Universiteit Utrecht

Bachelor Thesis

7.5 ECTS

Correlation between memory recall and word similarity in sentence processing

Author:	Joep Nouwens
Student Number:	6537618
First Supervisor:	dr. J. Dotlačil
Second Supervisor:	dr. C. Strauch
Degree:	Bsc Artificial Intelligence

July 2, 2021

Abstract

Regarding investigations on sentence processing, research is often done using complex models that use various parameters, which leads to opacity of models of human language. In addition, this causes the models to be extremely difficult to compare with other models from earlier experiments. With the intention of finding alternative features for computational models and to encourage further research, we aim to find a correlation between memory recall and word similarity in sentence processing. To test this theory, the outlines of a similar research of Smith & Vasishth (2020) were implemented, to investigate if reading times were affected by different word similarities regarding presupposition resolution. Reading times were subtracted from a memory recall experiment carried out by Jan Winkowski (currently unpublished) to compare reading data to cosine similarities between certain words in each sentence. Results indicate an unforeseen effect of word similarity on reading times, since reading times increased when word similarities became greater.

Keywords: *memory recall, sentence processing, word similarity, reading times*

Table of contents

1. Introduction

In the last couple of years, researchers have done many investigations on sentence processing to create computational models that generate predictions on human language behaviour. To increase the completeness and accuracy of these models, much more investigations have to be done in terms of sentence processing. However, many computational models created in investigations on sentence processing rely on a large amount of different parameters (Smith & Vasishth, 2020): models based on cue-based memory retrieval (Engelmann et al., 2019), self-organization (Smith et al., 2018) and expectation-based parsing (Futrell & Levy, 2017) aim to explain sources of human sentence processing. These models all work with many linguistic features that are used to create an optimal model, but they lack transparency and consistency. This is why computational models do not have any similarity at all, which leads to the fact that the investigations cannot be compared to each other very easily, because every created model relies on different features.

As already discussed, various theories and parameters have been used in order to create different computational models that represent human sentence processing. These models increasingly rely on competing psycholinguistic theories, which causes the models to take a huge variety of forms. The goal of these models is to explain parts of sentence processing done by humans. However, the linguistic features used to create such a model are not pre-arranged parameters: all modelers use their own set of linguistic features. Researchers prefer building an optimal model over a transparent one but this results in opacity, due to the fact that there is a large number of hidden degrees of freedom that go into the models. Parameters of the models are hidden and therefore cannot be explained, which leads to a great weakness of each model. So if people want to create a strong, complete computational model that perfectly represents people's behaviour, they must work with the same parameters.

To avoid the usage of a huge, complex set of hidden parameters to make a computational model, this experiment will only focus on one parameter to compare sentence processing done by humans. The main goal is to investigate whether computational models can be used to predict and describe reading times with resolution of presuppositions by only using word similarity. An investigation by Mitchell et al. (2010) tells us that research suggests that language comprehension can be highly predictive, which means that people are able to anticipate upcoming material. In this thesis, word similarity will be varied to measure whether humans really anticipate quicker on words later in the sentence, if words are more similar. This will be done by comparing memory recall to word similarities between certain words connected by presuppositions. Reading times will be used to represent memory recall, since a strong connection between these two components are found in a recent study of Smith and Vasishth (2020). In addition, vector space semantics will be used to calculate cosine similarities between words in a sentence. All these elements combined lead us to our main question: how do memory recall and word similarity interact in sentence processing?

To answer this question, findings of related papers will be discussed and applied to this thesis. A detailed set of 32 sets of sentences together with reading times for each sentence from an experiment of Winkowski (n.d.) will be used to provide an answer to the research question. The relevance of this thesis towards Artificial Intelligence will be debated in chapter 2. Furthermore, the exact definition of presuppositions will be explained in chapter 3, followed by our hypothesis in chapter 4. In chapter 5, the complete methodology will be presented. The results will be given in chapter 6 and will be analysed and discussed in chapter 7. And finally, a complete conclusion will be given in chapter 8.

2. Relevance for AI

This thesis is very relevant to Artificial Intelligence, as we try to improve currently existing computational models on human language processing. These types of computational models are used in mobile phones and other electronic devices. Since an enormous part of everyday communication is happening online nowadays, it is important to recreate human language behaviour as well as possible in computers. In addition, one of the goals of Artificial Intelligence is the realization of natural dialogues between humans and machines (Këpuska & Bohouta, 2018). We try to create a computational model that represents human behaviour in sentence processing based on word similarity, so that we are able to predict human reading times by only using this feature. Outcomes of this small model could be implemented in currently existing models, to improve the representation of human linguistic behaviour in a computer. If we prove that word similarity has an effect on reading times, it will help researchers create even better linguistic models, which then can be implemented into computers to generate more human-like behaviour in electronic devices.

3. Presupposition

Presupposition is a phenomenon whereby speakers mark linguistically information as being taken for granted (Stanford Encyclopedia of Philosophy, 2021). Expressions that imply presuppositions are called “presupposition triggers” and form a large class that include definites and factive verbs. Presuppositions have become inevitable in human language: they are used in almost every sentence. An example will be given to clarify this phenomenon in more detail:

(1) *Yesterday, I drove past my favourite store and today I passed it too.*

In the sentence above, the word ‘it’ suggests that people already know what the word refers to, otherwise the writer would have written ‘my favourite store’ instead of ‘it’. This is one of many examples of a presupposition. The word ‘it’ is called the *presupposition trigger*: this word indicates that one’s memory has to be recalled to something earlier in that sentence, because some information is being taken for granted. Often one’s memory has to be used to recall previously read words so that they understand what the presupposition trigger refers to. These types of presuppositions behave the same as anaphors, because they just refer to another word (combination). This was also pointed out by Van der Sandt (1992).

Furthermore, the word ‘too’ is another presupposition trigger in this sentence: our memory has to recall that a similar word (combination) to ‘passed’ appeared earlier in the sentence, which is true: ‘drove past’. If this was not the case, the word ‘too’ would not have been there, since there was no similar word in the first part of the sentence. In this sentence, the word combination ‘drove past’ is called the *target* and the word ‘passed’ is referred to as the *antecedent.* This presupposition trigger, ‘too’ is at our main interest in this investigation, which will be explained more later in the thesis.

Even though two small examples were discussed regarding presuppositions, the use of presuppositions is very common in human language and is used far more frequent than one could imagine. The domain of presupposition theory has expanded enormously since 1969 from definite descriptions to other trigger types (Stanford Encyclopedia of Philosophy, 2021) and this is why it is such an important part of language. In addition, research regarding this subject is also necessary to be able to make computational models of language as complete as possible. For this reason, this thesis aims to find out whether humans are able to expect the presupposition trigger more quickly in sentences where the target and antecedent are very similar. To make this investigation not too detailed, we only looked at sentences in which the presupposition trigger ‘too’ is used. It will be experimented if the word similarity between the target and the antecedent has an influence on the reading time of that sentence. Further details of this research will be explained in the methodology section.

4. Hypothesis

In the sentences where presuppositions appear, people have to recall the target and the antecedent when they read the presupposition trigger ‘too’. The recall should be affected by the properties of the target: if the target is very similar to the antecedent, so the target has a high cosine similarity with the antecedent, then the antecedent should be easier recalled. According to an investigation on anticipating on upcoming words by Van Berkum et al. (2005), people were able to predict unread words that appeared later in a sentence, resulting in higher reading times. Words that were more unlikely to appear next were read more slowly. Thus, we expect that higher word similarity will be reflected in faster reading times. In other words, the higher the cosine similarity, the bigger the speedup will be when reading the presupposition trigger ‘too’. A bigger speedup at the end will lead to a greater difference in reading times between the sentence with presupposition and the sentence without a presupposition, because the sentence with the presupposition ‘too’ will be read more quickly when readers predict this word. This will especially be the case for shorter sentences, where the target word and the antecedent lie very close to each other. The target will then be more easily remembered, so the presupposition trigger ‘too’ will probably be expected in most cases where the target is very similar to the antecedent.

Similarly, a high cosine similarity between the target word and antecedent will also result in a bigger speedup for longer sentences. These longer sentences contain extra words between the target and antecedent to see if people are still able to recall the target word when reading the antecedent. This distance should make it more difficult for people to recall their memory when reading the antecedent, so the effect of memory recall will be lowered since the target word might sometimes be forgotten by the participants. Still, if the cosine similarity between the target word and the antecedent is high, the presupposition trigger ‘too’ will be read more quickly than in the case of a lower cosine similarity. However, the effect on longer sentences might be less than the short sentences, because participants might not remember reading the target word which takes away the expectation of the word ‘too’.

To sum up, we expect the difference in reading times to be the biggest for the short sentences including the presupposition ‘too’ and a high cosine similarity between the target and antecedent, compared to the short sentences without presupposition. Due to the fact that the target word and the antecedent are close, people will expect the word ‘too’ after reading a similar word to the target, which causes a speedup at the end of the sentence. Similarly, reading times of the longer sentences with a presupposition will be read more quickly than the sentences without, when the cosine similarity between the target and antecedent is higher. However, the effect of word similarity will be lower for longer sentences, since distractor words can cause people to forget about the target word. As a consequence, the word ‘too’ might not be expected and the reading times of this sentence will approach the same reading times as the sentence without presupposition.

5. Methodology

5.1 Word similarity

In order to compare the reading times of the experiment of Winkowski to memory retrieval, word embeddings were used to calculate the word similarity between each target word and antecedent in all sentences. Word embeddings are vector representations of word meanings deduced from enormous language corpora. The word similarity that was used in this experiment is equal to the cosine similarity, which uses two vectors of the same length and returns a value between -1 and 1. Two vectors with exactly the same orientation have a cosine similarity of 1, two vectors orientated at an angle of 90 degrees have a similarity of 0, and two vectors orientated in the exact opposite direction have a similarity of -1. A detailed description of the formula of the cosine similarity can be found in figure 1.

Figure 1
Formula of cosine similarity

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Note: the formula for cosine similarity, where A and B stand for vectors. Taken from Neo4J

To calculate the cosine similarity between certain words, the target words and antecedents first had to be represented as vectors. These vectors are the word embeddings, subtracted from a large library called FastText (Mikolov et al., 2020). FastText is a large, open-source set of libraries for word embeddings for approximately 294 languages, made by Facebook's AI Research (FAIR) lab. Each library contains pre-trained word vectors which can be used for research regarding word embeddings. For this experiment, a data-set of 2 million pre-trained word vectors for English language was downloaded to make use of these English word embeddings. Python code had to be written in order to subtract the word vectors for all target words and antecedents, and can be found in Appendix Chapter 3.

However, not every target or antecedent consisted of only one word (e.g. 'shut down'). For this reason, a combination of two words had to be applied to some of the target words and antecedents. Due to the fact that word combinations did not appear in the library at all (e.g. 'shut_down', 'shut-down' or 'shutdown' were not found), a combination of the vectors 'shut' and 'down' had to be used to represent the word combination. According to Mitchell et al. (2010), vector addition has shown to be the best possible method of combining vectors in linguistic literature, because it can be very useful to represent multiple vectors as one singular vector. Applying vector addition, the representative vector for a word combination (A, B) is calculated by adding every index *i* of vector A to the index *i* of vector B. The output is a vector with the same length as the input vectors A and B, which then can be used to calculate cosine similarities.

5.2 Experiment of Smith and Vasishth and modifications

Before discussing the main elements of this thesis, the paper called 'Feature selection' of Smith and Vasishth (2020) will be discussed shortly, since this experiment is a great inspiration for our own investigation. Similar to our experiment, sentences containing a trigger word at the end of a sentence were created, which would activate one's memory recall. Though, subclauses were added in the middle of each sentence, containing a distractor word which could lead to confusion in recalling the correct word. The paper was subdivided into four different experiments, and since the third experiment also made use of word similarity and reading times, this experiment will be discussed.

In the third experiment, sets of four different sentences were presented. Each sentence differed in one word or two words from the other sentences. The reading times were measured on the verb 'shattered', for which memory recall had to be used to remember the object that was shattered. In one sentence, both underlined words are 'shatterable' objects, in two sentences only one of the nouns is 'shatterable', and in one sentence none of the underlined nouns are 'shatterable'. Example sentences will clarify this description a little more:

1. Sue remembered the plate that the butler with the tie accidentally *shattered* today in the dining room.
2. Sue remembered the plate that the butler with the cup accidentally *shattered* today in the dining room.
3. Sue remembered the letter that the butler with the tie accidentally *shattered* today in the dining room.
4. Sue remembered the letter that the butler with the cup accidentally *shattered* today in the dining room.

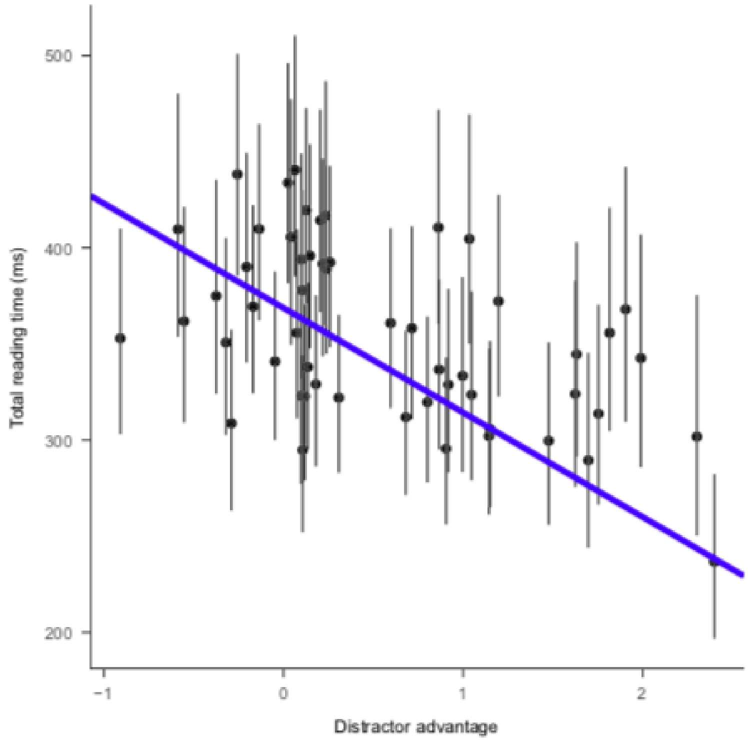
Vector space semantics was applied to calculate cosine similarities between the target noun and possible objects of the verb, and between the distractor noun and possible objects of the verb. This cosine similarity was referred to as the 'plausibility'. Difference in plausibility between the target and the distractor was calculated, and is referred to as the 'distractor advantage'. Positive values indicate that the distractor was a more plausible fit to the verb than the target and negative values state that the target was more plausible than the distractor.

Two different predictions were made, depending on whether the target matched the retrieval cues or not. For the target match condition, they expected slower reading times if the feature match of the distractor increased. For target the target mismatch condition, expectations were that the reading times at the verb were faster when the distractor advantage increased.

Results can be observed in Figure 2. In this graph, the distractor advantage is plotted against the reading times. Results do confirm their hypothesis, so the distractor advantage resulted in quicker reading times. In other words, reading times decreased when the distractor advantage increased.

Figure 2

Distractor advantage plotted against the total reading times



Note: a graph that illustrates the influence of the distractor advantage (x-axis) on the total reading times (y-axis)

Even though the aim of this experiment was a little different, the word similarity was calculated for other words, and no presuppositions were used, the same outlines of this experiment can be used in our own trial. From the observation that an increase of distractor advantage resulted in shorter reading times, we can expect our reading times to also decrease when the similarity between the target and antecedent increases. Since the distractor word in this experiment was very close to the verb, people their memory was recalled quicker if the distractor was a plausible object to the verb. The 'plausible' variable of this experiment can be seen as the cosine similarity in our experiment: if two words have a higher similarity, memory recall is accessed easier and the prediction of the word 'too' will be made more quickly. This will cause a speedup when reading the word 'too', since the word is already expected if the cosine similarity is high.

5.3 Data

The data used for our own experiment is subtracted from an unpublished experiment by Jan Winkowski (n.d.). In his experiment, 32 sets of 8 sentences were used to gather reading times for each sentence, to investigate whether the distance between a target word and an antecedent could influence the reading times on the presupposition trigger (or words following the trigger). Self-paced reading was used to collect reading times for all sentences. For our own thesis, the 32 sentence sets were taken from Winkowski's experiment, but we only use 4 sentences of each set, because the other 4 sentences are redundant. A file containing all 32 sets of sentences is be attached to this thesis, in which each sentence used in our own experiment can be found. Though, two sets of sentences will be discussed to create an idea of the framework of our experiment.

1. a) (too) The cook is a dancer and the waiter dances too, I have been told recently.
   b) (nil) The cook is a swimmer and the waiter dances often, I have been told recently.
   c) (too2) The cook is a dancer and the waiter, *who is a boxer*, dances too, I have been told recently.
   d) (nil2) The cook is a swimmer and the waiter, *who is a boxer*, dances often, I have been told recently.

2. a) (too) The bakery has shut down and the cafe is closed too, so the streets are empty.
   b) (nil) The bakery is dull and the cafe is closed down, so the streets are empty.
   c) (too2) The bakery has shut down and the cafe, *which was built nearby*, is closed too, but the streets are empty.
   d) (nil2) The bakery is dull and the cafe, *which was built nearby*, is closed down, so the streets are empty.

Winkowski measured reading times for each sentence from the beginning of each sentence to the end of the trigger word. In the (too) and (too2) sentences, the trigger word was 'too'. In the (nil) and (nil2) sentences, the word 'too' was replaced by a word with similar length, e.g. 'often' in the first and 'down' in the second sentence. The logarithm of the reading time in milliseconds was used as a measure of the reading times.

Notice that the sentences (too) and (too2) contain red and blue words, as well as the word 'too'. Each colored word refers to a certain element of the sentence: the green word 'too' is the presupposition trigger, which suggests that there must also be a target word and an antecedent earlier in the sentence: the red word represents a target word, and the blue word stands for the antecedent. Sentence (nil) has no presupposition at all, and is only created to compare the reading time of sentence (too) to a sentence with similar length. The same holds for sentence (nil2), which is also created to compare the reading times from sentence (too2) to a sentence of similar length. Sentences (too2) and (nil2) contain subclause with 4 words in every sentence, added to measure the effect of a distractor sentence on the memory recall. Results of Winkowski's experiment are not relevant, though the sentences and reading times serve as a base for this thesis.

The experiment of Winkowski was performed to investigate whether the addition of words between the trigger and the antecedent resulted in an increase in reading times (corresponding with slower reading) of the presupposition trigger. This delay could be observed on the presupposition trigger ‘too’. The reading times were gathered by collecting eye-tracking data of 36 participants. The reading times of the first 10 sets of sentences used in this experiment can be found in Table 1. The last two columns show the variables ‘difference_short’ and ‘difference_long’. ‘Difference_short’ indicates the difference between the short sentence with presupposition and the short sentence without presupposition. ‘Difference_long’ stands for the difference in reading times between the longer sentence with presupposition and the longer sentence without presupposition. However, these reading times are not measured in seconds or milliseconds, but in the difference in logarithm of both reading times in milliseconds (see (2) for the exact formula). This means that the difference between the too-sentences and nil-sentences are equal to the difference in the logarithm of the reading times for those sentences. These outcomes are directly taken from the experiment of Winkowski.

(2) Difference (too, nil) = log(reading time too) – log(reading time nil)
‘Reading time too’ stands for the reading time of the sentence with ‘too’ in milliseconds

Table 1
Differences in reading times for the first 10 sentences

Sentence	difference_short	difference_long
1	0.118983539	0.180339809
2	-0.15542066	-0.14962206
3	-0.07049593	0.002274407
4	-0.20720057	-0.251434402
5	0.158217101	0.206527654
6	-0.36999178	-0.310994353
7	0.215910282	0.169515742
8	-0.15840461	-0.131870023
9	0.052030483	-0.042361557
10	-0.08306979	0.0309871592

Note: differences in logarithmic reading times for short sentences ((too) – (nil)) and for longer sentences ((too2) – (nil2))

In this thesis, we will use ‘short sentences’ to refer to the sentences where no distracting words were added between the target and the antecedent. In contrast, ‘longer sentences’ will be used as another term for sentences where a subclause with distraction words was added between the target and antecedent.

5.4 Method

Even though Jan Winkowski had a different purpose for constructing his experiment, the sentences that were written are also really helpful for this experiment. As already mentioned, this thesis aims to find out whether the word similarity between the trigger and the antecedent could have an effect on the reading times of each sentence. Although the same sentences have been used, different elements of the created sentences had to be compared in order to get an answer to our main question. Differences in logarithmic reading times between sentences of a short length were compared to the word similarity between the target word and the antecedent in that sentence (for all 32 different sentence sets). The same applies to the longer sentences, where the reading times were also compared to the word similarity between the target word and the antecedent. These differences were calculated from the mean reading times of 36 participants who participated in Winkowski's experiment.

In order to be able to calculate correlations between word similarities and reading times, target words and antecedents had to be converted into vectors to obtain the cosine similarity between them. As previously addressed, these vectors were acquired from a pre-trained set with word vectors called FastText.

Pearson's correlation tests were conducted to determine the correlation between the word similarity and the reading times. These tests were executed for both the short and the long sentences. In both cases, the dependent variable was equal to the difference in logarithmic reading times, and the independent variable was equal to the cosine similarity between the target word and the antecedent.

6. Results

In Table 2, some examples of the results are listed for the first 10 sentences to give an outline of the results. The column 'add_target' contains the second word of a word combination, if the target consisted of two words. In addition, the column 'add_antecedent' stands for the second word of a word combination if the antecedent consisted of two words. Note that target words and antecedents with a high cosine similarity (0.5-0.7) often resulted in a negative difference in reading times. This means that the sentence without presupposition was read more quickly than the sentence with the presupposition 'too'. In contrast, targets and antecedents with a low cosine similarity (0.1-0.3) resulted in quicker reading times for sentences with the presupposition, compared to the sentence without. The whole table containing all cosine similarities and corresponding differences in reading times can be found in Appendix Chapter 2.

Table 2

Cosine similarities and differences in reading times

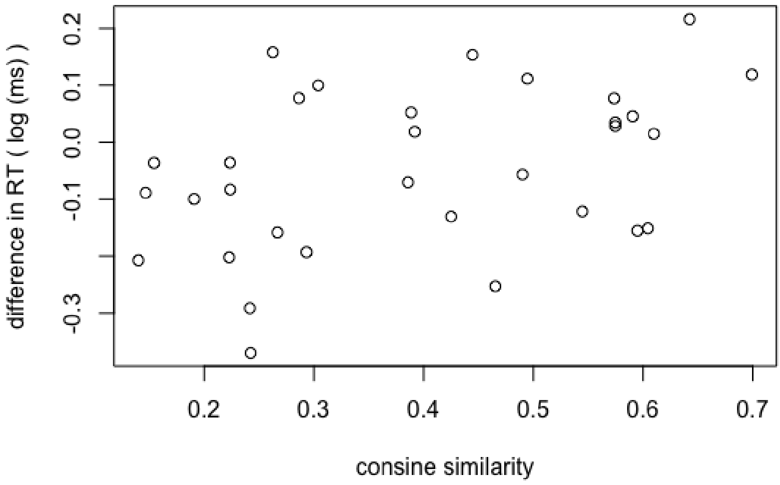
nr	target	add_target	antecedent	add_antecedent	cos_sim	difference_short	difference_long
1	dancer		dance		0.6992321	0.118983539	0.180339809
2	painter		paint		0.5950181	-0.15542066	-0.14962206
3	shut	down	close		0.3858766	-0.07049593	0.002274407
4	eat		have	food	0.1401565	-0.20720057	-0.251434402
5	predator		hunt	prey	0.2626999	0.158217101	0.206527654
6	secrete	poison	poisonous		0.2424746	-0.36999178	-0.310994353
7	hunter		hunt		0.6425202	0.215910282	0.169515742
8	extinct		die	out	0.2669547	-0.15840461	-0.131870023
9	acquire		buy	out	0.3888077	0.052030483	-0.042361557
10	herbivores		eat	plants	0.2238608	-0.08306979	0.0309871592

Note: results of cosine similarities together with the difference in mean reading times (in log(ms)) for short sentences (too - nil) and difference in mean reading times (in log(ms)) for the longer sentences (too2 – nil2)

Dot charts for the short sentences be found in Figure 3. This graph does not show a clear trend, since the data points are spread out over the whole graph. However, it does show a vague imaginary line from the bottom left to the top right of the graph, indicating that reading times increased when cosine similarities became greater. Pearson's correlation tests show a value of 0.4034 for the correlation between word similarity and reading times for short sentences. In addition, the influence of word similarity on reading times turned out to be a significant one, $t(30) = 2.4146$, $p = .02206$.

Figure 3

Comparison reading times (short sentences) and cosine similarities

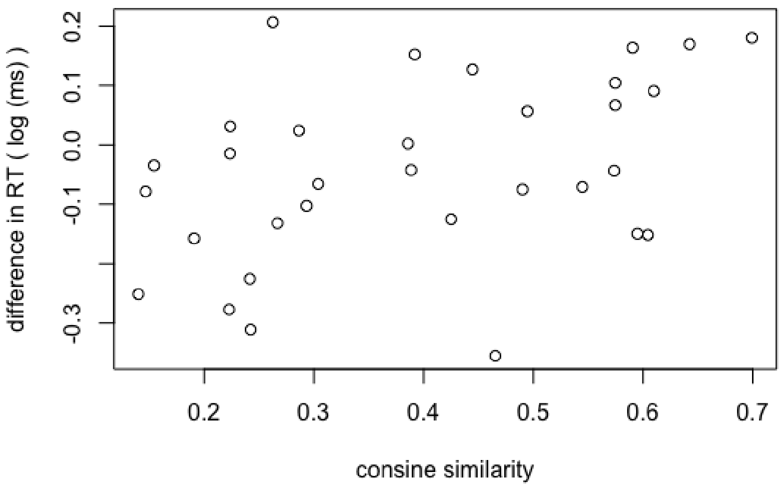


Note: plot containing difference in reading times for the short sentences on the y-axis and the cosine similarity on the x-axis

The influence of the word similarity on the reading times of longer sentences can be observed in Figure 4. Note that this graph shows the same ascending trend as the shorter sentences in Figure 3. This shows that the word similarity also had influence on the reading times, even though extra words were added between the target word and the antecedent. Here, an even greater correlation can be observed. Pearson's correlation test returned a value of 0.4189, indicating that reading times became greater when the word similarity increased. This result also turned out to be significant by a two-sided t-test, $t(30) = 2.527$, $p = .01701$.

Figure 4

Comparison reading times (long) and cosine similarities



Note: another plot with difference in mean reading times between the long sentences on the y-axis, and cosine similarity on the x-axis

7. Discussion
7.1 Results

The discussion of the results will be divided into three parts. The first part will provide an insight in the findings of the short presupposition sentences compared to the nil-sentences, and the second part will be used to discuss the results of the longer sentences, with words between the target and the antecedent, compared to the corresponding nil-sentences of the same length. Lastly, the first two subdiscussions will be added together for one complete, comprehensive discussion.

After comparing the cosine similarities to the difference in reading times between the short sentences with and without presupposition, an unexpected result was found. The hypothesis stated that higher cosine similarities would result in lower reading times for each sentence, since the reader would already predict the word 'too' from reading a similar antecedent to the target word. In contrast to our hypothesis, the results showed that difference in reading times was higher when the cosine similarity was also higher, which is exactly the opposite effect of what was foreseen. Pearson's correlation test indicated a value of 0.4034. This value turned out to be significant, which tells us that the word similarity indeed had an influence on the rate at which the sentences were read. However, the main point that can be taken from these results is that the reading times became slower when the target word and antecedent were very similar. Even though the article of Van Berkum et al. (2005) found that anticipation of words resulted in quicker reading times, outcomes of this research show exactly the opposite.

Likewise, an equally unforeseen outcome was found for the longer sentences. As a matter of fact, these results demonstrated an even stronger effect of the word similarity on the reading times. Pearson's correlation tests resulted in a value of 0.4189, a larger number compared to the shorter sentences. T-tests also confirmed this as a significant effect.

Various reasons can be given to explain the positive effect of word similarity on reading times. Firstly, reading data is extremely sensitive to noise due to a large number of factors determining reading times for each sentence. Even the slightest distraction could lead to a loss of attention to the sentences, causing the reading times to be completely different. Furthermore, target words and antecedents were manually drawn from the sentences used in Winkowski's experiment. Some of these target words and antecedents consisted of a combination of two words, e.g. 'have food'. Vector addition was applied to get a singular vector for this combination in order to compute the word similarity between the target and antecedent. However, this might have resulted in non-representative cosine similarities, since words like 'have food' and 'eat' are very similar words, but the cosine similarity between these two words is only 0.1402. This might suggest that vector addition is not the perfect solution for representing word combinations. Additionally, the reading times of the longer sentences could have been affected by another factor. In these sentences, four additional words were added between each target word and antecedent in sentences containing a presupposition. An example is given in (3). These distracting words, and especially the word 'boxer' might have had an influence on the reading times of the presupposition, even though this sentence only served as a distance between the target and the antecedent.

(3) The cook is a dancer and the waiter, who is a boxer, dances too, I have been told recently.

7.2 Future research

The results do not correspond with our hypothesized results, which means either the design of the used method was not perfect, or the presuppositions indeed have a completely other influence than we expected. In this section, certain aspects of this thesis will be discussed for improvements of future research regarding the topic of this thesis.

For the longer sentences, in which additional words were added between the target and the antecedent, the distractor words could have had an influence on the reading times. To verify that this was not the case, the cosine similarity between the main distractor word and the antecedent could be calculated. These additional words might have caused trouble in one's memory, resulting in not expecting the trigger word 'too', or expecting the word 'too' more than usual. One example of these sentences is shown in (2) in chapter 7.1. Here, the word 'boxer' and 'dances' are not really similar, but this word could have had an effect on the reading times of that sentence, because the target word might have been forgotten.

Certainly, other improvements can be made as well to create an even better and more trustworthy experiment in the future. For instance, the same sentence could be used multiple times changing only the antecedent to create just one difference in each sentence. As a consequence, only the word similarity between the target word and antecedent changes, and the measured reading times can be compared without unwanted errors of other influences. Three example sentences will be presented to generate an idea of this possible investigation:

1. The teacher cried out in fear and the policeman yelled too, my mother told us.
2. The teacher cried out in fear and the policeman screamed too, my mother told us.
3. The teacher cried out in fear and the policeman shouted too, my mother told us.

In these sentences, only one word was varied so that the other factors could not have an impact on the results, which would improve the reliability of the experiment. Only the word similarity of the target and the antecedent is varied in this example, which could lead to more trustworthy results.

Apart from improvements on our investigation, further experiments could approach other types of presuppositions. Only the presupposition 'too' was covered in this research, but in order to gather information of human behaviour on other presuppositions, other types also have to be investigated as well, since presuppositions appear very often.

8. Conclusion

In sentences containing a presupposition ‘too’ at the end of a sentence, word similarity between the target word and the antecedent had a positive influence on the reading times of this sentence. A correlation of 0.4034 was found for sentences in which the target and antecedent were located very close to each other. This suggests a moderate connection between word similarity and reading times. The correlation value was a little higher for sentences with additional words between the target and antecedent: 0.4189.

These values show that word similarity has a complete different influence on the resolution of the presupposition trigger, since reading times increase when target words and antecedents are more similar. This holds not only for sentences where the target and the antecedent lie close to each other, but also when more words are added in between the target and the antecedent. Other features of the target and antecedent could have a bigger influence on the prediction of the presupposition trigger ‘too’ than the word similarity, which could be investigated in other investigations.

To sum up, reading times increased when cosine similarity between the target word and antecedent became higher. This means that one’s memory was indeed affected by the cosine similarity, because it seems that the less similar two words were, the more people used their memory to recall the target word and expect the word ‘too’. This resulted in faster reading times for words with less similarity. The outcomes of this experiment can be used for future investigations on word similarity and reading times, which have to be computed before applying these results to other computational models on human language processing. This thesis, together with further investigation regarding word similarity and reading times, can improve intelligent systems of human language behaviour, which is why this thesis was related to Artificial Intelligence.

9. Literature

Beaver, D.I., Geurts B., Denlinger, K. (2021, 7-1) *Presupposition.* Edward N. Zalta (ed.), https://plato.stanford.edu/entries/presupposition/

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). *The effect of prominence and cue association in retrieval processes: A computational account*. Cognitive Science.

Futrell, R., & Levy, R. (2017). *Noisy-context surprisal as a human sentence processing cost model*. In M. Lapata, B. Phil & K. Alexander (Eds.), Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (pp. 688–698). Association for Computational Linguistics.

Jäger, Engelmann, & Vasishth, (2017). *Similarity based interference in sentence comprehension: Literature review and Bayesian meta-analysis.* (pp. 316-339), Elsevier.

Këpuska, V., Bohouta, G. (2018) *Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home*), (pp. 99-103) IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)

Mannes, J. (2017). *Facebook's fastText library is now optimized for mobile.* https://techcrunch.com/2017/05/02/facebooks-fasttext-library-is-now-optimized-for-mobile/

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., Joulin, A. (2017). *Advances in Pre-Training Distributed Word Representations*

Neo4J Labs Team (n.d.). *Cosine similarity* [Picture]. Neo4J. https://neo4j.com/docs/graph-algorithms/current/labs-algorithms/cosine/

Smith, G. & Vasishth, S. (2020). *A Principled Approach to Feature Selection in Models of Sentence Processing.* Cognitive Science Society.

Smith, G., Franck, J., & Tabor, W. (2018). *A self-organizing approach to subject-verb number agreement.* (pp. 1043–1074) Cognitive Science.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P. (2005). *Upcoming Words in Discourse: Evidence From ERP's and Reading Times*. American Psychological Association.

Van der Sandt, R.A. (1992). *Presupposition Projection as Anaphora Resolution.* (pp. 333 – 377). N.I.S. Foundation.