 1986; Schilling, Rayner & Chumbley, 1998; Pollatsek, Juhasz, Reichle, Machacek & Rayner, 2008; Just & Carpenter, 1980; Rayner et al., 2004a). Longer reading times have been argued to reflect longer processing time (e.g. Rayner, 1998; Just & Carpenter, 1980; Rayner & Duffy, 1986). Thus, high-frequency words are processed faster than long-frequency words, as they require a shorter time to read. In addition, Rayner and Duffy (1986) have shown that gaze duration on the word following the target word is also longer when the target has a low frequency of occurrence. The authors argue that this indicates that it does not only take longer to process low-frequency words, but also to integrate them in the sentence. This suggests that the meaning of low-frequency words is retrieved later. As in LDT and word naming tasks, word length also affects gaze durations (Pollatsek et al., 2008), but word frequency has also been found to influence reading times low- and high-frequency words are matched on length (e.g. Schilling, Rayner & Chumbley, 1998). Other evidence for the faster processing of high-frequency words comes from EEG experiments (Van Petten & Kutas, 1990; Dambacher, Kliegl, Jacobs & Hofmann, 2006). In these experiments, event-related potentials were recorded while the participants silently read sentences. Similarly to semantic violations (e.g. Kutas & Hillyard, 1983), Van Petten and Kutas (1990) discovered that low-frequency words caused larger N400 amplitudes compared to high-frequency words. However, this effect was only established if the word occurred at the beginning of the sentence, but not sentence-finally. They therefore suggest that contextual information can diminish effects of word frequency. Dambacher et al. (2006) found a similar effect of word frequency on the P200 component of the ERP: The P200 amplitude was larger for low-frequency words than for high-frequency words. This component decreased when the target word was further in the sentence. In addition, the N400 was affected by predictability, more for low-frequency than for high-frequency words. Since there was only an interaction between predictability for the N400, but not for the P200, the authors suggest that low-frequency words are accessed later than high-frequency words. To conclude, both eye-tracking experiments and EEG experiments show that word frequency affects lexical access when reading sentences.

In the light of these findings, the results of Van den Hazelkamp (2018) are surprising. Whereas previous research showed a robust negative relation between word frequency and processing cost, her data suggest that word frequency has a positive effect on reaction time in a violation detection experiment. As described earlier, participants were asked to read sentences, which were presented word by word, and press a button when they detected a - grammatical or semantic - anomaly. Although no effect of frequency was found for the syntactic violations, participants were faster to detect a semantic anomaly with low-frequency words than with high-frequency words. The effect was less strong for violations of animacy constraints. However, the relation between word frequency and the detection of these types of violations were not statistically tested, as the number of data did not allow for statistical analysis. Since earlier research intimates that high-frequency words are recognized and accessed faster, an opposite pattern was expected. If high-frequency words are recognized faster, the contexts in which they can and cannot occur might also be retrieved sooner. Van den Hazelkamp suggests that the requirements for high-frequency words are relatively broad, which is why they can occur in many contexts. This could also explain why these words are so frequent. Low-frequency words, on the other hand, have very specific requirements and can only occur in a small number of contexts. Since their occurrence is so constrained, they are low in frequency. According to Van den Hazelkamp (2018), semantic anomalies with a low-frequency word can be detected faster, because the set of contexts in which it can be used correctly is smaller compared to that of a high-frequency word. Likewise, it takes more time to check whether a high-frequency word is semantically anomalous, since there are many possible correct situations: “to identify something as a violation, it is necessary to check all these possibly correct uses, which takes time” (p. 156). Thus, the opposite effect of frequency in a violation detection task could be explained by a difference in the number of SR.

3. The present study

Previous research suggests that violations of selectional restrictions increase processing cost. Semantic anomalies elicit a large N400 (e.g. Kutas & Hillyard, 1983), extend reading times (e.g. Rayner et al., 2004a) and delay response time in monitoring (Marslen-Wilson et al., 1980) and lexical decision tasks (Fodor et al., 1996). Similarly, word frequency also affects the N400 (Van Petten & Kutas, 1990) and low-frequency words require a longer time to be read, recognized and named (e.g. Schilling, Rayner & Chumbley, 1998) This indicates that word frequency facilitates lexical access. Interestingly, in a study investigating the detection of different violation types, Van den Hazelkamp (2018) found an opposite effect of word frequency on response times to semantic anomalies. In her study, semantic violations were detected faster if the target word was low in frequency. Van den Hazelkamp suggests that this is due to a difference in the number of constraints of high- and low-frequency words. However, the results of her experiment were not statistically tested. Therefore, the present study will attempt to replicate the findings of Van den Hazelkamp and investigate her assumptions regarding the detection of semantic violations by researching the effect of word frequency in a similar experiment. Since the mechanisms of detecting semantic anomalies and verifying correct words are hypothesized to be different, the current study includes two tasks: a negative semantic judgment task, in which participants have to detect semantic anomalies, and a positive semantic judgment task, in which they are asked to confirm the appropriate use of a word. The main research question that will be investigated in the current study is the following:

RQ: Is the effect of word frequency on response time in semantic judgment experiments task-dependent?

Two sub-questions were formulated to answer the main research question:

RQ1: How does word frequency affect the response time in a negative semantic judgment task?

RQ2: How does word frequency affect the response time in a positive semantic judgment task?

Regarding RQ1, it is hypothesized that low-frequency words have many specific requirements for their environment, resulting in a small number of contexts in which they can appear. As soon as one of these requirements is violated, a semantic anomaly is detected. On the other hand, the requirements for the contexts in which more frequent words can occur are less specific and these words can appear in many situations. Therefore, violation detection with more frequent words is expected to be relatively slow. Earlier research has shown that word frequency correlates positively with response times when participants have to detect semantic anomalies (Van den Hazelkamp, 2018). The hypothesis for RQ1 is therefore as follows:

H1: Word frequency positively affects the response time in a negative semantic judgment task.

It is expected that response times to semantically anomalous less frequent words are shorter than those to high-frequency words.

For RQ2, it is hypothesized that word frequency facilitates fast response times in this task. Evidence from LDT, word-naming tasks and eye-tracking experiments indicate that high-frequency words are processed faster than low-frequency words. In addition, Rayner and Duffy (1986) suggest that it takes longer to integrate a low-frequency word in a sentence. Since the requirements for high-frequency words are hypothesized to be less specific, they will be satisfied more easily, leading to a faster response time to the correct use of more frequent words. For low-frequency words, there are more requirements, as they can occur in relatively few contexts. Checking a larger number of requirements is expected to take longer. Thus, the following hypothesis is proposed for RQ2:

H2: Word frequency negatively affects the response time in a positive semantic judgment task.

We expect that response times to correct uses of high frequent words will be shorter than those to low frequent words.

From the previous hypotheses, it follows that the hypothesis for the main research question is the following:

H: The effect of word frequency on response time in semantic judgment experiments is task-dependent.

It is expected that word frequency positively affects response times in a negative semantic judgment task, but negatively in a positive semantic judgment task.

4. Method

4.1 Participants

Twenty-three adult native speakers of Dutch, who were unaware of the purposes of the study, participated in the experiment. They were on average 26 years old (SD = 11). Four of them were male, the others were female. None of the participants had dyslexia. All participants signed an informed consent before participating in the experiment.

4.2 Materials

Sixteen word pairs were created for this study, resulting in a total of 32 target items. The words in each word pair were of the same grammatical category, had the same length in letters and were matched in frequency. The latter means that the log frequency of the word form, based on the SUBTLEX-NL database (Keuleers, Brysbaert & New, 2010), did not differ more than 0.10 between the items in a word pair. For example, the items in word pair *fruit – kroon* (EN: *fruit – crown)* are both nouns of five letters and have a log word frequency of 2.80 and 2.84 respectively. The word pairs varied in frequency, ranging from a log frequency of 1.04 to 3.88. The target items were four (n=3), five (n=8) or six (n=5) letters long. They were evenly distributed among the various frequencies in such a way that there is no significant difference in word length between low- and high-frequency words. There were eight noun word pairs and eight verb word pairs. All items were singular and the verbs were in present tense. An overview of the word pairs, their log frequency and length can be found in appendix A.

For each item in each word pair both a semantically anomalous and a semantically correct sentence was created. Thus, for every word pair there were four sentences, resulting in a total of 64 sentences. In the semantically anomalous sentences, the target item is the only anomalous word. An example can be found in (a) and (b), with the target item printed in italics:

(a) Het zoete *fruit* past goed bij de zure smaak van de rabarber.

‘The sweet *fruit* goes well with the sour taste of the rhubarb.

(b) ! Het drukke *fruit* leidt tot lange rijen op het voetpad.

‘The busy *fruit* leads to long queues on the sidewalk.

The stimuli were inspired by the material used in Van den Hazelkamp (2018), but too many adjustments were made to allow for a direct comparison. The sentences followed the template det-noun-verb […] for the verb items and det-adj-noun-verb […] for the noun items. In both types of stimuli the target item was the third word of the sentence. This is also the first position where a semantic violation can occur, which was done to prevent any other effects of context. The sentences were ten to twelve words long. To rule out any effects of repetition, the start of each sentence was different. Since the words preceding the target word could affect the judgment of the following word, the adjective or noun preceding the target item were matched on frequency in the semantically anomalous and correct sentence. Also, to minimize other predictability effects, efforts were made to prevent the target item from being too predictable from the preceding context. Although no cloze probability could be established due to time limitations, participants indicated in an exit-interview that the pre-target word did not predict the third word.

Like the target items, the subject of each sentence is singular and all sentences are in present tense. The anomalies will never be due to a violation in animacy. Previous research (Van den Hazelkamp, 2018) has shown that response times to animacy violations differ from semantic violations, since many speakers process animacy violations as syntactical errors. Besides the target stimuli 32 filler sentences were created. Half of them were semantically anomalous and the other half were semantically correct. These fillers followed the same format as the sentences with the noun and the verb items. These were included to prevent that participants could always press the button as soon as the third word appeared. The materials were piloted with a few participants to test whether the stimuli sentences were judged as expected and modifications were made accordingly.

4.3 Procedure

The procedure was similar to that used by Van den Hazelkamp for the violation detection experiment (2018, p. 169). The participants were seated in front of a laptop in a quiet room. The words of each sentence were presented one by one on the screen. Unlike in Van den Hazelkamp's study, every word was visible for 500 ms followed by 500 ms of blank screen (instead of 300 ms) due to a technical mistake. Since this was only discovered after some participants had already completed the experiment, this was not corrected. The target word was printed in capital letters and the final word in each sentence was immediately followed by a period. The experiment was run using the software ZEP version 1.12.

The participants' task was to identify as quickly as possible whether the third word fits in the sentence in terms of meaning. Unlike in Van den Hazelkamp (2018), there were two types of tasks in the current study: a negative and a positive semantic judgment task. In the negative semantic judgment task, the semantically anomalous target sentences and the semantically correct fillers were used. The participants were instructed to press the space bar if the third word was semantically incorrect. In the positive semantic judgment task, the subjects were asked to confirm if the third word was semantically correct by pressing the space bar. The stimuli for this task were the semantically correct sentences and the incorrect fillers. The online response is the response time to the target words measured in milliseconds, starting from the moment that the word was presented. Previous research has shown that repetition affects reaction time in a lexical decision task and can also reduce the word frequency effect under certain conditions (Scarborough et al., 1977). The stimuli were therefore distributed among two versions. Although each participant evaluated each item in the word pair, no participant saw the same target item twice. For example, participant A saw *fruit* in a semantically correct sentence in the positive judgment task and *kroon* in a semantically anomalous sentence in the negative judgment task. Participant B saw a correct use of *kroon* in the positive task and an incorrect use of *fruit* in the negative task. The two versions can be found in appendix B.

After each sentence, the participants were asked to indicate again whether the sentence was semantically correct or not using the shift-keys. This will be referred to as the offline response. In each task, participants saw 32 sentences, excluding four practice items. Due to a technical error, the first item of version 1 in the positive task was excluded for a few participants. The order of the stimuli for each task was randomized for each participant and the order of the tasks was balanced. After having finished one task, they continued with the other task. Each task took ten to twelve minutes to complete. In total, the session lasted for approximately 25 minutes.

5. Results

Both offline judgments and online response times of each participant were recorded. Despite piloting the material, the offline responses to two sentences were judged differently from expected in more than 20% of the cases. For example, a target sentence from the positive task was rejected in more than 20% of the offline judgments. All response times to these sentences were therefore removed. On average, the other sentences were judged like expected in 98% and 96% of the time in the positive and negative task respectively. If a participant´s online response to a target item did not match their offline response, then their response time to that item was deleted. There were four participants for whom online and offline judgments did not systematically (>80%) match in a certain task. These participants possibly did not understand the task or were not concentrated. The online response times of these participants were therefore removed. Finally, the resulting data of each task was checked for outliers per participant and per sentence. Although none of the sentences revealed an abnormal pattern, the mean response time of one participant in the positive judgment task (6870 ms) was more than two standard deviations removed from the mean in that task (2112 ms; SD = 1621). The mean response time indicates that he or she only judged the appropriateness of the third word at the (one but) last word of the sentence, instead of assessing the compatibility of the adjective and noun or subject and verb. Therefore, these data were also left out of the analysis.

The mean response time to target words was 2055 ms (SD = 1560). For the positive judgment task, participants responded on average after 2112 ms (SD = 1621), while in the negative task the mean response time was 1999 (SD = 1498). A multi-factor analysis of variance (ANOVA) was performed to investigate if the effect of task was significant. The ANOVA also tested for effects of stimuli type (noun vs. verb), version, order (whether or not it was the first task) and word frequency category. For this last variable, the stimuli were divided into four categories ranging from low to high frequency, each category consisting of two noun pairs and two verb pairs. For each task and stimuli type, the average response time was calculated per participant. The difference in response times between the two tasks was not significant [$F(1, 261) = 0.16$, $p = .69$]. No main effect of type of stimuli [$F(1, 261) = 0.15$, $p = .90$], version [$F(1, 261) = 0.03$, $p = .87$] or order [$F(1, 261) = 0.04$, $p = .85$] was found. Word frequency did not significantly affect response time [$F(3, 261) = 1.83$, $p = .14$], nor did word frequency and task interact [$F(3, 261) = 0.93$, $p = .43$]. There was also no significant interaction between word frequency and stimuli type [$F(3, 261) = .23$, $p = .88$], order [$F(3, 261) = .14$, $p = .94$], or version [$F(3, 216) = .60$, $p = .62$]. However, the interaction between task and order was significant [$F(1, 576) = 9.33$, $p = .002$]. As can be seen in Figure 1, participants in the positive judgment task were faster when that task was their second task. The opposite pattern is true for the negative judgment task, in which participants are faster when it is their first task. This suggests that response times are faster when participants start with the negative judgment task. Although this result is unexpected, it is assumed that this will not harm the results, because half of the participants started with the negative judgment task and the other half with the positive judgment task. None of the other two-way interactions were significant: task x stimuli type [$F(1, 216) = 1.28$, $p = .26$], task x version [$F(1, 216) = 0.29$, $p = .59$], order x stimuli type [$F(1, 216) = 0.19$, $p = .66$], nor order x version [$F(1, 576) = 2.62$, $p = .11$]. There were also no significant three-, four- or five-way interactions ($p > .05$).

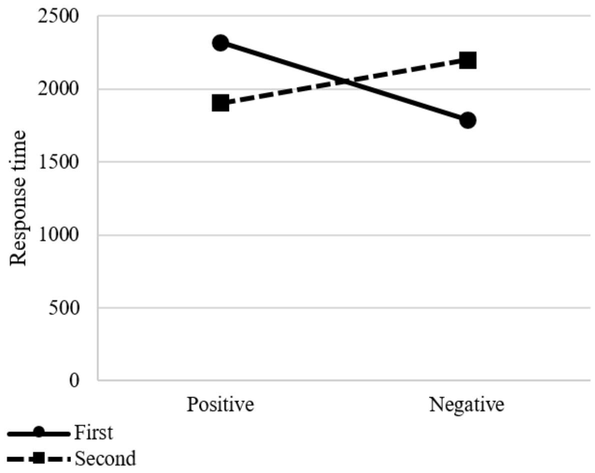

Figure 1. Mean response time (ms) per task for each order

To investigate the effect of word frequency more closely, a Pearson correlation test was performed with the mean response time per word pair. The slightly negative effect of word frequency on response time was not significant (*r* = -.18, *p* = .51). In Figure 2, the mean response time per word pair can be seen for each task. Pearson correlation analyses revealed no significant effect of word frequency on response time in neither the positive (*r* = -.19, *p* = .48) nor in the negative task (*r* = -.06, *p* = .83). This confirms that there is no significant effect of word frequency on response time in this study.

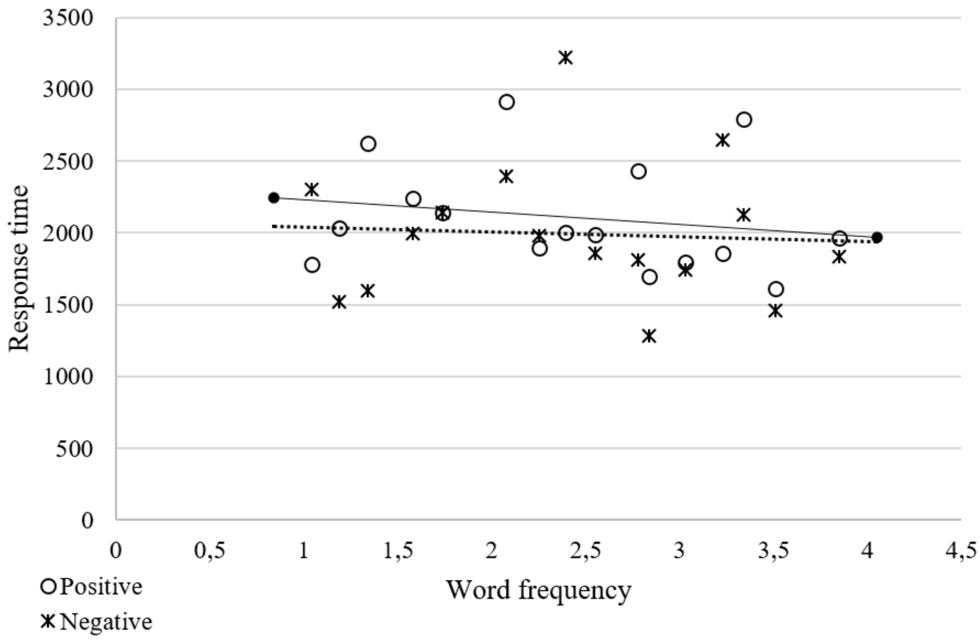

Figure 2. Relation between log word frequency and mean response time (ms) per task

6. Discussion and conclusion

The aim of the present study was to investigate the effect of word frequency in two tasks: a positive and a negative semantic judgment task. Participants were presented with sentences, appearing word by word, and were asked to press a button when the third word was used anomalously or correctly, depending on the task. It was hypothesized that word frequency would affect response time differently in the two tasks. Following the findings of Van den Hazelkamp (2018), semantic violations with low-frequency words were expected to be detected earlier than high-frequency words in the negative judgment task. The results did not confirm this hypothesis, since no effect of word frequency on the detection of semantic anomalies was found. In the positive judgment task, in which participants had to confirm the correct use of a word, an opposite pattern was expected, since high-frequency words are hypothesized to have broader SR and earlier research suggests that these words are processed faster. However, word frequency did not affect response time in this task either. Thus, the hypothesis that the effect of word frequency in semantic judgment experiments is task-dependent cannot be confirmed.

Unlike in Van den Hazelkamp (2018), no positive relation between word frequency and the detection of semantic anomalies was found in this study. This could be due to a variety of reasons. First of all, it is possible that the effects of frequency and the relative difficulty of detecting semantic anomalies with high-frequency words cancel each other out. Whereas frequency is assumed to facilitate processing, Van den Hazelkamp suggests that it might complicate anomaly detection. If these effects are equally strong, this could explain why no effect of word frequency was found in this study. Secondly, since the results of Van den Hazelkamp were not statistically tested, it is also possible that the effect she found was a coincidence, resulting in a Type I error. Also, the participants in this study knew where the violation might occur, unlike in Van den Hazelkamp's experiment. This might have confounded the results, since participants knew when to pay attention.

The results of the current study are also not in line with previous research on the effect of word frequency, since no negative effect of word frequency was found in either the positive or the negative task. This is surprising, since earlier studies have consistently shown that high-frequency words are processed faster. If high-frequency words are read and recognized faster, participants can start the process of checking its correct use earlier. This suggests that some effect of frequency could be expected. It could be argued that there is no relation between word frequency and the relative ease of detecting a semantic violation. Like word frequency has been shown not to affect category verification Balota & Chumbley, 1984), it is possible that it also does not affect the detection of semantic anomalies. Thus, frequency might only facilitate partial lexical access. However, the lack of a frequency effect in Balota and Chumbley's study (1984), might be due to a similarity in the mechanisms of category verification and the detection of semantic violations. Recall that Van den Hazelkamp suggests that high-frequency words have less requirements, which results in that they can appear in more contexts. Similarly, it could be assumed that high-frequency words are part of more semantic categories. For example, the frequent word *table* belongs to categories like *furniture,* but also to *things in a classroom* or *things in an operating room.* Low-frequency words (e.g. *stall*) only belong to a small number of categories (e.g. *things on a market*). When a participant has to detect a mismatch between a category and a target word, it is easier to do so if there are only a few categories to which an item can belong. This might explain why low-frequency words are responded to faster than high-frequency words, since possible categories are eliminated faster. Other effects of word frequency on processing might thus have been cancelled out in Balota and Chumbley's category verification task.

One explanation for the results of the present study might be that it is not the frequency or the number of constraints of the target word that predicts the time it takes to detect an anomaly, but that of the preceding word. More specifically, upon encountering a semantically odd combination (e.g. *sweet crown*), participants can adopt two different strategies. Since there is a violation if one of the two words is used incorrectly, they only have to prove the incorrect use of one of these words. They can either check whether the target word, *crown,* is used correctly, or whether the preceding word, *sweet,* is appropriate. If they find that one of these words does not satisfy the constraints of the other, a violation is detected. It could be argued that speakers choose the strategy which requires the least effort, which in some cases will be to check the constraints of the preceding word. However, this suggests that participants in some cases did not judge the appropriateness of the target word, but of the preceding word.

The lack of significant results could also be due to a task-inherent problem threatening the validity of the experiment. Response times to semantic violations in this study are even slower than in Fodor et al. (1996), which could be due to a difference in presentation rate. Fodor et al. used compressed speech twice as fast as normal speaking rates, while in this study one word was presented every second. More importantly though, the authors of that study suggest that semantic anomalies are actually detected earlier than the actual button-press, but that participants postpone the button-press to see if the problem caused by the semantic violation is resolved in the remainder of the sentence. Evidence for an earlier detection of semantic violations than the button-press comes from their effect on the N400 (Kutas & Hillyard, 1983), processing cost (Fodor et al., 1996) and longer reading times (Rayner et al., 2004b). Violations of SR can be adjusted by the discourse (Nieuwland & Van Berkum, 2006), which might explain why the actual response time is delayed, because the problem might be solved by the context in the rest of the sentence. Also, longer reading times of words following a semantically anomalous word (De Vincenzi et al., 2003), suggest that readers try to integrate the word in the sentence. Fodor et al. (1996) suggest that participants wait for the next words before making a definite decision. A similar process might have affected the response time in the current study. Although participants were encouraged to press the button as soon as possible and could correct their decision afterwards with their offline response, they did see on average at least one more word before making a decision. In this analysis, the task does not assess how quickly SR violations are detected, but how fatal they are. This delay might explain why no effect of word frequency was found in this study.

The only effect that was found to be significant in this study was the interaction between order and task. This result indicates that participants who started with the negative task had shorter response times in both tasks. This could be due to a learning effect. In the negative judgment task, the participants might have noticed that an initial violation was not solved by the rest of the sentence. Therefore, they learned that they could press the button as soon as they noticed an anomaly. Continuing with the positive task, they assumed that a correct combination of the subject-verb or adjective-noun pair was not violated in the remainder of the sentence. It might have been less clear for participants starting with the positive task that initially correct combinations did not become anomalous due to the context in the remainder of the sentence, explaining why their judgment was delayed in the positive task. In the negative task, they might postpone their judgment as well, anticipating that it might be resolved in the context. Another possibility is that participants' first task affected their strategy for the second task. After completing for example the positive judgment task, participants' strategy for the negative judgment task might be to press, unless the target word is semantically correct. Participants starting with the negative judgment task, on the other hand, would concentrate on detecting semantic violations, instead of on the correct use of the target word. However, in this analysis, some delay in the second task could be expected, but this was not found. Future studies could solve this problem, by using a between-participants design. If one half of the participants performs the negative task and the other half does the positive task, there will be no effect of learning. However, it is still possible that the strategies in the two tasks differ, since participants in the positive task can also choose to press the button unless they detect an anomaly.

A number of limitations of the current study need to be considered. First, the stimuli were not controlled for cloze probability, due to time constraints. Earlier research (Rayner et al., 2004a; Dambacher et al., 2006) indicated that predictability also facilitates faster processing, though the effect is larger for low-frequency words. This is why previous studies often controlled for cloze probability by asking participants in a pre-test to finish sentences in order to assess how likely it is for the target word to follow the preceding words. Although cloze probability was not established in the current study, in an exit-interview participants stated that they did not predict the upcoming word from the first two words they read. However, some word combinations could be considered more logical or frequent than others. For example, compare *the clothes hang* to *the sand slips*. This might have confounded the results and should therefore be taken into account in follow-up studies. Also, unlike in Van den Hazelkamp’s study (2018), words were presented for 500 ms, with a 500 ms break between words. This presentation rate is relatively slow, which might have affected the result in several ways. Some participants indicated afterwards that the task was quite tedious, because of the slow presentation rate. This could have resulted in a decrease in attention at the end of the task. Also, the slow presentation does not reflect natural reading very well, since in normal reading words are fixated on for around 250 ms and many words are skipped (Rayner, 1998). Finally, the slow presentation rate might have given participants more time to predict the upcoming word. As explained above, this could also have influenced response time.

The experiment described in this study might not have reflected participants' immediate judgment on the violations of SR. As discussed earlier, participants might have postponed their response to see if the context resolved the violation in the earlier part of the sentence. On the one hand, this could be solved by adjusting the material. Similarly to Hagoort et al. (2004), predicate adjectives could be used (e.g. *The fruit is sweet/!busy*). However, this complicates a comparison of the two tasks. In the current study, the target word was always the third word of the sentence, so the participants knew which word they had to assess. If the structure of the stimuli is changed, it will become unclear for the participant what the target word is. Although it could be presented in capital letters, like in the present study, this might delay judgments for the positive task. The participants would then first have to make out that the word they see is the target word, before they can decide whether it is appropriate in the sentence. For the negative task, they can press as soon as they detect an anomaly. Caution is therefore required when response times in these two tasks are compared.

Since the present experiment might not have assessed the initial detection of semantic violations, it would be interesting to investigate the effect of word frequency on the processing of semantic anomalies more indirectly. Several researchers argue that indirect measures should be used to investigate the processing of language (Fodor et al., 1996; De Vincenzi et al., 2003). Future research could therefore investigate how word frequency affects the detection of semantically anomalous sentences using different methods. Since previous research has shown that both low-frequency words and semantic anomalies lengthen reading times (Rayner et al., 2004a; 2004b), it would be interesting to see whether these factors interact in eye-tracking experiments with sentences containing semantic violations. Also, earlier studies demonstrated larger N400s for low-frequency words and semantic violations (Van Petten & Kutas, 1990; Kutas & Hillyard, 1983). Possibly, this effect is strengthened for semantic violations with low-frequency words. Eye-tracking experiments and EEG studies are only an example of other ways in which the detection of semantic violations could be examined. However, combining direct and indirect assessments of the detection of semantic violations could give us a more complete picture of how they are identified and how word frequency affects this process.

Future research should also examine the relation between word frequency and the number of SR more closely. However, SR have been argued to be difficult to operationalise (Jackendoff, 2002). Some studies (e.g. Keuleers et al., 2010) indicate that context diversity (CD) is also a good predictor of reaction time in LDT's. This could be seen as a proxy for the number of SR, because Van den Hazelkamp (2018) suggests that the small number of SR of high-frequency words allow for them to appear in more contexts. However, in these studies, context diversity is defined as the number of texts or movies a word occurs in (Keuleers et al., 2010) and does not provide information about the direct context of a word. More accurate measures of assessing the possible combinations of words are therefore needed. This might give us a deeper insight into the nature of word frequency and semantic violations. If the SR of a word can be quantified, this might also predict the detection of semantic anomalies more accurately. This would help us to understand how semantic violations are detected and give us more insight into the processing of language.

The current study investigated the effect of word frequency on the detection of SR violations and verification of the correct semantic content of a target word. It was hypothesized that this effect would be task-dependent, with word frequency facilitating positive judgments, but delaying negative judgments. However, no relation between word frequency and response time in either task was found. The findings were compared to previous research about the effect of word frequency and the processing of semantic violations and possible explanations for the results were discussed. Further research is necessary to investigate the effect of word frequency on SR violation detection more closely.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and performance, 10,* 340-57.

Balota, D. A., Cortese, M. J., Sergent-Marschall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283-316.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge: MIT Press.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084,* 89-103.

Dannenbring, G. L, & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology, 36,* 435-44.

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A. Penolazzi, B., Ciccarelli, L. & Vespignanie, F. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language, 85,* 280-96.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 1–36

Fodor, J. D., Ni, W., Crain, S. & Shankweiler, D. (1996). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research, 25,* 25-57.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behaviour, 12*, 627-35.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–41.

Hazelkamp, C. van den (2018). *Individual differences in processing syntactic and semantic information: Looking for subpopulations* (Doctoral dissertation, LOT).

Hudson, P. T., & Bergman, M. W. (1985). Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. *Journal of Memory and Language, 24,* 46-58.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87,* 329-54.

Katz & Fodor (1963). The structure of semantic theory. *Language, 39,* 170-210.

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior research methods, 42,* 643-650.

Kutas, M. & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies.

Marslen-Wilson, W., Brown, C. M. & Tyler, L. K. (1980). Lexical representations in spoken language comprehension. *Language and Cognitive Processes, 3,* 1-16.

McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language, 29,* 43-65.

Nieuwland, M. S. & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18,* 1098-1111.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language, 67,* 426-48.

Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 726-50.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191-201.

Rayner, K., Ashby, J., Pollatsek, A. & Reichle, E. D. (2004a). The effects of frequency and predictability on eye fixations in reading: Implications for the EZ Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720-32.

Rayner, K., Warren T., Juhasz, B. J. & Liversedge, S. P. (2004b). The effect of plausibility on eye-movements in reading. *Journal of Experimental Psychology: Language, Memory, and Cognition, 30,* 1290-1301.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of verbal learning and verbal behavior, 9,* 487-94.

Scarborough, D. L, Cortese, C. & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experiment Psychology: Human Perception and Performance, 3,* 1-17.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual difference. *Memory & Cognition, 26,* 1270-81.

Warren, T. & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14,* 770-75.

Warren, T., Milburn, E., Patson, N. D. & Dickey, M. W. (2015). Comprehending the impossible: What role do selectional restriction violations play? *Language, Cognition and Neuroscience, 30,* 932-39.

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brainpotentials. *Memory & Cognition, 18,* 380-93.

Appendix A – Overview of the word pairs

Overview of the noun (N) and verb (V) word pairs, their log frequency (F) and length (L).

	Type	L	Mean F	Version 1	F	Version 2	F
low	V	5	1.04	ketst	1.04	suist	1.04
	N	4	1.19	leus	1.18	vijl	1.20
	N	5	1.34	kegel	1.30	fabel	1.38
	V	6	1.58	grenst	1.59	splijt	1.57
mid-low	N	5	1.74	sluis	1.76	kraam	1.71
	V	6	2.08	zweeft	2.07	glijdt	2.09
	V	4	2.25	lekt	2.26	tikt	2.23
	N	5	2.39	debat	2.40	plein	2.38
mid-high	V	5	2.55	weegt	2.55	deugt	2.55
	N	5	2.78	kroon	2.80	fruit	2.75
	N	6	2.84	schuur	2.84	heuvel	2.84
	V	6	3.03	stinkt	3.02	breekt	3.03
high	V	5	3.23	stopt	3.22	trekt	3.24
	N	6	3.34	geluid	3.32	eiland	3.35
	V	5	3.51	duurt	3.55	hangt	3.47
	N	4	3.85	boek	3.82	film	3.88

Appendix B – Stimuli sentences

Overview of the stimuli sentences that were used in the experiment with the target words in capital letters. The practice items appeared in the same order, but the order of the test items were randomized for each participant. Participants would see a different version in the positive and negative task. Stimuli which were not judged as expected in less than 80% of the time, are underlined. These were not included in the analysis.

Positive semantic judgment task

Practice

1. De hortensia BLOEIT dit jaar erg mooi dankzij het warme weer.
2. Het tropische MEISJE wordt in de lange vakanties vaak bezocht.
3. De onbeleefde JONGEN beledigt de juffrouw die vervolgens onthutst achterblijft.
4. De ezel ONTWERPT koffers die bijna niets wegen maar wel groot zijn.

Version 1

1. De steen KETST tegen de muur van de kleine fietsenstalling.
2. De duidelijke LEUS staat op de deur van het gemeentehuis.
3. De bovenste KEGEL is oranje, de andere kegels zijn bruin.
4. De tuin GRENST aan het mooie natuurgebied dat beschermd is.
5. De dichte SLUIS moet eerst worden geopend voordat de sloep verder kan.
6. Het papier ZWEEFT door de lucht nadat de secretaresse is gestruikeld.
7. De theepot LEKT omdat er een kleine barst in zit.
8. Het felle DEBAT voor veel onrust bij het publiek.
9. De stoel WEEGT meer dan de verhuizer vooraf had ingeschat.
10. De zilveren KROON ligt in een vitrine en wordt goed bewaakt.
11. De vieze SCHUUR staat op het erf van de oude boerderij.

12. De kaas STINKT volgens het meisje maar haar vader vindt dat lekker.
13. De taxi STOPT niet voor het verkeerslicht dat op rood springt.
14. Het zachte GELUID is bijna niet hoorbaar maar wel irritant.
15. De bijeenkomst DUURT niet lang als de meningen niet veel verschillen.
16. Het voordelige BOEK ligt in veel winkels in het kleine dorp.

Version 2

1. De pijl SUIST door de lucht rakelings langs de muur.
2. De platte VIJL ligt op de werkbank van de monteur.
3. De interessante FABEL wordt al generaties overgedragen van vader op zoon.
4. De boom SPLIJT na de harde inslag van de bliksem.
5. De ouderwetse KRAAM staat op de oprit van de boerderij.
6. Het zand GLIJDT door de vingers van het schattige kind..
7. De bom TIKT al een uur maar is nog niet ontploft.
8. <u>Het levendige PLEIN ligt in het centrum van de metropool.</u>
9. De vertaling DEUGT niet volgens de ervaren docent die Duits geeft.
10. Het zoete FRUIT past goed bij de zoute smaak van de haring.
11. De zonnige HEUVEL ligt in het glooiende landschap van Limburg.
12. Het touw BREEKT als de groep jongens er aan trekt.
13. De show TREKT zowel in Europa als Amerika duizenden bezoekers.
14. Het particuliere EILAND is niet toegankelijk voor toeristen en dagjesmensen.
15. De kleding HANGT in het magazijn van de grote winkel.
16. De historische FILM wordt een groot succes in de Verenigde Staten.

Fillers

1. De sok KOLKT ontzettend zacht maar past niet bij de broek.

2. De wolk GIST hoog in de lucht maar is goed te zien.
3. De kroeg SMELT erg vol vandaag door de belangrijke wedstrijd.
4. De kruik REMT in de wieg van de pasgeboren baby.
5. De tomaat KLINKT in de fruitschaal tussen de bananen en appels.
6. Het ei PRIKT naast de bak met het voer in de ren.
7. De put ZINKT in het bos dat tussen de dorpen in ligt.
8. De parfum SCHEURT snel door de fout in de formule.
9. De vergeefse DRANK betekent de zoveelste teleurstelling voor de vrouwelijke minister.
10. De ronde GEUR komt uit de collectie van de jonge kunstenaar.
11. De mogelijke SLEE is volgens de directeur niet goed genoeg.
12. De smeuïge TEGEL ligt op het bord van de kok.
13. De comfortabele LIJST staat in de hoek van de studentenkamer.
14. De zuivere TAART is erg kostbaar en ligt in een kluis.
15. De interactieve NEUS gaat over de geschiedenis van de wetenschap.
16. De kalme HAMER ligt in de zon tussen de reusachtige bergen.

Negative semantic judgment task

Practice

1. De tafel LOOPT in de grote zaal naast de ingang.
2. De stoere AGENT sluit de inbreker op in de cel.
3. Het meisje ONTWIJKT het liefst de vragen van de leraar.
4. De drachtige ATLEET staat voorlopig nog in de ruime stal.

Version 1

1. De kelder KETST nog steeds niet vol met voedsel en dekens.
2. De biologische LEUS is erg populair bij jonge stellen met kinderen.

3. De interne KEGEL blijkt toch levensgevaarlijk voor de zwakke bejaarde.
4. De hoed GRENST volgens de dame op de onderste plank.
5. De rauwe SLUIS moet eerst worden gebakken voordat die gegeten kan worden.
6. De schade ZWEEFT toch nog erger dan de huisbaas had gedacht.
7. De scherf LEKT op de vloer als de man het aanrecht afneemt.
8. Het maximale DEBAT is niet voldoende om deze locatie te kunnen huren.
9. De rivier WEEGT eerst door de bergen voor ze uitmondt in zee.
10. De verse KROON ligt in mooie manden in de winkel.
11. De politieke SCHUUR staat hoog in de peilingen van de radio-omroep.
12. Het lied STINKT over het terras waar de toeristen een ijsje eten.
13. De richting STOPT niet op het navigatiesysteem vanwege een storing.
14. Het rechte GELUID past precies tussen de twee overgebleven bochten.
15. De uitgang DUURT niet aan het einde van deze gang.
16. Het akoestische BOEK staat tussen de elektrische gitaar en het keyboard.

Version 2

1. De kermis SUIST in het weiland naast het Groningse dorp.
2. De constante VIJL maakt dit voertuig een aangenaam en prettig vervoersmiddel.
3. De koude FABEL is nog lekkerder op een warme zomerdag.
4. Het bier SPLIJT als de jonge barman het in het glas schenkt.
5. De morele KRAAM wordt door de progressieve filosoof anders geïnterpreteerd.
6. De aanklacht GLIJDT niet in de documenten die de advocaat heeft verstuurd.
7. De kus TIKT volgens het bruidsmeisje het mooiste moment van de bruiloft.
8. Het landelijke PLEIN is na twee jaar weer gehouden in het stadion.
9. De periode DEUGT niet aan het begin van het jaar.
10. Het drukke FRUIT leidt tot lange rijen op het voetpad.

11. De gezamenlijke HEUVEL lijkt hun vriendschap op de proef te stellen.
12. De suiker BREEKT in een schaaltje op de tafel naast de bank.
13. Het feit TREKT de gezinsleden die het nieuws aanvankelijk niet geloven.
14. Het eenmalige EILAND hoeft allen betaald te worden bij het aanmelden.
15. De bibliotheek HANGT in het centrum van de mooie stad.
16. De afgelegen FILM staat niet eens op de uitgebreide kaart.

Fillers

1. De nieuwe AUTO staat voor de deur van de buurman.
2. De hevige REGEN is een verkoeling na de hete dag.
3. De zure SAUS past goed bij de bittere groenten in het gerecht.
4. De gele MANTEL valt erg op tussen de zwarte jassen.
5. Het verlaten STRAND wordt door de talentvolle fotograaf prachtig vastgelegd.
6. De antieke LAMP staat in het bekende museum in Amsterdam.
7. De ondiepe SLOOT ligt verborgen tussen de weilanden waar koeien grazen.
8. Het steile PAADJE leidt naar een vervallen klooster uit de middeleeuwen
9. De schutting ZWIEPT bij harde wind maar bij rustig weer niet.
10. De trui PLUIST nadat de vrouw hem met de hand heeft gewassen.
11. Het overleg VERGT veel concentratie als we de notulist moeten geloven.
12. De rok PLOOIT zo mooi dat de moeder hem meteen koopt.
13. De slagroom SCHIFT als de assistent het probeert te kloppen.
14. De olie SIST in de pan die op het vuur staat.
15. De voorraad SLINKT na de verkeerde levering van de slager.
16. De pen VLEKT in het schrift van de ijverige leerling.