Radboud University Nijmegen

Bachelor's thesis in artificial intelligence

INFLUENCING THE N400

A STUDY ON FACTORS THAT AFFECT THE N400 ERP COMPONENT DURING SENTENCE COMPREHENSION

Author:
Erik Eppenhof
Student number:
s4150007

Supervisors:
Prof. Dr. A.F.J. Dijkstra
Dr. S.L. Frank

August 29, 2014

Abstract

Effects of sentence position, word position, length, frequency, surprisal, LSA, concreteness, valence and arousal of nouns on the amplitude of the N400 ERP component were assessed during word-by-word rapid serial visual presentation of 205 sentences in 24 subjects. Linear mixed effects regression analyses revealed significant main effects for word position, length, surprisal, LSA, concreteness and valence. Interactions between length and frequency and between word position and surprisal revealed to have significantly influence on the size of the amplitude of the N400 component. The main effect of valence showed that nouns characterized as unhappy evokes smaller N400 amplitudes than nouns that are characterized as happy. The main effect of concreteness demonstrated that nouns with an increased concreteness produce bigger N400s than abstract words. The main effect of LSA showed that nouns that have a strong semantic relation with the preceding context evokes a drop in the amplitude of the N400. This last finding lead to the thought that word recognition takes place in a relatively late stage of the language processing stream.

Contents

1 Introduction

1.1 Event Related Potentials

Many EEG studies have shown that the electrical activity of the human brain can be directly related to events or stimuli which causes specific electrophysiological responses. These electrical potentials, that are a result of synchronized activation of neuronal networks in the brain caused by stimuli, are referred to as Event Related Potentials (ERPs). ERPs can be recorded at the scalp by means of electroencephalography (EEG). With EEG you can measure voltages evoked by groups of neurons at different spots on the head with respect to a reference. The voltage differences between each electrode on the head and the reference electrode is measured over time and this results in positive or negative peaks (voltage differences) in the EEG recording. This means that ERPs consist of series of positive and negative peaks, which we can time lock to stimulus material to see how our brains react to those specific stimuli. To extract the ERPs from the EEG we need to average over trials. We assume that the brain responds practically the same every time to a specific stimulus, while the random brain activity (noise) differs from time to time. When a stimulus is presented, an ERP on this stimulus can be measured by looking at small sections, which we call trials, in the EEG at a particular time after stimulus presentation. In this study I looked at an ERP component in a duration of 100 milliseconds

Extracting ERPs is a widely accepted technique in cognitive neuroscience to study mental processes which are a response on sensory, motor or cognitive stimuli. In psycholinguistics ERPs are used to study language processing. A great advantage of ERPs is that it has a high temporal resolution; there exists a temporal accuracy of a few milliseconds [13]. In this manner visually or auditory presented words in language experiments can be linked to the evoked potentials which can be derived from the data at the time the word was presented. There has been extensive research on different ERP components involved in language processing, like the P600, N200 and N400 [13, 25, 27]. The ERP component that is studied in this thesis is the N400.

1.2 N400 Component

The N400 component is an ERP that is known to be involved in lexical-semantic integration. More than thirty years ago, in 1980, Kutas and Hillyard reported the N400-effect for the first time [16, 17, 18]. They investigated ERPs that occur when an inappropriate word occurs unexpectedly at the end of a sentence. An example sentence in their paper was “I take coffee with cream and ...”. Most people would agree on completing this sentence with the word “sugar”. Kutas and Hillyard based their experiment on this human expectancy and the findings of Schuberth and Eimas (1977). Schuberth and Eimas stated that words that are expected in a given context are recognized more accurately and faster than the same words presented independently or in a semantic inappropriate context [28]. In the experiment of Kutas and Hillyard, a random 25 percent of the 160 sentences that were

presented to the subjects ended with an unexpected stimulus. They expected to see a P3b in the EEG, a positive peak around 300 milliseconds after the sentence final word [29]. Squires et al. described that what is now known as the P3b would occur after an unexpected or surprising stimulus. Donchin [6] later added to that fact that the P3b increases as the item is less predictable. The P3b was found in the experiment of Kutas and Hillyard when the sentence final word was physically unexpected, for example when the stimulus was in capitals: "I take coffee with cream and SUGAR". They expected the P3b to occur in each testing condition, so also when the sentence final word was semantic-inappropriate relative to the context. This semantic-inappropriate word could be the word "wood" in the example sentence: "I take coffee with cream and wood". Instead of a P3b on those stimuli, they discovered a large, robust peak around 400 milliseconds after stimulus onset in each participant. They called it the N400. This N400 effect was substantially different from the P3b and obviously due to semantic deviation. Kutas and Hillyard attributed therefore that the N400 wave could reveal more useful information about the processes involved in natural language comprehension.

1.3 The N400 in more detail

So what is it what Kutas and Hillyard found exactly? The N400 is a negative going deflection in a time-locked EEG signal (event related potential) that peaks around 400 milliseconds after stimulus onset. It starts around 200 milliseconds after the stimulus word and has a latency between 350 and 400 milliseconds [13]. The N400-response is mainly known for its involvement in lexical-semantic integration, as Kutas and Hillyard discovered in 1980 [16]. When presenting a semantic deviant word (another word than expected) in a sentence, there typically occurs an N400 effect in the EEG-signal. In a sentence with a syntactic failure, there does not occur an N400, so the N400 is not prone to any syntactic deviations [13].

There is shown that there is a link between the amplitude of the N400 effect and the difficulty of fitting the word in the sentence (Figure 1). Kutas and Hillyard found out that the more deviant the last word in the sentence was with respect to the participants expectations, the more negative the peak in the EEG around 400 milliseconds became. This resulted in seeing the N400 as a semantic violation detector. Kutas and Hillyard (1983) discovered later that in fact every potential meaningful word in a sentence elicits an N400 [9, 26]. They noticed that the amplitude of the N400 even decreases as the phrase progresses. Their explanation for this fact was that you are continuously testing and updating hypotheses about words that could occur next, while integrating every word in the context. When reading the sentence, you get more and more restricted in words that can follow in the accompanying context. They concluded that there is a negative correlation between the amplitude of the N400 and the predictability of the upcoming word in a sentence [9, 22, 26].

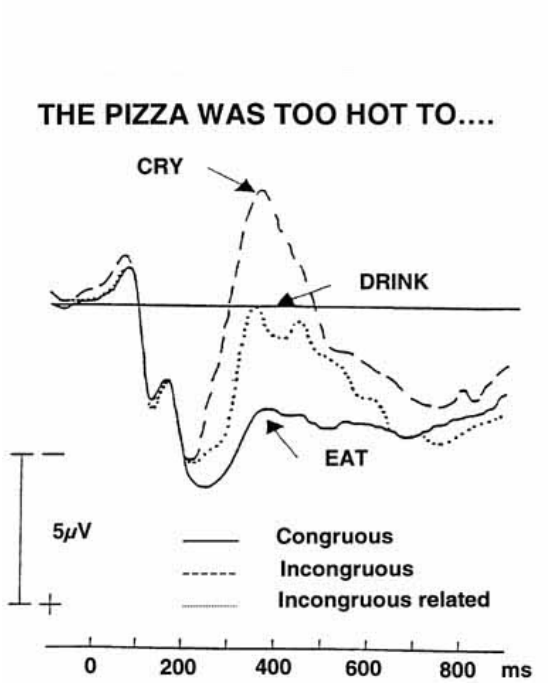

Figure 1: Grand average ERPs to three different sentence completions shows different effects on the N400. Kutas and Hillyard found out that the more deviant the last word in the sentence was, the more negative the peak became [16]. The word you would most likely expect in the sentence "The pizza was too hot to ...." is the word "eat". The word "drink" (which is incongruous and related to "eat") would elicit a N400 in the EEG, but evokes a smaller peak than the word "cry", which is incongruous and unrelated to "eat". The N400 is reduced when anomalous sentence completions are related to the most expecting ending [15].
N.B. Negative polarity is plotted up.

The N400-effect has been recorded to both spoken and written words. The largest N400s are elicited by open class words (nouns, verbs, adjectives, adverbs) that are semantically anomalous within the accompanying context. Closed class words (determiners, pronouns, conjunctions, prepositions, etc.) elicit considerably smaller N400s. These findings holds independently of the position of the word in the sentence [20].

The N400-effect is measurable over the entire scalp, but the wave is maximal over centro-parietal placed electrodes. In addition, the N400 wave tends to be slightly larger over the right than the left hemisphere [20]. This observation seems a bit paradoxical, knowing that for most people, the left hemisphere is the part which can process language independently of the right hemisphere. In this case the resulting electrical flow generated by a large number of neurons does not provide us information about the location of those neurons. Research on split-brain patients has shown that in most cases the N400 component arises from the left hemisphere, where the language processing takes place [19].

1.4 Factors that influence the N400

The N400 is often used in studies of language and memory as dependent variable [2, 3, 5, 14, 16, 17, 32, 34]. These studies have found several factors affecting the N400 amplitude. Some of them are already described in the preceding sections. The factors that occurred most frequently in these studies, will be briefly highlighted in the next sections, i.e. word-position, frequency, concreteness and priming.

1.4.1 Word position

As mentioned in section 1.3 every word elicits in fact an N400, albeit a small effect. In natural language sentences the N400 amplitude decreases along the sentence. There is a strong inverse correlation between position of the word in the sentence and the amplitude of the N400 [16, 17]. Kutas and Van Petten attribute this effect to the fact that every word along the sentence, restricts you more and more in predicting which word can follow after the preceding context [20]. More on the effect of word position is mentioned in the next section.

1.4.2 Frequency

Words that are used more often in a language and thus are high-frequent in the used language, have been found to evoke N400s with smaller (less negative) amplitudes than low-frequent words [32]. Kutas and Van Petten, however, found out that this view is only applicable when the eliciting words occurred early in sentences [32]. Based on this interaction between sentence position and word frequency, they concluded that frequency plays a subordinate role in recognition of words. The frequency effect can thus be overridden by contextual constraints provided by the sentence [32].

1.4.3 Concreteness

Concrete words are known to evoke faster response times and larger N400 responses than abstract words [2]. These effects are explained by stating that the links to associated semantic information of concrete words are denser. However, the N400 concreteness effect projects mostly to the frontal electrodes, while the classical N400 effect is strongest at the centro-parietal electrodes. Therefore it is not clear whether the N400 concreteness effect shares the same neural generator as the classical N400 effect [12].

1.4.4 Priming

The amplitude of the N400 becomes more positive on a stimulus when some of that stimulus features are already activated by a prior stimulus (the prime) or when the prior stimulus creates the effect of the current stimulus being more predictable [3, 34]. This could be the case when there are associations between the prime and the current stimulus. The priming effect on the N400 is shown across different modalities. This means the prime and the stimulus the prime has effect on, can be pictures, gestures, words, complete sentences, etc. [9, 22, 26].

1.5 Research question

Since the discovery of the N400 in 1980 there has been many publications on this subject. This has lead to multiple theories about the N400-effect. On the one end there is the theory (i.a. Hagoort et al., 2009, [11]) that positions the N400 late (post item recognition) in the processing stream of language, after a stimulus has been recognized. In this view the N400 reflects the proces of linking up the semantic information accessed from the current word [14]. This means that integrating the meaning of a word in the preceding context is preceded by complex perceptual analyses that lead to recognition, which in turn leads to semantic access to the word.

On the other end there is the theory that states that the N400 occurs early in the stream of language processing. Researchers who believe in this theory base it on the fact that the N400 can also be elicited by pseudowords (non-existing words), which show little resemblance to known words. Pseudowords are not in the mental lexicon (mental dictionary with word's meaning, pronunciation, etc.) and therefore they presumably have no associated semantics. The supporters of this theory belief that the N400 occurs before the word is even recognized and before semantic acces can take place. They conclude that the N400 resembles orthographic and/or phonological analysis [14].

In this thesis I will research several factors that affect the N400-component during sentence comprehension. The question that will be at the heart is which properties of nouns in natural language sentences influence the N400-component? I will discuss what we can draw from these findings regarding sentence processing and in which theory we can place these findings. Hopefully this thesis will contribute to the ongoing debate about the underlying processes the N400 indexes and provide more insight in human language comprehension.

2 Method

The data that is used in the experiment, is obtained by Frank et al. and is used in previous experiments [8]. In the next sections I will elaborate more on the participants, the EEG recordings, procedure of the experiment and the sentence material that is used.

2.1 Participants

Twenty-four healthy, adult volunteers participated in the experiment. They were drawn from the UCL Psychology subject pool. All the participants were native speakers of English and most of them (19 out of 24) also master at least a second language. These second languages were varying from German and French to Japanese and Mandarin. The group of participants consisted out of 10 females and 14 men. The age of the participants varied from 20 to 64, with an average age of 27.96 years old. One of the participants was left handed, all the others were right handed.

2.2 Electroencephalography recordings

EEG data was collected by an electrode cap on the participant's head which continuously measured brain activity from 32 channels, during the presentation of the stimulus material. The electrodes used in the electrode cap were silver/silver-chloride (Ag/AgCl) electrodes. The electrodes were arranged according to the Montage No. 10 system (M10). This system describes the placement of the electrodes on the scalp. The electrodes are placed in triangles around Cz (central electrode on top of the scalp) and the distance between neighboring electrodes is 37 $\pm$3 millimeters. The electrodes on the central line are placed using the 10 percent division [23]. A detailed picture of the placement of the electrodes is seen in Figure 2.

2.3 Procedure and stimulus material

Each participant had to read 205 sentences, which were presented one word at a time on a computer screen. We call this way of presentation RSVP (rapid serial visual presentation). The participants could practice with 5 example sentences first, before the experiment began. These 210 sentences in total were selected from three novels, which were not published yet at that moment. The stimulus material therefore consists of natural semantic and syntactic correct sentences. The sentences, which are selected and are previously used in reading time studies by Frank et al. [7], contain only the words that are in the 10.000 high-frequency word list from the British National Corpus (BNC). The BNC contains 100 million words, which are extracted from samples of written and spoken language and is known to be a cross-section of current British English [31]. The sentences used in the experiment are listed in Appendix A. The subjects were instructed to minimise head movements and eye blinking during the presentation of the sentences.

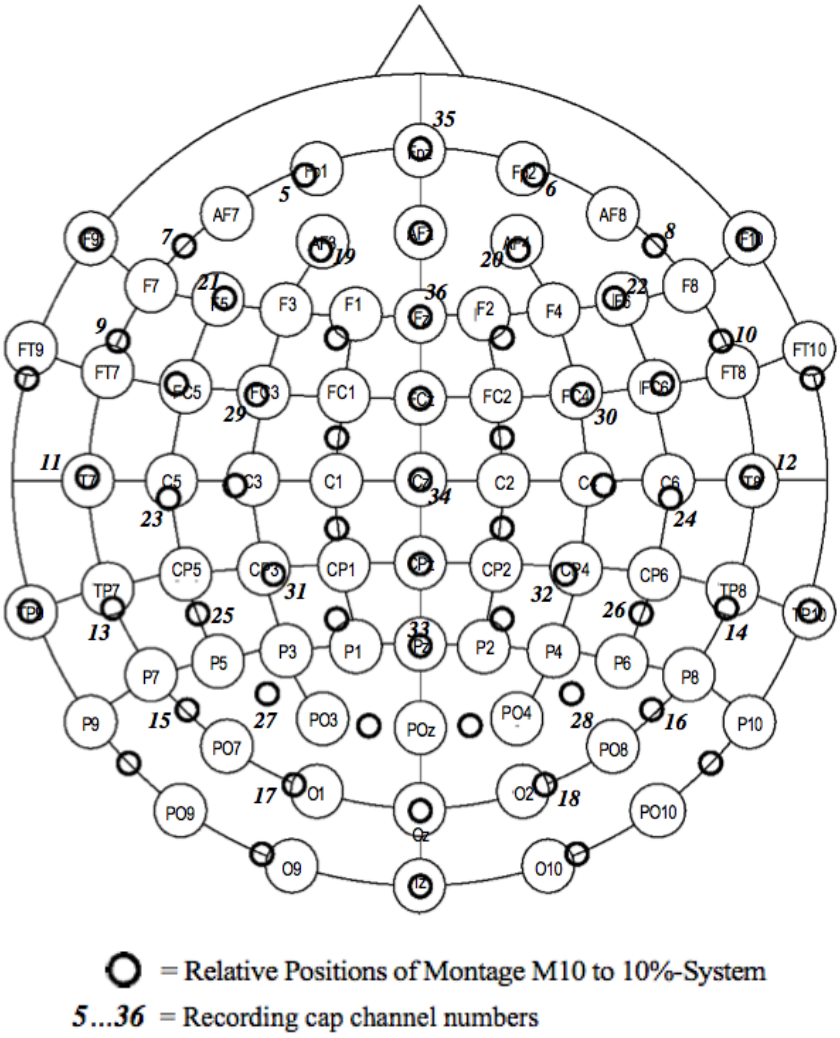

Figure 2: Detailed view on the positions of the electrodes in the electrode cap placed on the scalp according to the M10 system. The electrodes that are used are marked with a black circle and a number from 5 to 36.

Each sentence was preceded by a fixation cross, on which the subjects had to focus. Whenever they were ready they pressed a key on the keyboard which resulted in the presentation of the first word of the sentence in the middle of the screen. Every next word following in that sentence appeared automatically on the screen replacing the last word. The interval between two words was 390 milliseconds and every word was presented for 190 + 20n milliseconds, where n is the number of letters in the word. 110 of the 205 sentences had a yes/no comprehension question which immediately appeared after the last word of the sentence. This question was asked to be sure of the fact the subjects understood the sentences. All subjects answered at least 80 percent of these questions correctly. In figure 3 there is an example timeline of the sentence “Anne lost control.”.

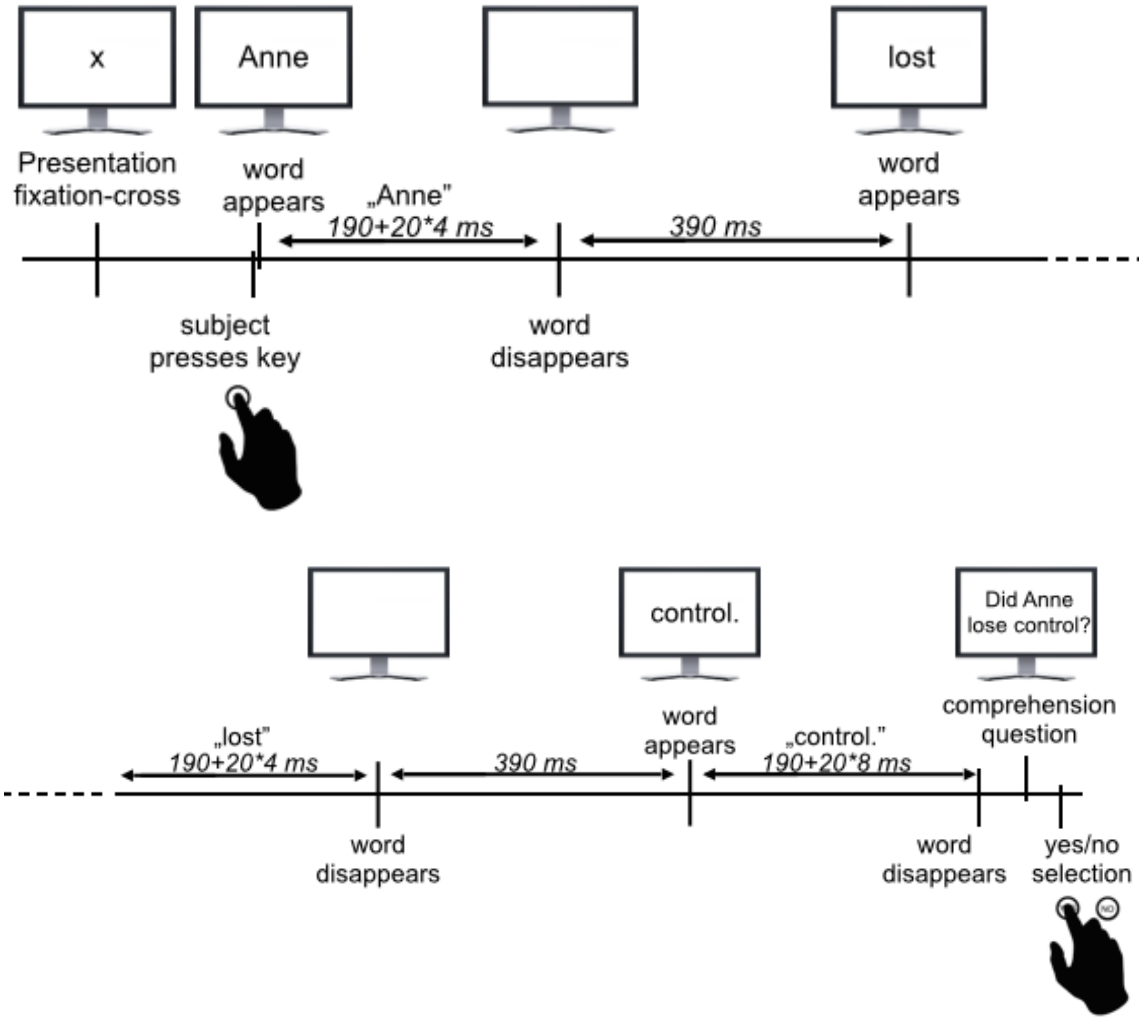

Figure 3: An example timeline of how the sentence "Anne lost control." would be presented in the experiment. After the last word of the sentence the subjects could have to answer a comprehension question about the content of the preceding sentence. In this case the question could be: "Did Anne lose control?"

2.4 Analyses

To analyze the relationship between several properties of nouns in natural language sentences and the amplitude of the N400 component, I have used R (R Core Team, 2014) and the accompanying lme4 package (Bates, Maechler, Bolker and Walker, R package version 1.1.6., 2014) to perform a linear mixed effects regression analysis. The definition of the N400 as used in these analyses is the average EEG potential between 300 and 500 milliseconds after word onset on the selected electrode sites (1, 14, 24, 25, 26, 29, 30, 31, 41 42, 44, 45) based on the approach of Dambacher et al. [5]. EEG data with artifacts (mostly due

to eye blinks) were identified and rejected. Because the largest and most robust N400s are elicited by open class words and because most of the factors that are involved in my analyses are only available on nouns, only the data on nouns in the experimental sentences will be analyzed. ERP-studies have proved that sentence-final words often appear to elicit more positive going waves than sentence-initial or sentence-intermediate words [24], probably due to the sentence wrap-up effect [10]. Data on sentence-final words are therefore also removed. This leaves us with a total of 6375 data points.

2.4.1 Linear mixed effects regression analysis

The contribution of a number of factors will be assessed by performing a linear mixed effects regression analysis, using the amplitude of the N400-component as the dependent variable and the features of the nouns as independent variables. A mixed effects model is a statistical model which describes a mixture of fixed effects and random effects and is commonly used in the field of statistical linguistic analyses. In such an analysis the fixed effects are the independent variables. These fixed effects are the things we understand and are in a way systematic. The factors we do not understand and we can not control experimentally are called random effects. Random effects could be variables which are not in the analysis, because you did not know about their existence for example. But a random factor could also be a subject itself. Every subject reacts differently to a certain stimulus and could also have a totally different basic brain response. In this way the model considers variation which can not be attributed to the fixed effects. The random effects are there for the variables that are specific to the used data sample (e.g. subject, item). Another advantage of mixed effects models is that any missing values are not a problem to the model analyses [1]. In the next section I will shortly discuss the fixed effects that are involved in my analyses.

2.4.2 Fixed effects

The fixed effects that are included in my analyses are: sentence position in the experiment, word position in the sentence, word length (number of characters including punctuation), log-transformed word frequency, surprisal, LSA, concreteness, valence and arousal. Where the first four are clear by name, the definitions for the latter five predictors require more elaboration, which will follow next.

Surprisal
The surprisal of every word is measured according to the 4-gram model, that is trained on the BNC. An n-gram model predicts a word based on the previous words [30]. With taking care of the independence assumptions, every word depends only on the last n-1 words. The value of surprisal of the word is the negative log-probability of a word given the context. In this case this context incorporates the 3 last words [30].

LSA
LSA is an acronym for latent semantic analysis and is a method to determine the similarity in meaning between units of text. The idea of LSA is that a value can be assigned to (groups of) words that do or do not occur together in sets of texts. This value resembles to what extent these (groups of) words match in meaning. LSA represents the meaning of the word as the average meaning of all the text passages in which the word occurs. The LSA value of a word in a context is determined by means of statistical calculations applied to a large corpus of text [21]. To construct the LSA values, the pairwise comparison on the LSA Website developed by Darrell Laham, is used (URL: http://www.lsa.colorado.edu) [21]. The LSA value of the noun is based here on the comparison between the preceding context thus far and the noun itself.

Concreteness
"Concreteness evaluates the degree to which the concept indicated by a word refers to a perceptible entity" [4]. Concrete words are therefore easier to remember than abstract words. Concreteness values have been established by a norming study asking over four thousand participants how easy it is to think of a context in which the presented word can be used. The participants had to score the presented words on a five point rating scale going from abstract (1) to concrete (5) [4].

Valence and arousal
Valence and arousal are both describing emotions provoked by stimuli. Valence evaluates the pleasantness of the stimulus and arousal evaluates the intensity of emotion provoked by the stimulus Values for valence and arousal have been established by a norming study where 1872 participants rated words on a scale from 1 to 9. In the case of valence they rated the emotion provoked by the word from happy (1) to unhappy (9). For arousal the rating scales goes from excited (1) to calm (9) [33].

3 Results and discussion

I performed multiple linear mixed effects analyses of the relationship between the amplitude of the N400 on nouns in the experimental sentences, described as average EEG potential between 300 and 500 milliseconds after word onset, and several features of these nouns. In the analyses I entered all the fixed effects (sentence position, word position, word length, frequency, LSA, concreteness, valence and arousal) as described in section 2.4.2 into the model. As random effects there were intercepts for subjects and items, to take by-subject and by-item variability into account. By centering all the fixed effects, any collinearity between the predictors with the intercept were reduced. After centering the fixed effects, the correlation between frequency and length was r = -0.69, which means they are not independent.

3.1 Reduced model

The first analysis without any interactions between fixed effects added to the model, yielded to the results in table 1, based on 6375 data points.

	mean estimate	standard error	t value	p value
sentence position	<0.0007	<0.001	0.728	0.467
word position	0.151	0.022	6.952	<0.0001
length	0.311	0.044	7.065	<0.0001
frequency	0.046	0.061	0.756	0.450
surprisal	-0.144	0.026	-5.581	<0.0001
LSA	1.632	0.502	3.251	0.001
concreteness	-0.197	0.085	-2.331	0.020
valence	0.124	0.063	1.984	0.047
arousal	-0.152	0.080	-1.907	0.057

Table 1: Mean estimates, standard errors, t values and associated P values for fixed effects in the reduced model. Statistics are based on 24 subjects and 6375 data points.

This first model revealed very strong statistically significant results for word position, number of characters, surprisal and LSA. Concreteness also revealed a significant effect and valence was considered to be weakly significant. From these values you can conclude that nouns that occur later in the sentence, nouns that have a higher number of characters, nouns that have a bigger LSA value and nouns that have a big valence value, all have the effect of the N400 amplitude becoming significantly smaller (more positive). When the values for surprisal and concreteness of nouns increase, the amplitude of the N400 becomes significantly larger. As you would expect the position of the sentence in the experiment is not a significant predictor for the N400. This model will be called the reduced model in the following parts.

3.2 Interactions between length, frequency and surprisal

To test whether interactions between some of the fixed effects are significant, I performed likelihood ratio tests using ANOVAs comparing the reduced model to the model with the interaction of interest. The interactions tested are of interest for my research question because of the explanation value they can have for the understanding of language processing and the competing theories as described in section 1.5.

While keeping all the other factors in the model, the interaction between the length and the frequency of the word revealed to be significant. The interaction between length and frequency affected the N400 (χ^2= 9.691, p <0.002), raising it by about 0.11 ±0.036 (standard errors). All the main effects that were significant in the reduced model were still reliably significant in the model with this interaction of length and frequency. But what was strongly noticeable is that the main effect of length dropped in t-value from 7.065 to 3.745, while the other factors almost did not change on their t-value, except for surprisal. The variance difference was mainly absorbed by the interaction effect and the main effect of surprisal. The interaction between word position and surprisal revealed, while keeping again the other factors in the model, to be marginally significant (χ^2= 3.453, p = 0.053), lowering the N400 by about 0.013 ±0.007 (standard errors). The interaction between surprisal and frequency did not reveal a significant effect (χ^2= 0.131, p = 0.71), just as the interaction between word position and frequency (χ^2= 0.555, p = 0.46). From these four models follows that length and frequency are significantly interdependent on each other and have an additive effect on the N400. Furthermore word position and surprisal are slightly interdependent on each other. The other significant main effects which occurred in the reduced model were still reliably significant in these three last models.

These findings are in line with the findings of Dambacher et al. [5]. Dambacher found a strong effect of predictability on the N400. Van Petten and Kutas already stated that the N400 amplitude was inversely correlated with predictability. The effect in this research which we could compare to that of Dambacher et al. is the effect of surprisal. The significant interactions that they found were interactions of predictability and frequency as well as of position and frequency, which pointed to an interplay of frequency and context in the sentence. They stated that these findings accounted for the same effect: The frequency effect decreased when context information increased. The found interaction of word position and surprisal as described in this section supports this view. In addition, surprisal absorbs a big part of variance from the effect of length and it was still one of the strongest effects on the N400. This makes surprisal a very appropriate predictor for the N400 amplitude.

3.3 Main significant effects

Besides the effect of the main effect of surprisal there were other main effects which have shown to influence the N400. In the previous constructed models the significant effects of LSA and concreteness were always present and reliable. The effect of valence was also always marginally significant present.

3.3.1 Valence

From the reduced model follows that when the valence value increases, the N400 amplitude decreases. This means that when a noun is rated more unhappy than happy, there is a smaller N400 amplitude and therefore it is easier to fit happy words in a sentence than unhappy words. This could point out that it is easier to acces positive words from the mental lexicon than negative words.

3.3.2 Concreteness

LSA and concreteness have shown to be very stable predictors for the N400. The concreteness effect has been shown before to be a significant predictor for the N400, as was mentioned in section 1.4.3. The analyzed models prove again that concrete words evoke larger N400 amplitudes than abstract words. Because of the fact that concrete words are easier to remember than abstract words, they would likely to be more easily accessible.

3.3.3 LSA

The LSA effect as found in these data (natural language sentences) has never before been demonstrated. The significant main effect of LSA shows that when the LSA value increases, the N400 amplitude decreases. This means that when the noun of interest has a strong semantic relation with the preceding context, the N400 amplitude drops. This is easy to explain by the fact that when there is a high semantic constraint of which word will follow next in the sentence, it is not difficult to predict which word will follow next and you will not easily be surprised by a word that you would not expect. This significant main effect of LSA on the N400 provides evidence for the theory of i.a. Hagoort. This theory believes that the N400 is evoked relatively late in the processing stream of word recognition. The N400 is evoked when the preceding context is already understood to integrate the following word into this context. The integration of the meaning of the upcoming word is already elicited before the word is recognized and a semantic representation is already constructed.

4 Conclusions

The purpose of this study was to investigate factors that affect the amplitude of the N400 ERP component during sentence comprehension. Since the discovery of the N400 component in 1980, there has been extensive research on this topic. This research lead to a variety of theories and a set of factors that directly influence the N400. This study supported significant results for the main effects of position of the word in the sentence, length of the word and concreteness on the N400 amplitude. Interactions between length and frequency, and position and surprisal revealed to have a significant influence on the N400 amplitude. These interactions were already expected on the basis of findings of Dambacher et al. (2006). Significant robust main effects of surprisal, LSA and valence resulted in seeing these factors as appropriate predictors of the N400 amplitude. The N400 amplitude decreases along with increasing LSA values or increasing valence (nouns rated as unhappy) value of a noun. When the value of concreteness or surprisal increases, the N400 amplitude increases along. Surprisal was presented as a very strong predictor of the N400 amplitude. Due to the found effect of LSA it is more likely to believe in the theory of Hagoort et al. [14] than in other theories: the N400 is elicited relatively late in the processing stream of word recognition.

5 Acknowledgments

I would like to thank both of my supervisors, Dr. S.L. Frank, for the knowledge he has taught me about mixed effects modeling and for providing and filtering the EEG data that is used in this thesis project, and Professor Dr. A.F.J. Dijkstra, for his advice and his many valuable suggestions during my thesis project.

References

[1] Baayen, R.H., Davidson, D.J., Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

[2] Barber, H.A., Otten L.J., Kousta, S.T., Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, 125, 47-53.

[3] Bentin, S., McCarthy, G., Wood, C.C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography & Clinical Neurophysiology*, 60, 353-355.

[4] Brysbaert, M., Warriner, A.B., Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 1-8.

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

[6] Donchin E. (1981). Surprise! ... Surprise? *Psychophysiology*, 18, 493-513.

[7] Frank, S.L., Fernandez Monsalve, I., Thompson, R.L., Vigliocco, G. (2013). Reading-time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, 45, 1182-1190.

[8] Frank, S.L., Otten, L.J., Galli, G., Vigliocco, G. (2013). Word surprisal predicts N400 amplitude during reading. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 878-883.

[9] Ganis, G., Kutas, M., Sereno, M.I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8, 89-106.

[10] Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 838?899.

[11] Hagoort, P., Baggio, G., Willems, R.M. (2008). Semantic unification. *The New Cognitive Neurosciences*, 4, 819-836.

[12] Huang, H., Lee, C., Federmeier, K.D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *Neuroimage*, 49, 1116-1123.

[13] Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, volume 1, issue 6, 571-591.

[14] Kutas, M., Federmeier, K.D. (2011). The N400. *Annual Review of Psychology*, 62, 621-647.

[15] Kutas, M., Hillyard, S.A. (1988). Contextual Effects in Language Comprehension: Studies Using Event-Related Brain Potentials. In: F. Plum (Ed.), *Language, Communication and Brain*, 87-100.

[16] Kutas, M., Hillyard, S.A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, 11, 99-116.

[17] Kutas, M., Hillyard, S.A. (1980). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, 11, 354-373.

[18] Kutas, M., Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.

[19] Kutas, M., Hillyard, S.A., Gazzaniga, M. S. (1988). Processing of semantic anomaly by right and left hemispheres of commissurotomy patients. *Brain*, 111, 553-576.

[20] Kutas, M., Van Petten, C.K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. *Handbook of Psycholinguistics*, 83-143.

[21] Landauer, T.K., Foltz, P.W., Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

[22] McPherson, W.B., Holcomb, P.J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, 36, 53-65.

[23] Oostenveld, R., Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112, 713-719.

[24] Osterhout, L., Holcomb, P.J. (1995). Event-related brain potentials and language comprehension. *Electrophysiology of Mind*, 171-215.

[25] Osterhout, L., Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

[26] Pratarelli, M.E. (1994). Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain and Cognition*, 24, 137-157

[27] Schmitt, B.M., Münte, T.F., Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, 37, 473-484.

[28] Schuberth, R.E., Elimas P.D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 27-36.

[29] Squires, N.K., Squires, K.C., Hillyard, S.A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography & Clinical Neurophysiology*, 38, 387-401.

[30] Steffen, J. (2004). N-Gram Language Modeling for Robust Multi-Lingual Document Classification. *The 4th International Conference on Language Resources and Evaluation.*

[31] *The British National Corpus*, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

[32] Van Petten, C.K., Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18, 380-393.

[33] Warriner, A.B., Kuperman, V., Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207.

[34] Zhang, Q., Lawson, A., Guo, C., Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin*, 71, 316-323.

6 Appendix A

Table A shows all 205 sentences used in the reading study experiment.

Table A

Anne lost control and laughed.
Billy wrote on the envelope.
He called over his shoulder.
He stayed against the wall.
Helen ran to the toilet.
I cannot tell you more.
Michael gave Adam two pounds.
Roger smiled and sat down.
Things were looking up again.
This man's hair is black.
Turn your hearing aid up.
You'll demand nothing of me.
A horse has thrown a shoe.
A tube train took his legs.
Five will be listening for sure.
Gordon felt his anger build up.
Harry smiled at his little brother.
He needed putting in his place.
He shouted and the calling ceased.
He uses his belt on you.
He's broken his wrists very badly.
I am touched deeply in places.
I don't want to hear it.
I knew I heard your voice.
I wouldn't want to miss this.
Jack walked up to the door.
Ross sat back down and shivered.
She couldn't help feel for Alex.
The bored looking soldier just pointed.

Continued on next page

Table A - *Continued from previous page*

The campaign will pay for itself.
They're riding out to meet them.
We barely escaped with our lives.
We've been travelling almost three days.
Alice is looking for somewhere to live.
Alan forgot to shave again this morning.
Have a carriage sent for us immediately.
He probably doesn't want people to know.
He put the fork on the table.
I'll give him more than a slap.
Andrew nodded and stood against the counter.
Looks like a walk did the trick.
Sam felt proud when he went home.
Scott got up and washed his plate.
Simon nodded and walked into the room.
Steve sat down to eat his lunch.
Ted smiled and bit his bottom lip.
Thomas stepped back and shook his head.
Tom stood up and wiped his jeans.
Tony stopped off at his first delivery.
Normally they want to cut things up.
Powerful in the wrong hands, is that.
She had refused to leave her home.
She hid the boxes and matches well.
The bored looking soldier continued looking bored.
The dead must not kill the living.
The prison guard walked along the row.
The van is your lead, follow that.
You're getting up early in the morning.
After lunch he was told he had visitors.
Albert spoke before Barbara had a chance to.
Come within the walls to escape the enemy.
Follow me closely and watch for their leader.

Continued on next page

Table A - *Continued from previous page*

He called out but she did not move.
He sat down and began to eat breakfast.
He slammed the knife down on the lamb.
He staggered into the chair at his desk.
I do not spare the birds a killing.
I think our friend's demands are well justified.
Let us leave these poor people in peace.
Walter closed the door and took a breath.
William sat down and tucked into his breakfast.
Alec stood against the wall to steady himself.
No-one else had dared step within the ring.
None of the city people dress like that.
She saw the throwing knife in his hand.
She was smiling in the dark, teeth flashing.
Shut your smart foul mouth and sit down.
Tell me what really happened to the truck.
The door was opened and Alexander walked out.
There were choked screams and shouts of dismay.
There will be men to kill either way.
They were definitely holding hands underneath the table.
Claudia still teaches and still claims to hate it.
Burn him now, don't wait for him to escape.
Don't die of cold before you die in battle.
Good God, if you are watching, give me strength.
He heaved his shoulders and swung the axe again.
He is after glory and heads for his hall.
He just wished he could talk to his dad.
He lay his sword aside and extended his hands.
He sighed and walked back to the wood shop.
He smiled when she let out a big yawn.
He was leaning against his truck smoking a joint.
I was thinking of getting the band together again.
It is a narrow ship, much like our own.

Continued on next page

Table A - *Continued from previous page*

Arthur placed the bars of chocolate on the counter.
John gave Bernard a hard slap on the head.
Let's get out and I'll buy you a drink.
Let's have the gates closed and go home early.
Andrew closed the office door on the way out.
Charles felt sick to the pit of his stomach.
Christopher groaned and lay his hands across his stomach.
Dave kicked him in the stomach and Paul groaned.
Dexter opened his letter and found a photograph inside.
Duncan took a breath and wanted to be sick.
Edward washed his hands and flicked water at Bob.
She turned and saw him standing in her path.
They turned round just as Frankie walked toward them.
Underneath the table her hand was gripping the chair.
Whether Peter heard him or not Gabriel didn't care.
You give me a lead and I'll follow it.
All the bowls from the empty tables are cleared up.
His dark hair was just visible under his white hat.
His men will not follow him and they are few.
I'll break your neck if you don't fix that water.
Ian gave Geoffrey a hard slap across his right ear.
Jake smiled again at his baby daughter, she was perfect.
Jimmy liked the fact they were calling him Jim.
Brian sniffed and wiped his grubby hand across his nose.
Phil let him go and Charlie fell to the ground.
Let me through and you'll hear nothing more of it.
Men could be paid to do all sorts of things.
Gilbert sat as still as he could nursing his lip.
Jacques sighed and tried to move out of the way.
Jamie took a cigarette and Andy gave him a light.
Joseph walked into the office and removed his flat cap.
She would soon be found if she tried to hide.
The brothers stood up and came down from the platform.

Continued on next page

Table A - *Continued from previous page*

Anthony was approached, just as he had predicted, in the park.
Despite what Frank had told her Ellen thanked the young soldier.
He called earlier and I already arranged to go swimming tomorrow.
He drove his truck home and walked down to the village.
He held his mother tight and she cried into his chest.
He went to the door and walked out of the room.
His helmet was lost but he could still do some damage.
Ken just hoped he wasn't taking his temper out on Chris.
Lawrence rubbed the back of his neck and hung his head.
Lewis sat down and Lucy laid her hand on his cheek.
Matthew stood against the wall with his hands behind his back.
Run along home now and bring us some men to kill.
She just kept telling herself he had to learn a lesson.
She wound up talking to Barry about her time in India.
The gate was double, one huge mass of wood following another.
The gathered villagers took a step back, almost as one man.
We shall arrive properly or you will never live this down.
While they sat drinking their tea, Mick dug them a trench.
You know better than to talk to your mother like that.
You know that I will do whatever you ask of me.
You shouldn't have been racing my truck in the first place.
A smile came to his lips as he thought of an idea.
All last week there was a white van parked across the street.
Finally Maria sat down with a cup of tea and a sandwich.
He didn't seem to mind that she was screaming in his ear.
He drank too much at times and washed too little at times.
He folded his arms across his chest and put out his lip.
He had another shock when they told him he was moving cells.
He put his hands in his pockets and leaned against the truck.
He was built with muscle and liked to throw his weight around.
I am trying, old man, but I would rather kill the ship.
I shall have a victory and I want it to be remembered.
In the morning Mickey stood straight with his hand behind his back.

Continued on next page

Table A - *Continued from previous page*

In the morning at breakfast Mike smiled at the letter he received.
It had been two weeks since Philip had been sent to prison.
It is not the most fragrant of places but we make do.
James walked in to wake Samuel at four thirty in the morning.
Let us give battle to these foreign snakes once and for all.
Tommy came out of the staff room toilet looking a little pale.
Michael knew they never hit him and that was half the problem.
Alan told Jason he would just sit outside to have his lunch.
Now the man is coming who will give it back to me.
She threatened to leave before and this time she just had enough.
The first knife caught in her skirt and she shook it free.
The last thing Alex felt like doing after work was more work.
The truck was an older version of the ones the others drove.
They stared at the black wall and now this offered no comfort.
We should talk about the things people talk about on first dates.
Why don't you tell me what happened on Saturday, step by step.
Emma put her mug of tea down on the table with a bang.
Bruce is looking for someone to take the spare room in his flat.
From the moment we leave this spot be very careful what you say.
He knocked on the door so hard that he almost broke it down.
He let Donald go and he fell to the floor with a thud.
He smiled again and felt like a man and not just the kid.
He's got the ability to work but he just gets bored too easy.
I can't see any amount of talk getting you out of this mess.
I flung out my hands and they said I had broken my wrists.
I mostly just pick a dry white and drink it to get drunk.
Andy felt good when he got ready to go out on Saturday night.
Billy knocked on the door and waited till he was told to enter.
Bob sat at the entrance to the warehouse and made up a cigarette.
Brian sat down at an empty table and began to eat his breakfast.
One by one she placed her teeth about their necks and killed them.
The shout caught him off guard and he fell back a little way.
He rose from his seat and stopped mid way when Joe glared at him.

Continued on next page

Table A - *Continued from previous page*

He sat his son up on the table and laid his belt beside him.
If I have time at the end I'll fill you in on what happened.
Henry washed his hands and sat on the seat in front of the stove.
The police officer got out of the van and picked up the two cans.
The speaker turned to her with a smile and a bow of his head.
They say he can breathe fire and kill a man with a single word.
He came back to his senses when he heard his dad stand up to him.
He was already up and dressed and invited us in for a cup of tea.
I know Richard drinks a little too much and has a bit of a temper.
If this were a movie instead of a book this would be a good bit.
James was only in the cell for an hour but it felt like a week.