Anna Giskes

The On-Line Processing of Epistemic Modal Verbs During Language Comprehension

An ERP Study

Master's thesis in English Linguistics and Language Acquisition
ENG3920
Trondheim, May 2018

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

NTNU
Norwegian University of
Science and Technology

Abstract

Epistemic modal verbs enable the speaker to express attitudes about the factuality of a proposition: they signal a specific degree of likelihood that the proposition in their scope is true, as evaluated by the speaker. In this ERP experiment, we investigated the degree and timing of processing of the Norwegian epistemic modal verbs *å tro* (to believe) and *å tvile på* (to doubt). In sentences presented word for word (RSVP paradigm), we recorded ERPs of nouns that rendered propositions in the scope of these verbs (3rd person singular) either true/typical or false/atypical, as verifiable through general world knowledge, for example *Tora believes/doubts that birds have wings/gills*. As a control condition, we embedded the same clauses in matrix clauses with the non-modal factive verb *å vite* (to know).

We assumed that if the verbs are processed fully at the moment of encountering the critical word, then the N400 of the critical word, being sensitive to modulations of plausibility and predictability, would be modulated by the epistemic modal values of the matrix verbs. This modulation would transpire as an attenuation of the N400 typicality effect for *believe* compared to *know*, and a further attenuation or even reversal of the effect for *doubt*.

We observed an N400 typicality effect in both the *know*- and the *believe*-condition, with a larger and more wide-spread effect for *believe* than for *know*. In the *doubt*-condition, we observed no N400 typicality effect, but sustained anterior negativity (more negative-going for typical words) and a central-parietal typicality effect in the same direction in the time window 500-800 ms after onset.

Although additional data collection should be carried out in order to confirm the robustness of the effects in this small data set (N=12), this study indicates distinct ERP modulations by the main verbs, suggesting that at least parts of the lexical semantics of *believe* and *doubt* are processed before encountering the critical word.

Acknowledgments

Constructing an experiment from scratch was exactly as much fun as I had imagined. It was also much more difficult. I am grateful to the many people who have helped me along the way.

First of all, I would like to thank Prof. Giosuè Baggio for his supervision, advice, and for giving me the great opportunity to do an EEG experiment as a thesis project. I also want to thank my partner-in-panic Lia Calinescu, together with whom I created and executed the experiment. Our endless conversations about modality and typicality were invaluable, and so was having a friend to work with in the lab.

Eskil Wendel, Mats Øien, Erlend Granøien, Ingrid Bondevik and Sunniva Briså Strætkvern have been our Norwegian informants, and played an important role in constructing the stimuli. I want to thank them for their time and input: without them, this experiment would have been impossible.

A big hug to my library pals Camilla and Sharmyn for all the shared coffee, Kvikk Lunsj, sashimi, skiing, jokes, and sometimes despair: you made my thesis writing days something to look forward to, no matter what.

Finally, I want to thank my amazing husband Jeroen for his love, patience, and programming skills.

Contents

Chapter 1

Introduction

Language comprehension involves a human processing device that decodes detailed linguistic input, and makes sense of it with an astonishing speed and accuracy. One element involved in this interpretation task is the mapping of the decoded message onto world knowledge. Language provides the means to express situations that do not concern the *hic et nunc* of the speaker, such as situations far away in time and space, but also hypothetical, possibly factual, and counterfactual situations. As such, the recruitment of world knowledge for interpreting and evaluating linguistic input requires a more sophisticated mechanism than a simple 'fact-finder': before sensibly relating a complete message-level representation to world knowledge, linguistic cues about the speaker's attitude towards the factuality of the proposition must be interpreted.

A central question in research on language processing concerns the nature and time course of processing different elements of meaning. In order to understand an utterance of *Maybe John bought a bar* or *John did not buy a bar*, do we construct the proposition *John bought a bar* before adding factuality modifications such as *maybe* and *not*? Some models of language processing suggest such a step-wise procedure, in which the message-level representation of a sentence (or constituent) is built only after all syntactic and semantic information has become available. Opposing this view, there are models that assume full and immediate processing of every new piece of information as soon as it is encountered. Between these extremes, there are models that assume that some elements are processed immediately, but that other elements are delayed, producing intermediate processing stages with underspecified or partial representations.

By recording the electricity brought about by brain activity at the scalp through electroencephalography (EEG), we are able to observe a reflection of

language processing in real-time. Studying event-related potentials (ERPs), electro-physiological reactions to external stimuli such as linguistic input, has contributed to increasingly detailed processing models. In particular, research on the N400 component has proven to be a useful tool for investigating the time course of on-line language processing. The amplitude of this negative waveform, which is elicited by all content words, is sensitive to the degree of semantic fit of a word in context, and can, as such, be used to investigate the state of that context when a word is processed.

1.1 The present study

In this project, we investigated the timing and degree of processing of intensional verbs during on-line language comprehension. The project contained two related parts. The part described in this thesis concerns the processing and downstream effects of the Norwegian epistemic modal verbs *tro* (believe) and *tvile på* (doubt). During the same experiment, the 'imaginary-world evoking' verbs *drømme* (dream) and *innbille seg* (imagine) were investigated. The results of the imaginary-world counterpart to this thesis are described in the Master's thesis of Lia Calinescu (2018). The construction of the stimuli, the cloze probability test, and the organization and execution of the experiments are the result of a joint effort of Lia Calinescu and me.

The research question of this part of the project is:

- Are the lexical semantics of epistemic modal verbs processed fully and immediately during on-line processing?

We conducted an ERP experiment, in which we visually presented participants with sentences that denoted typical or atypical propositions, such as *birds have wings* and *birds have gills*, embedded in matrix clauses with the verbs under investigation (e.g., *Tora believes that birds have gills*). An example set of stimuli is displayed in table 4.1 on page 40; a list of all experimental stimuli can be found in appendix A.

As follows from the definitions of these verbs discussed in chapter 3, they express an attitude towards the factuality of the embedded proposition: *know* expresses certainty, and *believe* and *doubt* express different degrees of uncertainty and likelihood with respect to the factuality of a proposition. These expressed attitudes influence the plausibility of sentences containing typical and atypical complements, and, thereby, comprehenders' expectations of the typicality of a complement.

If the lexical semantics of these verbs are processed fully and immediately during on-line language comprehension, their modulations of plausibility

and expectedness should affect processing immediately, and, consequently, influence N400 amplitudes for words following the verbs that render the embedded propositions typical or atypical (e.g., *wings* or *gills* following *birds have...*). A detailed account of the hypotheses and predictions can be found in section 4.1 (p. 39).

1.2 Overview

In chapter 2, I will review relevant parts of the vast body of research on the N400 component. I will give an overview of linguistic manipulations that the N400 is sensitive to, and discuss the processes that are assumed to underlie the N400. In section 2.3, I will discuss the hypothesis of full incrementality in language processing, and how the N400 has been used in order to investigate this hypothesis.

Chapter 3 provides an introduction of the phenomenon of epistemic modality, and a definition of the lexical semantics of the verbs *know*, *believe*, and *doubt*. I will discuss how these verbs modulate the meaning of their complements. My account of modality is very selective, and, in fact, barely scratches the surface of the theories that form the foundations of these definitions. However, they suffice to support my predictions regarding the verbs' modulation of the N400.

Chapters 4-7 describe the current study and discuss the results.

Conventions

In most chapters, I use the English translations *know*, *believe*, and *doubt* of the verbs *vite*, *tro*, and *tvile på* as if they are fully interchangeable. Unless stated otherwise, I assume that my statements about *know*, *believe*, and *doubt* to be valid for the Norwegian verbs as well.

When displaying stimuli of this and other ERP studies, I indicate critical words, for which ERPs are measured, by underlining them. I indicate different words in different conditions with a '/'. For example: *Tora knows/believes/doubts that birds have wings/gills*.

Chapter 2

The N400

2.1 A measure of semantic fit

The N400 is negative waveform that is visible in ERPs elicited by all content words, peaking around 400 ms after the word is presented (visually or auditorily). In a famous ERP study, Kutas and Hillyard (1980) discovered the N400, and observed that the amplitude of the N400 waveform was larger for semantically anomalous sentence-final words, such as *socks* in (1), than for semantically non-anomalous words. Furthermore, they discovered that the magnitude of this N400 effect was correlated to the degree of violation of "semantic expectancy": the N400 effect was larger for strong violations (2c) than for moderate violations (2b), compared to semantically non-anomalous control conditions (2a).

(1) He spread the warm bread with socks.

(2) a. He took a sip from the bottle.
    b. He took a sip from the waterfall.
    c. He took a sip from the transmitter.

Stimuli from Kutas and Hillyard (1980).

Examining the effect of cloze probability and degree of contextual constraint, Kutas and Hillyard (1984) found a correlation between the N400 amplitude and cloze probability: although all critical words in their experiment resulted in meaningful and plausible sentences, words with a high cloze probability elicited smaller N400s than words in less constraining sentences.

Kutas and Hillyard (1984) further observed in a *post hoc* analysis that the N400 amplitude of words with a low cloze probability in highly constraining

contexts was inversely correlated to the degree of semantic relatedness to the word with the highest cloze probability, the 'best completion' word, regardless of plausibility. This result was corroborated by Federmeier and Kutas (1999), who compared ERPs of implausible words with different degrees of semantic relatedness to the best completion word. They observed reduced N400 effects for implausible words that were of the same semantic category as the best completion word, as compared to implausible words that were of a different semantic category. (3) is an example of the stimuli from this study:

(3) The tourist in Holland stared in awe at the rows and rows of colour. She wished she lived in a place where they grew *tulips/roses/pines*.

From Federmeier and Kutas (1999, p. 473).

The critical word *roses* is, compared to *pines*, semantically more closely related to the best completion word *tulips* (i.e., shares one more semantic category), and elicited a smaller N400 effect than *pines*.

In addition to semantic relatedness, it has been observed that implausible words that are lexicographically or phonologically related to a best completion word of high cloze probability can elicit a reduced N400 effect as compared to unrelated but equally implausible words (Ito et al., 2016).

Similar observations have been made in ERP studies of a different paradigm: in lexical priming studies, words were not presented in a sentence context, but following one or more semantically related or unrelated words. The N400 is generally smaller when words are primed semantically, associatively, phonologically or orthographically (see Kutas and Van Petten (1988) for a review of early ERP priming studies).

The N400 amplitude has been shown to be sensitive to a very wide variety of factors that modulate the 'semantic fit' of a word. In a Dutch ERP study, Hagoort et al. (2004) showed that the N400 effect is indistinguishable for semantic violations depending on world knowledge (*Dutch trains are white and crowded*[1]) and semantic violations depending on lexico-semantic constraints (*Dutch trains are sour and crowded*).

Furthermore, the N400 is sensitive to plausibility and expectedness modulated by discourse context, regardless of whether this modulation takes place in the sentence of the critical word or in the wider discourse context (Van Berkum et al., 1999, 2003b; see section 2.3.1). Also speaker information belongs to this 'wider discourse context', and can modulate the N400: Van Berkum et al. (2008) found N400 effects for degree of consistency with

[1] They are actually yellow.

inferences about the speaker. In example (4) from their experiment, the word *wine* elicited a larger N400 when spoken by a child's voice than when spoken by an adult's voice.

(4) Every evening I drink some wine before I go to sleep.

From Van Berkum et al. (2008, p. 281).

2.2 Processes underlying the N400 component

Since its discovery, researchers have attempted to pinpoint which processing-related mechanisms are reflected in the N400. In order to do so, models of language processing break down the overall process into distinct steps or sub-processes, which—in an ideal world—can be isolated and manipulated in an experimental environment, so that possible correlations between the N400 and distinct sub-processes can be observed. Different models distinguish different steps, but any model of language processing needs to account for both non-combinatorial processes, that concern activation and retrieval of lexical and semantic representations in the brain upon encountering a specific string, and 'higher-level' processes, that integrate the recognized and retrieved lexico-semantic element into the unfolding sentential and wider discourse context, and update the mental representation of the communicated content. Almost 40 years of ERP studies on the behavior of the N400 has not led to consensus as to the nature of the processes reflected in the N400. Both non-combinatorial processes and processes of semantic integration have been argued to underlie the N400, an issue that is often referred to as the 'access-integration' debate. In section 2.2.1, I will discuss the 'traditional' viewpoints in this debate, and in section 2.2.2, I will describe a third approach that is currently gaining more and more support, viewing the N400 as a summation of activity generated by multiple combined processes.

2.2.1 The access-integration debate

The integration view assumes that the N400 reflects processes that integrate a word into the preceding discourse context (Hagoort et al., 2004; Van Berkum et al., 1999). In this view, the N400 effect elicited by semantic anomalies is ascribed to the fact that these anomalies result in an implausible or semantically incoherent message, which causes relative difficulty of integration. This integration process occurs post-lexically, and is directly affected by plausibility formed by both world knowledge and the specific discourse context. As such, the integration view can account very

well for the widely observed fact that the N400 amplitude is correlated with plausibility, regardless of whether it is brought about by world knowledge, lexico-semantic constraints, discourse context, or pragmatics.

A challenge for the integration view is the sensitivity of the N400 to other factors than plausibility, that do not affect ease of semantic integration as clearly or directly: the sensitivity of the N400 to cloze probability, independent of plausibility (Kutas and Hillyard, 1984), and the N400 attenuations caused by non-semantic priming are not naturally explained within the integration view.

Opposing the integration view, there are models that take the N400 to reflect processes of lexical activation and/or retrieval (Chow and Phillips, 2013; Brouwer et al., 2012; Kutas and Federmeier, 2011; Van Berkum, 2009). In this view, attenuation of the N400 is assumed to be caused by lexical pre-activation, which results in a less effortful activation and retrieval process. Whereas the integration view considers the N400 and semantic fit into the preceding discourse to be directly related, the lexical access view proposes a more indirect effect of preceding discourse: words that are predictable or plausible in a certain discourse context are assumed to be pre-activated in the brain, which leads to facilitation of the retrieval process. Thus, crucially, this view proposes that non-combinatorial processes underlie the N400. All higher-level processes that modulate the N400 are taken to do so indirectly by affecting these non-combinatorial processes.

The lexical access view can account for a variety of seemingly unrelated factors that have been shown to affect the N400 amplitude. It assumes that the commonality of factors such as frequency, repetition, phonological, orthographic and semantic similarity to a predictable word, is that they affect ease of lexical access.

Additionally, the lexical access view is supported by a group of studies on the 'semantic P600 phenomenon'.

(5) a. De speer heeft de atleten <u>geworpen</u>.
The javelin has the athletes <u>thrown</u>.

b. De speer heeft de atleten <u>opgesomd</u>.
The javelin has the athletes <u>summarized</u>.

From Hoeks et al. (2004, p. 62).

Hoeks et al. (2004) found an expected N400 effect for sentence-final semantically anomalous verbs in highly constraining sentences such as (5b), but did not observe an N400 effect for verbs in the same position that were only implausible because of the verb's thematic role assignment, as in (5a). Instead, a P600 effect occurred in this condition: a late positive wave that

is associated with morphosyntactic processing and integration (Kutas and Federmeier, 2011).

Subsequent studies have found a similar absence of an N400 effect for verbs that were lexico-semantically associated with the preceding discourse, but that did not result in a semantically coherent message-level representation, for example through violated animacy constraints or reversed thematic roles (e.g., *for breakfast, the eggs were eating...*, see Kuperberg et al. (2003); *the murder was witnessing*, see Kim and Osterhout (2005)). In a review of studies on the semantic P600 phenomenon, Brouwer et al. (2012) argued that the results of these studies support a processing model that takes the N400 to reflect lexical retrieval. They suggest that the lexico-semantic association between words in the preceding context and the critical word pre-activates the latter, causing facilitated lexical retrieval, reflected in an attenuated N400 amplitude compared to equally implausible or unexpected words that are not lexico-semantically related. This suggestion also offers a good explanation for the findings of Nieuwland and Van Berkum (2005): they observed no N400 effect when they replaced a character in a short story with an object that was central in the story (e.g., *the woman told the tourist/suitcase...* in a story featuring a tourist with a suitcase at an airport counter). Here, lexical access is facilitated by the fact that the anomalous object is an activated entity in the discourse context[2].

The lexical access view can also account for the results from N400 priming studies. Whereas the integration view must assume that primes function as an abstract form of discourse context in which target words are integrated, the lexical access view has a more straightforward explanation for N400 priming effects: relatedness to previous words has a priming effect, resulting in activation and facilitated retrieval of the target word.

Inconsistent with the lexical access view is the fact that N400-like effects also show up for non-linguistic semantic anomalies, such as the appearance of anomalous objects in short movie scenes (Sitnikova et al., 2003), mismatches between visual and auditory information, and gestures, albeit with a more frontal scalp distribution than N400s elicited by linguistic stimuli (see Sitnikova et al. (2008) of a review). Another finding that is difficult to reconcile with the lexical access view is that the N400 can be modulated by information structure. Li et al. (2008) observed effects of topic/focus

[2]Additionally, the anomaly was presented out of discourse focus. Phenomena such as the Moses illusion indicate that we might not process certain words out of discourse focus fully and immediately during comprehension. Moreover, as briefly discussed further on in this section, it seems that the N400 can be manipulated by information structure in itself.

marking in Chinese. In their study, new information that was (appropriately) focus-marked elicited larger N400s than topic-marked new information. Van Berkum (2009) accounted for such findings by extending the lexical access view, and proposed that the N400 is not only modified by facilitation through pre-activation, but also by 'intensity' of retrieval: if words are presented and/or interpreted as salient or important, the retrieval process is intensified and uses more resources. Within a more strict lexical access view, however, these findings pose a challenge.

2.2.2 The multiple process view

Recent research gives rise to a third approach, in which the N400 component is assumed to reflect more than one process, consisting of the result of several interdependent processes or processing stages. Lau et al. (2008) presented a meta-analysis of a large number of EEG, MEG, and fMRI studies and suggested that activity in multiple brain regions might contribute to the N400 wave. They observed that the brain region that is associated with lexical storage and access, the posterior middle temporal cortex, consistently shows N400-related activity in the MEG and fMRI studies reviewed. The authors concluded that "these data strongly suggest that at least some substantial part of the N400 effect reflects facilitated lexical access, and thus that the N400 effect cannot be attributed only to post-access processes." (p. 928) Thus, they did not rule out the possibility that the observed activity in other brain regions contributes to the N400 as well, and they suggested that the N400 wave might not have a one-to-one relationship with a single underlying process.

Baggio and Hagoort (2011) proposed that the N400 indeed reflects multiple processes. To begin with, they called into question whether we are able at all to sufficiently isolate lexical and combinatorial processes in experimental settings in order to observe their unique contribution to sentence processing. But even if we can, they argued, the distinction is a theoretical one: processing is not a unidirectional procedure. Processing of words in (discourse) context is more accurately described as a cyclic interaction of intertwined lexical and combinatorial processes. Supporting their model with data from neurobiological studies on localization of processes, they argued that the N400 reflects a cycle of mutually connected processes of both lexical (pre-)activation and integration ('unification').

The findings of an ERP study by Otten and Van Berkum (2007) fit this model well. They reported that discourse contexts with different degrees of constrainingness yielded N400-like deflections with different distributions

and shapes, suggesting that the N400 is formed by different processes. Furthermore, a recent *post hoc* analysis (Nieuwland et al., 2018) of a large-scale EEG experiment (Nieuwland et al., 2017, see section 2.3.1), presented evidence for the view that the N400 reflects multiple processes. They assessed the predictability (through cloze probability) and plausibility (through plausibility judgment tests) of each of the critical words in the experimental sentences, and disentangled the effects on the N400 of predictability and plausibility, which showed clear differences in shape, timing, and distribution. The effect of predictability occurred relatively early and was more negative-going; the effect of plausibility manifested itself in a later and more flattened negative waveform. These findings support the hypothesis that multiple processes generate subcomponents that combine into the N400.

2.3 N400 effects as evidence for incrementality of language processing

Having discussed the types of manipulations the N400 amplitude is sensitive to, and which processes have been proposed to be reflected in the N400, I will, in this section, discuss ERP studies that have utilized the behavior of the N400 to establish whether certain elements in the context preceding a critical word are processed fully at the moment of processing the critical word.

2.3.1 Incrementality of language processing

The meaning of an utterance, i.e. a concrete occurrence of a specific sentence, uttered by a speaker in a certain discourse situation, is determined by different aspects of meaning. To begin with, it contains word meaning: the relatively stable range of meanings that is stored for each lexical item in our mental lexicon. Secondly, syntactic rules enable us to construct compositional sentence meaning: meaning that contains a combination of lexical items engaging in a specific relationship that is dictated by syntax, resulting in the representation of a state of affairs. For example, the sentences *the dog persuaded the man to go for a walk* and *the man persuaded the dog to go for a walk* are made up of the same lexical material, but have different sentence meanings, as determined by their syntactic structure. A third element that determines sentence meaning concerns integration in the discourse context in the broadest sense of the word. Speaker information and preceding discourse are needed to establish, for example, which unique entities are referred to by *the man* and *the dog*, and what the deictic center is to relate the tense

of *persuaded* to. World knowledge helps the speaker determine that *to go for a walk* is an action that probably involves both the man and dog, and that the act of persuasion probably involves different actions depending on whether the dog or the man is the Agent.

The observation that a sentence has lexical and syntax-dependent meaning elements on the one hand and context-dependent meaning elements on the other, has given rise to the idea that these meaning elements are computed during distinct and sequential processing stages. In line with the influential Chomskyan focus on sentential syntax as the central element of language, determining and dominating semantics, some models (e.g., Carpenter and Just 1975; Just and Carpenter 1980; Fodor et al. 1974; Frazier and Fodor 1978) view processing as a bottom-up, stage-like process, in which the context-free, propositional meaning of a sentence (or constituent) is first constructed step-by-step on the basis of syntactic structure and lexical semantics (in that order); integration into the discourse context follows after this construction has been completed.

However, more and more research seems to contradict the idea that interpretation nicely waits for its turn during on-line processing. Instead, following results from studies using eye-tracking, EEG, and MEG, there is a growing consensus that language processing is largely incremental. The human parser seems able to process and integrate many aspects of linguistic input 'on the fly' as a sentence unfolds, without waiting for a (sub-)sentential syntactic structure to be completed.

EEG-research on the N400 has contributed greatly to our knowledge of on-line processing. An important pair of studies supporting an incremental view of processing are the ERP studies by Van Berkum et al. (1999, 2003b). They presented their participants with sentences such as (6) with two alternative critical words that were equally plausible in the sentential context, but differed in plausibility when looking at the discourse context the sentences were embedded in.

(6) As agreed upon, Jane was to wake her sister and her brother at five o' clock in the morning. But the sister had already washed herself, and the brother had even got dressed. *Jane told the brother that he was exceptionally quick/slow.*

Example stimulus from Van Berkum et al. (2003b, p. 703), experiment 1. Translated from Dutch by the authors.

In this experiment, the N400 for the critical words was modulated by plausibility brought about by discourse context. When the same sentences were

presented without discourse context, there was no significant difference between the critical words, showing that the N400 is just as sensitive to extra-sentential discourse coherence as to local coherence. The N400 elicited by discourse-anomalous words was not distinguishable in latency and distribution from N400 effects induced by sentence-level anomalies, indicating that the processes reflected in the N400 do not distinguish the two. Chwilla and Kolk (2005) came to similar conclusions in an ERP priming study. They observed N400 effects when words were primed by word pairs that were not directly related to the target word through world knowledge, but together formed a plausible 'script' (e.g., *move—piano—backache*). They concluded that incoming information is immediately combined with the widest possible discourse context and common ground: no evidence was found for a distinction between world knowledge and discourse-specific 'scenario knowledge'.

Subsequent N400 studies have given ample evidence for the idea that words are immediately maximally integrated in the wider context (see Hagoort and Van Berkum (2007); Van Berkum (2009) for reviews); at least many elements of language processing seem to happen in a largely incremental fashion.

Beyond incremental processing: the role of prediction

In addition to processing upcoming words incrementally, comprehenders seem to engage in forms of prediction during comprehension. Although the N400 amplitude is clearly correlated to predictability, this is, in itself, not evidence of prediction: proponents of the integration view would argue that this correlation can be explained by the fact that both predictability and the N400 amplitude are correlated to ease of integration. However, when assuming that the N400 at least partly reflects lexical pre-activation, the correlation between the N400 amplitude and predictability indicates some form of prediction. Evidence from behavioral studies shows that there are many aspects of language we predict. The phenomenon of active gap-filling indicates that comprehenders predict syntactic structure (Boland et al., 1995), and eye-tracking experiments using the Visual World Paradigm have observed anticipatory eye movements to arguments that fit lexico-semantic constraints (Altmann and Kamide, 1999) and plausibility (Kamide et al., 2003).

A long-standing debate concerns the extent to which we actively predict upcoming words: do we only predict semantic fields and senses, or also lexical items, grammatical features, specific word forms, and even lexicographic and phonological form? ERP studies have been influential in forming a pic-

ture of the role of prediction during processing. The fact that the N400 is smaller for words that are incoherent but semantically related to a coherent completion (Federmeier and Kutas, 1999, see section 2.1, p. 12) indicates that we at least predict senses of upcoming words. N400 effects for phonological/orthographic relatedness to high cloze probability words show us that highly constraining sentences can also lead to prediction of lexical items. Nevertheless, prediction is a tricky issue to investigate, since most experimental designs only allow us to measure indirect effects of prediction of a word *after* it is presented. With these measures, it is difficult to disentangle effects of prediction from effects of, for example, ease of integration.

Cleverly making use of number and gender agreement features on determiners and adjectives, a number of ERP studies have investigated lexical prediction by looking at the N400s of words preceding a predictable noun in order to see whether the noun was actively predicted by comprehenders. For example, Van Berkum et al. (2005) measured ERPs of adjectives that were morphologically consistent or inconsistent with highly predictable nouns following the adjective, and found an N400 effect of consistency, indicating that the morphosyntactic features of the predictable noun might have been activated at the moment of processing the adjective. An impactful study by DeLong et al. (2005) went a step further, and investigated ERPs of the English indefinite articles *a* and *an* when they were consistent or inconsistent with predictable following words, for example *the day was breezy so the boy went out to fly a kite/an airplane*. This study found an N400 effect on the determiner for consistency with the expected noun, and has long been a central piece of evidence for a strong form of prediction, including pre-activation of a specific word form and its phonological form.

However, as Kochari and Flecken (2018) pointed out, the results from studies measuring ERPs before the occurrence of the predictable word are far from uniform. The measured effects differ across studies in latency, distribution, and even polarity; something that is at odds with the general uniformity of N400s that are elicited by nouns. Moreover, recent attempts to replicate these studies have not been successful. The attempt of direct replication of DeLong et al.'s experiment by Nieuwland et al. (2017) has failed quite spectacularly. Presenting the same experimental materials to no less than 334 participants across 9 labs, only the N400 effect on the noun was replicated, but no effect was observed on the determiner.

It seems, then, as Nieuwland et al. concluded, that there is at this moment no indisputable evidence for a strong form of lexical prediction.

2.3.2 The N400 as index of degree and timing of integration

Assuming that the N400 reflects processes that are affected—directly or indirectly—by available discourse context, the sensitivity of the N400 is often used as a measure of degree of integration of certain elements in the preceding context. As such, the behavior of the N400 can provide us with a detailed picture of language processing. When the manipulation of an element causes an N400 effect, this element can be assumed to be processed and integrated fully into the discourse context at the moment the critical word is encountered, adding to the evidence for incrementality of processing. The absence of an N400 effect could, provided that confounding factors are controlled for, be caused by incomplete processing or integration of the manipulated element when encountering the critical word, which would be compatible with bottom-up, multiple-stage models of (elements of) language processing.

There are, however, alternative explanations for insensitivity of the N400 to certain manipulations. It could be the case that the manipulated element is an exception to the assumption that all available discourse context affects the process underlying the N400; the manipulated element would in that case have some special (semantic) characteristic that causes the N400 process not to utilize it, although it is available. A more mundane type of explanation could be that the manipulated element does affect the N400, but that the effect is obscured by confounding factors, that are not controlled for in the research design. As we have seen, the N400 seems to be the result of a continuously communicating complex of intertwined processes, that may never be isolated completely. Consequently, results of ERP studies are more often than not susceptible to multiple interpretations.

In the remainder of this section, I will discuss a number of studies that investigated degree and timing of processing and integration using the N400. I will discuss both studies that find N400 effects modulated by preceding context and studies that do not find context effects. As we will see, when taking all evidence into account, a general picture emerges of an incremental processing system, of which the consequences for processing a word in context *can* be visible in the N400, under the right circumstances. Furthermore, these studies illustrate how research design can affect results and obscure effects.

Processing negation

Although many studies found an N400 effect when manipulating the truth value of a proposition, the effect of truth value manipulations is not always

visible as an N400 effect. Fischler et al. (1983) observed that altering the truth value of simple statements such as *a robin is a* bird/tree by adding *not*, did not alter the N400 effect for the sentence-final critical words. The affirmative and negating sentences elicited significant N400-effects of the same magnitude and in the same direction, namely with more negative-going N400s for *tree* than for *bird*, regardless of the presence of *not*. Fischler et al. concluded that the negations did not affect the N400s of the critical words because negation is not integrated immediately and fully during on-line processing. They took their findings as support for a non-incremental model of language processing, and considered the insensitivity of the N400 to the negation to bear witness to a pre-final processing stage in the negating sentences such as *a robin is not a bird*, in which the proposition *a robin is a bird* is constructed before being negated.

Our current knowledge about the influence of lexico-semantic association on the N400 renders these conclusions invalid, since they hinge on the assumption that the N400 amplitude is only, or at least mainly, related to sentence congruity. The current view that the N400 reflects processes of pre-activation, which can be triggered by lexico-semantic association only, can explain the results without rejecting immediate integration of the negation: in Fischler et al.'s sentences, close lexico-semantic association with context words (*robin* in the example sentence) could pre-activate the critical word (*bird*), explaining the attenuated N400 for both affirmative and negating sentences.

In line with this suggestion, Nieuwland and Kuperberg (2008) argued that the absence of an observable N400 effect of negation has to do with the fact that the sentences used by Fischler et al. (1983) were presented without discourse context. Stating a triviality such as *a robin is a bird* out of the blue is an infelicitous speech act. When not embedded in a pragmatically licensing discourse context, such an utterance violates Grice's maxim of relevance (Grice, 1975). Since the N400 is sensitive to pragmatic violations (Van Berkum, 2009), the effect of pragmatic infelicity might have overshadowed any effect of congruence or truth value. Nieuwland and Kuperberg tested this hypothesis by comparing ERPs in true and false negating sentences that were either pragmatically felicitous (7b) or infelicitous (8b)[3].

(7) a. With proper equipment, scuba-diving is very dangerous/safe...
b. With proper equipment, scuba-diving isn't very dangerous/safe...

[3]Isolated sentences outside a discourse situation are, of course, never completely felicitous. Nevertheless, negations in isolated true sentences can be rendered more felicitous by making them part of propositions that are not trivially true, but could plausibly be believed to be false, as in example (7b, true condition).

(8) a. Bullet-proof vests are very dangerous/safe...
b. Bullet-proof vests aren't very dangerous/safe...

Set of example sentences from Nieuwland and Kuperberg (2008, p. 2018).

Just like Fischler et al. (1983), Nieuwland and Kuperberg found no effect of negation in the infelicitous sentences; however, in the pragmatically licensed sentences, the N400 of the critical word was modulated by truth value, indicating that negation is processed and integrated immediately, or, in any case, fast enough to affect processing of the critical word.

Processing quantification

A similar development of insights can be seen in studies on quantifier processing. Kounios and Holcomb (1992) manipulated the truth value of sentences by changing quantifiers (e.g., *all/some/no rubies are gems/spruces*). They observed no N400 effect of quantifier, and concluded that quantifiers are not processed immediately, in the same vein as Fischler et al. (1983) did for negation. However, Urbach and Kutas (2010) did find an N400 effect using a similar design. In a 2x2 design, they fully crossed quantifier type (*most* and *few*) and typicality (typical and atypical), resulting in items such as *most/few farmers grow crops/worms*. Urbach and Kutas found an asymmetric result. In addition to a main effect of typicality (atypical words elicited a larger N400), they found that the N400 elicited by the typical words was modulated by the quantifier type in the expected direction, i.e. a larger N400 for *few*-type quantifiers than for *most*-type quantifiers. The N400 of the atypical words, on the other hand, was not significantly affected by the quantifier type. The authors explained the asymmetry by proposing that, initially, quantifiers may only be interpreted partially, resulting in an incomplete or underspecified representation available at the time of processing the critical word.

However, when assuming that the N400 reflects processes concerning lexical access, the absence of an N400 effect for the atypical words can also be explained as a 'ceiling performance' (or, rather, 'floor performance'): whereas the cloze probability of typical words drops significantly when changing *most* to *few*, the cloze probability of the atypical words is extremely low for both the *most*-type and the *few*-type sentences. Although the *few*-type quantifier sets the reader up to expect something atypical, it usually does not give any indication of which of the vast number of atypical continuations to expect, and does not enable the reader to predict or pre-activate specific continuations.

Urbach et al. (2015) followed up this study, and investigated the influence of the participants' task and of a pragmatically licensing discourse context on the same set of stimuli. They found that without a specific task for participants, and with an appropriate preceding discourse context, the typicality effect was fully and symmetrically reversed by the quantifiers: for typical words, the *few*-type quantifiers elicited the largest N400, but for atypical words, *most*-type quantifiers did. This result shows that quantifiers are in principle processed fully and immediately, and integrated rapidly enough to affect lexical retrieval.

Processing counterfactuals

The studies on negation and quantifiers discussed above all investigated the processing of simple indicative sentences that represent well-known facts of the world, that the participants were assumed to know. We have seen that comprehenders recruit their world knowledge to process, pre-activate, and even predict during reading or listening. Studies such as Van Berkum et al. (2003b) (see section 2.3.1, p. 18) show us that in addition to world knowledge, comprehenders utilize the preceding discourse context in the same way. This inspired the question of the relative weight of congruity dictated by world knowledge and by discourse context. What happens when world knowledge and discourse context contradict each other? Nieuwland and Van Berkum (2006) investigated this issue by presenting comprehenders with short fictive stories featuring inanimate objects as animate characters, such as the dancing peanut in (9):

(9) A woman saw a dancing peanut who had a big smile on his face. The peanut was singing about a girl he had just met. And judging from the song, the peanut was totally crazy about her. The woman thought it was really cute to see the peanut singing and dancing like that. The peanut was salted/in love, and by the sound of it, this was definitely mutual. He was seeing a little almond.

Example story from Nieuwland and Van Berkum (2006, p. 1106).
Translated from Dutch by the authors.

The extensive discourse context reversed the N400 effect that would be expected of a sentence as *the peanut was salted/in love* in isolation. In stories like (9), world-knowledge-coherent critical words (*salted*) elicited a larger N400 than discourse-coherent critical words (*in love*), showing that discourse context can outweigh world knowledge. Comprehenders are able to create

a fictional world on the spot, and utilize all consequences of this world's elements that differ from the real world, such as the animacy of peanuts.

Nieuwland (2013) pointed out that the results of Nieuwland and Van Berkum (2006) could be dependent on, or in any case aided by, the fact that the animacy alteration was repeated many times in the short story before the occurrence of the critical word. Moreover, genre conventions of cartoonesque fiction, in which objects regularly have human characteristics, might have encouraged the participants to interpret the objects as animate characters. Ruling out this alternative explanation, results similar to Nieuwland and Van Berkum's have been found in ERP studies on the behavior of the N400 in counterfactual reasoning, conveyed in sentences of the form *if...then...* (e.g.,*if ducks were fluorescent, they would be easy to spot at night*), in which the antecedent *if...* introduces a counterfactual situation, and the consequent *then...* contains a proposition that is true of that counterfactual situation (see Kulakova and Nieuwland (2016b) for a review). ERP studies have shown that counterfactual antecedents are rapidly integrated, and counterfactual discourse context created by them can prevail over world knowledge during processing: N400 amplitudes are smaller for critical words in the consequent that render a proposition factually false but coherent with the counterfactual antecedent, than for words that render a proposition factually true but incoherent with the antecedent (Nieuwland, 2013; Nieuwland and Martin, 2012). What is more, even within the antecedent the N400 is already modulated by the expectation of counterfactuality. Kulakova and Nieuwland (2016a) investigated N400 effects in antecedents such as *if sweets/words were made out of sugar...*, and found a larger N400 for the factual condition (*sweets*). This result shows us that the linguistic cues for counterfactuality (*if...* and subjunctive mood) are processed immediately and affect expectations about upcoming words.

2.4 Summary

I have discussed the long-standing debate about the functional significance of the N400, and suggested, in line with recent research, that the behavior of the N400 can best be explained by assuming that it is a reflection of multiple processes, which involve, and are facilitated by, lexical pre-activation. Lexical pre-activation can be brought about by both plausibility and predictability set up by the preceding context. A fruitful research design to investigate incrementality of language processing rests on this assumption that the N400 is modulated by plausibility and predictability; many studies have shown that the N400 can be modulated by a large variety of plausibil-

ity and predictability manipulations. I have also illustrated that this wide variety of possible influences introduces a large number of possibly confounding factors in N400 studies. Along the way, I have given an overview of relevant ERP studies on processing negation, quantification, and counterfactuals. These studies form an important background for the hypotheses of the current study, and the interpretation of the results.

Chapter 3

Epistemic modality

Although modality is a fundamental and omnipresent property of language, it proves to be very hard to capture in a single, concrete, and unambiguous definition. A common core that seems to be part of most definitions is the linguistic expression of things that are not necessarily real, or not sensibly evaluable in terms of truth.

One way of expressing things that do not need to be factual is through epistemic modality: this sub-category concerns expressions that contain a degree of uncertainty about whether something is true. The lexical semantics of the Norwegian verbs under investigation in this study, *å tro* and *å tvile på*, are of this category. In this chapter, I will provide a definition of the modal mechanisms that the lexical semantics of these verbs bring about. I base most of my description and terminology in the following sections on the framework described in the boldly titled chapter 'The definition of modality' by Declerck (2011). Unless stated otherwise, the terminology and definitions in this chapter (printed in italics when introduced) are in accordance with Declerck's terminology. In section 3.4, I will zoom in on the semantics of *doubt*, and highlight a number of properties of the verb that distinguish it from the other verbs in this experiment.

3.1 Possible worlds

A useful way of analyzing verbs like *know*, *believe*, and *doubt* is to describe them in terms of possible world semantics. A possible world is a way the world could be. The world we live in, the *factual world*, is one possible world. The factual world contains all facts of the world, regardless of distance in time and space: *Beethoven wrote 9 symphonies*, *the melting point of gold is 1064 degrees Celsius*, *my neighbors own a black cat.* All states of affairs

that take place, that are the case, or that have done so in the past, are part of the factual world: they *actualize* in the factual world. But this world could have been different. Beethoven could have died before he wrote his ninth symphony, or the melting point of gold could have been 1063 degrees Celsius. The world could be different in much more drastic ways as well: one can imagine a world in which humans did not develop language, or a world that solely consists of a floating grid of uninhabited cubes of chocolate pudding. As there are infinitely many ways the world could be, there are infinitely many possible worlds. To use the imagery of Lewis (1986, p. 2), these possible worlds are like gigantic, complete and inclusive planets, or rather whole universes, that exist beside the factual world, but that do not have any spatial or temporal relation to the factual world, or to any other possible world.

Reflecting the human ability to engage in non-factual reasoning, language allows us to express things that are not necessarily real, in other words, that do not actualize in the factual world.

(10) John and Mary should buy a bar!

(11) I believe that an elk ate the supplies.

Example (10) contains the proposition *John and Mary buy a bar*, but does not represent an event in the factual world: it is irrelevant for understanding this utterance whether John and Mary will buy a bar in the factual world at some point. Similarly, (11) can be a valid statement regardless of whether the proposition *an elk ate the supplies* is true of the factual world. *Modalizers* such as the modal verbs *should* and *believe* in (10) and (11) are elements of language that signal that an expressed proposition actualizes in a possible world that is not the factual world: they are *world-evoking* devices.

Before taking a closer look at the modal verbs used in this study, I should say something about the nature of possible worlds. Using the concept of possible worlds begs the question of what they are, and whether and how they exist. Lewis (1986, p. 2-3) advocated the view that all possible worlds are as 'real' as the factual world:

> [The factual world does not differ from the others] in its manner of existing. I do not have the slightest idea what a difference in manner of existing is supposed to be. Some things exist here on earth, other things exist extraterrestrially, perhaps some things exist no place in particular; but that is no difference in manner of existing, merely a difference in location or lack of it between things that exist.

The existential status one assigns to possible worlds has consequences for the description of the operations executed by modalizers. When following Lewis' concretism, a modalizer merely selects or accesses a set of existing possible worlds in which the proposition in its scope actualizes (see Portner (2009, p. 29 et seq.) for an account of epistemic modalizers along these lines). A more abstract approach to possible worlds might allow a view in which a modalizer rather creates a specific possible world in which a proposition actualizes; this seems to be the assumption behind the framework of Declerck (2011). For the purposes of describing the modal verbs in this study, I (fortunately) need not and will not take a stance in this philosophical debate, and I will use terms such as 'evoke', 'create', or 'refer to' possible worlds to describe modalizers' mechanisms rather carelessly in this respect.

3.2 Epistemic modalizers

Although the factual world might not be special or more real than other possible worlds, it certainly has a special status to us, inhabitants of the factual world, and consequently, to our language. Modalizers do not only locate propositions in a non-factual possible world, but also place this possible world in a certain relation to the factual world. Different 'flavors' of modality denote different types of relations. One flavor, epistemic modality, as Declerck described, creates a possible world in which the proposition under its scope actualizes, and specifies that this possible world, with a certain likelihood, may or may not coincide with the factual world (or its future extension[1]). For example, the epistemic modalizer *believe* in (11) creates a possible world in which the proposition *an elk ate the supplies* actualizes. The lexical semantics of *believe* further signal that this world is likely to coincide with the factual world, as evaluated by the subject of *believe*.

Defining the likelihood of coincidence of worlds that epistemic modalizers establish, Declerck placed epistemic modalizers on a scale of *specified factuality values* (see table 3.1).

Highest on the scale we find factuality. Factuality is non-modal: it means that there is no doubt about the actualization of the proposition in the factual world. On the other extreme, we find counterfactuality. Modalizers denoting counterfactuality signal that the evoked possible world in which the

[1]Declerck noted that reference to the future of the factual world, or, in his words, *not-yet-factuality*, has received relatively little attention in the literature. In the current study, we only used sentences that refer to present and past states of affairs. Here, I will just consider the future extension of the factual world as part of the factual world, and disregard the notion of not-yet-factuality.

	factuality	*John is ill.*
relative factuality values	strong necessity	[*John is not here.*] *He must be ill.*
	weak necessity	[*The money is not in the till, so*] *it should already be in the safe.*
	probability	*John may well be ill.*
	possibility	*John may be ill./ It might be true.*
	improbability	*It should not be difficult to find his address.*
	impossibility	*It cannot be true./ You cannot be serious!*
	counterfactuality	[*If they would have been rich*] *they would have bought a bar.*

Table 3.1: Epistemic scale of specified factuality values. Modified from Declerck (2011, p. 36). Example sentences partly from Declerck.

proposition actualizes necessarily does not coincide with the factual world. Consequently, the proposition in the scope of the modalizer is necessarily false of (i.e., not actualizing in) the factual world. Between factuality and counterfactuality we find *relative factuality values*. All relative factuality values signal that the evoked possible world may or may not coincide with the factual world, but they differ in degree of likelihood they assign to whether the worlds coincide, from '99,9% certain that they coincide' to '99,9% certain that they do not coincide'. This is a continuous scale: the relative factuality categories as distinguished in Declerck's table are arbitrary cutoff points.

3.2.1 *Å vite*, *å tro* and *å tvile på*

The Norwegian verbs investigated in the current study, *å tro* and *å tvile på*, translatable to *to believe* and *to doubt*, are both epistemic modal verbs that assign relative factuality values to their complements, since they both create a possible world that may or may not coincide with the factual world. *Tro* can be located higher on the scale than *tvile på*: it signals that its Agent finds it likely that its complement actualizes in the factual world. The Agent of *tvile på*, on the other hand, will not find it likely that the complement actualizes in the factual world. Although both verbs have a range of different meanings in different contexts, it is safe to say that all meanings of *tro* can be placed above the neutral mid-line of the scale (*possibility* in table 3.1), and that the meanings of *tvile på* reside below the mid-line. However, as our Norwegian informants pointed out, *tro* and *tvile på* are not each other's

exact counterparts, neatly mirrored along the mid-line axis. *Tro* is very frequently used to express near-certainty, and even as a polite alternative for factives like *å vite* (to know). In this sense, *tro* is comparable to the English *think*, which, in many cases, would actually be a better translation for *tro* than *believe*. As such, *tro* spans a wider range on the epistemic scale than *tvile på*, and has its most frequent use further away from the mid-line. As I will suggest in the Discussion (chapter 6), this asymmetry might have had implications for the current study.

The control condition, *vite*, is a non-modal verb expressing the value *factuality*: when a speaker says that he knows something, he expresses that he is certain that a proposition actualizes in the factual world.

3.3 Differentiating speaker and evaluator

So far, I have treated the factual world as if all information about the factual world is accessible to speakers, so that they can directly compare possible worlds with the objective factual world. This is, of course, a simplification; before Galileo realized that the earth orbits around the sun, it was not considered a fact of the factual world. Every speaker has an internal model of the world, which contains everything that he assumes to be real, based on his perception and experience: everything that he would classify as 'knowledge'. It is this continuously updated and altered *subjective* possible world[2], that a speaker relates other possible worlds to. Although the content of individuals' subjective worlds overlap to a very large extent, each individual's subjective world is slightly different. Speakers are aware of this. The discourse model by Verhagen (2005) captures this awareness by assuming an *intersubjective* common ground that speakers make use of when they engage in conversation. This common ground contains the part of the speaker's subjective world that he believes to overlap with the addressee's subjective world: knowledge that he assumes to be mutual.

In this study, we have used propositions of which the factuality status is stable and widely shared. They are either obviously true/typical or obviously false/atypical (e.g., *birds have wings* vs. *birds have gills*). However, as I will briefly discuss in section 3.5, the distinction between objective and subjective possible worlds is useful to make with regard to the pragmatics involved in interpreting the stimuli of this study.

[2]The terms *subjective* and *intersubjective* are defined as in Verhagen (2005).

3.3.1 Third person epistemic modal verbs

When using epistemic modal verbs in the third person, the speaker is no longer the Agent of the verb, resulting in a situation with two distinct subjective factual worlds: the speaker's (Mary in (12)) and the Agent's (Peter in (12)).

(12) a. Mary: «Peter believes that John owns a bar.»
b. Mary: «Peter mistakenly believes that John owns a bar.»

In their default use, epistemic modal verbs denote a relation as evaluated by their Agents, not necessarily the speaker. This means that the subjective factive world of the speaker is not relevant for a semantic representation of sentences like (12a). As visualized in figure 3.1, we cannot deduce from sentence (12a) whether the proposition *John owns a bar* actualizes in Mary's subjective factive world: the epistemic relation between the evoked possible world and Mary's subjective factive world is not specified. Sentence (12b)

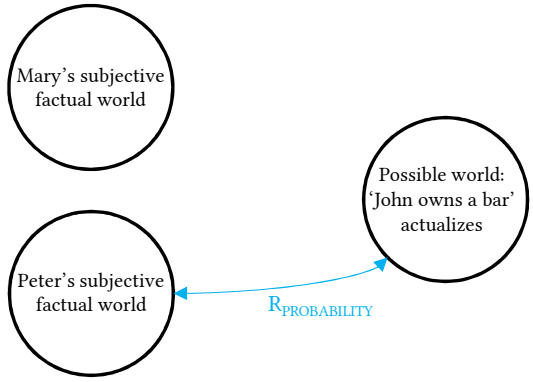

Figure 3.1: Visualization of sentence (12a), where R stands for the epistemic relation that is established by *believes*. Types of relations in this and subsequent figures in accordance with Declerck's tentative categories (see table 3.1, p. 30).

proves that we can indeed separately specify the relation between the evoked possible world and Mary's subjective factual world, for example through the adverb *mistakenly*, without changing the relation that *believes* establishes (figure 3.2). The verb *doubt* works in the same way: it establishes a specified epistemic relation between the same possible worlds (figure 3.3).

Applying a similar analysis to the verb *to know* reveals how factive verbs differ from epistemic modal verbs. Compare the following sentence pair with the pair in (12):

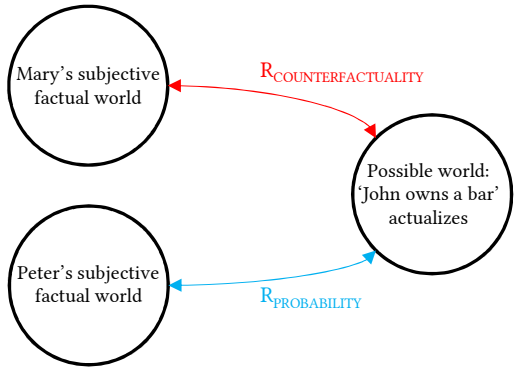

Figure 3.2: Visualization of the epistemic relations R established by *mistakenly* and *doubts* in sentence (12b).

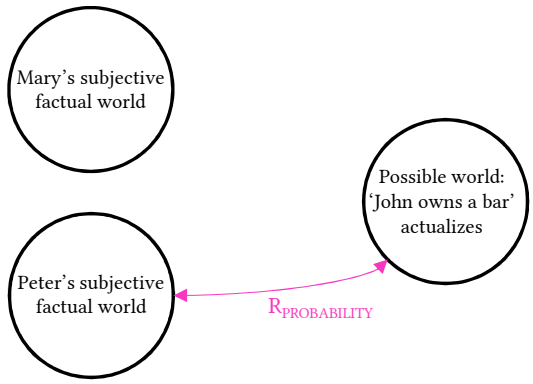

Figure 3.3: Visualization of the epistemic relation R established by *doubts*.

(13) a. Mary: «Peter knows that John owns a bar.»

b. # Mary: «Peter mistakenly knows that John owns a bar.»

The lexical semantics of *know* make (13b) incomprehensible. *Know* establishes an epistemic relation of factivity to both the subjective factual world of the Agent and of the speaker, as displayed in figure 3.4. As such, *know* is a suitable control condition in this study: just like non-embedded, non-modal, indicative sentences, such as the ones used in the study by Urbach and Kutas (2010) and Urbach et al. (2015), sentences with *know* signal factivity as evaluated by the speaker.

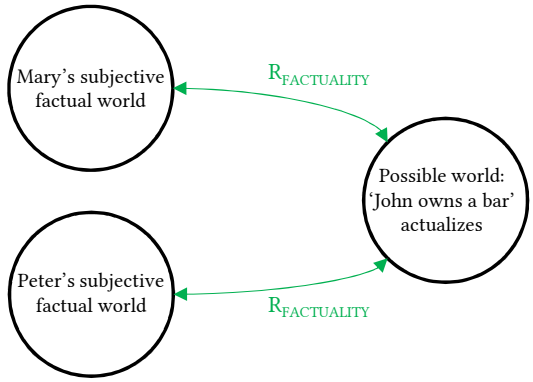

Figure 3.4: Visualization of the epistemic relation R established by *knows* in sentence (13a).

3.4 Qualitative differences between *believe* and *doubt*

In the previous section, I have described *believe* and *doubt* as quantitatively distinct: I positioned them on the same scale, and attributed different quantificational values to them. However, a more formal semantic approach exhibits also qualitative differences between *believe* and *doubt*. In anticipation of the results of the current study, which show effects of different latency and distribution for *doubt* as compared to the other main verbs, I will now briefly discuss qualitative differences between *doubt* on the one hand and *believe*, *know*, and the verbs in Lia Calinescu's counterpart to this study on the other.

In his typology of attitude-reporting verbs, Asher (1987) classified attitude expressions according to a number of distinctions: factive/non-factive, positive/negative, and definite/indefinite. The factive/non-factive distinction is comparable to my modal/non-modal distinction, and sets *know* apart from *believe* and *doubt*. The positive/negative distinction has to do with the presence of a negation operator: negative attitude verbs, such as *doubt*, bring about a negation operator. The last distinction, definite/indefinite, concerns the necessity of "ancillary attitudes" that are already present in the discourse context (Asher, 1987, p. 130) . Without a basis of justification made up by such presupposed attitudes, definite verbs are infelicitous; indefinite verbs are not. Asher classified *believe* as indefinite, and *doubt* as both definite and indefinite: *doubt* is sometimes indefinite, but "one can make a case that *doubt* [...] expresses at least on many occasions definite attitudes demanding a background justification." (p. 130)

That *doubt* is a negative verb is most clearly shown by the fact that it is a trigger for negative polarity items (NPIs):

(14) a. Mary doubts that John will **ever** learn how to make a decent Tequila Sunrise.

b. * Mary believes that John will **ever** learn how to make a decent Tequila Sunrise.

Also with regard to anaphoric relations, *believe* and *doubt* behave differently. Whereas the indefinite NP *a bar* (interpreted with narrow scope) is accessible for anaphoric reference when embedded in a report of belief (15a), it is not when embedded in a report of doubt (15b):

(15) a. * Peter doubts that John bought a bar. Mary hopes that **it** has copper taps.

b. Peter believes that John bought a bar. Mary hopes that **it** has copper taps.

Asher (1987) suggested that negative attitude verbs such as *doubt* activate a negation operator, and that this is the element that blocks the establishment of a relation between the anaphor *it* and the antecedent *a bar* in (15b).

Containing a negation operator, triggers of NPIs such as *doubt* have been proposed to evoke, at some level, both an affirmative and a negative version of the expressed proposition. In his 'mental model' framework, Verhagen (2005) observed that triggers for NPIs can indeed cause both the affirmative and negating version of a proposition (*mental space*, Verhagen) to be 'in the air', and illustrated this point with the following fragment:

(16) This time, there was no such communication [about the plans]. **It**'s a pity because **it** could have resulted in greater participation by employers.

From Verhagen (2005, p. 29).

Verhagen pointed out that the first *it* refers to the negative mental space, whereas the second *it* refers to the affirmative mental space: apparently, both are available.

Asher's observation that *doubt* can be definite has to do with the fact that it is a negative verb and thus involves both a negative and an affirmative mental space. In order for a definite attitude verb to be felicitous, the affirmative mental space must be retroactively added to the common ground if it is not already present in the preceding discourse (*accommodation*, Asher (1987, p. 130).

3.5 A word on pragmatics

Although the sentences in this study are presented without any discourse context, it is still useful to consider the possible influence of pragmatics. Following only the semantic definitions, we hypothesize that comprehenders will expect a true or typical proposition as the complement of *believe*, and a false or atypical proposition as the complement of *doubt*. However, the forces of pragmatics seem to work in the opposite direction. One of the main ideas of relevance theory (Sperber and Wilson, 1986) is that comprehenders assume contributions to discourse to be maximally relevant. They expect speakers to further the conversation, for example by altering or specifying the common ground. From a relevance-theoretic point of view, reporting that a third person's beliefs are as one would expect is not very informative *an sich*, just like specifying that some person has two legs, and lives in a house: as comprehenders, we constantly presuppose everything that is not specified to be in accordance with some default expectation. Thus, when presented with a sentence fragment like *John believes that...*, a comprehender might expect a complement denoting some proposition that is not trivially true, but something that alters his subjective factual world. Conversely, *John doubts that...* might not be expected to be followed by something that is trivially false.

Since the sentences in this study are presented without discourse context, and the 'speakers' of the sentences are very distant and completely implicit, the force of pragmatics might be minimal. Nevertheless, pragmatics might attenuate the hypothesized differences in expectation between sentences with *believe* and sentences with *doubt*.

3.6 Summary

In this chapter, I have defined the verbs under investigation as follows:

- *Know* signals a relation of factuality between the evoked possible world and the factual world: it expresses certainty that its complement actualizes in the factual world. This evaluation is presented as shared by the speaker and the Agent of the verb.

- *Believe* signals that the evoked possible world may of may not coincide with the factual world, and specifies that actualization is more probable than non-actualization, as evaluated by the Agent of the verb.

- *Doubt* signals that the evoked possible world may of may not coincide

with the factual world, and specifies that non-actualization is more probable than actualization, as evaluated by the Agent of the verb.

- *Doubt* is a negative verb, evoking an affirmative and a negating mental space. *Believe* is positive, and evokes only an affirmative mental space.

Chapter 4

Methods

4.1 Design and aims of the present study

In this project, we investigated the processing of intensional verbs during on-line processing.

The research question of the present study is:

- Are the lexical semantics of epistemic modal verbs processed fully and immediately during on-line processing?

We embedded sentences that denoted typical or atypical propositions in matrix clauses consisting of a proper name as subject, and a main verb (*know, believe, doubt*) in the present tense (see table 4.1).

Following the definitions of the verbs as given in section 3.6, we assume that the plausibility of the embedded sentences is modulated by the epistemic modal verbs. The non-modal verb *to know* sets the reader up to expect an embedded proposition that is intersubjectively evaluated as factive. Modal verbs can be combined with both typical/true and atypical/false propositions, but, adhering to Declerck's epistemic scale of factuality values (table 3.1, p. 30), we assume that in a neutral discourse situation with a 'neutral' speaker, a typical proposition is more likely to occur after *believe*, and an atypical proposition is more likely to occur after *doubt*.

In order to test whether these expectations evoked by the modal verbs are immediately available during on-line processing, we measured the ERPs elicited by critical words that made the embedded propositions typical or atypical. Since the N400 is sensitive to plausibility and predictability, we hypothesize that the modulation of plausibility by the modal verbs affects the N400 amplitude, if they are processed and integrated immediately and fully:

Matrix verb	**Typicality**	**RSVP sentence**
know	Typical	Tora vet at fugler har vinger
(control cond.)		*Tora knows that birds have wings*
know	Atypical	Tora vet at fugler har gjeller
(control cond.)		*Tora knows that birds have gills*
believe	Typical	Tora tror at fugler har vinger
		Tora believes that birds have wings
believe	Atypical	Tora tror at fugler har gjeller
		Tora believes that birds have gills
doubt	Typical	Tora tviler på at fugler har vinger
		Tora doubts that birds have wings
doubt	Atypical	Tora tviler på at fugler har gjeller
		Tora doubts that birds have gills
dream	Typical	Tora drømmer at fugler har vinger
		Tora dreams that birds have wings
dream	Atypical	Tora drømmer at fugler har gjeller
		Tora dreams that birds have gills
imagine	Typical	Tora innbiller seg at fugler har vinger
		Tora imagines that birds have wings
imagine	Atypical	Tora innbiller seg at fugler har gjeller
		Tora imagines that birds have gills

Table 4.1: Example stimuli of the control condition and the experimental conditions. The conditions with the verbs *dream* and *imagine* were used in Lia Calinescu's thesis project.

- H_0: The lexical semantics of the modal verbs *believe* and *doubt* are not processed fully at the moment the critical word is presented, and do not influence the processes reflected in the N400 elicited by the critical word. The N400 typicality effect should not differ depending on the matrix verb.

- H_1:The lexical semantics of the modal verbs *believe* and *doubt* are processed fully at the moment the critical word is presented, and influence the processes reflected in the N400 elicited by the critical word. The N400 typicality effect should be modulated by the matrix verbs in the following way:

  - For the verb *know*: the amplitude of the N400 elicited by the atypical word is larger than the amplitude of the N400 elicited by the typical word.
  - For the verb *believe*: the amplitude of the N400 elicited by the atypical word is larger than the amplitude of the N400 elicited by the typical word, but the effect is reduced as compared to *know*.
  - For the verb *doubt*: the typicality effect is even further reduced, disappears, or is even reversed. Since factors that are not plausibility-related also affect the N400, most importantly lexico-semantic association, the plausibility reversal of *doubt* might not result in a reversal of the N400 effect under H_1.

The predictions under H_1 are visualized in figure 4.1.

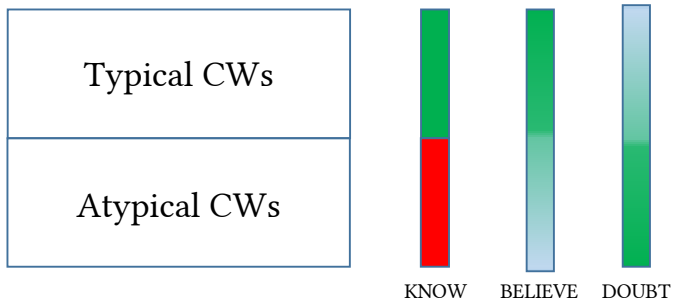

Figure 4.1: Hypothesized plausibility of typical and atypical complements for each main verb (green denotes 'plausible', red 'implausible') under H_1.

4.2 Off-line data: cloze probability test

In order to compare the on-line EEG data with off-line expectation, we are currently running a cloze probability test. Since the amount of data needed is quite extensive (1500 sentences x 20 answers in total for this study and Lia Calinescu's study), we will not finish the data collection within the scope of this thesis project. The results of this cloze probability data will eventually enable us to carry out a number of comparisons between on-line and off-line behavior. Specifically, we will be able to investigate whether the main verbs modulated cloze probability to different degrees for possible subcategories in our stimuli (see section 6.2.1 in the Discussion). Furthermore, the preliminary results have helped us while making final adjustments to the stimuli.

We are running the cloze probability test as an on-line experiment with the software of Ibex Farm, and made it available through an on-line link. The experiment is hosted on the Ibex Farm platform. We present the participants of the cloze probability test with a random set of 75 of the experimental sentences, truncated before the critical word, and ask them to complete the sentence with a word that comes to mind as a plausible continuation of the sentence. The test in this form takes 10-15 minutes. None of the participants of the cloze probability test were a subject in the EEG study.

At the moment of writing, the cloze probability data for this part of the study are close to complete (845/900 sentences). Appendix B displays the preliminary results of this partial data set. For incomplete sentences (i,e, less than 20 answers), the cloze probability is calculated based on the available results. This means that the data of the incomplete items are less reliable: the number of answers for the incomplete 55 sentences ranges from 7 to 19 (average: 15.4). The distribution of the preliminary cloze probabilities shows expected tendencies when taking into account the definitions of the epistemic modal verbs: the cloze probability of the typical critical words is highest for the *know*-sentences (0,346), slightly lower in the *believe*-sentences (0,273), and even lower in the *doubt*-sentences (0,192).

4.3 Materials

The stimuli were constructed in Norwegian (Bokmål) in collaboration with five native speakers, who all were students or employees in the Linguistics department of NTNU. They helped us construct sentences that sounded natural, and they functioned as informants by informally assessing the typicality of the critical words, and suggesting alternatives for critical words or

formulations that they perceived as not particularly typical or atypical. For part of the sentences, we drew inspiration from the stimuli used by Urbach and Kutas (2010), Urbach et al. (2015), and Nieuwland (2013).

The experimental stimuli consisted of 300 pairs of subordinated sentences, denoting typical or true propositions in one version and atypical or false propositions in the other version. The subordinated sentences all started with a subject NP and a verb. The subject was either a generic plural noun or a well-known unique entity (*Beyoncé, the Constitution*). The critical word was in most cases the direct object of a transitive verb or the complement of a particle verb. In ten percent of the sentences, the critical word was in an adverbial phrase specifying location (e.g. *boxers fight in the ring/cockpit*) or instrument (e.g., *fencers duel with swords/teaspoons*). In one sentence, the critical word was preceded by an adjective (*spies have fake identities/artworks*), and in another sentence by a quantifier (*laws apply to all citizens/ghosts*). As displayed in table 4.1, each sentence had ten versions, of which six are relevant for this thesis: the conditions with the verbs *dream* and *imagine* were used in Lia Calinescu's related thesis project. Appendix A shows all experimental sentences and English translations.

We created ten lists, each consisting of 250 fillers and 300 of the experimental sentences, such that each list contained one version of each experimental sentence. Thus, each participant read 150 typical and 150 atypical experimental sentences (in both conditions 30 per main verb). We randomized the order of the lists, but made sure that fillers and experimental sentences were more or less evenly distributed throughout each list.

Spelling and morphological variation

The Norwegian spelling situation is quite unique. There are two official written versions of Norwegian, Bokmål and Nynorsk, and within Bokmål, many words have multiple official spellings and inflections. When there were multiple options, we took the most widely used spelling and form, based on numbers from the NoWaC corpus (see section 4.3.2), and as indicated by our informants.

4.3.1 Sentence-final position of the critical words: wrap-up effects

It seems to be common practice in the field of language ERP studies to go to great lengths to avoid the sentence-final position for critical words, in order to avoid wrap-up effects obscuring the ERPs elicited by the critical

words. Wrap-up effects reflect exclusively sentence- or clause-final syntactic or integrative processes. However, there is no consensus about what these processes might be, and, if they take place at all, how they manifest themselves. In a comprehensive review on (alleged) wrap-up effects, Stowe et al. (2018) reconstructed how a 'wrap-up dogma' came into existence, and showed that the evidence for general wrap-up effects[1] is not very strong, and can be reduced to observations in a minimal number of studies, that are, moreover, not free of possible confounds. Stowe et al. argued that the little evidence there is for a special status of the sentence-final position regarding processing is not sufficient to avoid this position at all costs.

Like the example sentences in table 4.1, most of our critical words occurred sentence-finally. In 61 sentence pairs, the critical word was followed by a phrase between one and six words long (median = 2, mode = 2). The continuations were neutral additions, that were meant not to change the overall plausibility of the sentence. Most of the continuations specified locations, times, instruments, or purposes (e.g., *sailing boats need wind/encouragement in order to move*). The purpose of the continuations was to introduce some variation in sentence length, which reduced (together with the variable sentence length of the fillers) the predictability of sentence length and thereby the risk of wrap-up effects. As another measure, we did not display a full stop after the final word of the sentence.

We did not add continuations to all sentences because there is no *specific* reason to expect confounding sentence-final effects, beyond the general nervousness that surrounds using the sentence-final position. Our experiment did not contain a decision task, and the experimental stimuli are grammatically well-formed. Adding continuations to all sentences would have made the experiment much longer. This would have meant mean pushing the required attention span of participants, which was, in the experiment's current version, already long for concentrated reading. Furthermore, not all sentences could be provided with natural-sounding continuations that would not affect the overall typicality of the sentence.

4.3.2 Critical words

Every subordinated sentence had a version with a typical critical word, resulting in a proposition that was typical and verifiable through world knowledge, and a version with an atypical critical word, resulting in a false or atypical proposition. Typicality was in the first place assessed by the ex-

[1] i.e., not related to tasks or specific types of syntactic violations; these factors are more robustly associated with sentence-final ERP differences.

perimenters and the five informants. In addition, preliminary results from the cloze probability test (see section 4.2) were used. Since only ca. half of the cloze probability data were collected when we constructed the stimuli, we only used them as an indication. When our initial typical critical word (or a synonym) did not show up in the results in the control condition, or when another critical word was overwhelmingly more frequent[2], we considered changing it. In the stimuli, each critical word appeared only once as a critical word.

No critical word or word in the preceding sentence exceeded a length of twelve characters, so that all words could be read without eye movements (means: typical words 6.5; atypical words 6.8). In the sentence parts following the critical words, some words were longer than twelve characters.

Furthermore, the critical words were controlled for a number of factors that are known to affect the N400. We avoided alliterations between the critical words and the preceding word, and between the typical and atypical word within each sentence pair. We also made sure that there were no words with the same stem in the preceding context, and we approximately matched the typical and atypical critical words for frequency.

Frequency

The frequency of the critical words was assessed using the NoWaC, a web-based Bokmål corpus developed by the Text Laboratory at the University of Oslo (see Guevara (2010)). This corpus contains a large collection of texts taken from websites in the .no domain, selected, filtered and POS-tagged automatically. The advantage of the NoWaC corpus is its size (700 million tokens) and the variety of texts it contains, including very informal texts from forums and comment sections. A disadvantage of this corpus is the inconsistent accuracy of the POS-tagging. Although the Text Laboratory measured a 96,5% overall accuracy of the automatic tagger (A. Nøklestad, personal communication, September 24, 2017), we noticed that there is quite some variability in its accuracy across lemmas. Searches in the corpus for nouns with homonymic forms contained large amounts of misclassified occurrences in some cases. To illustrate this with one of the more extreme cases: the results for the 'lemma search' (includes all inflections) for the noun *gir* ('gear') consisted for 91% of misclassified homonymic forms (*gir* as the indicative 'give', *gira* as the adjective 'excited').

[2]At the time we finalized the stimuli, some sentences already had all or almost all answers needed; others only had a few. We took into account the results for sentences which were complete or almost complete and that had one completion with a high cloze probability.

Lemma search *gir*(noun)	41414
Form search *gir*(noun)+*gira*(noun)	40974
Misclassified occurrences in sample (n=100)	91
Estimated number of misclassified non-target forms	0.91*40974 = 37286.34
Form search *gir*(verb)+*gira*(verb)	306350
Misclassified occurrences in sample (n=100)	0
Estimated number of misclassified target forms	0*306350 = 0
Corrected lemma frequency	41414-37286.34+0 = **4127.66**

Table 4.2: Example of the frequency correction for misclassified occurrences in the NoWaC corpus.

We corrected for the misclassifications in the following way: for each lemma with homonymic forms (about 20% of all critical words), we took a random sample of 100 occurrences from the search results of the homonymic form classified as the target form (e.g., *gir* classified as a noun), and counted the number of misclassifications. We subtracted the corresponding percentage of hits from the total number of hits for the form search. Then, we searched the homonymic form tagged as the non-target form (e.g., *gir* as a verb and *gira* as a verb[3]). We counted the number of misclassified target forms (the noun *gir*), and added the corresponding percentage of hits to the overall frequency of the lemma. An example of a calculation is displayed in table 4.2.

For words that were full homonyms (having different meanings but the same inflection, making all morphological forms of the word homonymous), we used a similar method: we counted the distribution of meanings in a random sample of 100, and adjusted the frequency accordingly. We considered a word to be a full homonym when its meanings occurred as separate lemmas in the on-line *Bokmålsordboka* of the University of Bergen and Språkrådet.

We encountered two other flaws in the NoWaC corpus. To begin with, the corpus contains some English, Nynorsk, and Danish texts, due to inaccuracy of the language identification filter that was used to construct the corpus. Furthermore, some texts occur multiple times. I suspect that the repeated texts come from threads on forums, where it is common for users to quote the text they are reacting on. Within the scope of this project, we were

[3]The word *gir* has in fact even more complications, since there is also the masculine word *gir* (strong desire), which makes also the form *girene* homonymous. However, none of the 128 occurrences of *girene* in the corpus was of this type. For clarity, it is left out in the calculation displayed in table 4.2. The same goes for *gir* as the imperative of the verb *å gire*.

not able to correct for these shortcomings of the corpus, and accepted them as factors that render the frequency data less reliable than they could have been. A more systematic investigation of the magnitude of NoWaC's flaws would be of great value for future research in Norwegian linguistics.

Since the frequencies of the critical words are not normally distributed (Shapiro-Wilk normality test $p < 2.2^{e-16}$), we assessed the significance of the distributional differences between typical critical words (mean=30847.44) and atypical critical words (mean=20707.63) with the non-parametric Wilcoxon rank sum test (non-matched, two-sided). On the basis of this test (W = 43678, p-value=0.5337), since the differences between typical and atypical critical words is not significant, we conclude that the groups are roughly matched.

4.3.3 Fillers

Just as the experimental sentences, the 250 filler sentences consisted of a main clause with a name and a matrix verb followed by a subordinated clause. We introduced some further variability in sentence length, which ranged from 5 to 16 words (average=7,9, median=8, mode=7). We used different attitude verbs (e.g., *hope*), intensional verbs (e.g., *claim*) and factive verbs (e.g., *forget*). About 45 % of the fillers contained subordinated sentences that expressed states of affairs that were not verifiable through general world knowledge, but that expressed some assertion about a specific fictive person or object (e.g., *Maria hopes that the secretary on the second floor loves her*) or a subjective assertion (e.g., *Tor claims that graffiti makes the city more beautiful*). 28 filler sentences contained syntactic violations in the form of word order violations or morphosyntactic violations.

4.4 Participants

We recruited 41 volunteers (18 male, mean age 24,2 years, range 19-41) at and around NTNU Trondheim, who received a cinema giftcard for their participation. They all had Norwegian as their only native language, and Bokmål as their preferred written language. All participants were right-handed, and all but one[4] had normal or corrected-to-normal vision. They reported no history of neurocognitive impairment, and were not using any medicines regularly or in the period prior to the experiment. Before the

[4]One participant reported less vision in the right eye due to retinal surgery. His nearby vision (incl. the distance to the screen) was, however, normal. This particular participant was one of the 12 participants included in the data analysis.

experiment, participants read a document describing the procedure, and signed an informed consent form.

Given the artifact detection and rejection criteria we adopted, the data quality of only twelve participants were good enough to include in the analysis. The mean age of this group was 25,5 (range 21-41), and ten of them were male. Eight of the ten versions of the experiment were done by one or two participants; for two versions, all participants were excluded, so that no data from the sentence versions in those lists were included in the analysis.

4.5 Procedure

The participants were seated in front of a screen (distance ca. 90-110 cm) in a normally lit, electrically shielded and noise reducing room. We ran the experiment using the software Presentation® (Neurobehavioral Systems). The participants were instructed to read the sentences for comprehension. The sentences were presented word-by-word (RSVP) in the middle of the screen, in a white font against a black background (SOA: 600 ms, word duration: 300 ms). Between the sentences, a fixation cross appeared in the middle of a screen for 2000 ms. The sentences were presented in blocks of 15 sentences (37 blocks in total). After each block, the word 'Break' appeared on the screen. The participants could take as long as they wanted, and signaled to the experimenters when they were ready to continue. The whole experiment without breaks would take 1:06 hour. Including instructions and breaks, most participants spent between 1:15 and 1:30 hour in front of the screen. After the experiments, the participants were briefly interviewed about the experience, in order to form an impression about how concentrated they had been during the experiment.

4.6 EEG recording and analysis

We recorded continuous scalp ERPs from 32 electrodes, placed according to the Extended 10/20-System in the Easycap 32 Channel Standard EEG Recording Cap (see figure 4.2). A common reference electrode was positioned on the mid-line between the Cz and Fz electrodes. The signal was amplified using a low pass filter with a 1000 Hz cutoff point and a high pass filter at 0,1 Hz, and digitized with a sampling rate of 1000 Hz.

Epochs spanning from 200 ms before to 800 ms after onset of the critical word were extracted from the EEG. For these segments, a baseline correction was done using the average of the interval 200-0 ms prior to stimulus onset. Artifact rejection in the extracted segments was based on two functions of the

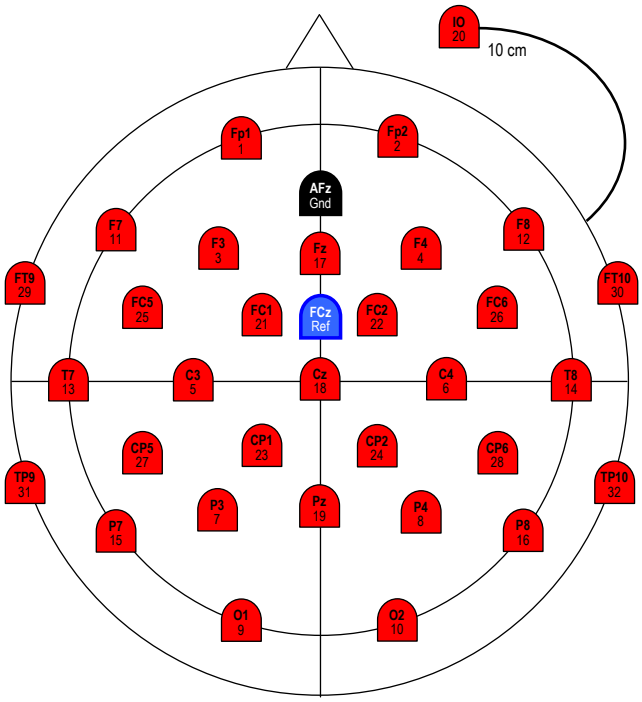

Figure 4.2: Electrode layout of the Easycap 32 Channel Standard EEG Recording Cap. Source: BrainProducts GmbH. Retrieved 1-5-2018 from *http://www.brainproducts.com/filedownload.php?path=products/brochures_material/TC32.pdf.*

MATLAB toolbox FieldTrip. Firstly, all trials where one or more amplitude values exceeded a threshold of ± 150 μv from the baseline were detected and rejected. Secondly, trials containing eye blink and eye movement artifacts were detected and rejected based on the data from the Fp1- and Fp2-channel: z-transformed values from these channels were pre-processed with a band-pass filter of 1-15 Hz. After artifact rejection, the cleaned segments were filtered with a digital low-pass filter at 30 Hz, in order to filter out muscle artifacts.

4.6.1 Statistical analyses

For the statistical analysis, we used the nonparametric statistical test procedure for multiple-subject EEG and MEG data described by Maris and Oostenveld (2007). As input for the test, we took the mean amplitudes in the time windows 300-500 ms after onset and 500-800 ms after onset, for

each channel averaged per subject per condition. For each sample (pairs with corresponding time window and channel), a t-test was done for the typicality conditions, producing a t-value. All samples for which the t-test yielded a t-value with a corresponding p-value smaller than 0.05 were identified, and clustered with adjacent samples for which $p<0.05$, if any. For each cluster (min. 1, max. 32 channels), the t-values were summed. A p-value of this cluster-level test statistic was estimated by comparing it to the distribution of the cluster-level test statistics for 1000 Monte Carlo simulations with the paired samples. Additionally, a 2 (Typicality) x 5 (Main verb) x 32 (Electrode) ANOVA was conducted.

Chapter 5

Results

Subjects with an average of at least 20 accepted trials per condition after filtering and artifact rejection were included in the data analysis. Unfortunately, the majority of subjects was excluded due to too many artifacts; they had either moved too much, or were not focused, and showed a lot of alpha activity. Additionally, a number of subjects was excluded from the analysis because there was too much other physiological noise in multiple channels. This left us with 12 subjects with sufficient data for the analysis. The consequence is that this study is likely underpowered in its current form; the results reported here should be taken as indications of effects, which deserve further data collection and testing, in order to see whether the trends observed in this small sample are robust. In the group of included participants, on average 9 percent of the trials were rejected per condition.

The ERP waveforms elicited by the critical words show an N1-P2 complex in the first 300 ms from onset, followed by a negative waveform in the N400 window. Figures 5.1, 5.2 and 5.3 show the grand averages per verb. A summary of the clusters resulting from the nonparametric test is displayed in table 5.1.

5.1 Time window 300-500 ms

In the *know*-conditions, a weak central-parietal typicality effect is visible in the time window 300-500 ms, with more negative-going waves for the atypical condition. The cluster-level statistics yield a p-value of 0.065, just above the 0.05 significance threshold. We interpret this effect as the expected N400 effect. For the *believe*-sentences, we found a significant and more widespread N400 effect in the same direction, spanning 5 electrodes (p=0.003). In contrast, the *doubt*-condition yielded no central-parietal typicality effects.

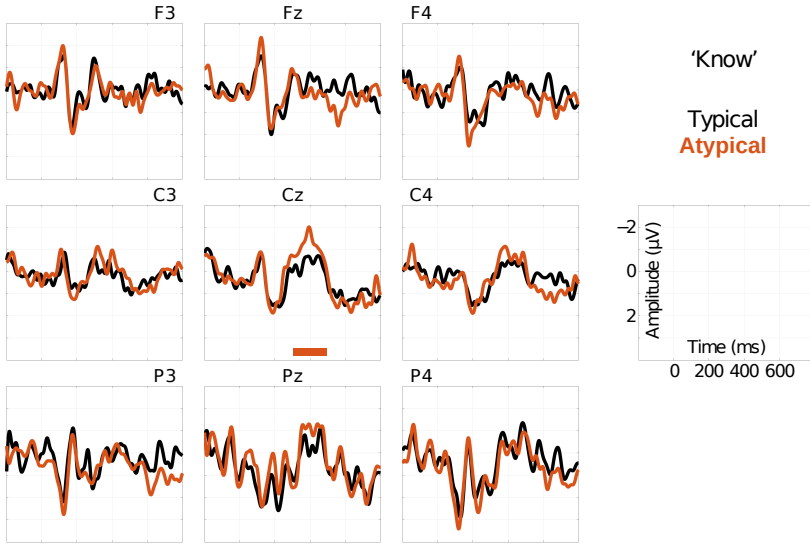

Figure 5.1: Grand average for *know*. Negative plotted upwards.

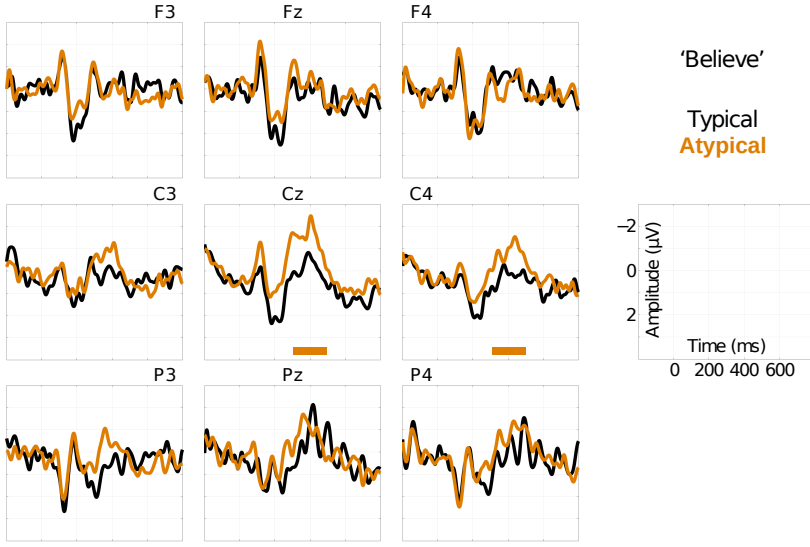

Figure 5.2: Grand average for *believe*. Negative plotted upwards.

Instead, two channels on the left frontal cortex showed a reversed typicality effect, with more negative-going waves for the typical condition (p=0.022).

The ANOVA for this time window showed a strong interaction effect of typicality and electrode (F=40.28, p=3.74^{e-15}).

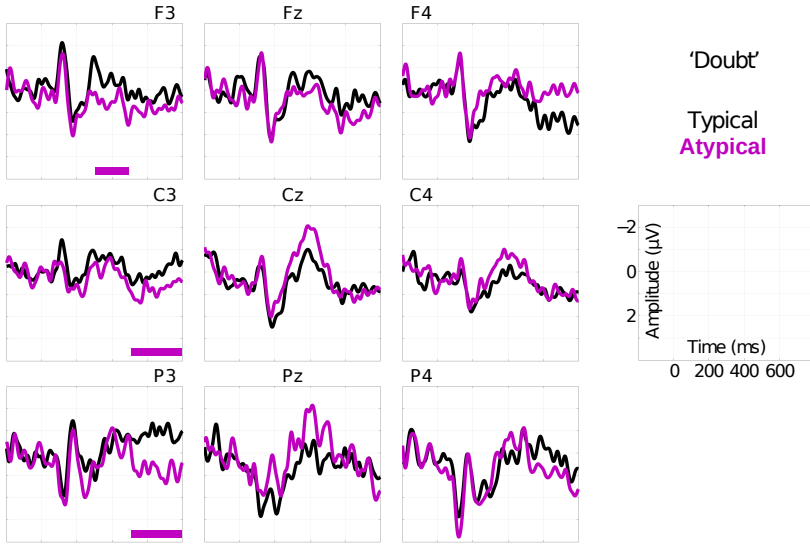

Figure 5.3: Grand average for *doubt*. Negative plotted upwards.

Main verb		300-500 ms	500-800 ms
Know	summed t-values	-4.7313	none
	p-value	0.0650	
	cluster size	2	
	channels	Cz, CP1	
Believe	summed t-values	-21.7182	none
	p-value	0.0030	
	size	5	
	channels	FC2, Cz, C4, CP1, CP2	
Doubt	summed t-values	8.6795	11.5607
	p-value	0.0220	0.0080
	size	2	4
	channels	F7, F3	C3, CP5, CP1, P3

Table 5.1: Overview of the spatially clustered samples yielding t-tests with $p<0.05$. The reported cluster-level p-values are Monte Carlo p-values based on 1000 draws.

5.2 Time window 500-800 ms

No typicality effects were found for *know* and *believe* in this time window. In the *doubt*-condition, a left-lateralized central-parietal negativity was observed for the typical condition as compared to the atypical condition ($p=0.008$).

Chapter 6

Discussion

Although the 3-way ANOVA only showed an interaction effect of typicality and electrode, which is visible as an overall N400 effect in some channels regardless of main verb, the nonparametric test, which is more sensitive to effects because of the incorporation of prior knowledge about the morphology of ERPs (Maris and Oostenveld, 2007, p. 187), shows a number of distinct differences between the main verb conditions. In this chapter, I will discuss the results per main verb.

6.1 *Know* as control condition

The *know*-condition showed a small N400 effect for typicality. The effect was only significant on two channels, and its cluster-level statistics have a p-value just above the 0.05 threshold. We will have to see whether the observed effect will be more pronounced and/or wide-spread with more data.

A factor that might have contributed to the relatively limited size of the typicality effect is the fact that for part of the stimuli, the atypical critical word was an antonym of, or shared very specific semantic categories with, the typical words or other words with a high cloze probability. Both these types of relations to a high cloze probability word have been shown to reduce the N400 amplitude (Kutas and Hillyard, 1984; Federmeier and Kutas, 1999).

A possible issue with using *know* as a control condition lies in the assumption that embedded propositions in the scope of *know* are 'neutral': that they create a discourse context that is interchangeable with non-embedded, non-modal (indicative) clauses with regard to the processes underlying the N400. As I discussed in chapter 3, the verb *know* is non-modal, i.e., it signals factuality of the embedded proposition as evaluated by the speaker, just like non-embedded non-modal sentences. Merely considering the lex-

ical semantics of *know*, we expected the N400 to behave exactly like in non-modal, non-embedded situations. However, pragmatics might influence these two non-modal types differentially, or with a different strength. The pragmatic forces described in section 3.5 for *believe* and *doubt* may also influence *know*-sentences: assuming maximal relevance, a comprehender might expect a speaker to report someone's knowledge of a fact of the world that is not assumed to be known by everyone. Such a pragmatic mechanism might have made the N400 amplitudes in the most trivial statements larger, possibly partly obscuring the typicality effect.

Arguably, the same pragmatic mechanisms are at work for non-embedded indicative non-modal sentences. However, the (magnitude of the) effect of pragmatics on reporting a non-embedded proposition and on reporting a third person's knowledge of a proposition is not necessarily equal in both cases: those are quite different speech acts. It would be useful to measure N400s in non-embedded versions of the stimuli of this study in order to establish the effect of embedding a proposition in a *know*-clause.

Nevertheless, the effect observed in this small data set indicates that the stimuli elicited an N400 typicality effect when embedded in a non-modal environment.

6.2 *Believe*

The N400 typicality effect observed in the *believe*-condition was larger than the effect observed for *know*, and it spanned more channels. The direction of the effect, more negative-going for atypical words than for typical words, was predicted, but its relative magnitude was not. Based on the compatibility of *believe* with both factual and non-factual propositions, an attenuation of the effect as compared to *know* was hypothesized. If the current results turn out to be robust with more data, they are not consistent with predictions following from either of the hypotheses in this study: assuming full and immediate incremental processing of the matrix verbs predicts an attenuated N400 typicality effect for *believe*, and assuming partial/delayed processing predicts no effect of modulating the matrix verb.

As pointed out in section 3.2.1, the Norwegian verb *tro* is very frequently used to express near-certainty in everyday speech. In this respect, it is not surprising that *tro* elicits a large N400 typicality effect. Additionally, as discussed above, the *know*-condition might not be as a neutral a baseline as its semantic definition suggests, and a direct comparison should, especially in light of the small number of participants in the current state of the study, be approached with caution.

This being said, I will, notwithstanding, in the next section discuss a property of the stimuli that may account for a larger N400 typicality effect for *believe* than for *doubt*. This property has to do with the distinction between 'true' and merely 'typical' propositions: the within-condition typicality variability.

6.2.1 Typicality variability

It is difficult to objectively assess or quantify the exact degree of typicality of propositions, or the extent to which propositions are considered common knowledge. When constructing the stimuli, we attempted to stay away from gray areas, and used propositions that we considered typical and atypical beyond doubt. Nevertheless, our conditions are clearly not monolithic: the critical words in each condition are not all equally typical or atypical. The typical critical words cover a range from extremely trivial (e.g., *turtles have shells*) or characteristic (*cats catch mice*) to propositions that are true but less trivial, such as *spinach contains iron*, or typical (sometimes even stereotypical) but not always true (*strategists form alliances*, *fortune tellers read coffee grounds*). Conversely, the atypical condition contains propositions that are downright false and impossible (*cobras spit white wine, soap removes memories*), but also propositions that are possible but atypical (*knights wear bikinis*) or perhaps not for everyone obviously false (*yoga improves your vision*).

It is the distinction between true/false and typical/atypical that might play a role here. Whereas true propositions denote facts, which are always true of the factual world for all denoted entities, typical propositions denote states and events that usually or often are the case, but not always for all denoted entities. It could be the case that the typical but not necessarily factual propositions are actually less plausible in the *know*-condition than in the *believe*-condition, since using the verb *know* explicitly draws attention to factuality, whereas *believe* does not, even though it might reside very close to *know* on Declerck's factuality scale.

However, as mentioned, it is tricky to objectively separate factuality and typicality; an absolute distinction involves questions about what properties define certain entities, such as the question whether a turtle is still a turtle if it does not have a shell. Easier, and perhaps more informative, would be to carry out a detailed plausibility rating test for the embedded propositions presented as non-embedded main clauses. If the resulting scale of such a test would correlate with the N400 amplitudes of the individual items in the *know*-condition, this would be an indication that the within-condition

typicality variability contributed to the observed difference in N400 effect size between *know* and *believe*. Assuming that this is the case, the results for *know* and *believe* could be compatible with the hypothesis that these verbs are processed fully and immediately.

6.3 *Doubt*

6.3.1 No classic N400 effect

In the *doubt*-condition, no N400 typicality effect was found on the central-parietal channels. As specified in my hypotheses, I consider the absence of a classical N400 effect in the *doubt*-condition as indication that (at least parts of) the meaning of this verb are processed before the critical word is reached. An even stronger indication would have been a reversal of the N400 effect, showing a larger amplitude for typical than for atypical words. There are, however, several factors that influence the N400 amplitudes of typical and atypical words in the opposite direction, which may explain why the semantics of *doubt* in this study do not cause a reversed N400 typicality effect, also under the assumption of full and immediate processing of the verb.

To begin with, there exists a strong lexical-semantic association between most typical critical words and the preceding context, while this is generally not the case for the atypical critical words. This forms a stable influence on the N400 amplitudes across verb conditions, attenuating the N400 for typical words. Whereas this enlarges the observed N400 effect for *know* and *believe*, it works against the influence of the semantics of *doubt*.

Furthermore, as discussed in section 3.5, and as illustrated by many of the studies described in section 2.3.2, pragmatics may attenuate the results for both *believe* and *doubt* in the isolated sentences in this study: we expect utterances to be maximally relevant, and doubts and beliefs that differ from people's general doubts and beliefs are more informative pieces of information about a third person than expected doubts and beliefs. At this point, a comparison with the results of the quantifier studies by Urbach and Kutas (2010) and Urbach et al. (2015) (see section 2.3.2) is useful. Regarding modulation of plausibility and expectedness of typical and atypical propositions, our *believe*- and *doubt*-conditions could be compared to respectively the *most*- and *few*-conditions from these quantifier studies. Similar to our results, when presented without discourse context, both quantifier studies reported an N400 typicality effect in the expected direction only for the *most*-type quantifiers. When preceded by a licensing discourse context,

Urbach et al. (2015) observed the reversed N400 typicality effect for the *few*-type quantifiers that they expected on the basis of the off-line data (cloze probability and plausibility tests). In further research, it would be interesting to investigate whether a similar full cross-over effect can be elicited by the stimuli of this study as well, when provided with discourse context.

As I mentioned in my discussion of the quantifier studies, another possible reason that an N400 typicality effect is not always visible in plausibility- or predictability-reversing conditions such as our *doubt*-condition, is a floor performance: although *doubt* might give the reader a clue about what *not* to expect, namely something typical or clearly true, it only marginally improves the predictability of a specific atypical word or semantic field, if there is no help from a discourse context. The fact that discourse context was a decisive element for eliciting a reversed N400 effect in Urbach et al. (2015) might also be related to this point.

Typicality variability: possible consequences for the *doubt*-condition

Another factor that may have influenced the N400 amplitudes for *doubt* is the typicality variability described in section 6.2.1. It has possibly affected the results of the *doubt*-condition more than the results of the *believe*-condition. Whereas a dominant use of *tro* expresses near-certainty, making it very compatible with both factual and typical complements, the uses of *tvile på* may be closer to the mid-line of Declerck's scale of factuality values (see section 3.2.1). Consequently, it might be just as odd or unexpected to doubt something that is obviously false as it is to doubt something that is obviously true. It could be the case that the most obviously false atypical stimuli were less plausible in the *doubt*-condition, decreasing the average plausibility of the critical words in the atypical condition, and thereby increasing the N400 amplitude. A rough sketch of how different degrees of typicality may have affected plausibility per condition is displayed in figure 6.1: this picture is much more complicated than our initial hypothesis, which assumed internally uniform typical and atypical categories (see figure 4.1 on p. 41). Lacking any measure of typicality of the individual stimuli and anything more than a general idea about the precise range and frequency of the uses of *doubt* as compared to *believe*, this figure remains a speculation. It does, however, invite to attempt classifying both elements in order to further investigate this possible influence on the observed results.

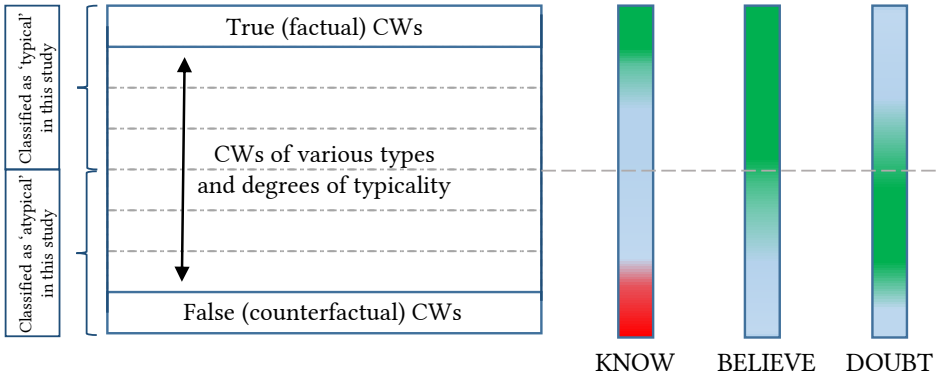

Figure 6.1: Possible distribution of degrees of typicality of critical words (CWs) in the 'typical' and 'atypical' conditions, and possible degrees of plausibility of complements for each main verb (green denotes 'plausible', red 'implausible').

6.3.2 Sustained anterior negativity (SAN)

Instead of a central-parietal N400-effect, the *doubt*-stimuli elicited an unexpected anterior typicality effect in the 300-500 ms time window: the typical words elicited a wave that was more negative-going than the wave elicited by atypical words from 300 ms after onset, and lasted throughout the whole time window.

Sustained anterior negativity (SAN) is not usually associated with plausibility manipulations. SAN effects similar to the effect observed here have been associated with referential ambiguity: Van Berkum et al. (2003a) observed SAN for definite nouns in short stories that were referentially ambiguous (e.g., *girl* in (17)). Van Berkum et al. suggested the SAN to be a reflection of processing an increased working memory load.

(17) David had told the two 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 told the girl that had been on the phone to hang up.

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

Nieuwland et al. (2007) showed that this SAN effect is sensitive to subtle manipulations of the discourse context. Their stimuli contained stories similar to the scenario displayed in (17), but in one condition, one of the possible referents could logically not be the antecedent, because he or she had left the scene in the preceding sentence. Nieuwland et al. found that SAN only appeared in discourse situations where both possible antecedents were logically available. Additionally, SAN has been observed for non-chronological

temporal adverbial clauses as opposed to chronological adverbial clauses (sentences starting with *Before...* as opposed to with *After...*, Münte et al. (1998)); when comprehenders reinterpret events denoted by durative, goal-oriented expressions (e.g., *write a letter*) from 'successfully completed' to 'terminated before attainment of the goal' (Baggio et al., 2008); and for iterative interpretations of punctive verbs (e.g., *for several minutes, the cat pounced on the rubber mouse*, Paczynski et al. (2014)). Also in situations in which a comprehender revises conditional inferences established in preceding discourse, SAN has been observed (Pijnacker et al., 2011).

As Baggio et al. (2008) pointed out, a commonality of the situations that elicit SAN is that they require extra effort to compute or revise a discourse model in order to fit an incoming piece of information. If the SAN observed in the *doubt*-condition of this study turns out to be a consistent effect with additional data, it could be a reflection of processing the negation that is contained in the lexical semantics of *doubt*, or its consequences. The SAN could reflect the extra processing involved in evoking mental spaces (an affirmative and a negated version) and selecting one of them. However, it would, in that case, not be clear why this process elicits a larger waveform for typical than for atypical words.

Alternatively, the SAN could reflect processes that have to do with accommodation. As the typology of Asher (1987) (section 3.4) suggests, reporting doubt may require a justification: the affirmative version of the doubted proposition must somehow be activated or assumed in the discourse context on order to sensibly doubt it. If this is not the case, as in the isolated sentences in this study, accommodation needs to take place. This accommodation involves, arguably, a more drastic alteration of the common ground for typical complements than for atypical complements of *doubt*. Ascribing the SAN typicality effect in the *doubt*-condition to accommodation would fit very well with the idea the SAN reflects (re)computation of a discourse representation. However, it also raises the question why SAN is not observed in many more ERP studies containing triggers for negative polarity, not in the least place studies on simple negation in main clauses such as Fischler et al. (1983).

6.3.3 Effects in the 500-800 ms time window

The *doubt*-condition is the only condition that elicited a typicality effect in the later time window: a left-lateralized centro-parietal and parietal effect in the same direction as the SAN, more negative-going for typical words than for atypical words. Just like the SAN in the 300-500 ms time window, this

later effect is an indication that some form of processing takes place that is qualitatively different from *believe* and *know*. In any case, the distinct differences can be seen as support for the hypothesis that at least parts of the semantics of the verb are processed at the moment of encountering the critical word. However, the functional significance of this effect, if it will prove robust, is not clear.

Slow-wave effects in Urbach and Kutas (2010) and Urbach et al. (2015)

Also Urbach and Kutas (2010), and Urbach et al. (2015) in some of the experiments, found effects in later time windows that exclusively showed up in the *few*-conditions. Urbach and Kutas (2010, experiment 2) reported that in their latest time window, 800-1300 ms after onset, the *few*-type quantifiers elicited a typicality effect in the same direction as the effect observed here for *doubt*, namely more positive for atypical than for typical critical words. For the *most*-type quantifiers, no such effect was observed. The same pattern was elicited by modulations with the adverbs *rarely* and *often* in their experiment 3. Unlike our 500-800 ms effect, however, the effect reported by Urbach and Kutas had a prefrontal focus.

In Urbach et al. (2015), a very small but similar slow-wave effect only for *few*-type quantifiers was observed again, but only in one of the four experiments, namely the version with no task for participants, but with a discourse context (experiment 2). Confusingly, in the version of the experiment with a discourse context and a plausibility judgment task (experiment 1), a (marginally significant) central-parietal slow-wave typicality effect was observed for the *few*-type quantifiers, but in the opposite direction of the direction of the effects reported so far: more positive for typical words. In the experiment versions without discourse context (3A and 3B), Urbach et al. observed that for *few*-type quantifiers, the typicality effects observed in the 300-500 ms window had a tendency to persist in the two subsequent windows.

Although the effects in the current study are generally different in both latency and distribution from the effects in Urbach and Kutas (2010), it is striking that in both their experiments and the current experiment, the conditions that are the 'odd one out' are similar: *to doubt*, *few*-type quantifiers and adverbs such as *rarely* are all triggers for negative polarity, and all reverse the plausibility and expectancy of the typicality of the propositions in their scope. It is also striking that the versions of the quantifier experiments that were more similar to the present study, elicited more similar effects

(albeit still quite different in latency).

If the slow-wave effects in the quantifier studies and the negativity observed here for typical words in the *doubt*-condition reflect activity that has to do with the processing of the negative polarity or its consequences, it is a process that is not very stable, and sensitive to, or easily confounded by, task-related processes.

Another commonality between the *few*-type quantifiers and *doubt* is that they create, as discussed above, less constraining contexts than the other conditions. It is also possible that this is the difference that elicits the deviant effects for *doubt* and *few*-type quantifiers. However, also this suggestion can, to my knowledge, not be supported by earlier studies finding similar effects for constrainingness.

Chapter 7

Conclusion and suggestions for further research

In this study, the processing of epistemic modal verbs was investigated by measuring their influence on ERPs in their complements. The rationale behind this design is that if epistemic modal verbs are processed fully and immediately, the possible worlds and relations created by them influence plausibility and comprehenders' expectations when processing the complements of the verbs, which affect the processes reflected in the N400.

Although the ANOVA only showed a main effect for typicality, the non-parametric test we executed (see Maris and Oostenveld (2007)) indicated modulations of the typicality effects by *know*, *believe* and *doubt*: *know* and *believe* elicited N400 typicality effects, but *doubt* did not; instead, *doubt* elicited a reversed effect in the form of sustained anterior negativity in the 300-500 ms time window, and a centro-parietal negativity in the 500-800 ms time window, both more negative-going for typical critical words.

These ERP differences related to main verb modulations indicate that at least some part of the lexical semantics of the verb is processed at the moment of encountering the critical words. However, the predictions as to the exact character of the differences between typicality effects for each main verb were not borne out: we observed a larger instead of smaller N400 effect for *believe* as compared to *know*. *Doubt* did show the predicted absence of an N400 typicality effect, but also elicited unexpected other typicality effects. As such, this study has not provided concrete positive evidence for strong incrementality of processing epistemic modal verbs.

However, I discussed a number of factors that could account for the results under the assumption of full incrementality. In the first place, the variability of typicality within categories might have had a different impact

on average N400 amplitudes for different main verbs: our assumptions about the expectations brought about by each verb might have been a too simplistic model of reality. A *post hoc* analysis of N400 effects in subcategories of our typical and atypical conditions (preferably in a larger data set) could be a way to investigate this theory, as well as further research with a more fine-grained set of conditions.

Additionally, I pointed out that the lack of a pragmatically licensing discourse context might have attenuated or altered the participants' expectations to a large extent, especially in the *doubt*-condition; I discussed a number of studies that have shown much clearer results in designs similar to this study when a discourse context was added.

With regard to sustained anterior negativity and later negativity elicited by typical conditions for *doubt*, I argued that, whichever processes are reflected in these waveforms, they are in any case incited by some aspect of *doubt* that qualitatively distinguishes it from *know* and *believe*. As such, these results provide some support for the rapid processing of elements of this verb. A further investigation of ERPs elicited by triggers of negative polarity items would be needed to better understand the results we obtained for *doubt*. It would in particular be interesting to investigate whether pragmatic accommodation plays a role in the emergence of the observed SAN.

Furthermore, a follow-up study should record an additional time window from 800 ms after onset; this would enable a further comparison with the puzzling slow-wave effects observed for the *few*-type quantifiers in the quantifier studies by Urbach and Kutas (2010) and Urbach et al. (2015). For the current study, we do not have data from later than 800 ms after onset.

Finally, another approach to further investigating the processing of *believe* and *doubt* would be to take as a starting point typical beliefs and doubts, instead of typical knowledge. The cloze probability test showed that for some items, *believe* and *doubt* have atypical completions with high cloze probabilities, for example *Sverre doubts that trees have <u>feelings</u>*. Typical or well-known doubts and beliefs might produce much clearer ERP effects, resulting in more clear-cut evidence regarding the incrementality of the processing of epistemic modal verbs.

Bibliography

Altmann, G. T. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Asher, N. (1987). A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy*, 10(2):125–197.

Baggio, G. and Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, 26(9):1338–1367.

Baggio, G., Van Lambalgen, M., and Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*, 59(1):36–53.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., and Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, 34(6):774–806.

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain research*, 1446:127–143.

Calinescu, L. (2018). The online processing of intensional modal verbs in Norwegian: An ERP study. Master's thesis, NTNU, Norway.

Carpenter, P. A. and Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological review*, 82(1):45.

Chow, W.-Y. and Phillips, C. (2013). No semantic illusions in the "semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, 1506:76–93.

Chwilla, D. J. and Kolk, H. H. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25(3):589–606.

Declerck, R. (2011). The definition of modality. In Patard, A. and Brisard, F., editors, *Cognitive approaches to tense, aspect, and epistemic modality*, pages 21–44. Amsterdam–Philadelphia: John Benjamins Publishing Company.

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word preactivation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8):1117.

Federmeier, K. D. and 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.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4):400–409.

Fodor, J., Bever, A., Garrett, T., et al. (1974). The psychology of language: An introduction to psycholinguistics and generative grammar.

Frazier, L. and Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4):291–325.

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics 3: Speech Acts*. New York: Academic Press.

Guevara, E. (2010). NoWaC: A large web-based corpus for norwegian. In *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*, pages 1–7. Association for Computational Linguistics.

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669):438–441.

Hagoort, P. and Van Berkum, J. J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):801–811.

Hoeks, J. C., Stowe, L. A., and Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1):59–73.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., and Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86:157–171.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87(4):329.

Kamide, Y., Altmann, G. T., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1):133–156.

Kim, A. and Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and language*, 52(2):205–225.

Kochari, A. and Flecken, M. (2018). Lexical prediction in language comprehension: a replication study of grammatical gender effects in Dutch. *Open PsyArXiv (preprint)*, 20.

Kounios, J. and Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, 121(4):459.

Kulakova, E. and Nieuwland, M. S. (2016a). Pragmatic skills predict online counterfactual comprehension: Evidence from the N400. *Cognitive, Affective, & Behavioral Neuroscience*, 16(5):814–824.

Kulakova, E. and Nieuwland, M. S. (2016b). Understanding counterfactuality: A review of experimental evidence for the dual meaning of counterfactuals. *Language and Linguistics Compass*, 10(2):49–65.

Kuperberg, G. R., Sitnikova, T., Caplan, D., and Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17(1):117–129.

Kutas, M. and 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*, 62:621–647.

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

Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947):161.

Kutas, M. and Van Petten, C. (1988). Event-related brain potential studies of language. *Advances in Psychophysiology*, 3:139–187.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12):920.

Lewis, D. (1986). *On the plurality of worlds*. Blackwell, Oxford.

Li, X., Hagoort, P., and Yang, Y. (2008). Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese. *Journal of Cognitive Neuroscience*, 20(5):906–915.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Münte, T. F., Schiltz, K., and Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395(6697):71.

Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Zu Wolfsthurn, S. V. G., Bartolozzi, F., Kogan, V., Ito, A., et al. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *BioRxiv*, page 111807.

Nieuwland, M. S. (2013). "If a lion could speak. . . ": Online sensitivity to propositional truth-value of unrealistic counterfactual sentences. *Journal of Memory and Language*, 68(1):54–67.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., et al. (2018). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *bioRxiv*, page 267815.

Nieuwland, M. S. and Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19(12):1213–1218.

Nieuwland, M. S. and Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, 122(1):102–109.

Nieuwland, M. S., Otten, M., and Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, 19(2):228–236.

Nieuwland, M. S. and Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3):691–701.

Nieuwland, M. S. and Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7):1098–1111.

Otten, M. and Van Berkum, J. J. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, 1153:166–177.

Paczynski, M., Jackendoff, R., and Kuperberg, G. (2014). When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of Cognitive Neuroscience*, 26(9):1905–1917.

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., and Hagoort, P. (2011). Reasoning with exceptions: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, 23(2):471–480.

Portner, P. (2009). *Modality*. Oxford University Press.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., and Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, 20(11):2037–2057.

Sitnikova, T., Kuperberg, G., and Holcomb, P. J. (2003). Semantic integration in videos of real–world events: An electrophysiological investigation. *Psychophysiology*, 40(1):160–164.

Sperber, D. and Wilson, D. (1986). *Relevance: Communication and cognition*. Basil Blackwell, Oxford.

Stowe, L. A., Kaan, E., Sabourin, L., and Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, 176:232–247.

Urbach, T. P., DeLong, K. A., and Kutas, M. (2015). Quantifiers are incrementally interpreted in context, more than less. *Journal of Memory and Language*, 83:79–96.

Urbach, T. P. and Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63(2):158–179.

Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory*, pages 276–316. Palgrave Macmillan.

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

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3):443.

Van Berkum, J. J., Hagoort, P., and 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., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4):580–591.

Van Berkum, J. J., Zwitserlood, P., Hagoort, P., and 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.

Verhagen, A. (2005). *Constructions of Intersubjectivity.* Oxford UK: Oxford University Press.

Appendices

Appendix A

Stimuli

Tora vet/tror/tviler på at...

1 rullebrett har hjul/gir
skateboards have wheels/gears

2 prinsesser går med kjoler/smoking på gallaer
princesses wear dresses/smokings at galas

3 geiter spiser gress/yoghurt for å overleve
goats eat grass/yoghurt to survive

4 terapeuter tar seg av pasienter/kaniner
therapists take care of patients/rabbits

5 planter danner oksygen/tid
plants produce oxygen/time

6 frisører klipper hår/billetter
hairdressers cut hair/tickets

7 rørleggere reparerer toalett/motorer
plumbers repair toilets/engines

8 pyramider har gravkammer/kontorer
pyramids have tombs/offices

9 murere setter op vegger/feller
bricklayers make walls/traps

10 vann løser opp salt/metall
water dissolves salt/metal

11 snekkere jobber med tre/krem
carpenters work with wood/cream

12 sjåfører kjører bil/slalom
drivers drive cars/go slalom skiing

13 lungene trenger luft/røyk for å puste
the lungs need air/smoke for breathing

14 ørene oppfatter lyd/smak
the ears perceive sound/taste
15 pianoer har tangenter/sjel av elfenben
pianos have keys/a soul made of ivory
16 fioliner har strenger/rør
violins have strings/pipes
17 briter snakker engelsk/latin med hverandre
Brits speak English/Latin with each other
18 bønder dyrker hvete/meitemark på åkere
farmers grow wheat/earthworms on fields
19 gallerier stiller ut malerier/dører
galleries exhibit pictures/doors
20 speidere sover i telt/reir
scouts sleep in tents/bird's nests
21 orkaner forårsaker ødeleggelser/hepatitt
hurricanes cause destructions/hepatitis
22 bakerier selger boller/hav
bakeries sell buns/seas
23 dyrleger behandler katter/spedbarn i klinikken
veterinarians treat cats/newborns in the clicic
24 konditorer baker kaker/fliser
confectioners bake cakes/tiles
25 bryggerier produserer øl/smør
breweries produce beer/butter
26 mygg lever av blod/vodka
mosquitos live off blood/vodka
27 dirigenter leder orkester/diskusjoner med en taktstokk
conductors lead orchestras/discussions with a baton
28 munker bor i kloster/dyreparker for å føle seg nær Gud
munks live in monasteries/zoo's to feel near to God
29 meglere planlegger visninger/farer for å selge boliger
real-estate agents plan viewings/dangers to sell real-estate
30 plastposer skader miljøet/kvelder
plastic bags damage the environment/evenings
31 geologer studerer jordlag/litteratur
geologists study earth layers/literature
32 briller korrigerer synsfeil/holdning
glasses correct visual defects/posture
33 appelsiner inneholder vitaminer/kull
oranges contain vitamins/charcoal
34 vepser elsker nektar/kål
wasps love nectar/cole

35 tog går på skinner/olivenolje
trains run on rails/olive oil

36 forfattere skriver romaner/bøter
authors write novels/fines

37 maur bygger tuer/kjøpesentre for å lagre forsyninger
ants build anthills/shopping centers to store supplies

38 tolker assisterer døve/hvaler
interpreters assist deaf people/whales

39 biologer dissekerer frosker/studenter for å studere anatomi
biologists dissect frogs/students to study anatomy

40 eskimoer bygger igloer/skyskrapere av is
Eskimos build igloos/skyscrapers of ice

41 pingviner spiser fisk/grøt i naturen
penguins eat fish/porridge in nature

42 inuitter jakter etter seler/pensjonister
Inuits hunt for seals/retirees

43 patologer undersøker lik/svindel for å bestemme dødsårsaken
pathologists investigate bodies/fraud to determine the cause of death

44 trær har grener/nebb
trees have branches/beaks

45 katter fanger mus/leger
cats catch mice/doctors

46 elger er pattedyr/insekt
moose are mammals/insects

47 bøker har sider/negler med tekst
books have pages/nails with text

48 rever har pels/rulleblad
foxes have fur/criminal records

49 sykler har bremser/sønner
bikes have brakes/sons

50 spinat inneholder jern/helium
spinach contains iron/helium

51 jordbær er frukt/leker
strawberries are fruit/toys

52 roser har torner/planer
roses have thorns/plans

53 aviser publiserer artikler/sanger
magazines publish articles/songs

54 politifolk arresterer kriminelle/prester
policemen arrest criminals/priests

55 dommere dømmer forbrytere/hjerter i rettssalen
judges convict criminals/hearts in the courtroom
56 komikere underholder publikum/babyer på scenen
comedians entertain audiences/babies on stage
57 lærere vurderer elever/foreldre
teachers evaluate students/parents
58 detektiver forhører mistenkte/gjenferd
detectives interrogate suspects/ghosts
59 foreldre oppdrar barn/uteliggere med kjærlighet og omsorg
parents raise children/hobos with love and care
60 selskaper ansetter revisorer/ministre for å håndtere budsjettet
companies employ auditors/ministers to manage the budget
61 barn vil ha godteri/press hele tiden
children want to have candy/pressure all the time
62 syklister drikker vann/maling
cyclists drink water/paint
63 klovner kaster paier/spyd
clowns throw pies/spears
64 biler trenger drivstoff/syltetøy
cars need fuel/marmelade
65 nygifte mottar presanger/trusler ved brylluppet sitt
newlyweds receive presents/threats at their wedding
66 rådgivere gir veiledning/immunitet
advisors give coaching/immunity
67 advokater hjelper klienter/elskere
lawyers help clients/lovers
68 ingeniører planlegger byggverk/fester
engineers plan buildings/parties
69 bakere skjærer opp brød/skinke
bakers cut bread/ham
70 atleter løfter vekter/peanøtter
athletes lift weights/peanuts
71 psykologer bruker hypnose/vold
psychologists use hypnosis/violence
72 astronomer observerer stjerner/kulturer med teleskop
astronomers observe stars/cultures with telescopes
73 pirater kaprer skip/banker
pirates hijack ships/banks
74 ansatte får lønn/plommer hver måned
employees get salary/plums every month
75 verter inviterer gjester/kritikere til middag
hosts invite guests/critics for dinner

76 kirurger bruker skalpell/gafler under operasjoner
surgeons use scalpels/forks during operations
77 sjimpanser skreller bananer/epler
chimpanzees peel bananas/apples
78 lover gjelder for alle borgere/spøkelser
laws apply to all citizens/ghosts
79 hunder gnager på kjøttbein/selleri
dogs gnaw on bones/celery
80 spioner har falske identiteter/kunstverk
spies have fake identities/artworks
81 slanger sluker egg/dinosaurer
snakes devour eggs/dinosaurs
82 Sherlock Holmes finner ledetråder/tyggegummi
Sherlock Holmes finds clues/chewing gum
83 familier planlegger ferier/rettssaker
families plan holidays/lawsuits
84 fastleger stiller [diagnoser]/[klokka tilbake]
GPs [make diagnoses]/[set back the clock]
85 baristaer brygger kaffe/svette
baristas brew coffee/sweat
86 badevakter redder druknende/klatrere
life guards save drowning people/climbers
87 tannleger borer i tenner/fingre
dentists drill in teeth/fingers
88 leietakere leier hus/stiger
tenants rent houses/ladders
89 skippere bruker kompass/ordbøker for å navigere
skippers use compasses/dictionaries to navigate
90 soldater bekjemper fiender/allierte på slagmarken
soldiers fight enemies/allies on the battlefield
91 leger redder liv/skudd
doctors save lifes/goals
92 sekretærer tar notater/kokain
secretaries take notes/cocain
93 monarker leder nasjoner/foreninger
monarchs lead nations/associations
94 puber serverer fatøl/hyller
pubs serve draft beer/shelves
95 bier samler pollen/diamanter
bees collect pollen/diamonds
96 slaktere selger kjøtt/parfyme i slakterbutikken
butchers sell meat/perfume in the butcher shop

97 akrobater opptrer på sirkus/høyskoler
acrobats perform in circuses/colleges

98 sovnløse teller sauer/kyllinger
insomniacs count sheep/chickens

99 surfere leter etter bølger/måker
surfers search waves/seagulls

100 reisende oppsøker turiststeder/hevn på reisene sine
travelers [visit touristic places]/[look for revenge] on their journeys

101 melk kommer fra kua/elver
milk comes from cows/rivers

102 mus spiser ost/elefanter
mice eat cheese/elephants

103 røyk signaliserer brann/jordskjelv
smoke signals fire/earthquakes

104 kjendiser skriver autografer/tall for fans
celebrities write autographs/numbers for fans

105 diabetikere trenger insulin/morfin
diabetics need insulin/morphine

106 banker tilbyr lån/suppe
banks offer loans/soup

107 griser ruller seg i gjørme/mel
pigs roll in mud/flour

108 musikere memorerer noter/rabatter
musicians memorize notes/discounts

109 kinoer selger popcorn/firma
cinemas sell popcorn/firms

110 setninger består av ord/sanger
*s*entences consist of words/songs

111 ekorn gjemmer nøtter/muffins
squirrels hide nuts/muffins

112 servitører bærer mat/terninger
waiters carry food/dice

113 pandaer tygger bambus/sko
pandas chew on bamboo/shoes

114 fektere duellerer med sverd/teskeer
fencers duel with swords/teaspoons

115 jordmødre jobber på sykehus/loftet
midwives work at the hospital/the attic

116 apotek selger medisin/brettspill ved disken
apothecaries sell medicine/board games at the counter

117 torden kommer etter lyn/valgkampen
thunder comes after lightning/the elections
118 fotografer tar bilder/lydopptak
photographers take photos/audio recordings
119 hester hopper over hinder/fjelltopper i konkurranser
horses jump over fences/mountain tops in competitions
120 jobbsøkere forbereder seg til intervju/ballspill
job seekers prepare for interviews/ball games
121 veganere spiser grønnsaker/larver
vegans eat vegetables/larvae
122 pistoler skyter kuler/rosiner
guns shoot bullets/raisins
123 grunnloven garanterer frihet/nedbør for alle
the constitution guarantees freedom/rainfall for everyone
124 selgere prøver å overtale kjøpere/søsken
sellers try to persuade buyers/siblings
125 fiskere bruker agn/sitater
fishermen use bait/quotes
126 firkløver betyr lykke/tragedie
clovers mean luck/tragedy
127 gjetere beskytter saueflokker/data
shepherds protect sheep herds/data
128 diktatorer straffer opprørere/fugler
dicators punish rebels/birds
129 massører masserer rygger/deig
masseurs massage backs/dough
130 iskrem ligger i fryseren/badstuen
icecream sits in the freezer/the bathroom
131 vinnere får medaljer/straff
winners receive medals/punishment
132 gartnere sår frø/piller om våren
gardeners sow seeds/pills in the spring
133 bevere bygger demninger/hotell
beavers build dams/hotels
134 kameler bor i orkenen/jungelen
camels live in the desert/the jungle
135 skihoppere bruker hjelm/slips
ski jumpers use helmets/ties
136 generaler gir ordre til underordede/vafler til underordnede
generals give orders/waffles to subordinates
137 anklagere anklager tiltalte/nonner
prosecutors prosecute accused people/nuns

138 groupier følger etter artister/forskere på turné
groupies follow artists/researchers on tour
139 trenere veileder utøvere/filosofer
trainers coach sportsmen/philosophers
140 boksere kjemper i ringen/cockpiten
boxers fight in the ring/the cockpit
141 arkeologer leter etter fossiler/blåbær
archaeologists look for fossils/blueberries
142 blinde går med stokk/flagg i byen
blind people walk with a stick/flag in the city
143 løver jager gaseller/professorer
lions hunt gazelles/professors
144 band spiller inn plater/dialoger i studioen
bands record albums/dialogues in the studio
145 kattunger klorer på møbler/tavler
kittens scratch furniture/blackboards
146 vaktmestre skifter lyspærer/bleier på jobb
janitors change lightbulbs/diapers at work
147 jegere skyter vilt/tulipaner
hunters shoot game/tulips
148 lommetyver stjeler mobiler/fjernsyn
pickpockets steal mobile phones/TVs
149 piloter lander fly/hauker
pilots land planes/hawks
150 lektorer underviser klasser/dyr
lecturers teach classes/animals
151 smuglere smugler narkotika/parasoller over grensen
smugglers smuggle drugs/parasols over the border
152 turister besøker museer/faresoner
tourists visit museums/danger zones
153 talere snakker i mikrofoner/kopper
speakers talk in microphones/cups
154 arkitekter designer bygninger/klær
architects design buildings/clothes
155 vitner beskriver ranere/grammatikk
witnesses describe robbers/grammar
156 postmenn leverer pakker/organer
postmen deliver packages/organs
157 investører investerer i aksjer/sokker
investors invest in stock/socks
158 cowboyer rir på hester/ulver
cowboys ride on horses/wolves

159 lagre oppbevarer produkter/hensikter
warehouses store products/intentions
160 fugler har vinger/gjeller
birds have wings/gills
161 Disney lager barnefilmer/nyheter
Disney makes children's films/news
162 spedbarn drikker morsmelk/brus
newborns drink breastmilk/soda
163 rømlinger flykter fra politiet/barnehagen
escapees flee from the police/kindergarten
164 lamper henger fra taket/himmelen
lamps hang from the ceiling/the sky
165 aper klatrer i trær/kraner
monkeys climb in trees/cranes
166 småbarn leker med dukker/gevær
little children play with dolls/guns
167 oppvaskere skyller glass/truser i kafeer
dishwashers rinse glasses/panties in cafes
168 gutter spanderer på jenter/lærere på date
boys treat girls/teachers on dates
169 intervjuere spør om meninger/sigaretter
interviewers ask [about opinions]/[for cigarettes]
170 huseiere betaler avgifter/leie
house owners pay fees/rent
171 brudepar gifter seg i kirken/lesesaler
couples marry in church/study rooms
172 bommer stenger veier/forhold
barriers close off roads/relations
173 paraplyer beskytter mot regn/ansvar
umbrellas protect againts rain/reponsibility
174 raketter reiser til månen/øyer
rockets travel to the moon/islands
175 bjørner sover i hi/senger
bears sleep in winter lairs/beds
176 riddere har rustninger/bikinier på
knights wear armors/bikinis
177 insatte rømmer til utlandet/fengsler
inmates flee [abroad]/[to prisons]
178 diagrammer viser statistikk/følelser
diagrams show statistics/feelings

179 skomakere jobber med lær/papir
shoemakers work with leather/paper
180 aktivister planlegger aksjoner/salg
activists plan actions/sales
181 ektepar feirer bryllupsdag/skilsmisser hvert tiår
married couples celebrate [their wedding day]/ separations every decennium
182 borgere velger politikere/rektorer hvert fjerde år
citizens elect politicians/headmasters every four years
183 gribber spiser åtsler/agurker
vultures eat carcasses/cucmbers
184 svamper absorberer væske/risikoer
sponges absorb liquids/risks
185 skilpadder har skall/briller
turtles have shells/glasses
186 møll spiser ull/betong
moths eat wool/concrete
187 bandasjer dekker sår/bark
band aids cover wounds/bark
188 grantrær har nåler/spiker
fir-trees have needles/nails
189 spåkoner leser i kaffegrut/føtter
fortune tellers read coffee grounds/feet
190 skulptører hogger ut statuer/desserter
sculpturers chisel out statues/desserts
191 huskyer trekker sleder/radiatorer
huskies pull sleds/radiatorer
192 kobraer spruter gift/hvitvin
cobras spit venom/white wine
193 undulater sitter i bur/kameraer
parakeets sit in cages/cameras
194 kameleoner skifter farge/navn
cameleons change color/name
195 tannkrem inneholder fluor/støv
toothpaste contains fluoride/dust
196 duer leverer brev/tips
pigeons deliver letters/tips
197 kidnappere krever løsepenger/nettsteder
kidnappers demand ransoms/websites
198 såpe fjerner skitt/minner fra klær
soap removes stains/memories from clothes
199 gardiner dekker vinduer/graver
curtains cover windows/graves

200 sakser har blader/venner
scissors have blades/friends

201 muldvarper graver tunneller/badekar under jorden
moles dig tunnels/bathtubs under the ground

202 sjørøvere graver ned skatter/hytter
pirates bury treasures/cabins

203 fabrikker avgir gasser/regnbuer
factories emit gases/rainbows

204 romerne bygde murer/roboter
the Romans built walls/robots

205 egypterne skrev hieroglyfer/bokstaver på papyrus
the Egyptians wrote hierglyphs/letters on papyrus

206 gjær består av sopp/gips
yeast consists of fungus/plaster

207 hackere lager virus/e-bøker
hackers make viruses/e-books

208 huggorm har skinn/bein
adders have scales/legs

209 flaggermus genererer ultralyd/inntekt
bats generate ultrasound/income

210 müsli inneholder korn/nikotin
muesli contains grains/nicotine

211 smeder lager hestesko/flasker av jern
smiths make horseshoes/bottles from iron

212 sjøforsvaret eier ubåter/kommuner
the navy owns submarines/municipalities

213 glasskår punkterer dekk/kajakker
glass fragments puncture tires/kayaks

214 haier angriper svømmere/løpere
sharks attack swimmers/runners

215 vakthunder skremmer bort fremmede/dykkere
watchdogs scare away strangers/divers

216 kaktuser tåler ekstrem tørke/kritikk og varme
cacti endure extreme draught/criticism and heat

217 krokodiller bor i sumper/rådhus
crocodiles live in swamps/city halls

218 låsesmeder kopierer nøkler/skjermer
locksmiths copy keys/screens

219 ugler har nattsyn/mareritt
owls have night vision/nightmares

220 sebraer har striper/prikker
zebras have stripes/dots

221 frosker spiser fluer/ordfører
frogs eat flies/mayors
222 fyrtårn veileder båter/delfiner om natten
lighthouses guide boats/dolphins at night
223 østers lager perler/mynter
oysters make pearls/coins
224 turgåere bærer ryggsekker/kjøleskap
hikers carry backpacks/fridges
225 revolusjoner feller regimer/kuratorer
revolutions fell regimes/curators
226 Stephen Hawking sitter i rullestol/arresten
Stephen Hawking sits in [a wheelchair]/[custody]
227 vaksiner forebygger sykdommer/mord
vaccines prevent diseases/murder
228 hvelv sikrer verdisaker/kvinner
safes secure valuables/women
229 magikere gjemmer kort/brunost i ermene
magicians hide cards/brown cheese in their sleeves
230 turbiner genererer strøm/universer
turbines generate power/universes
231 skreddere syr plagg/kutt
tailors sew clothes/cuts
232 strateger danner allianser/poteter
strategists form alliances/potatoes
233 barbere trimmer skjegg/busker
barbers trim beards/bushes
234 annonsører finner på slagord/nekrologer
advertisers create slogans/obituaries
235 foreldre ansetter barnepiker/trenere
parents hire nannies/trainers
236 oppfinnere tar [patent for oppfinnelsene sine]/[makten tilbake]
inventors take [patents for their inventions]/[back the power]
237 gitarister bruker plekter/hammere
guitarists use plectrums/hammers
238 vinbønder høster inn druer/erter
winegrowers harvest grapes/peas
239 gamblere spiller på kasino/hester
gamblers gamble [in the casino]/[on socialists]
240 skatere øver på triks/flyging på ramper
skaters practise tricks/flying on halfpipes
241 pikkoloer bærer kofferter/laks
piccolos carry suitcases/salmon

242 dagbøker har lås/tær for beskyttelse
diaries have locks/toes for protection

243 ordstyrere leder debatter/fotturer
chairmen lead debates/hikes

244 spisevaner påvirker helse/lufttrykk
eating habits influence health/air pressure

245 skiløpere går opp bakker/fosser
cross country skiers go up hills/waterfalls

246 bilulykker forårsåker dødsfall/sult
car accidents cause dødsfall/hunger

247 studenter pugger fakta/adresser
students learn by heart facts/addressess

248 peiser avgir varme/kulde i stua
fireplaces emit heat/cold in the living room

249 rengjørere vasker gulv/ører
cleaners clean floors/ears

250 røykere kjøper tobakk/sagflis
smokers buy tobacco/sawdust

251 bartendere blander drinker/lim
bartenders mix drinks/glue

252 kjøpmenn selger varer/bestemødre
merchants sell products/grandmothers

253 diplomater forhandler om fred/menyer
diplomats negotiate peace/menus

254 designere ansetter modeller/leiemordere
designers hire models/assassins

255 forlag gir ut bøker/steiner
publishing houses publish books/stones

256 yoga forbedrer balansen/synet
yoga improves balance/vision

257 pianister spiller konserter/biljard
pianists play concerts/billiards

258 fysikere studerer naturlover/økonomi
physicists study laws of nature/economics

259 seilbåter trenger vind/oppmuntring for å bevege seg
sailing boats need wind/encouragement in order to move

260 russetida foregår før eksamenstida/jul
the 'russetid' takes place before the exams/Christmas

261 kenguruer er pungdyr/muslimer
kangaroos are marsupials/muslims

262 Mozart komponerte symfonier/lister
Mozart composed symphonies/lists

263 Van Gogh malte selvportrett/elektroder
Van Gogh painted self-portraits/electrodes
264 Louis Armstrong spilte trompett/håndball
Louis Armstrong played the trumpet/handball
265 Beyoncé fikk priser/influensa for musikken sin
Beyoncé received prices/influenza for her music
266 MacDonalds selger burgere/tapeter og pommer frites
MacDonalds sells burgers/carpets and French fries
267 kokker arbeider på restauranter/likhus
cooks work in restaurants/morgues
268 keisere erobrer land/postkontor
emperors conquer lands/post offices
269 bueskyttere skyter piler/blyanter
archers shoot arrows/pencils
270 esler bærer sekker/puter
donkeys carry bags/pillows
271 falsknere kopierer penger/mennesker
forgers copy money/people
272 birøktere selger honning/kondomer
birøktere selger honey/condoms
273 Shakespeare skrev drama/kalendere
Shakespeare wrote drama/calendars
274 Lord Byron skrev dikt/blogger
Lord Byron wrote poems/blogs
275 sukkertøy forårsaker hull/kompetanse i tennene
candy causes cavities/competence in the teeth
276 Jimi Hendrix spilte gitar/Tetris
Jimi Hendrix played guitar/tetris
277 dansere lærer koreografi/algebra
dancers learn choreography/algebra
278 yogier lærer meditasjon/historie
yogis learn meditation/history
279 apspirin kurerer hodepine/demens
aspirin cures headaches/dementia
280 blomster trenger sollys/perspektiver
flowers need sunlight/perspectives
281 kardemomme er krydder/pølser
cardamoms are spices/sausages
282 Cristiano Ronaldo spiller fotball/bordtennis
Cristiano Ronaldo plays football/table tennis
283 ambulanser har blålys/lidelser
ambulances have flashing lights/disorders

284 geishaer har på kimono/romdrakt
geishas wear kimonos/space suits
285 toreadorer vifter med kapper/undertøy
toreadors wave with capes/underwear
286 Trump skriver mange meldinger på Twitter/datingsider
Trump writes many messages on Twitter/datingsites
287 vikingene plyndret landsbyer/drivhus
the Vikings plundered villages/greenhouses
288 Copernicus studerte astronomi/ballett
Copernicus studied astronomi/ballet
289 Marie Curie oppdaget radium/fløte
Marie Curie discovered radium/cream
290 salt dreper snegler/hummere
salt kills snails/lobsters
291 bestemødre strikker gensere/regnbukser
grandmothers knit sweaters/rain pants
292 vloggere legger ut videoer på YouTube/biblioteker
vloggers post videos [on YouTube]/[in libraries]
293 korker tetter vinflasker/porer
corks close wine bottles/pores
294 påfugler sprer sine fjær/doktriner
peacocks spread their feathers/doctrines
295 brannbiler har sirener/plener
fire engines have sirens/lawns
296 kontrakter stadfester avtaler/fantasier
contracts affirm agreements/fantasies
297 desertører flykter fra krig/internett
deserters flee from war/internet
298 dietter foreskriver helsekost/potetgull
diets prescribe healthy food/potato chips
299 edderkopper spinner nett/tepper
spiders spin webs/carpets
300 svenner utdanner lærlinger/senatorer
craftsmen educate apprectices/senators

Appendix B

Preliminary cloze probability data

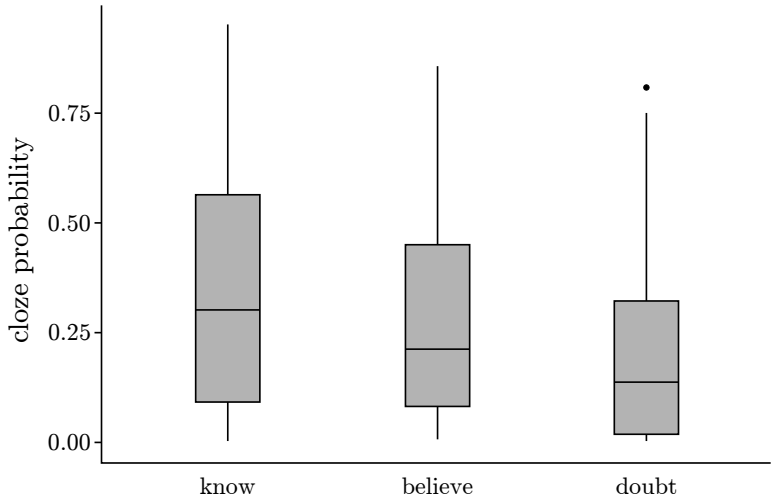

Figure B.1: Cloze probability of the typical critical words.

	know	believe	doubt
Mean	0.35	0.27	0.19
Minimum	0	0	0
Maximum	0.95	0.88	0.81
Standard deviation	0.27	0.23	0.19
No. of incomplete sentences	16	20	19