Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

Ane Søreide Grytten

Reference in sentence processing: An ERP study

Master's thesis in Language Studies with Teacher Education
Supervisor: Professor Giosuè Baggio
Co-supervisor: Doctor Isabella Fritz

June 2020

NTNU
Norwegian University of
Science and Technology

Ane Søreide Grytten

Reference in sentence processing: An ERP study

Master's thesis in Language Studies with Teacher Education
Supervisor: Professor Giosuè Baggio
Co-supervisor: Doctor Isabella Fritz
June 2020

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

Abstract

In this study, we investigated how the brain processes sentences with referring and non-referring indefinite noun phrases. For the current purpose, a referring noun phrase is defined as denoting a specific entity which can be identified, whereas a non-referring noun phrase has a generic interpretation. The two conditions (referring and non-referring) were examined by recording event-related brain potentials (ERPs) while participants read sentences and subsequently answered comprehension questions. The critical word in both conditions was a sentence-final noun, which was introduced by an indefinite article. The different meanings induced by the two conditions has been highlighted in the literature; therefore, we expected differences in ERPs when comparing the two conditions. Based on previous studies on referential processing, the hypothesis was that there would be a difference in ERPs when participants read sentences with a referring NP versus a sentence with a non-referring NP, and that they would easily be able to distinguish between them in the comprehension questions that followed.

However, our results showed that there was not a significant difference in brain activity for the two conditions. In the behavioral data we also found that participants were not able to easily distinguish between the two conditions. The accuracy in the referring condition was considerably higher than in the non-referring condition, which may indicate that they either answered by random, or preferred to interpret the sentence as referring. We did, however, see that the referring condition was more easily processed (shorter RTs), and as the accuracy was higher, may have been the preferred interpretation.

These findings are interesting since they highlight that, in the absence of a textual or conversational context, we are not that sensitive to specific versus non-specific objects. This is true even though the literature states that these expression types mean different things and should evoke different mental representations, and we could thus assume that they would be processed differently. The fact that almost half the participants did not get the distinction between the conditions even after the initial processing may imply that we settle on a "good-enough" approach to processing when multiple interpretations of a sentence were possible, like previous studies (e.g., Ferreira et al., 2002) have found.

Sammendrag

I denne studien har vi undersøkt hvordan hjernen prosesserer indefinitte nomenfraser med og uten referanse. En nomenfrase med referanse defineres i denne sammenhengen som noe som peker ut en spesifikk enhet som kan identifiseres. En nomenfrase uten referanse har en generisk betydning. De to betingelsene ble undersøkt gjennom hendelsesrelaterte potensialer (Event-Related Potentials, ERPs) mens deltakere leste setninger. Det kritiske ordet i begge betingelsene var nomenet, som alltid var plassert til sist i setningen. Nomenet ble introdusert av en ubestemt artikkel (*en*, *ei* eller *et*). Atferdsanalytisk datainnsamling ble også gjort gjennom ja/nei-spørsmål som ble etterfulgt setningene som deltakerne hadde lest. Svarene vi fikk på disse spørsmålene hjalp oss å forstå om deltakerne faktisk forstod forskjellen mellom de to betingelsene. De to konstruksjonene vi har undersøkt har forskjellige betydninger, noe som har blitt fremhevet i språkvitenskapen. Hypotesen vår var deretter at vi ville kunne observere forskjellige ERP-effekter når deltakere leste setninger med nomenfraser med referanse sammenlignet med nomenfraser uten referanse. Vi antok også at deltakerne ville klare å skille mellom de to betingelsene i spørsmålene som etterfulgte setningen.

Resultatene av datainnsamlingen tyder derimot på at det ikke er en signifikant forskjell i hjerneaktivitet for de to betingelsene, og ingen ERP-effekt ble utløst. Fra den atferdsanalytiske datainnsamlingen fant vi at deltakerne heller ikke klarte å skille godt mellom de to betingelsene. Derimot viser dataen på at betingelsen med referanse hadde høyere antall riktige svar, som kan indikere at deltakerne enten svarte tilfeldig, eller foretrakk analysen med referanse. Resultatene tydet også på at setningene med referanse var lettere prosessert (kortere responstid, RTs) og siden antall riktige svar var høyere for denne betingelsen kan det tyde på at dette var den foretrukne tolkningen.

Funnene er interessante da de fremhever at vi kanskje ikke er så sensitive til forskjellen mellom spesifikke og ikke-spesifikke enheter når de ikke er presentert som en del av en større kontekst. Dette, selv om disse betingelsene ifølge litteraturen betyr to forskjellige ting, og som man skulle anta blir prosessert forskjellig. Det faktum at nesten halvparten av deltakerne ikke så ut til å være klar over forskjellen selv når de ble spurt om betydningen av setningene, kan tyde på vi ofte finner oss i en «god-nok»-modell av setningsprosessering dersom flere analyser er mulige, slik som tidligere studier har tydet på, blant annet Ferreira et al. (2002).

Acknowledgements

When I started at NTNU as a teacher student 2015, I never had imagined ending up here five years later. The process behind this master's thesis has been quite a journey – the most challenging in my life so far. Not to mention the last semester, in the middle of a pandemic. Nevertheless, it has been a most rewarding process, and I have learned so many valuable skills that I will treasure forever.

This thesis is a result of a hugely co-operative project. First and foremost, I need to thank my supervisors: Professor Giosuè Baggio and Postdoctorate Candidate Isabella Fritz, for this unique opportunity. Thank you for taking me on as a student without knowing anything of my academic skills beforehand, and for allowing me to be a part of the Language Acquisition and Language Processing Lab. Thank you for your wisdom, your cooperation, and for your help in running and analyzing the data. This thesis wouldn't be possible without you! I also need to thank previous master student Anne Marte Haug Olstad, who along with Isabella created the stimuli and programmed the experiment.

Secondly, I want to thank all my participants. This thesis has required a large number of volunteers, whether it was for the norming tests or the experiments itself. I also need to thank co-students and members of the lab who have helped me out with all sorts of advice and technical aspects of the study.

I also need to thank my family for always believing in me, I needed it! Thank you for always wanting to help, whenever possible, and for showing interest in what I've been studying, even at the times when it didn't make sense. I would also like to thank my friends and co-students for inspiring me, supporting me, checking up on me, and making sure I was doing well.

A huge thank you must also be given to a specific few:

To Mary Chandler, for all your helpful feedback; to my wonderful roommate for the last five years, Hanna; to my sister and best friend, Eline; and last, but definitely not least, my boyfriend, Ole, the most supportive, generous, and kind soul.

I have loved being a student at NTNU, and in particular, being a master student.

Contents

1. Introduction

Language is one of the main capacities that define human cognition. In recent times, scientists have become interested in learning more about the unique language mechanisms in the brain. Language in the brain has thus become a popular area of study in a branch of research on the bridge between psychology and linguistics called *psycholinguistics.* Psycholinguistic researchers ask questions about *where*, *how* and *when* language is processed in the brain, and which neural mechanisms are involved in language comprehension. Research in the area can tell us which neural mechanisms take place when we process language. Language processes in the brain are rapid, complex and only partly and indirectly accessible (Van Berkum, Koornneef, Otten & Nieuwland, 2006) and many of the processes that take place are not available for introspection. However, with the help of experimental methods we can test various linguistic theories in language comprehension studies. Words that we acquire are stored in the mental lexicon of our brain. These lexical items are assumed to be stored blocks; e.g. phonological, morphological, and syntactic building blocks. Still, when we process language, we need to be able to retrieve more than the single objects mentioned, we also need to be able to combine elements from memory in novel ways (Hagoort, 2016). The processes and neural mechanisms behind language processing have been studied in innumerous ways, testing an abundance of linguistic theories and manipulations.

The current study investigates how the brain processes sentences containing referring and non-referring noun phrases. Saeed (2016) defines reference as a moment-to-moment relationship. This means that the entity someone refers to by using a word will depend on the context in which it is presented. Reference is thus linked to both world and word knowledge. Language interpretation is a complex business, and we need to ground our neural research into precise theories that take semantic, pragmatic, and behavioral findings into account (Van Berkum, 2012). In our study, we have used event-related potentials (ERPs) to find out which component(s) were elicited when participants read the stimuli containing referring and non-referring expressions. These online measures are combined with behavioral measures to help us interpret our results. Different ERP components are related to different semantic and/or syntactic processes, which give us information about how and when different specific linguistic features are processed and analyzed. Behavioral data collected in the form of comprehension questions will allow us to compare online and offline measures, as they both can provide insight to the processing of the other (Kaiser, 2013). Online measures provide insight to the earlier stages of sentence processing, whereas offline measures provide information about the processes that occur later in sentence processing (Warren, 2013). More about this will follow later in the thesis.

1.1 The present study

In this study, we investigated the processing of referring and non-referring indefinite noun phrases (NPs) during online sentence processing. This study will try to answer the following research question:

Does the brain distinguish between referring and non-referring indefinite noun phrases during online sentence processing?

To answer the question, we used a stimuli set consisting of 100 sentences. The stimuli functioned as filler data in a previous study by Anne Marte Haug Olstad (2019), and the study was expanded by using the previously obtained data as well additionally collecting data. For the present study, the stimuli consisted of sentences with either a referring or a non-referring indefinite NP as the critical word. Every indefinite NP was used in both a specific and a non-specific context, resulting in 50 pairs. In the experiment, participants were asked to read sentences, as well as answer comprehension questions about them afterwards. An example of sentences from each condition in the stimuli is presented below:

Condition	*Sentence*	*Det.*	*Critical word*
Referring	Ute på tunet vårt har vi ei	ei	ku
	Out in our field we have	*a*	*cow*
Non-referring	For å få melk må man melke	ei	ku
	To get milk one must milk	*a*	*cow*

Table 1.1: Example of the two conditions and the critical word appearing sentence-finally. English word-by-word translations in italics.

The comprehension questions following the sentence were the same for all sentences in the two conditions, and the correct answer would depend on condition:

		Answer
Question	Er det **ei** spesifikk **ku**?	**Ja/Nei**
	*Is it **a** specific **cow**?*	***Yes/No***

Table 1.2: Example of comprehension question. Words in bold would be replaced by the indefinite article *en*/*ei*/*et* (*a*/*an*) and the noun in question. The correct answer would always be *yes* for the referring trials and *no* for the non-referring trials. English word-by-word translations in italics.

The construction of the stimuli was done by previous master student Anne Marte Haug Olstad (2019) and co-supervisor Isabella Fritz. Some of the EEG data I used had already been collected, but several additional EEG experiments were conducted to get enough data for the condition that I was testing. Moreover, I collected additional behavioral data to norm the stimuli, which will be further explained in the Methods section.

The hypothesis was that there would be a visible difference in ERPs in the referring and non-referring conditions. In specific, we wanted to see whether the critical nouns would be processed differently when presented in a referring context and a non-referring context, meaning whether the noun denoted a specific entity or whether it had a generic interpretation. We also expected that we would find a different ERP component than the N400 effect, on the basis of previous studies looking at reference in language comprehension, that are in brief explained below. Moreover, we predicted that participants would easily be able to distinguish between the two conditions, something that would be visible in the behavioral data that we collected by asking comprehension questions during the online data collection. This would be visible in the results if the majority of the participants had a high accuracy rate in both conditions.

From previous research on referential processing, we know that the brain handles referentially complex situations in different ways. It has also been proven in several EEG experiments that we can use ERPs to selectively keep track of these referentially complex situations (Van Berkum et al., 2006). In earlier EEG experiments testing reference, referential critical items were found to elicit a P600 effect, in experiments with referential ambiguity, i.e. switching out the pronoun to one that does not match the gender of the antecedent. Another important finding was a sustained anterior negativity starting at around 300-400 ms. This component has been named the Nref effect, as it seems to be linked to language and memory in referentially ambiguous situations. It became visible for ERPs time-locked to a word that has more than one possible antecedent in a previously established discourse model, when compared to an unambiguous referent (Van Berkum, Brown, Hagoort & Zwitserlood, 2003a). The present study is different from most other studies looking at referential processing, as the

sentences tested are neither complex nor ambiguous or anomalous.

1.2 Overview

The following section will present the theoretical background for the present study. I will give an overview of what referring and non-referring expressions are, and how these are defined, and are distinguished from each other, in the literature. The second part of the theoretical background will present important features related to language processing, and the cognitive mechanisms underlying sentence interpretation during reading. This will help me analyze the findings of the present study. Following that, I will give an overview of EEG, the method used in this study, and how this method has been used to collect ERPs, resulting in innumerous discoveries on the factors underlying language processing. Some important ERP components and findings that are relevant to the present study will also be presented. Following that, I will present the methods used for the study, consisting of both online and offline data collections. This section will also present the stimuli that were tested, using both norming and EEG, and the reasoning for these tests will be thoroughly explained. Following that, the ERP results and the behavioral results will be presented, and afterwards, this will be discussed in the following section. Finally, I will conclude the thesis by arguing for what our findings suggest, as well as mentioning what further research in the area should investigate.

1.3 Conventions

In the following sections, some abbreviations and linguistic terms will be used. Firstly, I will use the abbreviation NP to indicate noun phrase. Other than that, EEG is short for electroencephalography, and ERPs is short for event-related potentials. Also, as the stimuli are in Norwegian, I have provided word-by-word translations of the sentences, which means that the word order will often not be in correct English, and some of the sentences will be ungrammatical or unacceptable. These will appear in cursive, under the sentence in question. In the presentation of the stimuli or other experiments, the critical words, from which ERPs were measured, will always be indicated in **bold** letters.

2. Defining reference

In this section, we will take a closer look at what *reference* means in terms of language and comprehension. Can an NP like, *a girl*, have different meaning? According to the literature, *a girl* can be used in both a specific situation, denoting a specific girl, or it can be used in a non-specific context, speaking of girls in general. This will be exemplified in the following section. The underlying idea is that there is a difference in *linguistic meaning* and *world knowledge* (Hagoort, Hald, Bastiaansen & Petersson, 2004). The terms *reference* and *specificity* have been widely discussed in the literature and are closely connected to each other. A recap of some popular theories and definitions will be given in this section, and following that, we will see how reference can be understood in terms of psycholinguistic language processing.

Already in 1892, the famous logician Gottlob Frege claimed that there was another side to meaning, involving reference. This involves finding out what referring expressions contribute semantically to the phrase, and to the sentences in which they occur. Knowing the meaning of a word is one thing, but understanding what that word means in the context in which it is presented, is just as important. Language users encounter these terms often, and it is therefore necessary to not only be able to understand the sense of the word, but also who or what it refers to (Van Berkum et al., 2006). We use words to mentally identify parts of the world and make statements about it, and we use nominals to refer to specific entities in the world. Nominals are often used in discussion of referential possibilities, as they are the linguistic unit which most clearly reveals this function in language (Saeed, 2016, pp. 11-24). The focus on nominals will be reflected in the stimuli for the present study, which will be presented in a later section. More specifically, we will focus on indefinites, as Norwegian indefinites behave like English indefinites with an indefinite article preceding the noun: *a boy* → *en gutt*. This is different from definites, as they in Norwegian require a suffix rather than a definite article: *the boy* → *gutten*. Referring and non-referring indefinites are therefore interesting from a crosslinguistic point of view, which is something we will look closer into in the following section.

2.1 Referring and non-referring expressions

As mentioned above, nominals are often used in discussions of referential possibilities as they are the linguistic unit which most clearly reveals this function in language. Saeed (2016) states that expressions can be either referring or non-referring, which means that they can either be used to identify an entity, or as a more generic expression where no specific entity is identified. Some sentences are also ambiguous and can be interpreted as both a referring and a non-referring expression.

Saeed (2016) explains the distinction between referring and non-referring as instances for when speakers use elements in the sentence to refer a specific entity, and instances when they do not. The same indefinite noun phrase (NP) can be used in a referring context:

(1) *They performed a cholecystectomy this morning.*

And in a non-referring context:

(2) *A cholecystectomy is a serious procedure.*

(Saeed, 2016, p. 25)

The difference between the two is that the NP in (1) refers to a specific procedure, whereas the NP in (2) has a generic interpretation that describes the procedure in general. Whereas some expressions have *constant reference*, like nominals such as *Barack Obama* or *the Eiffel Tower*, other expressions, such as *a girl* or *the President of the United States* have *variable reference* (Saeed, 2016), meaning that they are context dependent.

Referring expressions seem to denote a *specific* event or object. Von Heusinger (2007) uses the term specificity when defining reference. He argues that in terms of reference, indefinite NPs behave a lot like definite NPs in that they both have a *specific* and a *generic* reading. He presents the examples below (from p. 253):

(3) *A body was found in the river today*. (specific)
(4) *A tiger has stripes*. (generic)
(5) *The body was found in the river yesterday*. (specific)
(6) *The tiger has stripes*. (generic)

Whereas (4) and (6) have the same meaning (unless *the tiger* refers to a previously introduced discourse referent), the difference between (3) and (5) has to do with whether the body is known for the hearer. Otherwise, they both refer to a specific body. Thus, we can summarize that the term specificity is used to describe referring expressions, and non-specific or generic can be used to explain non-referring expressions.

Expressions with indefinite NPs can also have an ambiguous reading where two interpretations of the NP are possible. This is called the *lexical ambiguity approach*. In the example below, the indefinite NP *a student* can have either a specific interpretation, where the referent is determined (3a), or a non-specific, plain existential interpretation, like in 3b.

(7) *A student in Syntax 1 cheated on the exam.*

a. *His name is John.*
b. *We are all trying to figure out who it was.*

(Von Heusinger, 2007, p. 245)

Von Heusinger suggested a scale to measure the *identifiability*-criteria for definiteness and specificity. The practical overview is presented below:

identified by	definite (+ specific)	indefinite specific	indefinite non-specific
speaker	+	+	-
hearer	+	-	-

(Von Heusinger, 2007, p. 249)

In general, specific indefinite NPs are assumed to be speaker-known but hearer-unknown. Von Heusinger treats specificity as a referential property of the NP, so that we have a specific and a non-specific reading possible from each indefinite NP. A specific indefinite is assumed to have a wide scope and a referential reading - meaning that we have a specific referent in mind when encountering the NP. In terms of definiteness and specificity, Enç (1991) states that definites require that their discourse referents are linked to the previously established context, whereas for indefinites, this is not the case. An indefinite cannot be linked to the previously established context, but must be novel in the sentence, introducing discourse referents that have not previously been established. As seen in the scale above, this confirms that an indefinite is never hearer-known, but in specific-contexts, it is speaker-known.

2.1.1 Discourse referents

Karttunen's discussion on discourse referents is largely inspired by Noam Chomsky's discussion of *referential indices* in transformational grammar. Each noun phrase is associated with a referential index, and in his proposal, Chomsky wished to augment the notion of phrase identity. Adding to this, Karttunen argues that constraints on co-referentiality extend beyond the sentence level (1969).

Under what circumstances are discourse referents established? Karttunen asks the question: "*When is there supposed to be an individual associated with an indefinite NP?*" (Karttunen, 1969, p. 21) As we have seen, indefinite noun phrases have a specific and a non-specific interpretation. Examples (8) to (15) below are from Karttunen (1969, pp. 5-23). The

first example exemplifies the two possible interpretations of an indefinite NP with an ambiguous reading, similar to the ambiguous case presented by Von Heusinger above:

(8) *Bill didn't see a misprint.*
   a. *There is a misprint which Bill didn't see.* (specific reading)
   b. *Bill saw no misprints.* (non-specific reading)

The examples presented above represent the different interpretations of sentence (8). Sometimes, the verb preceding the NP can say something about the specificity of the object.

(9) *John tried to find a piano.* (…but he didn't succeed in finding one)

(10) *John tried to lift a piano.* (…but he didn't succeed in lifting it)

The sentences look almost identical, except for the verbs *to find* versus *to lift.* The latter suggests that a piano in (10) is a specific object, whereas sentence (9) only provides information about the kind of object John wants to find. Thus, we can argue that sentence (10) establishes a discourse referent, just as sentence (8a) did. Sentence (9) and (8b) fail to do so.

A group of verbs that presuppose the existence of the entity or event denoted by the following NP are called factive verbs, where the truth of the proposition is represented by the complement. Examples of factive verbs are *know*, *realize* and *regret.*

(11) *John knew that Mary had a car.*

The indefinite NP *a car* establishes a referent when it is the complement of a factive verb like *to know*. The truth of the proposition is true even if the verb is negated:

(12) *John didn't know that Mary had a car.*

In opposition to factive verbs there are non-factive verbs, such as *believe, think, say,* or *doubt*. In general, a nonfactive verb does not presuppose the truth of the embedded proposition. The speaker is not committed to the truth of the embedded proposition, but it represents his or her beliefs or attitudes to the utterance. An indefinite NP in the complement of a nonfactive verb implies a positive belief that establishes a referent or a peculiar sort.

(13) *I doubt that Mary has a car.*

However, a nonfactive verb that implies a positive belief, like *claim, think, believe, say* allows an indefinite NP in the complement to establish a referent of a peculiar sort, namely in the world of the subject person.

(14) *Bill says he saw a lion on the street.*

Nonfactive verbs seem to establish ambiguous discourse reference. The use of quantifiers can also lead to ambiguity. Indefinite NPs are generally ambiguous in sentences containing quantifiers, meaning that they have several interpretations.

(15) *Harvey courts a girl at every convention.*

This sentence can mean either that at every convention, there is a girl that Harvey courts, or that there is some girl that Harvey courts at every convention, thus giving both a non-specific and specific interpretation. In the non-specific reading, Harvey may court a different girl each time, whereas in the specific reading, he always courts the same girl.

2.1.2 Structural features: R-expressions

A referring expression, or an R-expression, is an NP that gets its meaning by referring to an entity in the world. This is true both for the actual world and a fictional world created by the speaker[1]. These NPs get their meaning from the context and discourse around them. The examples below are from Carnie (2007, pp. 135-143).

(16) *Felicia wrote a fine paper on Zapotec.*

The meaning of the word Felicia comes from the situation it is uttered in, and the NP "a fine paper on Zapotec" presupposes that there is some paper in the world, about Zapotec, that Felicia wrote. These two expressions get their meaning by referring to objects in the world. The majority of NPs are R-expressions, but not all. In some sentences, reflexive pronouns are used to refer back to an antecedent. These are called anaphors.

(17) *Heidi*(i) *bopped herself*(i) *on the head with a zucchini*[2].

Herself is used as an anaphor in this case, as it is pointing back to its referent, the R-expression Heidi. This NP obligatorily gets its meaning from another NP in the sentence and cannot refer to anyone other than Heidi (below, we will see some experiments that test this type of anaphor by changing out the referent to a pronoun that does not match the gender). The theory of the syntactic restrictions that decide where different types of NPs can appear in the sentence is called Binding Theory. An anaphor must refer back to an antecedent in the same sentence, as an anaphor must be bound in its binding domain. R-expressions receive their meaning from outside the sentence, which means that they cannot be bound at all. They cannot get their meaning from another word in the sentence via binding and do not seem to allow any instances to bind at all.

[1] This is relevant to mention, as some of the stimuli in this study include fictional world references. See the whole list of stimuli in Appendix B.

[2] The (i) is an indice. When two indices have the same letter, it indicates that the NP they refer to is the same.

(18) **Art*(i) *kissed Geoff*(i).

In sentence (18), Art and Geoff cannot co-refer to each other, but must refer to two different people, meaning that this sentence cannot have the implied reading as suggested above. This means that syntactically, R-expressions look similar. The critical words in all sentences are an R-expression, both the sentences with and without reference. For the sentences in our experiments, the structural relations should be similar across conditions.

2.1.3 Indefinite noun phrases in Norwegian

In Norwegian, indefinite articles mark the gender of the following noun, as opposed to English, which uses the neutral *a* and *an*. Just like in English, Norwegian indefinites have a referring and non-referring usage. In their discussion on reference grammar in Norwegian, Faarlund, Lie and Vannebo (1997) argue that different determining elements give noun phrases different kinds of referential properties. In Norwegian, indefinite articles are only used for countable nouns, like in English. However, for uncountable nouns, it is missing. The examples below are extracted from Faarlund et al. (1997, pp. 284-291), and English word-by-word translations are provided in italics.

(19) Vi kjøpte ei bok.
We bought a book.

(20) Vi kjøpte smør.
We bought butter.

Also, just like in English, indefinite NPs can have a generic interpretation.

(21) En ulv er et rovdyr.
A wolf is a predator.

The statement refers to the whole species, giving the expression a generic reference. Just like in English, these can come about using both definites and indefinites. Generally, indefinites seem to behave the same way. There are, however, some unique features in Norwegian, like the bare noun. This is unacceptable in English:

(22) Kari er lærer.
**Kari is teacher*[3].

(23) Han spilte klovn i stykket.
**He played clown in the play.*

Another Norwegian variation is using the indefinite article as a determiner for a non-specific time or place:

(24) En gang må det ta slutt.

[3] The star signals that the sentence is ungrammatical/unacceptable.

**A time it has to end. (Meaning some time or other)*[4]

(25) Et sted må vi kunne få fred.
**A place we have to get peace.* (*Meaning some time or other*)

2.2 Properties of sentence processing

The properties involved in sentence interpretation will be important to clarify in this study, as they are key in order to analyze our findings. This section will discuss sentence processing during reading, as it is the method used in the present study. How do we carry out the processing of sentences, and how do we build sentences using the input from each word?

When reading in our native language, we immediately begin to perceive meaning. Not only must we understand each individual word as we come across them, we must also work out their syntactic relations. Moreover, we need to build a conceptual analysis of the sentence in that specific context, in order to understand what is said. Thus, the meaning of a larger constituent, such as *a girl* are constructed from the meanings of its individual words (Van Berkum et al., 2003). It is also believed that as we come across these words and structures, we create a mental model representing the information in the input (Warren, 2013). In psycholinguistic research, these semantic, syntactic, and referential analyses can be studied using both online and offline measures.

To understand a sentence, we need to build a syntactic structure for it. This structure can be explicitly marked to facilitate processing, which will be especially helpful in more complex sentence structures (Warren, 2013). As we encounter a new word in a discourse, we immediately link it into a syntactic tree structure. For that to happen successfully, we need to know the type of word that we are dealing with at each stage. As we read, we try to link each word to a syntactic category, and this category can often be apparent from the structure of the tree which has been constructed up to that point (Warren, 2013). It seems then that during reading, it is more important to link words into the correct syntactic category, rather than considering the word's communicative function. Warren (2013) also discusses the alternative view; that the language processor builds different syntactic structures for the same clause immediately, which allows for multiple analyses of the same word, especially in ambiguous sentences. Only after the initial reading must the parser decide which usage of the word represents the sentence's intended meaning.

We will consider these approaches further in the following subsections, where certain features and findings that are relevant to language processing will be discussed.

2.2.1 Incrementality and syntactic complexity in language processing

During language comprehension, language users must recognize the signals that reach the brain, which can be either visual or auditory. They must also recognize whether the word is meaningful or not, and the word's meaning in the context in which it is presented (Warren, 2013). In classical models of sentence comprehension, the assumption has been that a full interpretation of the discourse is not made until the input is complete (Hagoort, 2016). However, after the discovery of experimental methods in language studies, we have access to how and to when the brain parses written or spoken material, and how this information is integrated into the context. These days, most linguists seem to agree that sentence processing is both continuous and incremental, meaning that we process sentences word-by-word, and that these words are immediately integrated into the sentence or discourse representation (Warren, 2013; DeLong, Urbach & Kutas, 2005). This does however vary depending on available knowledge about the discourse context. Studies from Van Berkum, Hagoort and Brown (1999) as well as Van Berkum, Zwitserlood, Hagoort and Brown (2003b) investigated when the language comprehension system relates the coming word into local sentences, as well as a wider discourse. Sentences like, *Jane told the brother that he was exceptionally* ***slow*** were presented in a discourse where he had previously been described as quick, and later, the same sentences were presented in isolation (Van Berkum et al., 1999, p. 659) Their findings confirmed that except in isolated sentences, words, especially referring expressions, are immediately interpreted at the discourse level.

Sentence constituents tend to get packaged up as we read, and we have preferred methods of packaging. This preference can be visualized in garden-path sentences, which are sentences where the interpreter initially induces an interpretation which turns out to be incorrect, often due to a misleading syntactic analysis (Warren, 2013). An example of this is from an experiment testing the garden-path effect by Ferreira, Christianson & Hollingsworth from 2001:

(26) *While Anna dressed the baby spit up on the bed.* (p. 3)

This sentence has been found to be particularly difficult to understand when presented visually and without punctuation, and many participants correctly believed that the baby spit up on the bed, but also often incorrectly believe that Anna dressed the baby. Not until encountering the VP *spit up* do the interpreter realize that they have parsed the sentence incorrectly. In other experiments by Ferreira and Henderson, they demonstrated that this reanalysis is even more complicated when the head of the misanalysed phrase (*the baby*) is distant from the error signal (*spit up*):

(27) *While Anna dressed the baby that was small and cute spit up on the bed*. (p. 4)

This effect comes about because readers prefer to assign a thematic role to a phrase only upon encountering its head. When the head is placed next to the error signal, this thematic role can easily be revised, but when those elements are far apart, the sentence comprehension system has more trouble abandoning this initial commitment.

Processing of sentences is also assumed to be helped by explicit markers of syntactic structure, as they help in preventing ambiguity and increase the efficiency of which readers can construct syntactic trees. Garden-path sentences can also be used to illustrate the preference of misleading syntactic analyses. Sentences (28) and (29) below are from Warren (2013, pp. 165-166).

(28) *The horse raced past the barn fell.*

The sentence, though well-formed, has caused much confusion for interpreters. This is due to the reduction of the relative clause, which makes the sentence more difficult to process. Readers tend to interpret the verb "raced" as past tense, rather than a past participle, and when they read the verb "fell", they realize that they have made the wrong analysis.

(29) *The horse which was raced past the barn fell*.

However, by the help of explicit markers of syntactic structures, the ambiguity reduces and increases the efficiency that readers use to construct meaning.

Another suggested approach to how sentences are analyzed is the *unrestricted race model*, which proposes that when various possible sentence analyses are possible in a "race", the winner will be the analysis which is built fastest (Warren, 2013). This model, like the garden-path model, proposes that the syntactic analysis is built before plausibility of the sentence is evaluated. Thus, this model proposes that when several readings are possible, these readings will "compete", and the one that is more easily processed will be the preferred analysis (Warren, 2013).

A way of measuring syntactic complexity and the preferred analysis is asking participants comprehension questions about the input they have been presented. It is assumed that the more time it takes to respond to a sentence, the more processing energy is required from that bit of linguistic information. Adding to that, findings from e.g. garden-path sentences show that frequent, grammatical and un-ambiguous sentence patterns are more easily processed than sentences involving complex structures, and/or ambiguous interpretation, which will require greater processing load (Altmann, 1998; Warren, 2013).

2.2.2 Attention control: the "good-enough" approach to language comprehension

Building on findings from garden-path sentences, Ferreira, Bailey, and Ferraro (2002) introduced the "good-enough" approach to language comprehension as an interpretation strategy used by listeners and readers when they parse language. Often, a reader or listener will interpret input on pieces of information that are only partially analyzed, and as consequence, may misinterpret semantic or syntactic information. They may even confidently believe that the correct analysis has taken place, as suggested by the illustrious example, "How many animals of each kind did Moses take on the ark?" to which people are likely to answer "Two" without realizing that it supposedly was Noah who saved the animals on the

ark, not Moses (Hagoort, 2016, p. 345). Thus, the meaning obtained from a sentence does not necessarily reflect the true content of it. Going back to the garden-path example above, Ferreira et al. (2002) discuss the effects of the common misanalysis of the sentence "While Anna dressed the baby played in the crib." (p. 12). When answering comprehension questions about the sentences, many inaccurately believed that Anna dressed the baby, even after reanalysis. This proves that the meaning obtained from a sentence is not always a reflection of its true content.

This model suggests that rather than following detailed parsing strategies, readers will construct a good-enough analysis based on available information (Warren, 2013). The fact that language processing is at times only partial and that incorrect interpretations may continue even after reanalysis, challenges the previously established view that the language processing mechanisms generate complete and accurate representations of the linguistic input (Ferreira et al., 2002). The good-enough approach highlights that the linguistics representation is not robust, and good-enough interpretations may be the case. The language comprehension system is not perfect: and in the real world, with natural language exposure, interpreters must process this imperfect language exposure somehow. Ferreira et al. (2002) argue that these interpretations are likely to be just "good-enough", as these interpretations help the language system coordinate all the information it receives during conversation, where comprehension and production processes intervene.

2.2.3 Preactivation, predictability and cloze probability

Some words are more easily retrieved from the mental lexicon than others. There is a semantic relationship between words, and word associations ease the predictability of upcoming words in the context (Warren, 2013).

Many studies confirm that readers use the words in a sentence as cues to their world knowledge to estimate the probability of the upcoming words, like DeLong et al., (2005). Also, several studies have found that predictable words are read faster and lead to reduced neural activation, which is visible as listeners move their eyes to the items that are predictable from the context, even before the item has been named (Fruchter, Linzen, Westerlund & Marantz, 2015). As we have seen above, ERPs can be used to track prediction in sentence processing due to the high temporal resolution.

Prediction about the coming word also closely relates to the N400 effect. This was established by Kutas and Hillyard already in 1984, in an experiment that found that unexpected words that were semantically related to highly expected words elicited lower N400 amplitudes. Federmeier and Kutas (1999) confirmed this correlation by comparing ERPs of words in sentences with different degrees of prediction and plausibility to the previous context. Words that were closely related to the context elicited a smaller N400 effect.

Relating to this, a correlation has been found between the N400 effect and cloze probability. Kutas and Hillyard (1984) found that words with high cloze probability elicited a *smaller* N400 effect than words in less constraining sentences. In their experiment, all critical words were part of plausible sentences. Rommers and Federmeier (2018) also concluded that if we get a significant N400 effect, we can go back to the cloze probability test results for that specific word and compare the on-line and off-line measures. Thus, the two methods seem to go hand-in-hand. The N400 effect has been found to decrease when there is a semantic relatedness to the preceding word in a list, and even semantic relatedness to the expected, but not actually present words.

High cloze probabilities lead to a reduced N400 effect, and vice versa (Kutas & Federmeier, 1980). In relation to these findings, DeLong et al. (2005) also found support in the preactivation account in an experiment on noun prediction. An N400 effect was found for the indefinite articles a and *an* based on whether the context would predict a noun that agreed with the article.

3. Event-Related Potentials (ERPs)

In the following section, I will present the method electroencephalography (EEG), which is used to derive event-related potentials (ERPs). I will then explain what ERPs are, and how researchers can use them to learn more about how language is processed. Following that, I will present some common ERP effects associated with language studies, the N400 and the P600, where the first is mainly associated with lexical semantic processing and the latter to grammatical processing and its relations to semantics and pragmatics. I will also introduce an ERP component that has been found in association with brain responses to reference; the so-called Nref effect.

The non-invasive measure of electrical brain activity generates some of the most direct evidence in terms of studying the processes that underlie language comprehension and production. The mechanisms involved in language processing take place extremely rapidly, and many of these processes are not available for introspection; the comprehender's ability to examine his or her own mental processes (Kaiser, 2013). Language studies like these therefore require methods that allow us to study these cognitive subprocesses with high temporal resolution. One of these methods is the recording of electrical brain activity through electroencephalography (EEG). The EEG signal is a continuous measure of brain activity, which is reported in milliseconds. The signal allows us to see how long it takes after a word is encountered until it is integrated into the context in which it is perceived. The method thus allows us to see which language component impacts late or early speech perception processes (Rommers & Federmeier, 2018).

3.1 EEG

Electroencephalography (EEG) is a non-invasive method that measures the brain's electric fields. The usage of EEG in language studies allows us to study both spoken and written language as it occurs naturally, though in an unnatural setting. The first human EEG recording was conducted in the 1920s by Hans Berger, who discovered that he could record the electrical activity from a patient with head injuries. The method was initially discarded by the scientific community, but from the 1950s and later, further discoveries were made, and the technique advanced with the discovery of event-related potentials. Today, using EEG to collect ERPs has been used for countless experiments to study all aspects of cognition, from perception to memory, and language (Biasiucci, Franceschiello & Murray, 2019; Rommers & Federmeier, 2018). The method has therefore shown that it can provide deep insight to how, and specifically *when* the brain processes language, either spoken or written, as it provides excellent temporal resolution. We can thus use the method to find out how long it takes after a

word's onset until it is integrated into its context. EEG is a direct and continuous measure of neural activity (Rommers & Federmeier, 2018), and the electrophysiological responses to the stimuli will depend on the nature of the information associated to that stimuli and its context (Kutas & Hillyard, 1980; Biasiucci et al. 2019). The signal is assumed to arise from potentials that are located close to the scalp, and by placing electrodes on the participant's scalp via a close-fitting cap, the connection can be established via conductive gel. 32 electrodes are typically used in studies looking at language processing (Rommers & Federmeier, 2018).

3.2 ERPs

Event-related brain potentials (ERPs) signifies the electrical activity recorded at the scalp that changes response to some experimentally induced event (Altmann, 1998). These small fluctuations in the EEG waveform are time-locked to an event of interest, such as the onset of a word. ERP responses are created by extracting average responses from the continuous EEG data. As EEG data is quite noisy, the signal is averaged around the critical event, typically a stimulus word, and the measures collected demonstrate activity to that event (Van Berkum, 2012; Kutas & Federmeier, 2007). ERPs reveal positive and negative deflections that can, typically with additional data or mathematical models, be linked to specific neural systems and functional processes. The peaks of a stimulus-locked ERP waveform allow us to see the cognitive processes as it unfolds during a trial. The timing and amplitude show us the sensitive indices of changes in specific cognitive processes to stimulus perception. These waveform features display both time and amplitude, and are referred to as ERP components. Most ERP components are typically titled according to their polarity (P=positive, N=negative) and their typical latency in milliseconds. Many components are also recognizable by their scalp distribution (Woodman, 2010).

One of the strengths of ERPs is their multidimensional nature: polarity, distribution, shape and timing. These can give us information about the identity of the processes engaged in processing language (Van Berkum, 2012). In the literature, there is a large set of many similar and many different ERP effects, with multiple competing functional accounts for each. Another strength of ERPs is that due to the precise temporal resolution of electrophysiological recordings, it is a valuable technique for tracking theories of perception and attention (Woodman, 2010). ERP effects are also often successfully combined with a method that can track their location and activation, such as fMRI, which tracks blood movement (Van Berkum, 2012). In many studies where the ERP component is already known, the research question is often something like: "Does component X differ in amplitude between conditions?". In studies where the effects are not known beforehand, the question will rather be something like, "Does the brain appreciate the difference between these conditions, and if so, how quickly?" (Rommers & Federmeier, 2018, pp. 254-255), which the researcher must then verify according to a statistical analysis of the averages of participants.

A disadvantage of using ERPs is that the data is usually noisy, and it is nearly impossible to filter ERP data and remove noise without distorting or removing the signal of interest.

Woodman (2010) suggests some methods for eliminating noise without causing unintended distortions of the amplitude. First, to shield the electrodes from environmental noise by recording in a radio-frequency shielded room. Secondly, he suggests randomizing the pattern in which the participant sees the stimuli to avoid alpha-wave activity to become phase locked with the stimulus presentation rate. A third recommendation is keeping participants alert and engaged throughout the recording, as alpha-band noise increases when participants are drowsy and bored (p. 2037). Lastly, Woodman (2010) also emphasizes the importance of encouraging participants not to blink and move their eyes during the trials, as this distorts the signal.

In the following subsections, I will present the most common ERP effects associated with language studies: the N400 and the P600. I will also introduce an ERP component that has been found in association with brain responses to reference: the so-called Nref or SAN (sustained anterior negativity) effect.

3.2.1 The N400

One of the ERP responses most commonly associated with language processing research is the N400 effect. Since its discovery in the late 1970s, it has become a well-known and dependent measure of semantic anomaly. The measure has led to milestones in terms of investigation of neural functions and representations of language processing and comprehension (Kutas & Federmeier, 2011). The N400 is a relative negativity that appears between 250 and 500 ms after the onset in a stimuli, providing evidence that the effects of semantic violations, or low expectation can be seen almost immediately in the processing of a critical word (Kutas & Federmeier, 2011). It is considered to be part of the default electrophysiological response to potentially meaningful items, linked to the processing of a word at the level of *meaning*, and is seen as a neural marker of lexical semantic processing (Kutas & Federmeier, 2007). The N400 gives us insights on how and when meaning is activated from words, as it is a reaction to several factors such as word frequency, sentence position, presence of semantically related words, plausibility, and predictability. The N400 effect has been found to be especially large for nouns that do not fit with their preceding contexts (DeLong, Urbach & Kutas, 2005). When something unexpected appears, we encounter a semantic violation and the N400 amplitude is higher. The stronger the semantic violation is, the higher the peak of the N400 effect will be. A wide range of manipulations such as repetitions and context, will affect the semantic processing of a stimuli and modulate the amplitude of the N400. The N400 amplitude will be reduced compared to the same stimuli when it is out of context, indicating that a reduction in the N400 effect is a sign of decreased demands on semantic processing (Wlotko & Federmeier, 2012).

One of the earliest studies that investigated the N400 effect on semantic violation was Kutas and Hillyard (1980), who tested whether the effect could be an electrophysiological sign of the processing of semantically ambiguous information. They investigated ERPs in response to sentence-final words according to their predictability within its context and found

that an occasional semantic deviation of a word in a sentence is followed by a negative brainwave, the N400. Sentences like the following were tested:

(1)	He took a sip from the **tap**.	(no semantic incongruity)
(2)	He took a sip from the **waterfall**.	(moderate semantic incongruity)
(3)	He took a sip from the **transmitter**.	(strong semantic incongruity)

(Stimuli from Kutas & Hillyard, 1980, p. 203; Rugg, 1992, p. 396)

The amplitude of the N400 was larger for sentence-final words that appeared out of context. A large N400 effect was found in sentences with highly unpredictable endings, such as sentence (3), where the critical word is violates the semantic expectancy. A moderate effect was elicited for sentences such as (2), where the degree of violation is smaller than for sentence (3). Sentence (1) represents the baseline control condition with highly predictable endings, such as sentence (1), that did not elicit a significant N400 effect. In comparing these results, the authors found that the amplitude of the N400 correlates to the degree of incongruity. They could then conclude that the amplitude of the N400 component provides useful information about temporal aspects of language processing, as well as the cognitive processes that underlie in natural language comprehension. They also concluded that words that are expected in a given context are recognized and remembered more quickly and accurately than when words are presented in a semantically inappropriate context (Kutas & Hillyard, 1980).

3.3 Language beyond the N400: Establishing reference in language processing

ERP signals can selectively keep track of particular aspects of real-time language comprehension. As seen above, the N400 effect is elicited when readers and listeners come across a semantic integration problem. Two additional ERP components are interesting to consider in terms of establishing referential processing, namely the P600 effect and the Nref/SAN-effect. These will be presented below.

3.3.1 The P600

The P600 is a positive deflection that starts at about 500 ms after word onset, and peaks at 600 ms (Altmann, 1998). The component is linked to problems with syntactic integration and ungrammaticality, and has been found in sentences such as: *The spoilt child **throw** the toy on the floor* (Van Berkum et al., 2003a, p. 236). Although more commonly associated with syntax-related processing, the effect has also been observed in studies testing referential

processing. In an experiment testing pronouns in which no referents were provided for the pronoun, Van Berkum et al. (2006) found a referentially induced P600 effect for sentences like, *Anna shot at Linda as **he** jumped over the fence* (p. 162). The authors concluded that the effect was elicited as comprehenders looked for a suitable referent in the immediate context, which they did not find, as the context only included female referents. Therefore, the sentences were both referentially and syntactically problematic, and readers saw it as a syntactic issue rather than a semantic issue (the pronoun *he* can typically not refer to a woman, and that could have led to a semantic N400 effect). The gender-marked pronoun failed to refer because the preferred referent was incorrect, and the result was a P600 effect (Van Berkum et al., 2006).

3.3.2 The Nref/SAN

Another ERP component that has been associated with referential ambiguity is the Nref or SAN-effect. Van Berkum et al. (2003a) conducted an ERP experiment where the number of candidate referents for a singular noun phrase was manipulated. Participants were asked to read ministories with referential ambiguity such as example (4) below:

(4) David had told the girls to clean up their room before lunchtime. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time. David **the girl** that had had been on the phone to hang up.

Van Berkum et al. (2003a, p. 236)

A sustained anterior negativity shift was elicited when participants encountered an anaphor in sentences that had two previously introduced antecedents. In the example above, the NP *the girl* is in a two-referent context, meaning that there were two possible referents, and the context in the story determines who the correct antecedent is. The fact that the referentially induced effect emerged at around 300 ms after noun onset suggests that language users can rapidly determine whether a singular noun or pronoun has a unique referent in the previously established discourse or not. This effect was interpreted to came about due to the referential ambiguity in the critical word, reflecting problems with establishing reference (Van Berkum et al., 2006; Van Berkum, 2012). However, Van Berkum et al. (2003a) also pointed out that the effect was very similar to the ERP effects observed for increased working memory load in language processing. Most importantly, the authors could conclude that the findings provide evidence that the neural mechanisms involved in establishing reference are qualitatively different from those that are involved in semantic and syntactic integration (Van Berkum et al., 2006).

4. The present study – *Reference in online sentence processing*

As established in the literature, there is a distinct difference between referring and non-referring NPs. A referring expression denotes a specific entity, whereas a non-referring expression has a general or abstract interpretation. This distinction is not easily interpreted in vague or ambiguous expressions (Von Heusinger, 2007), but in our stimuli, we have created non-ambiguous sentences, which will be further explained in the following section. Saeed (2016) distinguishes between the terms *constant reference* and *variable reference.* The nominals in the stimuli used in this study belong to the variable reference category, meaning that the context will decide the reference of the nominal, as they can appear in expressions with and without reference. When we consider the two sentences;

a. *Yesterday, I finished knitting a* ***sweater*****.** (referring, specific object)
b. *When fall arrives, most people wear a* ***sweater***. (non-referring, generic)

We expected that as the NPs were processed, the reader would access two sets of worlds, one in which a specific object would come to mind, such as *a sweater* in sentence a., and one in which no such object would come about, such as a sweater in sentence b., which should rather be seen as an indication of cold weather.

The choice of using indefinite NPs in the stimuli rather than definite NP was also explained above. As explained, in Norwegian, definite NPs look different than in English. Rather than having a definite article, the noun is suffixed, meaning that *the girl* becomes *jenta*. Indefinites, on the other hand, look similar in the two languages: *a girl* and *ei jente*. The present study tests Norwegian indefinites on Norwegian native speakers, but as we have seen in the section *Indefinites in Norwegian*, they seem to act very similar to English. Although Norwegian has some unique uses of indefinites, the sentences that are used in this experiment could be expected to be comparable across similarly structured languages.

The question asked in this study was whether the brain processes sentences with referring and non-referring indefinite NPs differently. The basis for our hypothesis was that participants would easily be able to distinguish between the two in the comprehensive questions followed by the stimuli. This should be visible in the following comprehension questions the participants answer; if they are able to answer them correctly, there is probably a distinction in ERP measures. However, if the participants do not answer the comprehension questions correctly, they probably processed the sentences the same way. Previous studies on reference in language comprehension studies should also be able to tell us something about what we could expect from testing the conditions of this experiment.

We have also seen that ERPs can be used to selectively keep track of major processes involved in language comprehension, and that combining that with comprehension questions about the linguistic input can tell us more about what the participant actually understands. In the present study, ERPs were used to collect online recordings of brain activity while participants read sentences. As found by Van Berkum et al. (2003a), reference in a discourse can be established very quickly. We have also seen that there is not one general ERP component that reflects reference in language comprehension. ERP effects seem to vary depending on context, ambiguity, and semantic prediction from the preceding words, like reflected upon in the studies by Van Berkum et al. (2003a; 2007). Again, the stimuli in the study is not created to test world knowledge violations, or semantic violations. Kutas and Hillyard (1980) found that words with high predictability elicited a *smaller* N400 effect than words in less constraining sentences. Therefore; we do not expect a distinct N400 effect from the referring-manipulation.

Additionally, studies on attention control have found that comprehenders are likely to settle on a good-enough model of processing, meaning that they do not reanalyze the sentence if they assumed that they got the meaning right. This could be interesting for this study, if our results find that participants are not sensitive to the difference between the two conditions, and settle on a good-enough interpretation in which they assume that they have analyzed the meaning of the sentence correctly.

Our study combines online sentence processing with a behavioral data collection through comprehension questions to clarify our findings. These comprehension questions will tell us something about the processes that occur later in sentence processing, whereas the ERPs will tell us something about the earlier processes of sentence processing. Kaiser (2013) argues that using a combination of behavioral and online measures are beneficial in successful experiments, as the two measures can provide insight that is not available from either method on its own, as behavioral approaches can provide crucial information about the final outcome of language processing as well. ERPs can tell us when certain features are processed, but the participant responses to the same sentence may tell us what the participant comprehends from what they have read. The comprehension questions will tell us whether the participants are able to distinguish between the two conditions, and if they do not, that should explain why the two conditions may be processed similarly.

Lastly, as our study deals with processing of referential expressions, we have seen from previous research on referential processing that the brain handles referentially complex situations in different ways (Van Berkum et al. 2003a; 2006). These studies have also proved that we can use ERPs to keep track of these referential situations. Other studies, such as Van Berkum 1999 and 2003b also found that the way that words are processed depend on the context in which they are presented. Words are immediately integrated into the context when comprehenders have some background knowledge about the situation in which the sentence is presented.

5. Methods

As discussed above, linking linguistic theories to experiments on language comprehension processing has led to interesting findings. For the present study, the research question is:

Does the brain distinguish between referring and non-referring indefinite noun phrases?

Like we have seen above, a combination of behavioral and online measures is beneficial in studies such as the present. Reference is an interesting topic both theoretically and in terms of sentence comprehension. It can help us see how ambiguity is processed, and to see how world knowledge and semantic knowledge is comprehended, as well as how memory and language are linked together.

The following section will present the data collection procedures for the present study, consisting of three parts: pre-experimental measures, as well as an online (EEG) and an offline (behavioral) data collection. The section will also present the stimuli that have been used in the experiment. As mentioned in the introduction, the current experiment was constructed by previous master student Haug-Olstad (2019), and as some of the previously collected data will be used in this study, the same procedures and stimuli will be used.

Before collecting any EEG data, several norming experiments were conducted. Norming tests are done to ensure that the sentences that were used in the experiment work well. These tests also allow us to make sure the observed EEG effects are driven by the controlled variables. These will be explained in subsections 5.2.5, 5.2.6, and 5.2.7.

5.1 Data collection

5.1.1 Participants

For the EEG study, 28 participants (seven men; mean age: 25.82; age range: 20-42) were recruited for the first experiment (Haug Olstad, 2019), followed by 12 participants (six men; mean age: 24.16; age range: 20-28) in the present experiment. Recordings from both samples were used in the data collection. Written informed consent was obtained from all participants, and experimental protocols were approved by the NSD (the Norwegian Centre for Research Data). All participants were right-handed and had normal or corrected-to-normal vision. They were also native speakers of Norwegian using *bokmål* as their written language, and were raised in monolingual homes. None reported problems with hearing or reading nor any neurological/psychological disorders, and none of the participants took part in the previous norming. Two of the participants were excluded from the final analyses due to too many artifacts in the EEG recordings. Participants were compensated with a gift card for their participation.

5.2 Materials

In the following section, I will present the stimuli and the comprehension questions used in the EEG experiment. Before the recording took place, three sets of norming tests were conducted to ensure that the conditions were balanced in cloze probabilities, naturalness and syntactic complexity. All norming sets were run using Wilcoxon rank sum test with continuity correction.

5.2.1 The stimuli

The stimuli used in the study were embedded in a previous experiment studying compositionality, by Haug Olstad (2019). In that study, the conditions that were tested in the present study functioned as filler sentences. In the current study, the stimuli consisted of 100 sentences consisting of 50 pairs. All sentences had a sentence-final indefinite noun phrase, which was the critical word. Each indefinite noun phrase was tested as a critical word twice, in both the referring and a non-referring condition (the difference between the conditions has been described above). The stimuli was in Norwegian *bokmål* (the dominant written language form in Norway) so that we could recruit native speakers for the experiment.

Six lists were created, so that the order of the sentences was randomized, and participants would not be presented the stimuli in the same order. All participants saw the whole stimuli and both versions of an NP (both referring and non-referring). Before the experiment started, participants were given practice trials, where each condition was presented two times. The practice sentences were the same across all participants.

Condition	*Sentence*	*Article*	*Critical word*
Non-referring	Når høsten kommer tar de fleste på seg	en	genser
	When fall arrives most (people) put on	*a*	*sweater*
Referring	I går ble jeg ferdig å strikke	en	genser
	Yesterday I finished knitting	*a*	*sweater*

Table 5.1: Example of stimuli in both relevant conditions. Each critical word was introduced by an indefinite article, and all critical words appeared sentence-finally. All critical words appeared twice, once in each condition. English word-by-word translations in italics.

5.2.2 Critical words

In the stimuli, each critical word appeared twice, once in both relevant conditions. The critical words were always sentence-final, and were always relevant in the comprehension question that followed the trial. The participants did not know when the critical word would appear as the stimuli included no punctuation, and as critical words appeared in sentence-medially in the filler conditions. All critical words were tested using cloze probability to make sure that the nouns were on average equally (un)predictable in the two conditions.

5.2.3 Fillers

In addition to the 100 referring/non-referring trials, 200 trials that were used to answer another research question for a previous study were included in the stimuli. For the present study, these 200 trials testing four conditions functioned as fillers. All 300 sentences were written out as part of an ERP study, so they were all similar in length, and consisted of six conditions in total. Just like in the referring and non-referring condition, the critical words were used twice; once in each condition. The sentences were used in a study testing compositionality, named *cut* and *compose* conditions. As they are not relevant to the current study, I will not discuss them in detail, but examples will be presented below. Table 5.2 shows examples from the sentences that were used as filler sentences.

Condition	*Sentence*	*Adj/V*	*Critical word*	
Adj Cut	[Havene i verden er	store]	[båter	kan være blå]
	[The oceans in the world are	*large]*	*[boats*	*can be blue]*
Adj Comp	På verdens hav seiler	[store	båter]	langt
	On the world's oceans sail	*[large*	*boats]*	*far*
Verb Cut	[Trommer er det han oftest	spiller]	[gitaren	står i hjørnet]
	[Drums are what he most often	*plays]*	*[the guitar*	*is in the corner]*
Verb Comp	Han har trommer men han	[spiller	gitaren]	sin hver dag
	He has drums but he	*[plays*	*the guitar]*	*every day*

Table 5.2: Example of stimuli functioning as filler sentences in the current study. The critical word is presented sentence-medially rather than sentence-finally. English word-by-word translations in italics. Examples from Haug Olstad (2019).

Although they were designed for another study, they functioned as filler sentences for this study. Filler sentences are necessary in experimental studies to not make the participants aware of a pattern in the sentences that they are reading. They are often used in experiments

to disguise regularities of the relevant condition so that participants do not use irrelevant cues in their responses (Warren, 2013). However, the amount of filler sentences could be unnecessarily tiresome as they made the experiment much longer. As suggested by Mackey and Gass (2015), it is beneficial to limit the number of sentences to eliminate fatigue and avoid compromising the reliability of the study. The advantage is, however, that the vast number of filler sentences made it less likely for participants to see a pattern in the condition that we were testing. Another advantage is that a vast number of trials is advantageous when extracting ERP averages (Woodman, 2010).

5.2.4 Comprehension questions

Behavioral data was collected by asking participants to answer yes/no-comprehension questions about the sentences they had read. There are two main benefits of collecting both online and offline measures in the same experiment. The first is that participants are likely to stay alert, as they know that they will be asked about what they have just read. When performing an EEG experiment, participants are in a dark room for a longer period, and it is likely that they will get tired or drowsy, and the questions will help them stay focused. The other, more important, reason is that it allows us to compare offline and online measures of the experiment. If participants fail to answer the questions correctly, we can assume that they have not actually understood the meaning of the sentence and are thus maybe not able to distinguish between the two conditions. This would later be reflected in the ERP results, and the behavioral measures can help explain why the ERP results look the way they do.

The behavioral data were collected during the EEG experiments. While the participants read sentences, they answered comprehension questions about what they had read. The comprehension questions were different depending on sentence type, but every question was asked twice (as they were the same for the same sentence pair). The correct answer would be *yes* for one condition and *no* for the other, within the same sentence types. Table 5.3 below presents the general form of the questions, without any actual nouns, adjectives or verbs.

Condition	*Question*	*Answer*
Adjective	Er det **noun** som er **adjective**?	
	*Is it the **noun** that is **adjective**?*	
Cut		Nei/No
Comp		*Ja/Yes*
Verb	Er det **noun** som **verb**?	
	*Is it the **noun** that **verb**?*	
Cut		Nei/No
Comp		Ja/*Yes*
Ref/Non-ref	Er det snakk om en spesifikk **noun**?	
	*Is it a specific **noun**?*	
Non-referring		Nei*/No*
Referring		Ja*/Yes*

Table 5.3: Construction of comprehension questions for the different sentence types. The actual adjectives, verbs and nouns are not included in the table, but for the relevant condition, an example was given in the introduction. The correct answer is indicated in the right-most column. English word-by-word translations in italics.

5.2.5 Cloze probability norming

Cloze probability tests, or "fill in the missing word"-tests were conducted to test lexical cloze probabilities and how likely participants were to predict the correct noun. All 100 sentences were tested in each data collection, and they were split into six blocks of 15-17 sentences. Sentences were randomized within the blocks so that they were tested in different orders. All sentences were truncated before the critical word, which was always a noun, so that the last word the participants read was an indefinite article (*en*, *ei* or *et* = *a* or *an*), which should predict that the missing word is a noun. Due to the gender marking on nouns used by articles in Norwegian[5], the predictability of the words should be increased. This is matched across conditions. Participants were asked to write down the first word that came to mind as a natural continuation of the sentence. Below is an example of what participants would see:

[5] *En*: singular, masculine; *ei*: singular, feminine; *et*: singular, neutral

93 for å forsterke lyd brukes ofte en _____

94 i går satt jeg i klasserommet og hadde en _____

15 Norwegian native speakers took part in the study and each test took 10-15 minutes to complete. All participants completed the whole stimuli set consisting of 100 sentences. The sentences were printed out on paper and the participants wrote down the missing word by hand. Written responses were coded based on whether the participant guessed the exact word (including words that had typos or the same exact meaning); 1 = match, 0 = no match.

Cloze probabilities in the referring condition were close to 0 (mean 0.37), compared to the non-referring condition (mean 0.43), which indicated a low predictability of the target noun. In other words, the test showed that there were no significant differences in cloze probabilities across the two conditions (W=1389.6; *p*=0.3329). This data suggests that the relative predictability of the target word was similar across conditions.

5.2.6 Naturalness norming

To verify the naturalness of the sentences, we also conducted a naturalness norming study, asking participants to rate the events taking place in the sentence. As we did not want the participants to rate the grammaticality of the sentences, the instructions were very specific, and they were asked to answer quickly and intuitively, so that they would not overthink the task. If the events of the sentences sounded unnatural or unusual, the wrong ERP component could be elicited. If the participants found the sentences to be unnatural or to have words that did not really fit into the context, we might get the N400 effect, which was not the goal of this study as the sentences were not constructed to be ambiguous.

In the study, participants were asked to judge how natural sounding the events taking place in the sentences were. The participants rated each sentence on a scale of 1 to 7, where 1 was *very unnatural* (*veldig unaturlig*) and 7 was *very natural* (*veldig naturlig*). The test was performed by pen and paper, and they were asked to circle the number they thought suited the sentence's naturalness. All participants rated all 100 sentences. Below is an example from the test:

99	*Av faren sin arvet hun en fiolin.*	**1 2 3 4 5 6 7** **Veldig unaturlig Veldig naturlig**
100	*Da hun vant løpet i går red hun på en hest.*	**1 2 3 4 5 6 7** **Veldig unaturlig Veldig naturlig**

Just like the cloze probability test, the sentences were tested by 15 native speakers of Norwegian, and each test took around 10 minutes to complete. The sentences were randomized so that the participants did not read the sentences in the exact same order. Written responses were coded based on the participant's rating of the sentence on numbers between 1 and 7.

The results showed no significant difference in naturalness ratings for the sentences in two conditions (W=275732; p=0.4789).

5.2.7 Syntactic complexity norming

Finally, to ensure that neither condition was more syntactically complex than the other, syntactic complexity norming was carried out as well. Certain structures are more difficult to process than others (Warren, 2013). If one condition proved to have more complex sentences than the other, that may be the reason why they were processed differently. All 100 sentences in the two relevant conditions were parsed using the online INESS syntactic parser based on an LFG grammar[6], provided by the University of Bergen.

In this test, the syntactic complexity of the stimuli was calculated based on the number of nodes as well as the number of words in each sentence. When added together, we got a score between 9-23[7] which allowed us to consider how complex each sentence would be when processed. All sentences are in a group of pairs (1-50), and sentences with the same critical word could thus be compared with each other, as each critical word appeared once in each condition. Below is an example of a parsed sentence pair:

[6] http://clarino.uib.no/iness/xle-web

[7] See Appendix B for a list of the whole stimuli

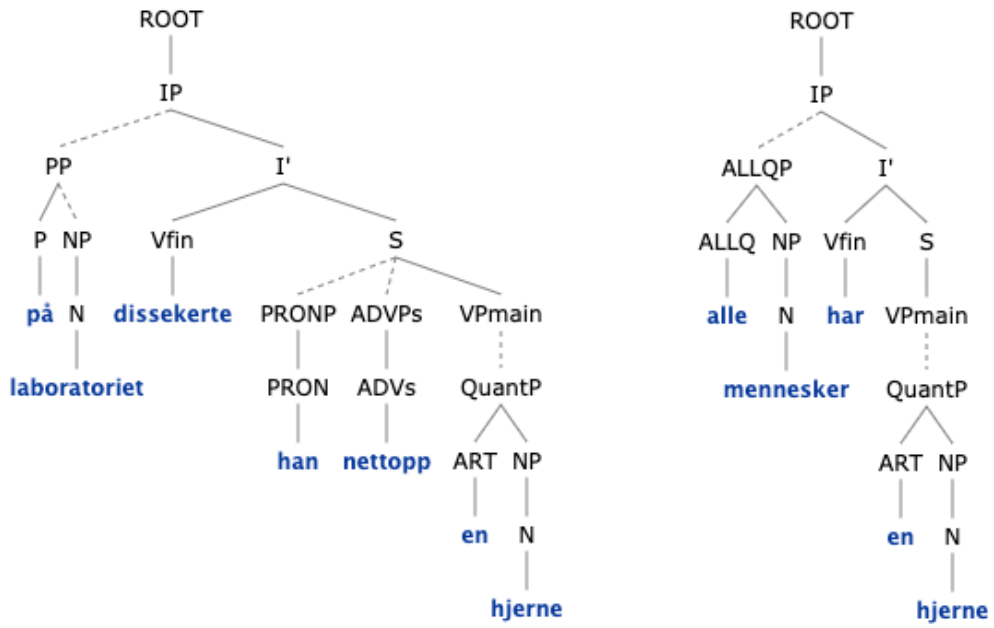


Fig. 5.1: These figures represent the syntactic complexity of the referring sentence *på laboratoriet dissekerte han nettopp en hjerne* and the non-referring sentence *alle mennesker har en hjerne*. The first one is rated to a complexity of 9, whereas the second is rated 13. In this case, the referring sentence has a more complex structure than the non-referring one, but we found that most sentence pairs did not have a remarkable difference in complexity.

Syntactic complexity in the referring condition came to an average rating of 16 (mean 16.04) and for the non-referring, the rating was close to 15 (mean 14.72). Referring sentences ranged from a rating of 10-23, whereas the non-referring sentences ranged from 9-22. Nevertheless, the Wilcoxon test showed that the sentences were not significantly more complex in one condition (W=979.5; p=0.06006).

5.3 Procedures

As some of the online data was already collected as part of a previous master's study (Haug-Olstad, 2019), the same protocols were followed in the current study to ensure accuracy in the data. The experiment was conducted in the Language Acquisition and Language Processing lab (LALP) at the Norwegian University of Science and Technology (NTNU).

After electrode application, participants were seated in a chair in a shielded, dimly lit booth, at about 90 cm distance from the computer screen. Before the experiment, participants were not given any indication of the goal of the experiment, but they were given a brief written instruction telling them to read the sentences in the experiment for comprehension and to answer questions about what they had read. The participants were instructed to answer by

using the keys F and J, where the two keys would be “yes” or “no”[8]. The response keys were counterbalanced across participants. The stimuli was presented word-by-word in a white 30-point Arial font on a dark screen using the program Presentation® by Neurobehavioral Systems.

A short practice was conducted before the recordings to ensure that the participant understood the task, and to ensure that the electrodes were well placed. Before each trial, a fixation cross appeared in the center of the screen for 500 ms to inform the participant that a new trial would appear. Participants were encouraged to avoid blinking and moving while the sentences were presented, and were encouraged to do so only after seeing the fixation. Each word was visible for 400 ms, with a 400 ms inter-word interval, and no punctuation was presented. The word before the critical noun was always a determiner (Norwegian: *en/ei/et*, English: *a/an*). After being presented each trial, the participants had 4000 ms to answer the following comprehension question, before the new trial would begin. Each round consisted of 25 trials between which the participant could take a break for however long they wanted to, and they pressed the ENTER-key when they wanted to proceed with the experiment. The breaks were also an opportunity for the experimenter to adjust electrodes when needed. Each recording would last between 1:15-2 hours, depending on how long the breaks the participants took were and how often it was necessary to adjust the electrodes.

5.4 EEG recording

ERPs were recorded continuously from the scalp from 31 active electrodes, mounted on an Easycap 32 Channel Standard EEG Recording Cap, using a 10-20 configuration. The figure below shows the layout of the electrodes. The reference electrode and the TP9 electrode (17 in the figure) were placed on the left and right mastoid respectively, for reference.

[8] See Appendix C for the instructions

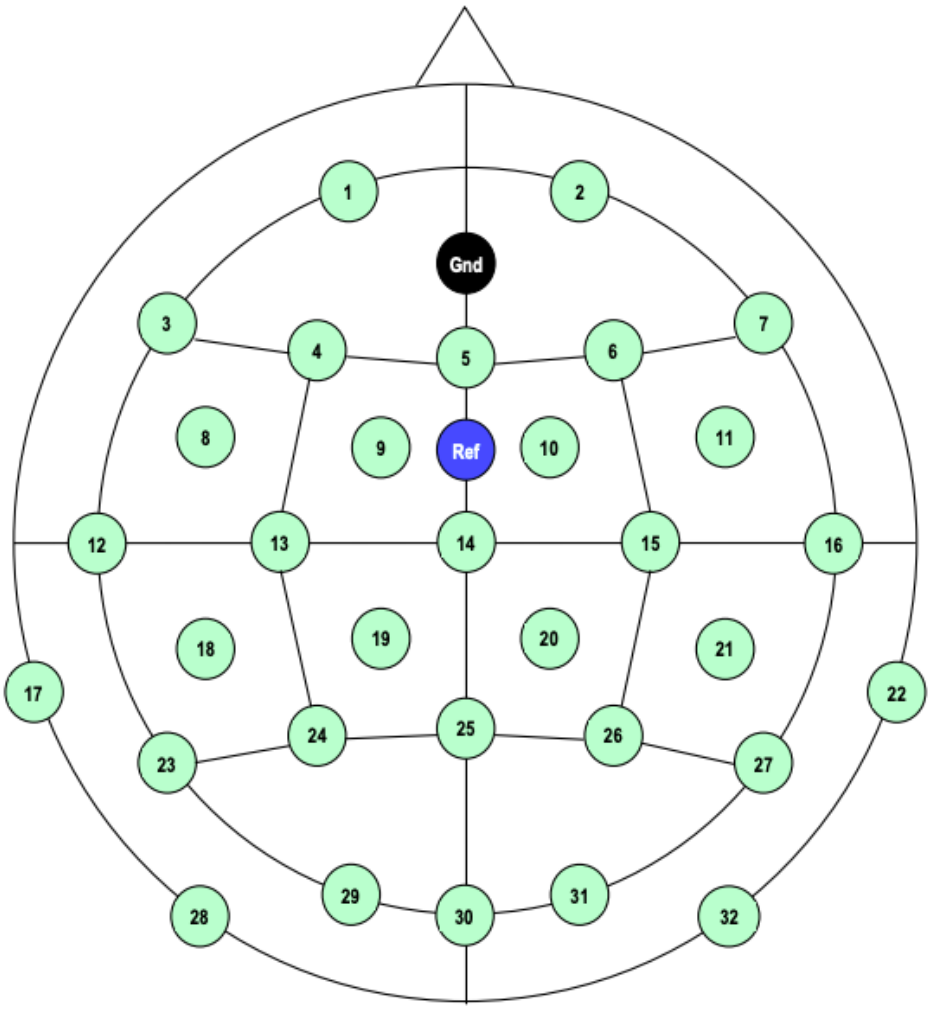


Figure 5.2: Model of electrode layout in the Easycap 32 Channel Standard EEG Recording Cap. Retrieved 13.05.20 from: https://www.brainproducts.com/files/public/downloads/actiCAP-32ch-Standard-2.pdf

The reference electrode should ideally be placed in a neutral location of the body, but since there is no location which is entirely neutral, it can be placed on the left mastoid, the thick bone structure between the electrode and the brain (Rommers & Federmeier, 2018). The EEG-data was recorded using BrainVision Recorder Professional V.1.20.0701. Electrode impedance were set to 0-10 $k\Omega$ for the ground node and reference node, and 0-5 $k\Omega$ for the other electrodes. The signal was amplified with a sampling rate of 1000 Hz, and re-referenced offline to the average of right and left mastoids. For each trial, EEG segments time locked to the critical word were extracted for each condition, set to a time segment of -200 to 800 ms in relation to onset of the target-word. Artifact rejection was done by detecting and rejecting trials with an amplitude exceeding +/- 150 microvolts from the baseline in any channels. Data from the Fp1 and Fp2 electrodes was also used to remove trials including eye movements, and trials with muscle movement were rejected by filtering the segment through a low-pass filter at 30 Hz. Signals were then averaged from the period 200 ms prior to the stimulus onset as a baseline correction.

6. Results

In this section, the results from the experimental study will be presented. First, I will present the behavioral results from the comprehension questions collected during the EEG recording. Afterwards, the ERP results will be presented. These results, as well as a general discussion, will take place in section 7.

6.1 Behavioral results

As mentioned above, the behavioral data was collected during the EEG experiments. On each trial, after reading the sentence, participants answered a comprehension question about the content in the sentence. Results from the behavioral data collection are presented in the figures below. In Figure 6.1 we see that the general accuracy is higher in the referring condition, whereas in the non-referring conditions, responses are more spread. This is furtherly underlined in Figure 6.2. Figure 6.3 shows the response time (RTs) across conditions, suggesting that participants responded faster overall in the referring condition. This figure also shows that some participants answered all questions wrong in one condition, and RTs are therefore missing for those participants. RTs that were below 200 ms were also excluded. In Figure 6.4 we can observe that when participants answered correctly, the RTs were faster in the non-referring conditions. Overall, participants responded faster when they answered correctly. Descriptive statistics describing the datasets are summarized in the table below:

Condition	*Accuracy (%)*			*RTs (ms)*		
	N		SD	N		SD
Referring	38	0.904	0.175	37	1563.850	423.096
Non-referring	38	0.667	0.380	32	1581.806	545.397

Table 6.1: Descriptive statistics summarized. Accuracy was on average higher in the referring condition. Some participants are missing in the RTs due to answering all questions incorrectly.

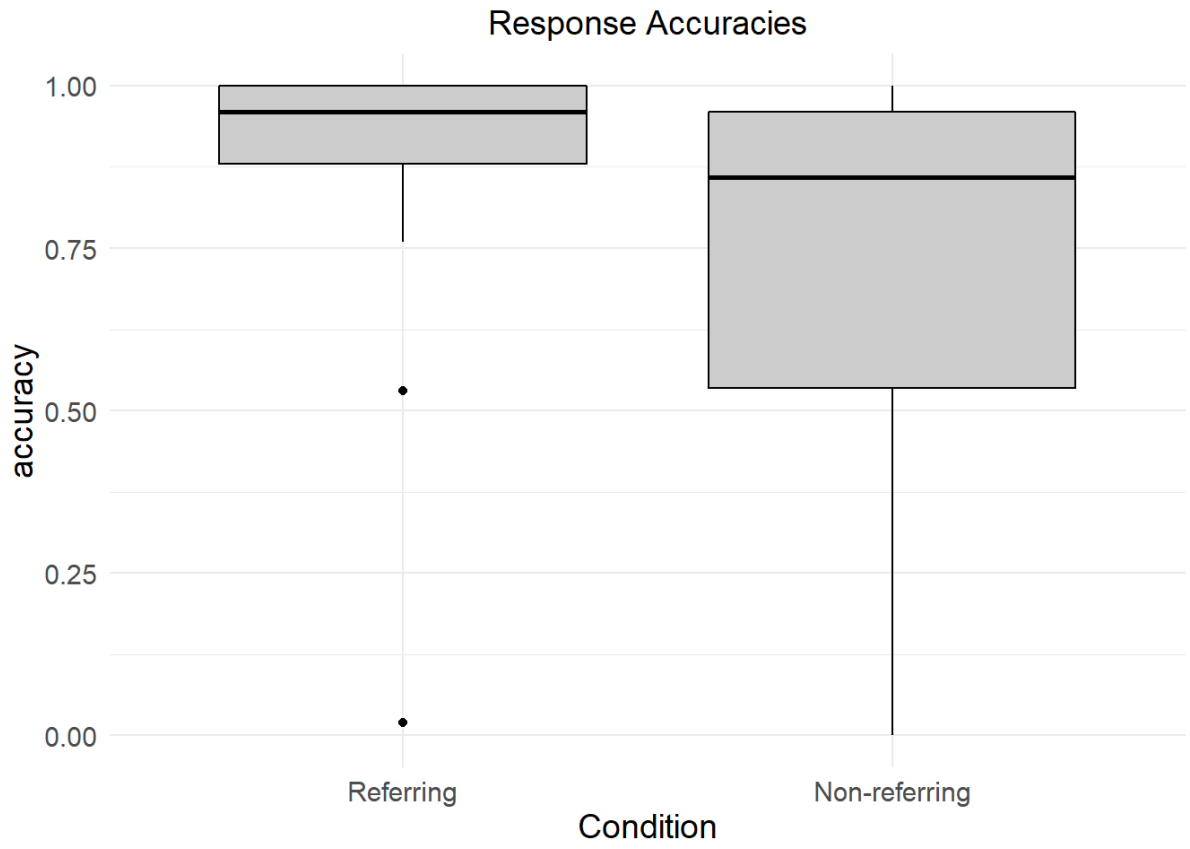


Figure 6.1: Distribution of accurate answers in the referring condition and the non-referring conditions. (N=38)

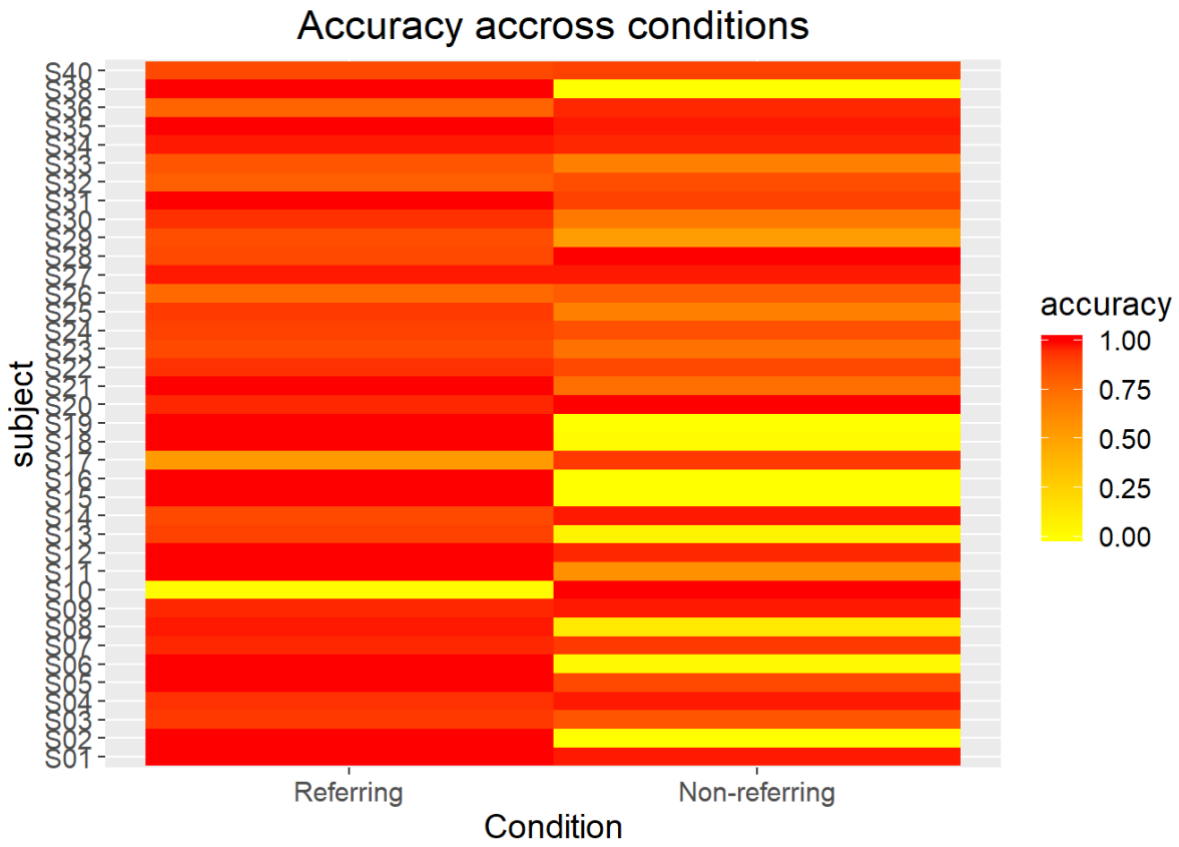


Figure 6.2: Accuracy of responses in the referring and non-referring condition. The darker color signifies higher accuracy levels. (N=38)

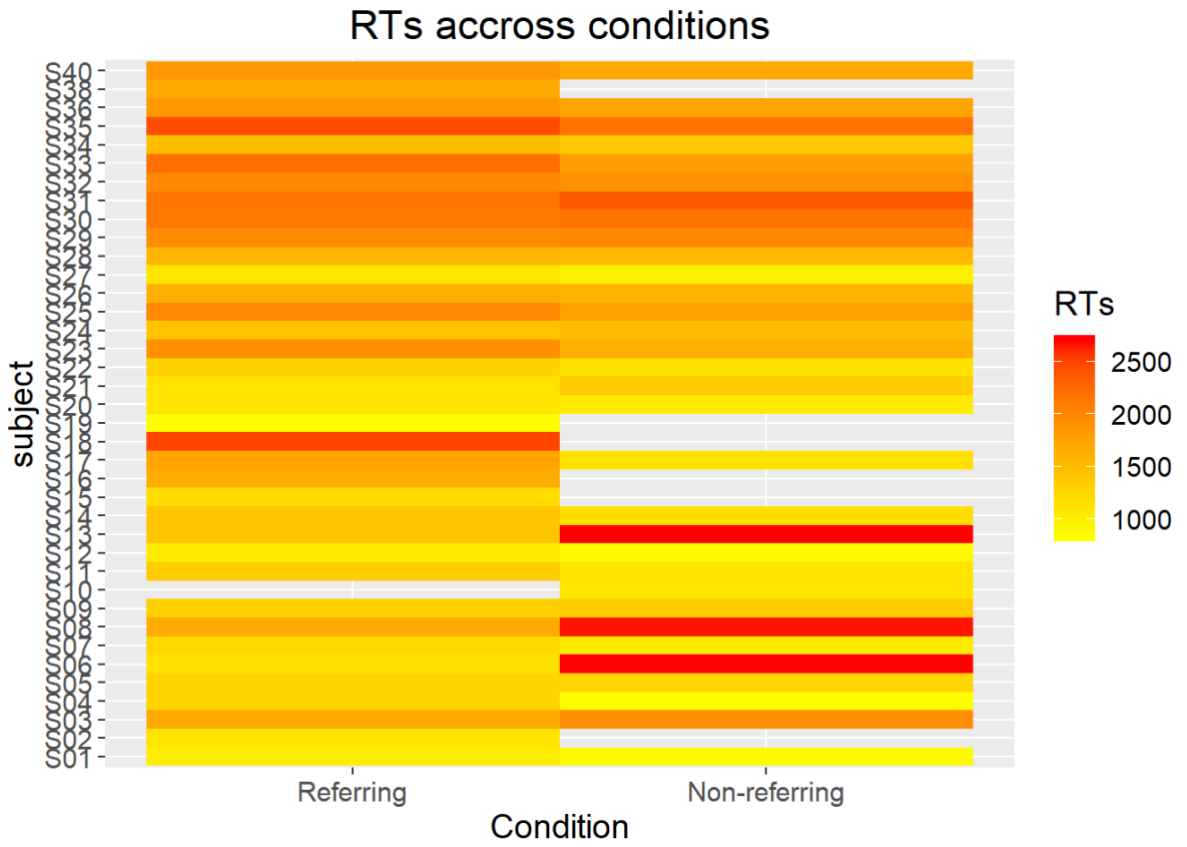


Figure 6.3: Overall response times across conditions in ms (1000-2500). The grey areas indicate that the participant answered all questions incorrectly in this condition, meaning that there was no data to plot.

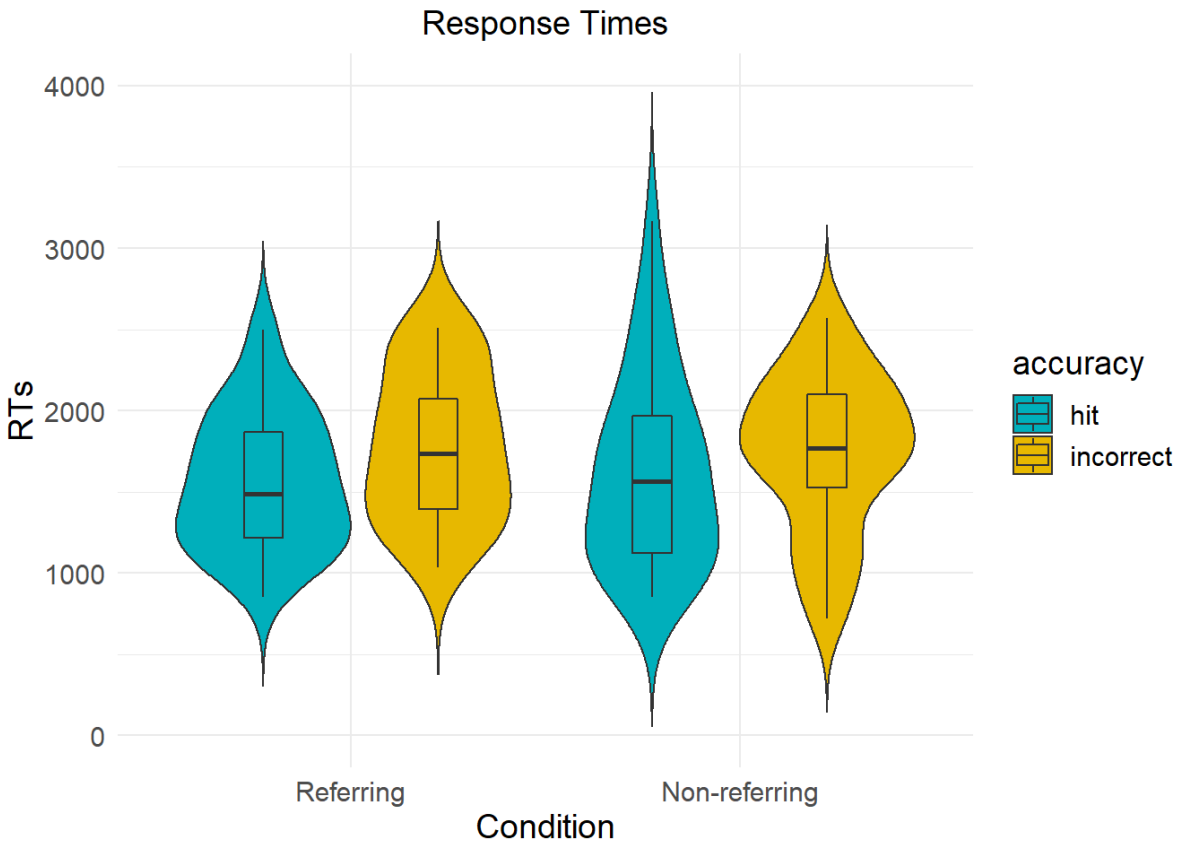
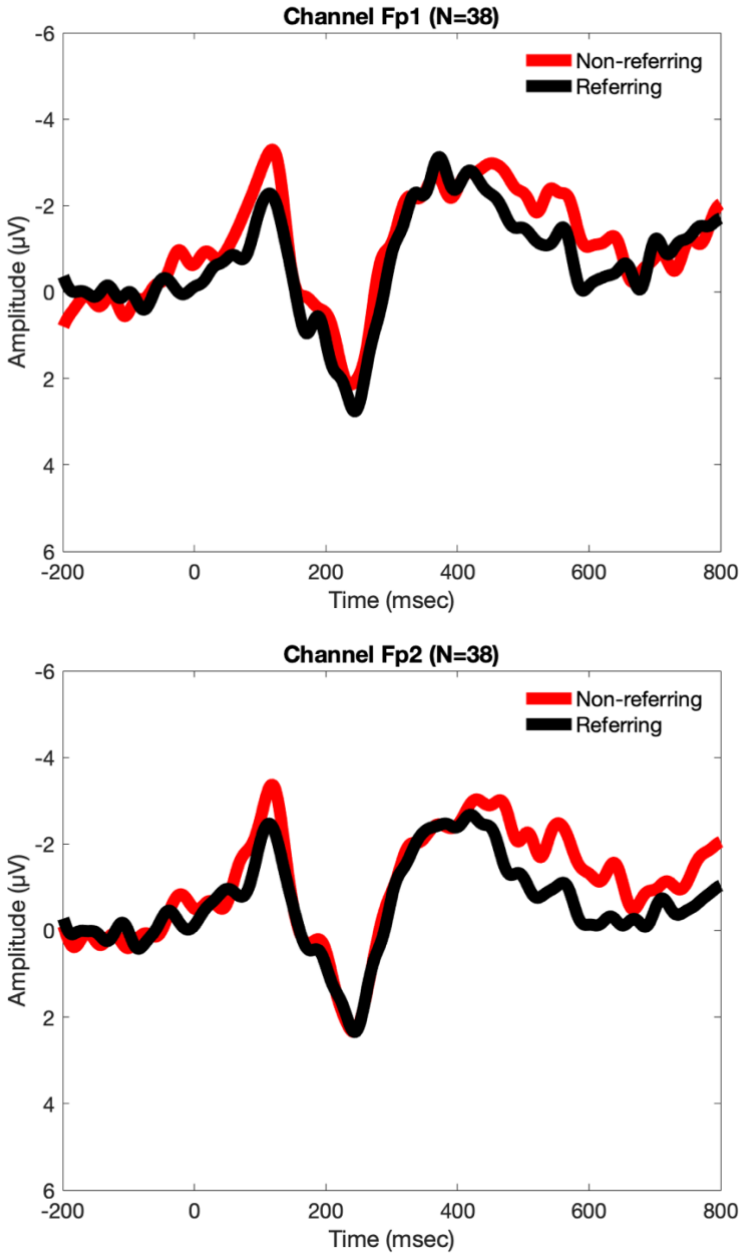


Figure 6.4: Response times according to accuracy. Blue bars are accurate responses whereas the yellow is inaccurate. The thickest area of the bars are where most of the responses are centered.

6.2 ERP results

Figure 6.5 below displays the grand average ERPs for electrodes FP1 and FP2, time-locked to the visual word representation of the critical noun, for all sentences. The averages are plotted from 200 ms prior to 800 ms after the sentence-final word onset. An example is given below the figure, with the critical word in bold, and English word-by-word translations in italics. The results show no group effect, and no significant difference across conditions. The waveforms show an early ERP component – not effect – to visual word representations, similar in both timing and polarity. The figure also shows that the N400 effect was not affected by the experimental manipulation. We do however see frontal changes to ERP amplitudes in a post-N400 time frame, which are not statistically significant, but yet they are observable.



Non-referring: Mange elsker lyden av en **fiolin.**
*Many love the sound of a **violin**.*

Referring: Av faren sin arvet hun en **fiolin**.
*From her father she inherited a **violin**.*

Figure 6.5: Grand average ERP waveforms (N=38) for electrode locations FP1 and FP2. ERPs were time-locked noun onset (0 ms), to the presentation of the critical, sentence-final word for both conditions. The non-referring condition is shown in red, and referring in black. The figure does not show a significantly different response for the two conditions. Negative polarity is plotted upwards.

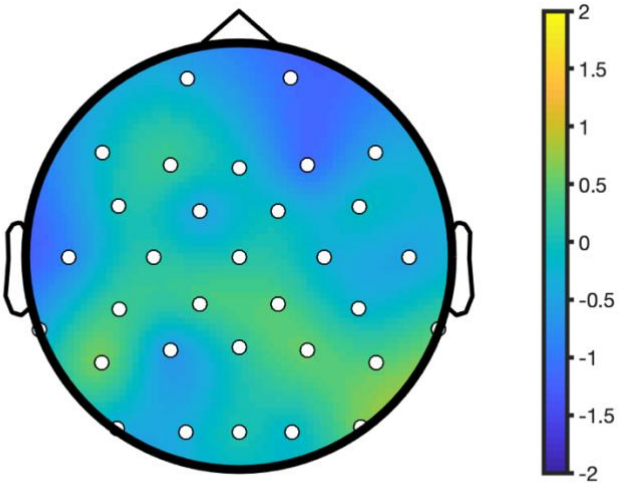


Figure 6.6: Spline isovoltage maps displaying the topographic distributions at time intervals from 500-700 ms.

7. Discussion

The current study investigated the online processing of expressions in two conditions; a referring condition and a non-referring condition. The results from our experiment were interesting, as they indicate that participants were not particularly sensitive to the distinction between the two conditions. As the data collected consists of both online measures, representing the initial stages of processing, and offline measures, representing the later stages of processing, it can provide us some interesting insight to how these sentences are interpreted. The participant group is also large, consisting of 38 participants after artifact rejection, meaning that we can interpret these results as representative rather than indicative.

With the previously discussed theoretical background in mind, what follows is a discussion and interpretation of the results we obtained from our referring-study. Due to the null-results in the ERPs, a brief overview will be given, before an interpretation of the behavioral results is presented. This will be followed by an overall discussion of what we can interpret of the results.

7.1 ERP results for the referring/non-referring conditions

The ERP results that were obtained for the stimuli containing expressions with referring/non-referring expressions indicate that these conditions are not processed online as we predicted. Unexpectedly, the conditions did not elicit any significant ERP effects time-locked to the target noun, and despite the different meaning underlying the words, the ERPs do not show an overall group effect among the participants. This suggests that the same brain mechanisms are active in the interpretation of both conditions. This consistency suggests that the differences in context are not enough information for participants to analyze the sentences differently.

Both our conditions included indefinite NPs, where the difference between these was that half of these consisted of referring expressions and the other half consisted of non-referring expressions. When an NP is referring, a specific entity is identified by the speaker, and when an NP is non-referring, it has a generic interpretation, like exemplified by Saeed (2016, p. 25) above.

*They performed **a cholecystectomy** this morning.*	(referring)
***A cholecystectomy** is a serious procedure.*	(non-referring)

We expected that as the NPs were processed, the reader would access two sets of worlds, one in which a specific object would come to mind, and one in which no such object would come about. The fact that no observable difference was elicited between the two conditions indicates that the expressions are processed similarly.

Via norming tests, we tested that the stimuli were balanced across conditions. Also, as mentioned, the critical words were not semantically incongruent and therefore not expected to cause processing ambiguity. This is visible in that there is no observable difference in the N400 between the two conditions. As the critical word appeared sentence-final in a fitting context, participants were prepared for the onset word. The fact that there is no significant difference in the N400 amplitude across conditions helps confirm that the processing system has at some level determined that the expressions do not need specific parsing strategies, at least when not placed in context. Also, as can be seen in figure 6.5, a frontal change in the ERP amplitude is observable in a post-N400 time frame. Although they are not statistically significant, they are observable, and similar for both conditions.

7.2 Behavioral results

As stated earlier, the behavioral data was collected during the EEG recording in the form of comprehension questions that participants answered after reading sentences. Comprehension questions were asked to verify whether the participants understood the difference between the conditions, which we in the hypothesis stated that we expected. As explained above, the comprehension questions were always the same for the two conditions; "*Er det snakk om et spesifikt* ***noun***?" (*Is is a specific* ***noun****?)*[9]*,* and the answer would always be *yes* for the referring condition, and *no* for the non-referring condition. Collecting these questions allowed us to compare their performance level across conditions. Overall, the accuracy in the referring condition is high (around 90% on average, Table 6.1), and in the referring-condition, the responses are more spread, and accuracy is lower (66%, Table 6.1). It is a large dataset, which gives us enough support to state that, overall, participants were not particularly sensitive to the difference between the two conditions. In addition to this, response times (RTs) were faster when participants got the question right. The fact that RTs were linked to accuracy may suggest that the participants felt more confident in some of the trials. Figure 6.4 highlights the vast spread of accuracy and RTs in the non-referring condition, whereas the referring-condition seems more stable.

At first glance, it may look like participants overall understood the referring condition better, as the accuracy is higher. However, if they only answered correctly in one condition, this still suggests that they did not really understand the difference, and may have answered "yes" to all questions, regardless of whether the object in question was specific or not.

Another visible factor from the behavioral results is that in the non-referring condition, many participants did not get a single question correct.

[9] Emphasis added to highlight the critical word, which would be different for all sentence pairs. The critical word was what we recorded ERPs from.

7.3 General discussion

In this study, we have tried to find out how sentences with referring and non-referring indefinite NPs are processed, and which processing strategies are used in distinguishing between them.

The results in this study are interesting, although not straightforward to interpret. Like we have seen, previous studies looking at reference in language comprehension have not established that there is one specific ERP component linked to reference in the brain. The ERP results from this study suggest that the online processing of the referring and non-referring sentences requires the same neural mechanisms, and that the processing of the sentence-final critical word is not immediately context-dependent. Not until after the initial processing, when considering the specificity of the word, does it seem like readers consider the word's overall meaning.

The results from section 6 suggests that readers are not particularly aware of a distinction between the referring and non-referring expressions. As mentioned in the previous sections, R-expressions have a similar syntactic structure. Neither the referring nor non-referring expression are bound to the previously established context, and they cannot have anaphoric reference. We also know that we use syntactic evidence available when we process the meaning of sentences. In terms of our stimuli, the two conditions were found to be more or less equally syntactically complex, which the similar processing seen in the ERP responses can confirm. Therefore, our findings suggest that referring or non-referring is not a default distinction we make during online processing. When comparing this to the behavioral data, we see that even during offline stages of sentence processing, almost half of the participants did not perceive this distinction.

In terms of our stimuli, norming tests ensured that the stimuli were balanced across conditions. This was the case for cloze probabilities, naturalness, and syntactic complexity. The sentences were not ambiguous, nor did they contain world knowledge violations.

The data in this thesis is not clear enough to suggest a distinct interpretation of the results, only suggestions. Our results may suggest that readers prefer to interpret the NPs as specific rather than non-specific, as the accuracy was higher in the referring condition, and that they tended to answer *yes* to whether the entity was specific, even when it was not. The reason behind this may be that specific contexts are somewhat slightly easier to process. This can also be indicated by the shorter RTs in the referring-condition. Previously in the thesis, the good-enough approach to language processing, as established by Ferreira et al. (2002) was also introduced. As the behavioral data suggests that participants were not overall particularly sensitive to the distinction between referring and non-referring sentences, this may indicate that participants believed that they had made the correct analysis of the sentence type, and that no reanalysis was necessary.

For our referring-manipulation, there appears to be a single operation that applies during online sentence processing. The fact that the brain responses were not significantly

different may indicate that the neural mechanisms engaged in processing these referring and non-referring expressions are the same, independent of the preceding information in the sentence. Only after the initial analysis has taken place did participants seem to notice the difference between the expressions according to referring or non-referring context.

8. Conclusion and suggestions for further research

In this study, we investigated how sentences with referring and non-referring expressions were processed during online sentence processing. To test this, we used a stimuli consisting of sentences constructed in two conditions: one containing referring indefinite NPs and the other containing non-referring indefinite NPs. These expressions always appeared sentence-finally, and in non-ambiguous contexts. Using EEG, we collected ERPs while participants read the sentences in the experiment, during which they also answered comprehension questions about the content in the sentence.

Our hypothesis for the study was twofold; firstly, that we expected that the participants would easily comprehend that the conditions were different from each other, so that they would answer the comprehension questions correctly most of the time. Secondly, as these two conditions are different from each other, and the distinction has been highlighted in the literature, we assumed that we would see a difference in ERPs when these two conditions were compared to each other.

The results we obtained were interesting, as they highlight that our brain is in fact not particularly sensitive to specific versus non-specific objects when these expressions are not presented in a clear context. The ERP results imply that the online processing of the two conditions involve the same neural mechanisms, as no significant differences were found in the averages between them. The behavioral data, on the other hand, imply that participants may be more sensitive to the differences between the conditions during offline sentence processing. In general, participants tended to answer more correctly in the referring-condition. These results may imply that this distinction is not overtly available to participants before they were asked to reconsider the meaning of the sentence, and during this reconsideration, participants were more likely to understand the difference. The results also suggest that participants had more problems with the non-referring condition, as the accuracy was lower, and response times were longer. This may have been because the context preceding the referring NPs helped the reader in the interpretation, e.g. by presupposition. This is not possible to conclude for the present, since the sentences varied too much to allow for an overall analysis. As mentioned in the previous section, these results may imply that the processing system has at some level determined that these two expressions do not need specific parsing strategies, at least when not placed in a larger discourse context.

As we know from previous studies on language comprehension, language is processed incrementally, and features that come in the way of the initial syntactic analysis of the sentence are visible almost immediately, such as semantic, syntactic, and referential ambiguity or violations (Warren, 2013; Van Berkum et al., 2003; 2006). Another interesting model of language processing is the good-enough approach, which highlights the fact that

when we parse language, the initial analysis is often only partially correct, and we may often settle on this analysis. Although these findings have previously been connected to structurally ambiguous sentences, like seen in garden-path sentences, our findings suggest that the good-enough approach happens for non-ambiguous sentences as well. This reflects how natural language is used in daily life; with constant interruptions, breaks, turn-taking, and so on, many of these containing semantic representations that are simply good enough, and not always completely correct (Ferreira et al., 2002).

Though our participants were not "led down the garden-path" like we saw in some examples from Warren, (2013), our results still suggest that participants were not overall sensitive to the referring and non-referring distinction, even after the syntactic reanalysis had taken place. These findings fit the good-enough approach to sentence processing, as participants seemed to parse the two conditions the same way. The behavioral data showed that many participants were did not understand the difference between the two conditions, and the ERPs confirmed this as the averages from the two conditions suggest that the online processing of the was very similar.

The results that have been presented in this thesis are a suggested analysis of results that were not straightforward to interpret. An approach to further investigate the processing of the two conditions could be to make the distinction more explicit for participants before starting the experiment. In the current study, participants were given limited information about the comprehension questions following each trial. The results may have been different if participants were given more detailed instructions, provided by examples that could exemplify the difference between them, e.g.:

Referring

For breakfast, I ate a waffle.

Is it a specific waffle? **Yes**, because the waffle can be identified, and does not refer to waffles in general.

Non-referring

During football matches, you can often buy a waffle.

Is it a specific waffle? **No**, because it does not point out a specific waffle, but refers to waffles in general.

An additional approach could have been to let participants see when they answered the questions wrong and adapt their interpretations of the sentences thereafter. For the current study, participants did not know whether they had answered the sentences correctly or not, and may therefore have confidently decided on one analysis, and followed the same interpretation strategy for the rest of the stimuli.

Bibliography

Altmann, G. T. M. (1998) Ambiguity in sentence processing. *Trends in Cognitive Sciences 2*(4), 146-152.

Biasiucci, A., Franceschiello, B. & Murray, M. M. (2019). Primer – Electroencephalography. *Current Biology, 29*(3), 80-85.

DeLong, K. A., Urbach, T. P. & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Enç, M. (1991). The Semantics of Specificity. *Linguistic Inquiry, 22*(1), 1-25.

Faarlund, J. T., Lie, S., Vannebo, K. I. (1997). *Norsk referansegrammatikk.* Oslo: Universitetsforlaget.

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language, 41(*4), 469-495.

Ferreira, F., Bailey, K. D G., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science 11*(1), 11-15.

Ferreira, F., Christianson, K. & Hollingworth, A. (2001). Misinterpretations of Garden-Path Sentences: Implications for Models of Sentence Processing and Reanalysis. *Journal of Psycholinguistic Research, 30*(1), 3-20.

Fruchter, J., Linzen, T., Westerlund, M. & Marantz, A. (2015). Lexical Preactivation in Basic Linguistic Phrases. *Journal of Cognitive Neuroscience, 27*(10), 1912-1935.

G. Frege, On Sense and Nominatum. Reprinted in Martinich, A. P. (1996). *The philosophy of language. 1892*, pp. 199-211. Oxford: Oxford University Press.

Hagoort, P. (2016). MUC (Memory, Unification, Control): A Model on the Neurobiology of Language Beyond Single Word Processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 339-337) Elsevier.

Haug Olstad, A. M. (2019). Composition in sentence processing: An ERP study. Master's thesis, NTNU, Norway.

Kaiser, F. (2013). Experimental paradigms in psycholinguistics. In: Podesva, R. J. & Sharma, D. (Eds), *Research Methods in Psycholinguistics.* (pp. 135-168) Cambridge, UK: Cambridge University Press.

Karttunen, L. (1969). *Discourse referents*. Accessed 15.03.20, retrieved from: https://web.stanford.edu/~laurik/publications/archive/discref.pdf

Kutas, M. & Federmeier, K. D. (2007). Event-related brain potential (ERP) studies of sentence processing. In: G. Gaskell (Ed.) *The Oxford Handbook of Psycholinguistics.* (pp. 385- 406) Oxford, UK: Oxford University Press Inc.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621-647.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science,* 207(4427), 203-205.

Mackey, A. & Gass, S. M. (2015). *Second language research: Methodology and Design* (2nd edition). New York, NY: Routledge.

Rommers, J. & Federmeier, K. D. (2018). Electrophysiological Methods. In A.M.B de Groot & P. Hagoort (Eds) *Research Methods in Psycholinguistics and the Neurobiology of Language*. (pp. 248-265). Hoboken, New Jersey: Wiley Blackwell.

Rosén, V., De Smedt, K. Meurer, P., & Dyvik, H. (2012). An open infrastructure for advanced treebanking. In J. Hajič, K. De Smedt, M. Tadić, & António Branco (Eds.) *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, pages 22–29, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA)

Rugg, M. D. (1992) Event-related Potentials in Clinical Neuropsychology. In: J. R. Crawford, D. M. Parker & W. W. McKinlay (*Eds) A Handbook of Neuropsychological Assessment*. (pp- 393-408). West Sussex: Woodfield Publishing Services

Saeed, J. I. (2016). *Semantics* (4th edition) West Sussex, UK: Wiley Blackwell.

Van Berkum, J. J. (2012). The Electrophysiology of Discourse and Conversation. In M. J. Spivey, K. McRae & M. F. Joanisse (Eds) *The Cambridge Handbook of Psycholinguistics*. (pp. 589-612). New York, NY: Cambridge University Press.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P. & Zwitserlood, P. (2003a). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*(2), 235-248.

Van Berkum, J. J., Hagoort, P. & Brown, C. M. (1999). Semantic integration in sentences and discourse.: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

Van Berkum, J. J., Koornneef, A. W., Otten, M. & Nieuwland, M. S. (2006). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146,* 158-171.

Van Berkum, J. J., Zwitserlood, P., Hagoort, P. & Brown, C. M. (2003b). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*(3), 701-718.

Von Heusinger, K. (2002). Specificity and Definiteness in Sentence and Discourse Structure. *Journal of Semantics, 19*(3), 245-274.

Warren, P. (2013). *Introducing Psycholinguistics*. Cambridge, UK: Cambridge University Press.

Wlotko, E. W. & Federmeier, K. D. (2012) So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage, 62*(1), 353-366.

Woodman, G. F. (2010). A Brief Introduction to the Use of Event-Related Potentials (ERPs) in Studies of Perception and Attention. *Attention, Perception, & Psychophysics, 72*(8), 2031-2056.

Appendices

Appendix A: Relevance for the teaching profession

Appendix B: Stimuli

Appendix C: Instructions for the participants

Appendix A

Relevance to the teaching profession

In this project, I have learned more about the cognitive processes that take place in different contexts during language interpretation. This thesis has challenged me in so many ways, and I have loved learning about so many new areas that are not necessarily a part of a degree in linguistics. A few years back, I never would have imagined that I could possibly write a thesis that looks so different from a typical English teacher's thesis. This project has prepared me for a lot more than being a teacher, but there is no doubt that this thesis has relevance for the teaching profession. First of all, being brave enough to step outside your comfort zone and believing in yourself is the first things that comes to mind when I reflect back on that specific question. Everybody needs a teacher who will encourage them to try out new things, giving good advice, and being supportive. Secondly, when students ask "Why do I have to learn this?" I will make sure to explain to them that skills from all areas can come in handy for their future studies, jobs, and so on.

In a classroom, I will be able to use knowledge I acquired from a cognitive context and see it in more of a sociocultural context. This research will be relevant in the classroom, where I will be able to see how these cognitive processes behind language comprehension take place in the "real" world, and how students' mindsets evolve as they acquire new knowledge. In preparing for this thesis, I have taken several courses that deal with language acquisition and language processing in the brain. I know so much more about what *actually* happens when we learn a language, whether it is our first, second, or third. What happens inside our heads is simply stated *amazing* - and I think that having a teacher who knows about the processes behind the language mechanisms in the brain can be advantageous.

This thesis has also taught me how to work in a team and cooperation. Everything from meetings with supervisors, planning and scheduling EEG-experiments in a lab shared with several others, presenting my work for others, and so on. These skills come in handy for teachers, - and I believe this thesis is just one out of several projects I will be part of where cooperation and communication is key. These are skills I will make sure that my students will work on developing in my future classrooms as well.

Appendix B

Stimuli: referring/non-referring sentences

Condition		Det	Noun
Nonref	i mange situasjoner må man møte	et	menneske
Referring	på bussen møtte jeg i går	et	menneske
Nonref	en kokk kan ofte være	en	mann
Referring	på lørdag giftet hun seg med	en	mann
Nonref	ved hver pult sitter det	et	barn
Referring	i går fødte moren min	et	barn
Nonref	en deltaker i en konkurranse kan være	ei	jente
Referring	på fjelltoppen møtte jeg i sommer	ei	jente
Nonref	noen ganger treffer jenter	en	gutt
Referring	i to år har hun vært kjæreste med	en	gutt
Nonref	for å kopiere trenger man	et	ark
Referring	på bordet la jeg i stad fra meg	et	ark
Nonref	mange som er syke må gjennomgå	en	operasjon
Referring	i stad gjennomførte legen	en	operasjon
Nonref	noen eventyr handler om	en	enhjørning
Referring	jenta i eventyret elsker	en	enhjørning
Nonref	når det er mørkt er det lurt å ha	en	lykt
Referring	i skuffen har jeg	en	lykt
Nonref	vann kan kjøpes i	en	flaske
Referring	i sekken min har jeg vann i	en	flaske
Nonref	de fleste mennesker eier	en	telefon
Referring	på kontorpulten hennes ligger det	en	telefon
Nonref	når høsten kommer tar mange på seg	en	genser
Referring	i går ble jeg ferdig med å strikke	en	genser
Nonref	det er ikke uvanlig at skjorter kan mangle	en	knapp
Referring	på jakka jeg strikket i stad sydde jeg på	en	knapp
Nonref	av ulla fra sauen kan man blant annet få	et	garnnøste
Referring	for å strikke den røde votten valgte jeg ut	et	garnnøste
Nonref	man kjøper sjelden bare	en	sokk
Referring	på foten min har jeg akkurat nå på meg	en	sokk
Nonref	mange forskjellige ting lages med	en	maskin
Referring	det viktigste redskapet hun eier er	en	maskin
Nonref	mange jenter prøver seg frem med	en	sminkekost

Referring	i sminkepungen sin la hun fra seg	en	sminkekost
Nonref	mange med langt hår har	en	hårbøyle
Referring	i skapet mitt har jeg	en	hårbøyle
Nonref	når man snekrer er det lurt å ha	en	hammer
Referring	i stad kjøpte hun endelig	en	hammer
Nonref	på veien kjører det nesten alltid	en	bil
Referring	i garasjen sin har han	en	bil
Nonref	det tar kort tid å slipe	en	nøkkel
Referring	for å komme inn i huset mitt bruker jeg	en	nøkkel
Nonref	mange som plukker sopp liker å bruke	en	kurv
Referring	foran på sykkelen min har jeg	en	kurv
Nonref	ved forskjellige anledninger er det noen som heiser	et	flagg
Referring	på bursdagen sin heiste han	et	flagg
Nonref	pendlere kan fylle	et	tog
Referring	på stuegulvet leker barnet mitt alltid med	et	tog
Nonref	mange syklister bruker	ei	vogn
Referring	i stad hvilte barnet hennes i	ei	vogn
Nonref	i noen lykter trenger man	et	batteri
Referring	fra kjøleskapet hentet jeg i stad	et	batteri
Nonref	noen mennesker har en gang vært i	en	ulykke
Referring	de to bilene var nettopp involvert i	en	ulykke
Nonref	biler kjører på	en	vei
Referring	hjem til huset mitt går det	en	vei
Nonref	alle mennesker har	en	hjerne
Referring	på laboratoriet dissekerte han nettopp	en	hjerne
Nonref	de fleste idrettsutøvere trener på	en	øvelse
Referring	på tirsdag gjennomførte jeg på fotballbanen	en	øvelse
Nonref	i et orkester spiller alle	et	instrument
Referring	hun har brukt mange år på å lære å spille	et	instrument
Nonref	på scener over hele verden står det	et	piano
Referring	på lageret oppdaget han i går	et	piano
Nonref	mange elsker lyden av	en	fiolin
Referring	av faren sin arvet hun	en	fiolin
Nonref	i noen fortellinger møter helten	et	troll
Referring	i eventyret måtte Askeladden overvinne	et	troll
Nonref	på de fleste arrangementer kan man kjøpe	en	vaffel
Referring	til frokost spiste jeg i stad	en	vaffel
Nonref	hvis man er sulten kan man spise	en	bolle
Referring	da han var på café i stad spiste han	en	bolle
Nonref	i selskaper får ofte gjestene	en	middag
Referring	i stad samlet vi familien og spiste	en	middag
Nonref	når arbeidstakere må lære noe nytt arrangeres ofte	et	kurs

Referring	på torsdag var jeg på et hotell og deltok på	et	kurs
Nonref	for å teste elevenes kunnskaper lager noen lærere	en	prøve
Referring	i går satt jeg i klasserommet og hadde	en	prøve
Nonref	for å undersøke forskjellige ting kan man ta	en	test
Referring	for å sjekke løpeformen min tok jeg i stad	en	test
Nonref	hvis man blir sint kan det hjelpe å gå seg	en	tur
Referring	vi koste oss i stad da vi gikk	en	tur
Nonref	for å få melk må man melke	ei	ku
Referring	ute på tunet vårt har vi	ei	ku
Nonref	mange vogner må dras av	en	hest
Referring	da hun vant løpet i går red hun på	en	hest
Nonref	på bygda bor mange på	en	gård
Referring	i helga var jeg hos foreldrene mine som bor på	en	gård
Nonref	jegere jakter forskjellige ting når de er på	en	jakt
Referring	da hunden min rømte i går måtte vi gjennomføre	en	jakt
Nonref	i alle skoger kan man finne	en	elg
Referring	i forrige uke fikk jaktlaget skutt	en	elg
Nonref	sportsbutikker kan ofte ha salg på	et	skipar
Referring	i går hentet jeg endelig fra boden	et	skipar
Nonref	brøytebiler samler snøen i	en	snøhaug
Referring	på biltaket mitt samlet det seg på mandag	en	snøhaug
Nonref	på taket til de fleste hus er det	en	pipe
Referring	til peisen vår bygde vi i sommer	en	pipe
Nonref	for å forsterke lyd brukes ofte	en	mikrofon
Referring	foran meg på bordet står det	en	mikrofon

Appendix C

Instructions for the participants[10]

EEG Eksperiment

I dette eksperimentet vil du lese setninger som vises ord for ord på en dataskjerm. Les hver setning nøye, men ikke høyt. Det er viktig at du følger med på betydningen av setningene, siden du etterpå vil bli spurt om hva du har lest. Setningene inneholder ingen store bokstaver eller tegnsetting slik du er vant med, det er *betydningen* av setningene du skal følge med på.

Du må svare på hvert spørsmål så fort og riktig som mulig. Du har bare en begrenset tid til å svare på hvert spørsmål.

Hvis svaret ditt er JA, trykket du på **J**, hvis NEI trykker du på **F**.

EEG Eksperiment

I dette eksperimentet vil du lese setninger som vises ord for ord på en dataskjerm. Les hver setning nøye, men ikke høyt. Det er viktig at du følger med på betydningen av setningene, siden du etterpå vil bli spurt om hva du har lest. Setningene inneholder ingen store bokstaver eller tegnsetting slik du er vant med, det er betydningen av setningene du skal følge med på.

Du må svare på hvert spørsmål så fort og riktig som mulig. Du har bare en begrenset tid til å svare på hvert spørsmål.

Hvis svaret ditt er JA, trykket du på **F**, hvis NEI trykker du på **J**.

[10] Half of the participants were given the first instruction, the other half was given the second.

NTNU

Norwegian University of
Science and Technology