Conversation Electrified

The Electrophysiology of Spoken Speech Act Recognition

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 2 november 2015 om 12.30 uur
precies

door

Rósa Signý Gísladóttir

geboren op 11 mei 1983
te Reykjavík, IJsland

Promotor:
Prof. dr. Stephen C. Levinson

Copromotor:
Dr. Dorothee J. Chwilla

Manuscriptcommissie:
Prof. dr. Peter Hagoort
Prof. dr. Gina Kuperberg (Tufts University en Harvard Medical School, Verenigde Staten)
Dr. Napoleon Katsos (University of Cambridge, Verenigd Koninkrijk)

ISBN: 978-90-76203-70-6

Omslagontwerp: Sighvatur Halldórsson

Het onderzoek in dit proefschrift werd mogelijk gemaakt door een beurs van de International Max Planck Research School for Language Sciences gefinancierd door de Max Planck Society, München, Duitsland.

Conversation Electrified

The Electrophysiology of Spoken Speech Act Recognition

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. Th.L.M. Engelen,
according to the decision of the Council of Deans
to be defended in public on Monday November 2 2015
at 12.30 hours

by

Rósa Signý Gísladóttir

Born on May 11, 1983
in Reykjavík, Iceland

Supervisor:
Prof. dr. Stephen C. Levinson

Co-supervisor:
Dr. Dorothee J. Chwilla

Doctoral Thesis Committee:
Prof. dr. Peter Hagoort
Prof. dr. Gina Kuperberg (Tufts University and Harvard Medical School, USA)
Dr. Napoleon Katsos (University of Cambridge, UK)

ISBN: 978-90-76203-70-6

Cover design: Sighvatur Halldórsson

The research reported in this dissertation was supported by an International Max Planck Research School fellowship funded by the Max Planck Society, Munich, Germany.

To my parents, Guðný and Gísli

Table of Contents

Acknowledgments

I had an incredible time at the MPI. During my four years in Nijmegen I had the opportunity to be involved in exciting research projects, interact with top academics, experience Dutch gezelligheid, learn resilience, and meet people from all over the world. My family, friends and colleagues in Nijmegen were certainly essential to this experience.

My deepest thanks go to Steve Levinson for bringing me over to the MPI and giving me the opportunity to do research in such an amazing intellectual environment. When I was first offered the IMPRS fellowship I felt a bit ambivalent, given my limited and inaccurate knowledge of the Netherlands (consisting mainly of Geert Wilders and the Red Light District). However, when I heard that Steve would be my supervisor the deal was sealed. Steve, words cannot describe how thankful I am for having the MPI experience and being able to work with you. It was truly a privilege. Thank you for everything!

I am also indebted to my second supervisor, Dorothee Chwilla. When the academic road was rough and I was running out of fuel, her office felt like a safe haven. Dorothee, thank you for your never-ending encouragement, insightful comments on countless paper drafts, and your inspiring calls to Iceland when I was on maternity leave!

In addition to my PhD work at the MPI, I was fortunate to be a part of a collaborative research project on conversational repair. I would like to thank Nick Enfield for involving me in the project, as well as for many helpful and fun discussions over the years. See you in Australia!

I also want to thank my manuscript committee – Peter Hagoort, Napoleon Katsos and Gina Kuperberg – for devoting their time to reading and evaluating my thesis. Thank you so much for your effort!

The experiments in this thesis would not have taken place without the support of the MPI staff and other PhD students. Special thanks go to Renske, Peter and Dirk, assistants in the Language and Cognition Department, and fellow PhD student Nienke for lending me their voices in the audio recordings. You did a great job! I would also like to thank the other student assistants in the department who helped me with the translation of the experimental materials and other tasks. Moreover, Ronald Fischer and Johan Weustink in the Technical Group helped me with the technicalities of programming and running EEG experiments. Thank you!

When it came to finishing the thesis I received help from several people. Sara Bögels gave me invaluable help with Fieldtrip and collaborated with me on the last experimental chapter. Milou Huijsmans, Sara Bögels and Julija Baranova assisted me with the Dutch translation of the thesis summary. My cousin Sighvatur Halldórsson designed the thesis cover. Thank you so much everyone!

I would like to thank Edith Sjoerdsma – the eyes and ears of the Language and Cognition Department – for helping with practical matters at the MPI. IMPRS staff Dirkje van der Aa, Rachel Sheer and Els den Os were also invaluable when it came to paperwork and gave me tips for adjusting to life in the Netherlands. Many thanks!

My time in the Netherlands would have been pretty lonely if it were not for my friends. Sho and Salomi – thank you for being there for me. I will always remember our cozy lunches and dinner parties, our trips to Amsterdam and Cologne, the Eurovision evenings...I expect a reunion sooner than later!

I would also like to thank my dear Ewelina for all the good times we spent together in our office (and for tolerating my mess!). Hartstikke bedankt, mevrouw, for your friendship!

Sarah, Wencui, Giovanni, Ely, Rebecca, Julija and all my other friends and colleagues at the MPI: thank you for the fun MPI discussions and nights out.

It was a pleasure to get to know all of you. Rebecca and Julija, thank you for being my paranymphs! It means a lot to me to have you with me at the defense.

This chapter of acknowledgments would not be complete without giving thanks to my family. I want to thank my partner, Bjarki, for his endless support over the years. Thank you for reminding me that "nobody said a PhD would be easy." Thank you for not giving up when I moved to the US, then the UK, then Netherlands... Takk ástin mín! Our wonderful daughter Saga Rós gets special thanks for keeping her mother entertained during the last stages of thesis writing – and for sleeping well during the day to allow me to work.

Last, but not least, I would like to thank my parents, Guðný Guðbjörnsdóttir and Gísli Pálsson. They deserve thanks at so many levels. Thank you for your advice when I didn't know what to do; for all your visits to Nijmegen; for helping with everything ranging from money to editing to babysitting; for tolerating my moodiness; for being my academic rolemodels; for teaching me that women can do everything; and for all your love and encouragement at every stage of my life. This thesis is dedicated to you.

1 Introduction

Imagine a world in which language serves no purpose; utterances are exchanged but lack functionality. Such a society – if one could call it such – would lack *speech acts* (Austin, 1976; Searle, 1969). Speech acts are verbal actions such as requests, teases and compliments. Participants in conversation have to be able to recognize these actions for communication to be successful. However, speech act recognition is far from straightforward, as utterances are frequently underspecified for the action level of meaning (Levinson, 2013; Searle, 1975). The challenge for listeners is further enhanced by the extraordinarily fast transitions between turns in conversation (De Ruiter, Mitterer, & Enfield, 2006; Heldner & Edlund, 2010; Stivers et al., 2009), allowing limited time to recognize the action and plan a relevant response (Levinson, 2013). How conversationalists manage to extract speech acts from the underspecified linguistic code and respond within these tight time constraints is a real puzzle. Can listeners recognize the action early on in the utterance, sidestepping the full propositional content, and thereby facilitate quick and efficient turn-taking? The aim of this thesis is to addresses this question by investigating the time-course of speech act recognition in spoken, naturalistic dialogues within a novel framework that bridges methods from cognitive neuroscience and research on conversational interaction.

In this introductory chapter I will first provide a short overview of theoretical approaches to speech acts and discuss the perspective adopted in this thesis. I will then outline the motivations for the research question. The reader is introduced to the principles of electroencephalography (EEG), the main methodology of the thesis. The chapter concludes with the research objectives and an overview of the thesis organization.

1.1 Theoretical background and approach

Research on how we do things with words has roots in two theoretical traditions, originating in philosophy and sociology. The former tradition started with the philosophy of Wittgenstein, Austin and Searle in the fifties and sixties. Moving away from the idea that language is mainly a system for representation, Wittgenstein placed language use in the spotlight by proposing that language is first and foremost a toolbox for social activities:

> *Think of the tools in a tool-box: there is a hammer, pliers, a saw, a screw-driver, a rule, a glue-pot, glue, nails and screws. – The functions of words are as diverse as the function of these objects. . . . Of course, what confuses us is the uniform appearance of words when we hear them spoken or meet them in script and print. For their application is not presented to us so clearly.* (Wittgenstein, 1953 para. 11)

Seen from this perspective, speech acts are the tools that allow us to get things done through language.

Austin (1976) systematized the study of verbal action under the rubric of what was later termed *speech act theory*. He argued that talking about "uses of language" was too vague and instead made a distinction between three different types of acts that are performed through language use. The *locutionary act* refers to the production of a sentence with some propositional content (roughly "meaning" in the traditional sense). The *illocutionary act* is what is being done through the utterance; the performing of actions such as requests, offers, and the like. The current thesis is concerned with this level of analysis. The term *speech act* usually refers to illocutionary acts or the related concept of *illocutionary force*, i.e., the pragmatic function or force of the illocutionary act. The third and final type of act in Austin's framework is the *perlocutionary act,* which involves bringing about some additional effects in the listener (e.g., making someone frightened).

Speech act theory was further developed by Searle (1969, 1976) who classified illocutionary acts into five fundamental types; representatives, directives, commissives, expressives and declarations. Another influential aspect of speech act theory is the discussion of *indirect speech acts* (see, for instance, Gordon & Lakoff, 1971; Searle, 1975), in which “one illocutionary act is performed indirectly by way of performing another,” (Searle, 1975, p. 60). A classic example of an indirect speech act is the sentence “can you reach the salt,” which is simultaneously a question about someone’s ability to reach for an object and a request to pass it on (the utterance has two illocutionary forces; Searle, 1975).

Speech act theory has inspired research in diverse fields such as linguistic pragmatics (e.g., Kissine, 2013; Purver, Gregoromichelaki, Meyer-Viol, & Cann, 2010), artificial intelligence (e.g., Traum, 1999), psychology (e.g., H. H. Clark, 1979; Gibbs, 1981; Holtgraves, 2008b), and to a limited extent cognitive neuroscience (Basnakova, Weber, Petersson, Van Berkum, & Hagoort, 2014; Egorova, Pulvermüller, & Shtyrov, 2013; Egorova, Shtyrov, & Pulvermüller, 2013). I will return to experimental approaches to speech act comprehension in section 1.3 below.

Although speech act theory was influential, the most developed understanding of speech acts is found within the framework of *conversation analysis* (CA), where the term *action* is used (Schegloff, 1996, 2007). Conversation analysis was developed within sociology by Sacks, Schegloff and Jefferson (e.g., Sacks, Schegloff, & Jefferson, 1974), who drew on the writings of Goffman and Garfinkel (Garfinkel, 1967; Goffman, 1967; see also Heritage, 2001). Conversation analysts record and study spontaneous conversation with the aim to “describe, analyze and understand talk as a basic and constitutive feature of human social life" (Sidnell, 2010, p. 1). Proponents of conversation analysis argue that speech act theory places too much emphasis on the single utterance as the fundamental unit of analysis (Schegloff, 1988):

> *What a rudimentary speech act theoretic analysis misses, and I suspect a sophisticated one will miss as well, is that parties to real conversations are always talking in some sequential context. I refer here not to social contexts like offices, classrooms or families, but sequential contexts formulated in terms of more or less proximately preceding talk and the real jobs of projecting further talk which utterances can do … Such prior and prospective contexts are inescapably implicated in the real life projects, however humble or exalted, which are being prosecuted through the talk. These real life projects, and the sequential infrastructure of talk-in-interaction, are involved in the production and analysis of talk by the parties in such intimate detail that we are only beginning to understand it. But it is clear that temporality and sequentiality are inescapable; utterances are in turns, and turns are parts of sequences; sequences and the projects done through them enter constitutively into utterances like the warp in a woven fabric.* (Schegloff, 1988, p. 61)

The study of *sequence organization* (Schegloff, 2007) – how actions are organized into larger sequences and what kind of contingencies operate between them – is a major contribution from CA in regards to speech acts. An important observation is that turns tend to come in pairs of actions and in such *adjacency pairs* (Schegloff, 2007) the first (pair) part sets up powerful constraints on what type of action can follow; questions are followed by answers, and invitations call for acceptances or rejections. The adjacency pair is one of the most basic action sequence types, attested in many languages (K. Kendrick et al., 2014). Adjacency pairs and other types of action sequences put constraints on speech acts, which can provide listeners with important information for the interpretation of talk.

CA findings have received little attention in the experimental sciences, particularly in cognitive neuroscience. In comparison to speech act theory, the conversation analysis tradition provides a much more detailed account of what verbal actions in natural conversation really look like (action formation; see, for instance, Levinson, 2013), how they pattern into larger sequences (sequence organization; Schegloff, 2007), and, more generally, their role in social interaction. As a consequence, the CA approach and findings form the theoretical foundation of the following chapters, although conceptual distinctions from speech act theory will also be drawn upon.

As for terminology, the term *speech act* will be retained here due to its familiarity. It is much more frequently used than alternative terms, including the new hybrid *speech actions* (used, for instance, by Sbisà & Turner, 2013). The terms *speech act*, *action* and *illocutionary act* will be used interchangeably throughout the thesis, unless otherwise noted, but should not be taken to reflect theoretical contrasts.

1.2 Why investigate the time-course of speech act recognition?

Speech acts get their tool-like power by virtue of being recognized by other participants in conversation. Indeed, Sacks wrote that "a culture is an apparatus for generating recognizable actions" (Sacks, 1992, p. 226), highlighting not only that action (non-verbal or verbal) is a defining characteristic of culture but also that actions must be recognizable to have an effect. *Speech act recognition* – the topic of this thesis – is the process of identifying or recognizing the action of an utterance in a given context. Speech act recognition is critical in conversation because the response to an utterance is dependent on the speech act being performed. Problems with speech act recognition can halt progressivity of the conversation or lead to misunderstanding.

There are two factors that make speech act recognition difficult from an individual's cognitive perspective; the lack of clear speech act marking on utterances and the tight time constraints in turn-taking (the back-and-forth exchange of talk). These factors will be discussed below in sections 1.2.1 and 1.2.2. Section 1.2.2 concludes with the introduction of an account of speech act recognition which will be tested in the experimental chapters of this thesis.

1.2.1 Underspecification at the action level

Speech act recognition would be an easy task if listeners could rely on turn design, i.e., the form of the utterance itself, to recognize the action. Searle noted that features such as word order, syntactic mood and intonation can provide clues about what type of speech act is being performed (Searle, 1969; Searle & Vanderveken, 1985); such speech act clues have been called *illocutionary force indicating devices* (Searle, 1969; Searle & Vanderveken, 1985). For example, interrogative word order, wh-question words (*who, when*) and rising intonation are associated with questions, while imperatives (*put, go*) are commonly used in orders or requests. However, listeners cannot always rely on such clues. Interrogatives and imperatives can perform many other actions than questioning and requests (see, for instance, Heritage, 2012). Similarly, rising intonation is not a good indicator of questionhood, as it is frequently absent from questions in English (Geluykens, 1988) and used in assertions serving a range of other actions. There is no evidence that prosody can, or normally does, fully disambiguate all the potential illocutionary forces associated with utterances. Moreover, illocutionary force indicating devices can be overriden by external factors such as epistemic status, i.e., which participants in conversation are considered to be most knowledgeable about a relevant domain of knowledge (Heritage, 2012). In fact, it has been noted that “it is possible to use nearly any sentence type with the effect of nearly any other, under appropriate circumstances” (Sadock & Zwicky, 1985, p. 191); there is no one-to-one correspondence between turn design and what speech act is being performed (Heritage, 2012; Levinson, 2013; Motsch, 1980; Schegloff, 2007).

The lack of discrete speech act marking leaves most utterances in conversation underspecified for the action level of meaning, such that the utterance is compatible with multiple speech acts. For instance, the declarative utterance *I have a car* could be used to offer somebody help with moving, to indirectly reject an offer for a ride, or to answer a question about commuting. This problem of *underspecification at the action level* is pervasive in everyday conversation, making speech act recognition far from straightforward.

1.2.2 Time constraints in turn-taking

The challenge for participants in conversation is further enhanced by tight time constraints in turn-taking. The seminal paper by Sacks, Schegloff and Jefferson (1974) noted that the turn-taking system is characterized by the minimization of overlapping talk and gaps between speakers' turns. Turn transitions with no gap or no overlap are common, and "together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions" (Sacks et al., 1974, p. 700; see also Levinson & Torreira, 2015). In recent years corpus research has provided important insights into the timing of turn transitions. Using a corpus of Dutch telephone conversations, De Ruiter, Mitterer and Enfield (2006) measured the time between the start of a turn and the end of the prior turn (*response offset* or *floor transfer offset*). The most frequent kind of turn transition interval, occuring in 45% of turn transitions, was a slight gap or slight overlap with a response offset between -250 and +250 ms (a negative value indicates an overlap with previous speaker while a positive value indicates a gap). In a study on Scottish English, Dutch and Swedish, the mode of the response offset distribution was centered around 200 ms in all corpora, indicating that a slight gap of 200 ms was most common, and 70–82% of all between-speaker gaps were shorter than 500 ms (Heldner & Edlund, 2010). A cross-linguistic study focusing on responses to questions revealed that in all 10 languages studied, the most frequent response offset was somewhere between 0 and +200 ms, depending on the language, with an astonishing overall mode of 0 ms (Stivers et al., 2009).

The emerging picture is that turn transitions between speakers in conversation are generally extremely quick, with gaps of roughly 200 ms being most frequent. From the perspective of a listener in conversation, a 200 ms gap does not leave much time for recognizing the action in the prior turn, planning a relevant response, and finally executing it. Indeed, as noted by Levinson (2013), research on the time course of speech production (e.g., Indefrey & Levelt, 2004; Levelt, 1989) indicates that planning and initiating a simple, one-word utterance takes considerably more time than most gaps in turn-taking provide. To highlight what this means in the context of conversation, these studies will be described in more detail below.

In a comprehensive meta-analysis of the relevant neuroimaging literature on word production, Indefrey and Levelt (2004) gave timing estimates for the core processes of word production, according to the theory of lexical access put forth by Levelt, Roelofs and Meyer (1999). The first core processing stage, *conceptual preparation*, involves accessing the lexical concept. In the second stage, *lemma retrieval* or *lexical selection*, the word's syntactical properties are retrieved, such as word category, gender of nouns, and syntactic argument structure in the case of verbs. The third stage, *form encoding,* is itself a staged process that involves accessing the word's phonological code, then clustering the segments into syllables, and finally creating motor action instructions to initiate articulation (phonetic encoding). The estimated time windows for these processes are as follows (see Table 1.1, based on Indefrey and Levelt (2004)):

Operation	Duration (ms)
Conceptual preparation (from picture onset to selecting the target concept)	175
Lemma retrieval	75
Form encoding:	
- Phonological code retrieval	80
- Syllabification	125
- Phonetic encoding (until initiation of articulation)	145
Total	600

Table 1.1: Estimated time-windows for the core processes of word production, as described by Indefrey and Levelt (2004, p. 108)

The core processes of word production together take around 600 ms. In other words, this is how long it takes to plan and initiate a simple, one-word utterance in a controlled experimental environment.

The question arises how these numbers translate into natural conversation. Turns in conversation usually contain more than one word. Picture naming studies have demonstrated that longer sentences take longer to plan; when participants are asked to describe a picture using a full sentence such as *the girl jumps*, containing a determiner + noun + verb, mean naming latency is 784 ms (Schnur, Costa, & Caramazza, 2006). When an adjective is also included, resulting in a four-word sentence, the latency is 857 ms (Schnur et al., 2006). Based on these findings, conversational utterances that contain more than one word should take even more than 600 ms to plan[1]. Moreover, in such picture

[1] It should be kept in mind, however, that language production is incremental; each processing component can start working on the still-incomplete output of the prior processing stage (Levelt, 1989). Thus “even though there can be no formulating without some conceptual planning, and there can be no articulating without a phonetic plan, message encoding, formulating and articulating can run in parallel” (Levelt, 1989, p. 24). Indeed, it is not the case that planning a five-word turn takes 5 x 600 milliseconds. It is nevertheless reasonable to assume that the basic process of “turning on”

naming studies the object to be named is often primed or participants have been familiarized with the pictures beforehand – a situation that is quite remote from natural conversation. A study on seven languages, in which participants were not shown the pictures to be named in advance, found longer reaction times (naming latencies), ranging from 1041 ms in English to 1254 in Bulgarian, for a single word (E. Bates et al., 2003).

The timing facts discussed in this section – that initiating a turn takes at least between 600 to 1200 ms, while most gaps in conversation are only 200 ms – suggest that listeners begin planning their responses before the prior speaker has finished speaking, and by extension that they can ascribe an action to the unfolding utterance well before the turn ends (Levinson, 2013). On this *early speech act recognition account*, recognition of the action is not made at the final stage in the comprehension process, occurring at the last word of incoming utterances, but takes place early on when the turn has only been partially processed. Early speech act recognition could be the key to efficient turn-taking, allowing listeners to plan their reply early and respond within the 200 ms time frame characteristic for turn-taking. The key question addressed in this thesis is whether this early recognition of action can be made, given that utterances are often underspecified for action and do not contain clear speech act clues to aid recognition.

1.3 Prior research on speech act comprehension

Several strands of experimental research have addressed speech act comprehension broadly construed, i.e., aspects of comprehension that do not necessarily involve the recognition of the action *per se*. In this section I will give

---

the speech production system in conversation takes at least 600 ms and probably longer when it comes to more complex turns in conversation.

a brief overview of the relevant literature, with a particular focus on eye-tracking and neuroimaging methods due to their edge in unraveling the time-course and neural substrates of cognitive processing.

One domain of research that has indirectly addressed speech act comprehension is the study of reference resolution and common ground using eye-tracking methodology. Experiments in this domain have demonstrated early sensitivity to speech act types in referential communication tasks, even in young children; listeners quickly interpret requests as asking about referents in common ground (known to both participants) (Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008), whereas when listening to questions they shift their attention to entities in privileged ground (only known to the addresses) (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Nurmsoo & Bloom, 2008). However, these eye-tracking paradigms are more informative about the on-line processing consequences of speech act recognition than how recognition of the action is made in the first place. The critical utterances are often all of the same or restricted type (e.g., only requests, or requests vs. questions) and contain clear speech act marking.[2] As a consequence, they do not address the key focus of the present thesis, namely how underspecified utterances (without illocutionary force indicators) are rapidly understood as performing certain actions.

Another strand of research has used functional magnetic resonance imaging (fMRI) to investigate the comprehension of indirect speech acts. An fMRI study on indirect replies in spoken dialogues (e.g., *Did you like my presentation? - It's hard to give a good presentation*) found that in comparison to direct replies, the indirect speech acts activated not only typical language

[2] In one experiment, for instance, all targets were requests with imperatives; *Pick up X and put it in area Y* (Hanna, Tanenhaus, &

regions, such as the bilateral inferior frontal gyrus, but also areas implicated in mentalizing and affective empathy, including the dorsomedial prefrontal cortex (dmPFC), right temporo-parietal junction (TPJ) and insula, as well as the right medial temporal gyrus (Basnakova et al., 2014). This was taken to indicate that listeners take the speaker's perspective both at cognitive and affective levels when they listen to speech acts (Basnakova et al., 2013). Similarly, a study on indirect requests (e.g., *it is very hot here* presented with a picture of a window) reported activations in theory of mind regions known to be involved in false belief tasks, including the medial prefrontal cortex (mPFC), the precuneus, and the bilateral TPJ (van Ackeren, Casasanto, Bekkering, Hagoort, & Rüschemeyer, 2012). The indirect requests also activated cortical motor areas associated with action planning and motor control (van Ackeren et al., 2012). These fMRI studies provide important insights into the neural substrates of indirect speech act comprehension, but due to the poor temporal resolution of fMRI cannot tell us much about the time-course of speech act recognition.

A number of studies have made use of the excellent temporal resolution of electroencephalography (EEG) and magnetoencephalography (MEG) to investigate the comprehension of speech acts. These methods are advantageous for tracking the time-course of cognitive processes, as discussed in section 1.4 below. Most of these studies throw only indirect light on speech act recognition, for instance by investigating the processing of irony (Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013) or pitch accents in question-answer dialogues (e.g., Dimitrova, Stowe, Redeker, & Hoeks, 2012; Magne et al., 2005; Wang, Bastiaansen, Yang, & Hagoort, 2011). More relevant for this thesis, an EEG study on visually presented (written) indirect requests (e.g., *My soup is too cold to eat* in a restaurant context) found differences between indirect requests and literal statements from the second word onwards, but no EEG differences at the final word (Coulson & Lovett, 2010). The results suggest that speech act recognition in the visual modality takes place relatively early in the sentence. Another EEG study investigated speech act recognition in written words (e.g., *plant*) that performed either a requesting or a naming speech act depending on a prior video-taped context sentence (*What are these called?/What can I get you? – PLANT*) (Egorova, Shtyrov, et al., 2013). The brain responses for the two speech acts diverged as early as 120 ms after the onset of the critical words *(PLANT)*. A follow-up MEG study (Egorova, Pulvermüller, et al., 2013) reported that the requests engaged comprehension systems in the right hemisphere within 100 ms after word onset, followed by theory of mind activations in the medial prefrontal and temporo-parietal areas from 200 to 300 ms. Naming speech acts, on the other hand, activated brain areas involved in lexico-semantic retrieval from 100 to 150 ms. These EEG and MEG findings provide some supportive evidence for an early speech act recognition account, at least in the visual modality. However, these studies are far removed from conversation, using written input instead of naturalistic, spoken dialogues. What is missing is an investigation of speech act recognition in the auditory modality that takes into account not only the temporal demands that characterize spoken conversation but also the sequential organization of action in turn-taking.

1.4 Methodology: Electroencephalography and language research

EEG is a neuroimaging method that is particularly well suited for investigating the time-course of cognitive processes, due to its fine temporal resolution. Since this is the main methodology used in this thesis I will describe it in some detail in the following section.

The German neurologist Hans Berger discovered in 1929 that the electrical activity of the human brain could be measured by placing an electrode on the scalp and plotting the voltage changes over time (Berger, 1929). This recording of electrical activity along the scalp is referred to as *electroencephalography* or simply *EEG*, and the output is the *electroencephalogram.* Electrical activity encodes information about brain states and processes, and by inference about mental states and processes involved in various cognitive tasks (Kutas & Dale, 1997). The recording of this activity through EEG has therefore proven very useful in both clinical and scientific contexts, particularly when information about the time-course of processing is needed.

Electrical activity is the result of electrochemical signaling between neurons. *Postsynaptic potentials* are voltages that arise when neurotransmitters from the presynaptic (sending) neuron bind to receptors on the postsynaptic (receiving) cell. This leads to net negativity in the region of the dendrite and a net positivity at the cell body, creating a small dipole (Luck, 2005a). The electric potential produced by a single dipole is too weak to be detected at the scalp. The voltage can only be measured with EEG when a large ensemble of neurons with a similar spatial orientation fires synchronously; if the dipoles from the individual neurons are not spatially aligned, they will cancel each other (Kutas & Dale, 1997; Luck, 2005a). The summation of postsynaptic potentials is most likely to occur in pyramidal cells, due to their alignment perpendicular to the surface of the cortex (Luck, 2005a). The EEG therefore reflects the summed postsynaptic potentials of thousands or millions of pyramidal neurons that fire synchronously and have a similar spatial orientation.

There are two main approaches to analyzing EEG data; event-related potentials and time-frequency analysis. These two approaches will be described below. The advantages and disadvantages of EEG relative to other neuroimaging methods will also be discussed. Section 1.4 ends with some practical challenges of investigating speech act recognition with EEG and how they can be addressed.

1.4.1 Event-related potentials

In experimental research, a large part of the raw EEG signal is not related to the manipulation of interest, but rather reflects background EEG activity (i.e., "noise" from an experimental point of view). *Evoked responses* – the electric potentials associated with the processing of a stimulus or an event – are much smaller in amplitude than the background EEG. One way to extract the evoked response from the background activity is to average the EEG across many trials that are *time-locked* to the stimulus, i.e., averaging the electric potentials that occur immediately before or after the event. Since the background EEG is assumed to be randomly distributed across trials, the averaging procedure reduces the noise to nearly zero (Bastiaansen, Mazaheri, & Jensen, 2008; Kutas & Dale, 1997). At the same time, the event-related response of interest, which should be invariant from trial to trial, is enhanced. The remaining, averaged signal is the *event-related potential* (*ERP*). More specifically, ERPs reflect brain responses that are both time-locked and *phase-locked* with respect to stimulus onset. The phase of an EEG wave is, roughly speaking, its slope or direction at a given point in time (Bastiaansen et al., 2008). The phase of an evoked response associated with a stimulus does not vary across trials (it is phase-locked), and as a consequence the evoked response of interest remains in the ERP average, while the phase of the background activity is random (non phase-locked) and is therefore averaged out (Bastiaansen et al., 2008; Pfurtscheller & Lopes da Silva, 1999). The ERP approach has been the dominant methodology to investigate the relationship between EEG and cognition (Bastiaansen et al., 2008). This approach is the foundation of this thesis and will therefore be discussed in some detail.

The ERPs elicited by a stimulus consist of negative and positive voltage fluctuations. By convention, negative voltages are normally plotted upwards. The fluctuations are frequently labeled according to polarity (i.e., negativity/positivity) and latency (measured from stimulus onset). As an example, the N400 component is a negative-going deflection in the EEG that peaks 400 ms post-stimulus onset. An ERP *component* refers to fluctuations in the EEG waveform that have certain functional or physiological characteristics. An important distinction has been made between endogenous and exogenous ERP components. Exogenous components occur early in the waveform and are obligatory, sensory-driven responses that are largely insensitive to cognitive factors (C. M. Brown & Hagoort, 2000; Luck, 2005a). In contrast, endogenous ERP components are not influenced by the physical characteristics of the stimulus, but vary as a function of the internal, cognitive operations engendered by it (Kutas, van Petten, & Kluender, 2006). The endogenous components, which usually occur with a latency beyond 100 ms after stimulus onset, are therefore more informative for psycholinguistic research (Kutas et al., 2006).

How can ERPs be used to investigate language comprehension? An influential study by Kutas and Hillyard (1980) marked the birth of the electrophysiology of language. Kutas and Hillyard investigated the ERP responses to semantically congrous versus incongruous sentences in a reading task. Words that were incongruous with the prior context (*He spread his warm bread with SOCKS*) elicited a larger negativity than congruous words (*He spread his warm bread with BUTTER*) between 250 and 400 ms after word onset. This ERP component has since been referred to as the N400. The N400 is thought to reflect the ease/difficulty of retrieving conceptual knowledge associated with a word or meaningful stimulus from semantic memory (e.g., Kutas & Federmeier, 2000; Kutas et al., 2006), or the subsequent semantic integration of that knowledge into context (e.g., D. J Chwilla, Brown, & Hagoort, 1995; Van Berkum, Hagoort, & Brown, 1999). The N400 has been found to be modulated by factors such as word frequency (Van Petten & Kutas, 1990) and semantic priming (Kutas & Federmeier, 2000). However, in recent years big strides have been made in research on how discourse and social context influence the N400 and other ERP components. As an example, an N400 effect is elicited when a target word does not fit into prior story context (Van Berkum, Zwitserlood, Hagoort, & Brown, 2003), violates world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004), or makes establishing causal coherence across sentences difficult (G. R Kuperberg, Paczynski, & Ditman, 2011). The N400 is also influenced by the personal values of experimental participants (Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009), and mismatches between a message and stereotypical expectations based on the voice of the person uttering it (Van Berkum, Van Den Brink, Tesink, Kos, & Hagoort, 2008), demonstrating that social context has an immediate impact on sentence interpretation.

Another language-related ERP component is the P600, also referred to as the late positive component. A positive-going deflection that peaks after 600 ms, the P600 was initially associated with syntactic reanalysis or repair following syntactic ambiguity or violations (Hagoort, Brown, & Osterhout, 1999). However, it was later found to be elicited in other situations involving semantic processing, such as in thematic violations (see Kolk & Chwilla, 2007 for a review) and has been argued to reflect domain-general reanalysis processes (Van de Meerendonk, Kolk, Vissers, & Chwilla, 2010). Most relevant for research on conversation, ironic utterances (e.g., *That's really BLAND* preceded by a description of a man tasting a very spicy dish) elicit a P600 and not an N400, both in the visual and auditory modality (Regel et al., 2011), suggesting that the P600 may reflect pragmatic interpretation processes.

Well-known ERP components such as the N400 and P600 provide important benchmarks for the interpretation of ERP data. However, inferences can be drawn from ERP data at several levels, even when familiar ERP components are not observed. At the most basic level, statistical differences between two conditions tells us that cognitive processing in the conditions differs in some way (Rugg & Coles, 1995). In the case of linguistic stimuli, for instance, such differences indicate that the comprehension system is sensitive to the experimental manipulation in question (sensitivity inferences; Van Berkum, 2004). A second type of inference concerns timing. If two waveforms diverge statistically at a certain point in time, one can infer that cognitive processing differs for the conditions at least at that moment (Rugg & Coles, 1995; Van Berkum, 2004). A third type of inference can be drawn based on the scalp distribution. If two experimental manipulations affect the ERPs in different ways in terms of the scalp distribution of effects, the inference can be made that two non-equivalent processes are involved (Rugg & Coles, 1995). Other aspects of the EEG waveform provide additional cues regarding whether there are qualitative or quantitative differences between conditions. Differences in polarity, morphology (wave shape) and/or scalp distribution are usually required to infer the presence of qualitatively different cognitive processes (Kutas, 1993). If, on the other hand, two ERP effects are similar in terms of these features, differences in the size of the effect or latency are interpreted as reflecting quantitative variations of the same cognitive process (Kutas, 1993). These dimensions highlight the wealth of information that can be obtained from event-related potentials.

1.4.2 Time-frequency analysis

ERPs only reflect a certain part of the event-related EEG signal. Raw EEG recordings are dominated by rhythmic oscillations of various frequencies. These oscillations are ongoing phenomena that occur even in the absence of an experimental task (Bastiaansen et al., 2008). The original report by Hans Berger in 1929 described two such oscillatory rhythms, alpha and beta (Berger, 1929), at a frequency around 8 – 13 Hz and 13 – 30 Hz respectively. Other rhythms include delta (1 – 4 Hz), theta (4 – 8 Hz) and gamma (30 – 70 Hz). Some of the oscillatory activity reflects "noise", i.e., background EEG activity that is uncorrelated with the experimental condition and is averaged out in the event-related potentials analysis. However, in addition to giving rise to evoked responses (ERPs), an experimental event can also modulate ongoing oscillatory activity, resulting in oscillations that may be of interest to the researcher. Such event-related changes in oscillatory activity are not phase-locked to stimulus onset, since the experimental event occurs at random phases of the ongoing oscillation (Bastiaansen et al., 2008). As a consequence, this type of event-related activity cancels out in the averaging procedure, in contrast to the phase-locked responses captured in the ERPs.

A different type of analysis is required to capture non-phase locked oscillatory activity. One method commonly used is *time-frequency analysis of power* which represents frequency-specific changes in EEG power (squared amplitude) over time at the single trial level. EEG power reflects the number of neurons that fire synchronously (Klimesch, 1999). The summation of post-synaptic potentials of a large number of neurons results in increased amplitude, and hence power, of the scalp-recorded EEG oscillations (Bastiaansen et al., 2008).

Time-frequency analysis of EEG data is complementary to ERP results in several respects. First, since the two methods capture different types of brain responses, combining them can result in a more complete picture of the time-course of cognitive processes in the post-stimulus interval. As an example, a recent study on oscillatory activity during comprehension of irony found evidence for integration operations (reflected by an increase in gamma band power) in an earlier time window than reported by ERP studies, challenging the view that irony processing takes place only at a later stage (Spotorno et al., 2013). Second, while ERPs in principle reflect the brain's phase-locked response after the stimulus, time-frequency analyses can be applied to the pre-stimulus interval, revealing power changes that may be of interest. This is an advantage, for instance, in paradigms investigating anticipatory attention – when attention is oriented towards an upcoming stimulus to facilitate its processing (Bastiaansen, Böcker, Brunia, de Munck, & Spekreijse, 2001; e.g., Bastiaansen & Brunia, 2001; Jones et al., 2010; van Ede, Szebényi, & Maris, 2014). The use of both ERP and time-frequency analyses can therefore "provide a unified temporal account (pre-stimulus and post) of how information is processed by the brain" (Bastiaansen et al., 2008, p. 38). Third, oscillatory activity provides additional insights into the cognitive processes of the phenomena under investigation. For instance, time-frequency analyses can dissociate superficially indistinguishable ERP components (see, for instance, Roehm, Schlesewsky, Bornkessel, Frisch, & Haider, 2004), providing further specification of the neuronal mechanisms involved. More generally, event-related oscillatory activity is informative about the functional network dynamics in the brain, since synchronization of oscillations is thought to play a key role in linking areas that are part of the same functional network (Bastiaansen et al., 2008).

In this thesis both event-related potentials and event-related oscillations are exploited to elucidate the time-course and nature of speech act recognition.

1.4.3 Practical aspects of investigating speech act recognition with EEG

EEG is a very useful technique for investigating language comprehension because it allows researchers to track brain responses to utterances as they unfold in real time, providing unique insights into the time-course of processing. There are nevertheless several constraints on the methodology that need to be considered. First, in addition to brain activity, EEG also records electrical activity arising from other sources, including the muscles of the body. For instance, movement of the head, clenching of the jaw, shoulder tension, eye movements and blinking can contaminate the EEG recording. Such *artifacts* are typically very large compared to the signals of interest and decrease the signal-to-noise ratio (Luck, 2005a). Although artifacts can be removed from the dataset by excluding contaminated trials according to certain criteria (*artifact rejection*) or correcting for their presence (*artifact correction*), it is necessary to instruct participants to move as little as possible. As a consequence, the researcher cannot simply place the EEG cap on participants' head and record their brain activity while they have a conversation; speaking and blinking would introduce too many artifacts into the signal.

Second, many trials per condition are needed to obtain a clear ERP signal. The signal-to-noise ratio increases as more trials are added to the ERP average; as a rule of thumb, at least 30 trials should be included for each condition, and for smaller components hundreds of trials are recommended (Luck, 2005b; Van Berkum, 2012). This puts constraints on the creation of stimuli, which are particularly relevant in the case of language research. Using lengthy narratives or dialogues is not feasible, keeping in mind the time required for the making of the stimuli as well as the length of the experimental session.

Third, the ERP method is sensitive to numerous factors, which may confound the manipulation of interest. These factors include, for instance, variation in word length, frequency of words, word class, semantic content and discourse context (Hagoort, 2008; Hauk & Pulvermüller, 2004; Kutas & Federmeier, 2000; Van Berkum, 2004; Van Petten & Kutas, 1990). Given the degree of variability in these factors in natural conversation, and the fact that many (similar) trials are needed for each condition, the use of recordings from natural conversation as stimuli for ERP research is very challenging, if not impossible.

How can we then investigate speech act recognition in a way that comes as close to language comprehension in conversation as possible? One solution is to use an *overhearing paradigm* in which participants listen to short, pre-designed spoken conversations (Van Berkum, 2012). This approach offers a good compromise between experimental control and ecological validity. An overhearing paradigm allows the researcher to use controlled materials, reducing noise from confounding factors. At the same time, the conversational exchanges can be modeled on natural conversation – based on findings from conversation analysis and other domains of research – to ensure a better approximation to everyday interaction. This is the approach adopted in this thesis.

1.5 Objectives of the thesis

The general aim of this thesis is to investigate the time-course of spoken speech act recognition – the core cognitive ability that gives language its basic functionality. More specifically, the main goal is to test the early speech act recognition account introduced in section 1.2.2, i.e., the hypothesis that speech acts are recognized early on in utterances, facilitating quick and efficient turn-taking. A second goal is to investigate the influence of the sequential organization of action on the time-course of speech act recognition.

The experimental paradigm developed to address these issues involves short, spoken dialogues containing target utterances that deliver three distinct speech acts depending on the prior turn. The target utterances are *identical* across conditions – and hence underspecified at the action level – they but differ in the type of speech act performed and how it fits into the larger action sequence. As a consequence, ERP differences between them can be attributed to their speech act function and how listeners arrive at that function.

The novelty of this thesis lies in two aspects. First, in contrast to the majority of EEG research on language (for overviews, see for instance Kolk & Chwilla, 2007; Kutas & Federmeier, 2000; Van Berkum, 2004), this study investigates speech act comprehension in spoken, naturalistic dialogues without any syntactic, semantic or pragmatic anomalies. Second, this thesis uses an interdisciplinary approach based on methods from cognitive neuroscience and findings from conversation analysis and other studies on talk in interaction. By modeling the experimental paradigm on everyday turn-taking, this investigation comes closer to capturing speech act recognition in its natural habitat – conversation.

1.6 Outline of the thesis

The next four chapters present the experimental research conducted for this thesis. Since the chapters are based on published articles or manuscripts in submission/preparation, some overlap between the chapters and this introduction is inevitable. However, when possible the reader is referred to prior chapters to minimize repetition.

Chapter 2 sets the stage by reporting behavioural findings from self-paced reading and comprehension tasks which form the foundation for the following EEG experiments. The aim of the behavioural experiment was twofold: to investigate how reliably participants can categorize the speech act of sentences that are underspecified at the action level, and to obtain a rough estimate of the time-course of speech act comprehension as reflected in self-paced reading times. The target stimuli were written versions of the spoken dialogues used in the remainder of the thesis. The results from an action categorization task demonstrate that participants can categorize the speech act in action-underspecified sentences with very high accuracy (95.8%), based only on the prior speech act. The reading time results indicate that speech act recognition in the visual modality begins early in the utterance, at the first word or the verb. The results demonstrate the feasibility of using the experimental paradigm in further research on speech act recognition.

Chapter 3 then proceeds to the core of the thesis, investigating the time-course of spoken speech act recognition using ERPs. How quickly can listeners recognize action-underspecified speech acts in the auditory modality? How does the sequential context influence this process? The ERP experiment presented in this chapter uses the spoken versions of the dialogues in Chapter 2 (briefly described in 1.5) and the same comprehension task (action categorization) to facilitate comparability with the behavioural results. Chapter 3 identifies three speech-act-related ERP effects. The results for an early utterance time-window, corresponding to the first word and the verb, indicate that spoken speech act recognition begins early in the turn when the utterance has only been partially processed. The results for a late utterance time-window, corresponding to the final word, demonstrate that the time-course of speech act recognition is influenced by the type of speech act and how it fits into the larger action sequence; recognition of the action can be made before the final word when the prior turn is highly constraining in terms of what action can follow, while additional processing, based on the complete utterance, is required in more complex actions.

Chapter 4 tests the robustness of the speech-act-related ERP effects that in Chapter 3 were observed using an action categorization task. In particular, it investigates whether the the ERP effects generalize to a more natural situation in which overt categorization is not required. The experiment in Chapter 4 includes a different task (true/false judgment) and additional filler dialogues, reducing strategic processing of the target speech acts. Chapter 4 replicates two of the three ERP effects reported in Chapter 3, demonstrating the robustness of these effects. However, it also reveals an influence of task and experimental design on the ERP results and thereby further refines the time-course of speech act recognition described in Chapter 3.

Chapter 5 investigates whether the pattern of results reported in Chapter 4 is supported by converging evidence from time-frequency analyses of the EEG data in the same experiment. As discussed in section 1.4.2, ERPs only reflect a certain part of the event-related EEG signal, namely phase-locked responses which remain after averaging in the time domain (Bastiaansen et al., 2008; Pfurtscheller & Lopes da Silva, 1999). By analyzing non-phase locked oscillatory activity, Chapter 5 aims to give a more complete picture of the time-course of speech act recognition. The more specific aim was to shed light on the role of anticipatory processes in speech act recognition. Overall, the results substantiate the time-course of speech act recognition as reported in Chapter 4 and provide evidence that early speech act recognition in highly-constraining contexts involves anticipatory attention before the speech act begins.

Finally, Chapter 6 presents a summary of the main empirical findings, discusses their theoretical implications and directions for future research. The thesis concludes with remaining questions that relate to speech act comprehension in a broader context.

2 Speech act recognition: Behavioural findings

This chapter is an extended and modified version of:
Gisladottir, R. S., Chwilla, D. J., Schriefers, H., & Levinson, S. C. (2012). Speech act recognition in conversation: Experimental evidence. In N. Miyake & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 1596–1601). Austin, Texas.

Abstract
Recognizing the speech acts in our interlocutors' utterances is a crucial prerequisite for conversation. However, speech act recognition can be a challenging task given that the form and content of utterances is frequently underspecified for this level of meaning. In the present study we investigate participants' competence in categorizing speech acts in such action-underspecified sentences and explore the time-course of speech act comprehension using a self-paced reading paradigm. The results demonstrate that participants are able to categorize the speech acts with very high accuracy, based on only limited context and without any prosodic information. Furthermore, the results show that the exact same sentence is processed differently depending on the speech act it performs, with reading times starting to differ already at the first word. The present results indicate that participants are very good at "getting" the speech acts, opening up a new arena for experimental research on action recognition in conversation.

2.1 Introduction

Knowing a language doesn't just require syntax or semantics, but the ability to extract speech acts from our interlocutors' utterances. This is crucial in conversation, since all actions – be they non-verbal or verbal – have implications for how we should respond (Schegloff, 2007): a greeting calls for another greeting, an offer is followed by an acceptance or declination. Scholars in Conversation Analysis were the first to reveal the systematicity of courses of action in turn-taking (e.g. Sacks et al., 1974; Schegloff, 2007). Moving away from the single act as the fundamental unit of analysis – the perspective of speech act theory (Austin, 1976; Searle, 1969) – conversation analysts pay attention to the *sequential context,* i.e., prior and upcoming turns in conversation. As discussed in section 1.1, one of the main observations from this literature is that turns tend to come in pairs of actions and in such *adjacency pairs* (Schegloff, 2007) the first (pair) part sets up powerful constraints on what type of action can follow. Moreover, not all actions have equal status. Reflecting an orientation towards *preference structure* (Schegloff, 2007), dispreferred actions such as rejections to invitations tend to be delivered with inter-turn gaps, turn-initial delay, hedges or other discourse markers (*uhm, well*), while preferred actions (e.g., acceptances) do not (Schegloff, 2007). Importantly, preference is calibrated to the specific action context; for instance, while rejections are usually dispreferred actions, they can in certain contexts be preferred over acceptances and are in such cases delivered without hedges or delays. This suggests that participants in conversation not only monitor their speech for actions but also orient to how they fit into the larger action sequence.

How do we map speech acts onto our interlocutors' utterances, bridging the gap between the literal meaning (sentence meaning; Grice, 1975) and action? In some cases this is a simple matter. In the utterance *please close the door*, for instance, the imperative mood of the verb and the adverb *please* function as "special markers" or "illocutionary force indicating devices" (H. H. Clark, 1979; Levinson, 1983; Schegloff, 2007) that clearly indicate this is a request. In most utterances, however, the absence of such dedicated vocabulary leaves the propositional content underspecified for the speech act level of meaning. As an example, the assertion *I have a credit card* can deliver different speech acts, depending on context. When responding to a question from our interlocutor (e.g., *How are you going to pay for the ticket?*), *I have a credit card* functions as an information-giving answer. If it follows an offer of payment (e.g., *I can lend you money for the ticket*), it is used to indirectly decline it. In this case it could be characterized as an indirect speech act, in which "one illocutionary act is performed indirectly by way of performing another" (Searle, 1975, p. 60). In yet another interchange, if our interlocutor has expressed the need or desire for some means of payment (e.g., *I don't have any money to pay for the ticket*), the same statement of ownership can function as a prelude to an offer, called a *pre-offer* in conversation analysis (Schegloff, 1988, 2007). In all three cases, the form and semantic content is underspecified for the action such that the full import of the utterance *I have a credit card* can only be determined relative to the context, in this case the prior speech act in the conversation.

There is some psychological evidence that people do extract speech act information online. Holtgraves (2008a) addressed whether the comprehension of a sentence like *Don't forget to go to your dentist* (an "implicit speech act") entails automatic activation of the speech act performed (reminding). Participants were presented with the critical sentence or a control, and then performed either a lexical decision task (is this string of letters a word?) or a recognition probe task (did this probe literally appear in the remark?). If, for instance, comprehension of the critical sentence does involve activation of the speech act (e.g., remind), then participants should respond faster in the lexical decision task when the target represents the speech act than when it does not. This is indeed what was found. Such priming studies indicate that people do extract speech act information from both written and spoken utterances (Holtgraves, 2008a). A further study (Holtgraves, 2008b) suggests that people recognize and retain in long-term memory the actions that people perform with their utterances. In line with speech act theory and conversation analysis, Holtgraves argues that “in conversation there is an action dimension, a dimension that does not exist for isolated sentences or texts. Speakers are usually constructing utterances with the intention to perform certain actions, and with the intention of having the recipient recognize those actions” (Holtgraves, 2008a, p. 640).

Clearly, action recognition crosscuts research on topics such as communicative intention and implicature in pragmatics (Grice, 1975; Levinson, 1983; D. Sperber & Wilson, 2004), the study of indirect speech acts (e.g., H. H. Clark, 1979; H.H. Clark, 1996; Coulson & Lovett, 2010; Gibbs, 1979) and discourse processing (e.g., Graesser, Singer, & Trabasso, 1994) in psycholinguistics, as well as research on non-verbal action understanding and theory of mind in the cognitive sciences (e.g., Baker, Saxe, & Tenenbaum, 2009; de Lange, Spronk, Willems, Toni, & Bekkering, 2008). There is limited experimental research, however, on speech act recognition in spoken dialogue. The experimental approach used by Holtgraves (2008a, 2008b) involves artificial tasks (lexical decision and recognition probe) and does not unravel the time-course of action recognition. The puzzle remains: how is it that we can extract speech acts from utterances so efficiently, as evidenced by extraordinarily fast turn transitions (Levinson, 2013; Sacks et al., 1974; Stivers et al., 2009)? Before investigating speech act recognition in spoken dialogues it is important to empirically assess whether, and if so how well, participants can extract the speech act from utterances that are underspecified at the action level.

2.2 The experiment

The aim of the present experiment was to investigate participants’ competence in identifying speech acts in action-underspecified sentences and explore the time-course of speech act comprehension. To do this we presented target sentences using the self-paced reading paradigm and asked participants to categorize the speech acts and rate how sure they were in the categorization. Self-paced reading has been used to investigate the processing of phenomena such as scalar implicatures in pragmatics (Breheny, Katsos, & Williams, 2006) and other types of inferences in text comprehension (Graesser, Swamer, & Baggett, 1996), with longer reading times (in comparison to a control) interpreted as indicating the generation of an inference. The self-paced reading paradigm allows us to obtain information on the word-by-word processing of action-underspecified utterances, thereby exploring the time-course of speech act inferencing.

The stimuli in our study consist of a context sentence which is presented auditorily, followed by a target sentence designed to be interpreted as an *Answer, Pre-Offer* or *Declination* depending on the context (see Table 2.1). These actions are commonly found in conversation and their form and function has been described in the conversation analytic literature.

The Answers in our study complete an adjacency pair by responding to a wh-question in the first turn. This condition serves as a benchmark for inferencing in the reading time analysis since the gap between literal (sentence) meaning and the action intended is the smallest. Moreover, since the other actions in the study can superficially be viewed as answers, because they respond to the prior turn, this condition provides a check on whether participants go beyond a simple characterization of the sentences as responses and identify the correct speech act.

The second action, a Declination, completes an adjacency pair by responding to a proposal (an offer or invitation) in the first turn. The Declinations in this study are relatively indirect responses, requiring a backward inference that bridges the response and the prior speech act. Conversation analysts have noted that, at least in English, indirect responses “need not be polite, nor unclear or obfuscatory. For certain activities, in specific sequential locations, responding indirectly may be the most efficient form of communication," (Walker, Drew, & Local, 2011, p. 17).

	Set 1		**Set 2**	
Condition	**Context**	**Target Utterance**	**Context**	**Target Utterance**
Answer	*Hoe ga je voor het ticket betalen?*	*Ik heb een creditcard.*	*Waar koop je je shampoo?*	*Ik ga naar de Kruidvat.*
	How are you going to pay for the ticket?	I have a credit-card.	Where do you buy your shampoo?	I go/am going to the Kruidvat [drugstore].
Declination	*Ik kan je wat geld lenen voor het ticket.*	*Ik heb een creditcard.*	*Ik kan wel shampoo voor je meenemen?*	*Ik ga naar de Kruidvat.*
	I can lend you money for the ticket.	I have a credit-card.	I can bring some shampoo for you?	I go/am going to the Kruidvat [drugstore].
Pre-offer	*Ik heb geen geld om het ticket te betalen.*	*Ik heb een creditcard.*	*Mijn shampoo is op.*	*Ik ga naar de Kruidvat.*
	I don't have any money to pay for the ticket.	I have a credit-card.	My shampoo is finished.	I go/am going to the Kruidvat [drugstore].

Table 2.1: Stimuli in Dutch and English translations.

The third action is the Pre-Offer, which belongs to a type of *pre-sequence* (Schegloff, 1988). Pre-sequences are preliminary to, or project, the main course of action – in this case an offer (Schegloff, 1988, 2007), as demonstrated in the following example:

Bookstore, 2.1: 107 (modified from Schegloff, 2007, p. 35)

1 A:	I'm gonna buy a thermometer though because I
2 B:	but
3 A:	think she's got a temperature
4 C: Pre-Offer	we have a thermometer
5 A: Go-ahead	you do?
6 C: *Offer*	want to use it?
7 A: *Acceptance*	yeah

Only if the response to the pre-offer is positive (line 5) is the offer put forward (line 6). This strategy allows conversationalists to check whether an offer would be welcome or not, preventing them from embarrassment that would arise if an offer were to be rejected.

Crucially, the Pre-Offers differ from the Answers and Declinations in that they do not complete an adjacency pair but rather open up or "project" a continuation of the sequence. Understanding Pre-offers involves knowing that a direct offer is underway; in this sense they involve a forward directed inference (projection). By including Pre-Offers in our study we can explore whether the distinction between projection (Pre-Offers) and a backward directed inference (Declinations) is borne out in reading times.

Given that the same sentence can be used as an Answer, Declination or Pre-offer depending on the sequential context, in this study we investigate: 1) Can participants reliably categorize action-underspecified speech acts? 2) Does the time-course of speech act comprehension differ for these actions as reflected in self-paced reading times? Due to lack of research in this area, in particular on Pre-Offers, we do not make specific predictions regarding reading times. However, we speculate that the reading time pattern of Pre-Offers and

Declinations will differ relative to Answers, based on the structural properties described above.

2.3 Methods

2.3.1 Participants

39 native speakers of Dutch were recruited from the student population in Nijmegen, Netherlands. Participants were paid 8 euros for participating.

2.3.2 Materials and Design

We created 378 two-sentence long, naturalistic dialogues in Dutch reflecting informal daily conversations between friends or relatives. The discourse topics include buying groceries, going out, and working/studying. Each dialogue consists of a target sentence and a preceding context sentence, which biases the interpretation of the target as an Answer, Declination or Pre-Offer (see examples in). The target sentence was presented visually, one word at a time in self-paced reading, while the context utterance was auditory. In total there were 126 target sentences, presented in three contexts (conditions).

To maintain a balance of variety and control in the stimulus materials, half of the target sentences started with “I have” (Dutch *ik heb*), e.g. “I have a credit card” (Set 1). The other half was more varied and included simple utterances like “I am going to the market” and “My brother is a mechanic” (Set 2). We varied the length of the sentences to make the stimuli as natural as possible, but constructed the target sentences such that the final word is critical for understanding the propositional content of the utterance (irrespective of speech act level meaning). In line with the reported characteristics of indirect replies (Walker et al., 2011), the target sentences do not involve ellipsis or pronominalization. To maintain consistency in the way the Declinations and Pre-offers are connected to their contexts, we ensured that there was at least one clear implicated premise and an implicated conclusion for each sentence pair: when presented with an utterance that is indirect, the hearer needs to access an implicated premise and combine it with the proposition expressed to derive the implicated conclusion (Blakemore, 1992). In the dialogue *(A) I can lend you money for the ticket. – (B) I have a credit card*, the implicated premise is that a credit card can pay for things, including tickets. The implicated conclusion is that speaker B does not need A's help with paying for the ticket.

In order to get a measure of the semantic relatedness between the context and the target sentence in each condition, Latent Semantic Analysis (LSA) values (Landauer, Foltz, & Laham, 1998) were computed for the English translation of each sentence pair using document-to-document mode with "General reading up to 1st year of college" as the semantic space. The average LSA values for each condition were: Answers 0.13, Declinations 0.33, and Pre-Offers 0.42 (the higher the value, the more semantic similarity).

The stimuli were translated from English into Dutch and checked by two native speakers of Dutch. The sentences were recorded by four native speakers, two male and two female. The recordings of the target sentences were not used in this experiment, since they are presented visually in self-paced reading. The stimuli were pseudo-randomized and balanced across three lists, such that participants saw each target sentence only once, in one context. After each trial (sentence pair), participants were given a comprehension and rating task. They were first asked to indicate what the second speaker was doing with his response and were given the options of Answering, Offering and Declining (in Dutch: *antwoorden*, *aanbieden*, *weigeren*). Since pre-offer is not a colloquial term, the broader term of offering was chosen. Participants were then asked to rate how sure they were in their categorization decision on a rating scale from 1 (very uncertain) to 7 (very certain). The purpose of the rating task was to assess the feasibility of using the items in future studies.

2.3.3 Procedure

Participants were given instructions that included one example of each action. They were instructed to imagine that they were listening to a conversation between friends or colleagues, and to read the sentences as quickly as possible, but not too quickly as they would have to "judge the underlying meaning" of the sentences. They were then seated in a chair in front of a monitor in a soundproof experimental booth. On each trial the context utterance was played while a small picture of a loudspeaker was presented in the middle of the screen. After the offset of the spoken sentence a blank screen was presented for 500 ms, followed by the presentation of the target sentence in a moving window self-paced reading format (Just, Carpenter, & Woolley, 1982). A series of lines appeared on the screen representing each word in the target sentence. When participants clicked on the mouse the first word appeared and upon subsequent button presses a new word was shown, while the previous word was again replaced by a line. When participants clicked the mouse after the last word had been shown, they were presented with the action categorization question, immediately followed by the certainty rating. There were 126 experimental trials, preceded by a brief practice session.

2.4 Results

2.4.1 Accuracy

Overall accuracy (number of correct responses in the action categorization task divided by the total number of responses) was very high, 95.8 percent. Accuracy percentages (summarized in Table 2.2) were very similar across conditions. The accuracy data were analyzed with mixed-effects logistic regression in the statistics software R (R Core Team, 2013) using the lme4 package (D. Bates, Maechler, & Bolker, 2012). Mixed-effects logistic regression is better suited for the analysis of categorical outcome variables (such as question-answer accuracy) than ANOVA and allows the inclusion of participants and items as random factors in a single analysis (for a discussion, see Jaeger, 2008). The fixed effects were Action and Set (for a description of the two stimulus sets, see 2.3.2). Set was included to check whether there is an interaction between the Action and the linguistic form of the sentence (Set). We used the most maximal random effects structure justified by the experimental design and for which convergence was reached (Barr, Levy, Scheepers, & Tily, 2013). This included random intercepts by participant and item, as well as by-participant random slopes for Action and Set and by-item random slopes for Action. For model comparison, see Appendix (section 8.1). The full model with Action, Set and the Action × Set interaction indicated that there were no differences in accuracy between the three actions ($ps > 0.14$). However, an Action × Set interaction was present. Pre-offers in Set 1 were categorized more accurately than those in Set 2 (Estimate: 1.15, SE: 0.28, $z = 4.06$, $p < .001$), while an opposite pattern was observed for Declinations and Answers; Answers in Set 1 were categorized slightly less accurately than in Set 2 (Estimate: -1.15, SE: 0.29, $z = -3.97$, $p < .001$), and Declinations in Set 1 were also categorized less accurately than in Set 2 (Estimate: -1.13, SE: 0.29, $z = -3.89$, $p < .001$). The Action × Set interaction suggests there are subtle differences in how salient the action of the target sentences is, depending on their linguistic form.

	Answer	Declination	Pre-offer
Overall Accuracy	96.0% (4.0%)	96.0% (6.1%)	95.6% (5.4%)
Set 1 Accuracy	95.2% (3.8%)	94.7% (7.8%)	98.0% (2.6%)
Set 2 Accuracy	96.7% (5.6%)	97.2% (5.1%)	93.2% (9.6%)
Overall Certainty	6.60 (0.36)	6.50 (0.39)	6.35 (0.51)
Set 1 Certainty	6.58 (0.37)	6.52 (0.42)	6.50 (0.45)
Set 2 Certainty	6.61 (0.43)	6.48 (0.41)	6.21 (0.62)

Table 2.2: Accuracy and mean certainty ratings. Standard deviations (in brackets) are based on data aggregated over participants for comparability between accuracy and certainty tasks.

2.4.2 Certainty ratings

Participants rated how certain they were in answering the action categorization question on a scale from 1 (very uncertain) to 7 (very certain). The overall mean certainty rating was 6.48 (SD 1.03). Mean certainty ratings for each condition and set are summarized in Table 2.2. A repeated-measures ANOVA on the mean ratings revealed a main effect of Action ($F(2, 76) = 11.20$, $p < .001$), a main effect of Set ($F(1, 38) = 9.66$, $p < .01$) and an Action $\times$ Set interaction ($F(2, 76) = 10.63$, $p < .001$). Pairwise comparisons between the actions indicated that certainty in categorizing the speech acts was lower for Pre-Offers than both Answers (Action; $F(1, 38) = 13.69$, $p < .01$) and Declinations (Action; $F(1, 38) = 14.98$, $p < .001$), while the difference between Declinations and Answers was not significant ($F(1, 38) = 4.07$, $p = .051$). A main effect of Set was present in the comparison between Pre-offers and Answers ($F(1, 38) = 9.89$, $p < .01$) and Pre-offers and Declinations ($F(1, 38) = 17.26$, $p < .001$), reflecting that certainty ratings were higher in Set 1. However, Action $\times$ Set interactions were also present in both comparisons (Pre-offers and Answers: $F(1, 38) = 16.66$, $p <$ .001, Pre-offers and Declinations: $F(1, 38) = 14.00$, $p < .01$). These interactions reflected that the main effect of Set was mainly driven by differences in certainty across Sets in Pre-offers; Pre-offers in Set 1 (M = 6.50, SD = 0.45) were rated higher than in Set 2 (M = 6.21, SD = 0.62) ($p < .001$).

2.4.3 Reading times

The time between button presses was recorded as the reading time for each word. Extreme values below 100 ms were excluded, as well as values above 1200 ms for non-final words and above 7000 ms for final words. In total 7 outliers (0,1%) were removed. Since online speech comprehension and the subsequent off-line categorization task tap different types of information, error trials were not excluded from the reading time analysis.

Mean reading times for the first word, the verb and the final word of the target sentences were used for the analysis, in addition to the mean reading time per word (sentence reading time divided by number of words) and the mean reading time of the entire sentence (see Table 2.3). Reading times were analyzed with repeated-measures ANOVA with Action and Set as factors. Since a main effect of Set is not of theoretical relevance, only main effects of Action and Action × Set interactions are reported in follow-up comparisons between the action conditions.

		Answer	Declination	Pre-Offer
First Word	Mean RT	251	252	259
	SD	56	57	65
Verb	Mean RT	260	267	265
	SD	60	69	70
Final Word	Mean RT	564	622	593
	SD	384	447	412
Word[3]	Mean RT	339	354	352
	SD	135	146	147
Sentence	Mean RT	1459	1528	1501
	SD	584	652	603

Table 2.3: Mean reading times in ms.

First word

Reading times started to differ already at the first word, reflected by a main effect of Action ($F(2,76) = 5.25$, $p < .01$). There was no main effect of Set or an Action × Set interaction (Fs < 0.68, ps $> .51$). First word reading times were slightly longer in Pre-Offers compared to Answers (Action; $F(1,38) = 7.07$, $p < .05$), and in Pre-offers compared to Declinations (Action; $F(1,38) = 5.84$, $p < .05$). The comparison between Answers and Declinations was not significant ($F(1,38) = 0.12$, $p = .73$).

Verb

There was a main effect of Action ($F(2, 76) = 3.41$, $p < .05$) and a main effect of Set ($F(1, 38) = 10.45$, $p < .01$), but no Action × Set interaction ($F(2,76) = 0.43$, $p = .65$)[4]. Pairwise comparisons revealed that verb RTs were longer in Declinations than in Answers (Action; $F(1,38) = 6.31$, $p < .05$). The

[3] RT of the entire sentence divided by number of words.

[4] The main effect of Set reflected that reading times were longer in Set 2 than in Set 1 at the verb, both in the overall analysis and in follow-up comparisons between the actions, where a main effect of Set was also present.

comparison between Pre-Offers and Answers was not significant ($F(1,38) = 2.97$, $p = .09$), nor between Pre-Offers and Declinations ($F(1,38) = 0.67$, $p = .42$). No Action $\times$ Set interactions were present in the pairwise comparisons (Fs < 0.8, ps $> .37$).

Final word

A main effect of Action was present in the overall analysis ($F(2,76) = 4.27$, $p < .05$), but there was no main effect of Set ($F(1,38) = 0.39$, $p = .54$) nor Action $\times$ Set interaction ($F(2,76) = 2.69$, $p = .09$). Pairwise comparisons revealed that final RTs were longer in Declinations than in Answers (Action; $F(1,38) = 6.16$, $p < .05$), and RTs in Pre-offers tended to be longer than in Answers (Action; $F(1,38) = 3.75$, $p = .06$). There were no differences between Pre-Offers and Declinations ($F(1,38) = 2.06$, $p = .16$).

Word

An ANOVA on mean RTs per word revealed a main effect of Action ($F(2, 76) = 5.73$, $p < .01$), but no main effect of Set ($F(1,38) = 1.50$, $p = .23$) nor Action $\times$ Set interaction ($F(2,76) = 2.60$, $p = .09$). Mean RTs were shorter in Answers than in Pre-offers (Action; $F(1,38) = 6.88$, $p < .05$), and shorter in Answers than in Declinations (Action; $F(1,38) = 8.30$, $p < .01$). The difference between Pre-Offers and Declinations was not significant ($F(1,38) = 0.25$, $p = .62$).

Sentence

Although the reading times for the entire sentence differed descriptively between action conditions, these differences were not significant ($F(2, 76) = 2.73$, $p = .09$). While there was a main effect of Set ($F(1,38) = 24.12$, $p < .01$), no Action × Set interaction was present ($F(2,76) = 2.15$, $p = .13$)[5].

2.5 Discussion

The present experiment demonstrates that participants categorize the speech acts of sentences whose form and semantic content is underspecified for action with very high accuracy (95.8%). They are able to do so based on limited context (the prior speech act) and without any prosodic information in the target sentence. Importantly, overall accuracy was the same for all three actions (Answer, Declination, Pre-Offer). If participants had processed the target sentences superficially, ignoring the speech act content, they could have categorized Declinations and Pre-Offers as Answers. This is the case since the Dutch term for answering (antwoorden) also means to respond and all three speech acts can superficially be seen as responses. This should have resulted in lower accuracy for Pre-Offers and Declinations vis-à-vis Answers. The high accuracy rate across actions shows that participants go beyond a simple characterization of the target sentences as responses and "get" the correct action. Participants were also very confident in categorizing all actions and rated the certainty of their categorizations on average 6.48 (out of 7). These results provide further support that participants are confident in orienting to the action content of sentences.

The reading time results demonstrate that the exact same sentence is processed differently depending on the speech act it performs. Declinations and

---

[5] The main effect of Set reflected that reading times were longer in Set 2 than in Set 1, both in the overall analysis and in follow-up comparisons between the actions, where a main effect of Set was also present.

Pre-Offers have different trajectories relative to Answers, which have shortest RTs on all measures. In the case of Pre-offers, reading times were longer than both Answers and Declinations at the first word. RTs also tended to be longer in Pre-offers relative to Answers at the final word. It should be pointed out that descriptively the difference between the means at the first word is small and standard deviation large. Event-related brain potentials might be a more sensitive measure to reveal processing differences at early positions in the sentence. In the case of Declinations, RTs were longer than in Answers at the verb and the final word, while there was no difference between Declinations and Answers at the first word.

Based on differences in sequence organization, we speculated that Declinations and Pre-Offers would exhibit different reading time patterns relative to Answers. This is indeed what we found; reliable differences were found at the first word for Pre-offers relative to Answers, but at the verb and the final word for Declinations. Recognizing the speech act in the Declinations requires a backward inference, connecting the incoming sentence to the prior proposal. Moreover, since Declinations close an action sequence (the adjacency pair), they do not heavily constrain the relevant next action. Declinations therefore do not invite strong predictions about the upcoming speech act. In contrast, Pre-offers initiate a new action sequence, containing a subsequent direct offer if the conversation were to continue. Thus although understanding Pre-offers inevitably requires integration with the prior turn, it also involves a forward directed inference about the continuation of the sequence. The distinction between an inference based on a backward bridge to the prior turn in an adjacency pair and an inference based on forward projection of a sequence is akin to the difference between causal antecedent and causal consequence inferences in text processing (Magliano, Baggett, Johnson, & Graesser, 1993). The reading time differences between Pre-Offers and Declinations provide some indication that the distinction between forward projection and a backward directed inference plays a role in the online processing of speech acts. However, it is not clear why the projective nature of Pre-Offers would call for more processing at the first word, whereas the backward inference in Declinations requires processing at the verb and the final word. More research is needed to investigate whether this finding holds for spoken language processing as well.

The early effects of action at the first word and the verb suggest that comprehension at the speech act level starts before the full propositional content has been presented. Familiarity with action sequences is one factor that could enable early speech act comprehension, making the fast transitions between turns (Levinson, 2013; Sacks et al., 1974; Stivers et al., 2009) possible. Neuroimaging studies suggest that people use their knowledge of the wider discourse context to predict specific upcoming words and that prediction is not the result of relatively low-level, word-based priming mechanisms, "but involves a more sophisticated message-level mechanism that can take into account the actual nuances of the preceding discourse," (Otten & Van Berkum, 2008, p. 485). This could be the case for sequential context as well. Whether and how implicit knowledge of the organization of actions guides the interpretation of utterances is a topic for further investigation.

An alternative explanation for the reading time results is that the experimental manipulation does not address speech act recognition per se, but some other confounding variable such as semantic priming from the context. Latent Semantic Analysis can be used to determine semantic relatedness of two texts and LSA similarity relations have been found to correspond well with the pattern of results in priming studies (Dorothee J. Chwilla & Kolk, 2002; Landauer et al., 1998). If semantic priming from the context is the main factor driving the reading times one would expect the condition with the lowest LSA value (least amount of priming) to have the longest mean reading times. A pattern opposite to what a priming account predicts was found: Answers had the lowest average LSA value but the shortest reading times on all measures. This suggests that the differences in reading times across conditions in our study were not due to lexico-semantic relationships between the content words of the context and the target sentences.

An additional finding from this study is that the linguistic form of the sentence (i.e., Set 1 vs. Set 2) seems to have a subtle impact on how accurately participants can categorize the speech acts and how confident they are in the categorization. This is not very surprising, since some formats may be more common than others in certain actions. However, the accuracy and certainty ratings were high across the board, with the lowest values 93.2% and 6.21 respectively (both for Pre-offers in Set 2). Moreover, the interaction between the action condition and linguistic form (Action by Set) was minimal in reading times. These results indicate that differences between the two stimulus sets are not substantial, which is important for future use of the experimental materials. Also relevant for further research is that comprehension of the speech acts seems to be more complex in Declinations and Pre-offers than in Answers, as reflected in longer reading times. This validates the use of the Answers as a control condition.

The self-paced reading paradigm used in this study has several limitations compared to other methods such as event-related brain potentials. First, while self-paced reading data only provide a rough measure of how long the processing of a particular word (e.g., the first word) takes, ERPs can indicate with much higher temporal precision, at the level of milliseconds, how soon differences between conditions emerge after a word is presented. Secondly, ERPs in contrast with self-paced reading can provide information about whether there are quantitative or qualitative processing differences involved (Kutas, 1993; Rugg, Schloerscheidt, & Mark, 1998). Most importantly, self-paced reading (and self-paced listening) cannot inform us about the realtime processing of continuous, spoken language, the prime modality of everyday conversation.

2.6 Conclusions

In this study on speech act comprehension we investigated the processing of sentences that perform different speech acts depending on prior context. In each case an assertion is used as a vehicle for some other action, and it is "part of competent membership in the society/culture and being a competent interactant to analyze assertions of this sort for what (else) they may be doing at this moment, at this juncture of the interaction, in this specific sequential context" (Schegloff, 2007, p. 35). Our study tapped into this competence by addressing two primary questions: how reliably participants can categorize action-underspecified speech acts, and whether the time-course of speech act comprehension differs for the actions as reflected in self-paced reading times. Participants in our study categorized the speech acts with very high accuracy, based only on limited context (the prior turn). This is striking given that the target speech acts did not contain any illocutionary force indicating devices nor prosodic information to aid recognition. Furthermore, the exact same sentence was processed differently depending on the speech act it performed, with reading times starting to differ already at the first word. These findings open up a new arena for experimental research on speech act recognition in conversation.

As a crucial component of social behavior, communication involves actions. Being a competent member of society must require a cognitive architecture that is oriented to speech acts. Having demonstrated that participants orient to the action content of sentences and can categorize speech acts with high accuracy, the next experimental step is to shed light on this ability in spoken dialogues – the foundation of doing things with words.

3 Conversation electrified: ERP correlates of speech act recognition

This chapter is a modified version of:
Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLOS ONE*, 10(3), e0120068.

Abstract
The ability to recognize speech acts (verbal actions) in conversation is critical for everyday interaction. However, utterances are often underspecified for the speech act they perform, requiring listeners to rely on the context to recognize the action. The goal of this study was to investigate the time-course of auditory speech act recognition in action-underspecified utterances and explore how sequential context (the prior action) impacts this process. We hypothesized that speech acts are recognized early in the utterance to allow for quick transitions between turns in conversation. Event-related potentials (ERPs) were recorded while participants listened to spoken dialogues and performed an action categorization task. The dialogues contained target utterances that each of which could deliver three distinct speech acts depending on the prior turn. The targets were identical across conditions, but differed in the type of speech act performed and how it fit into the larger action sequence. The ERP results show an early effect of action type, reflected by frontal positivities as early as 200 ms after target utterance onset. This indicates that speech act recognition begins early in the turn when the utterance has only been partially processed. Providing further support for early speech act recognition, actions in highly-constraining contexts

did not elicit an ERP effect to the utterance-final word. We take this to show that listeners can recognize the action before the final word through predictions at the speech act level. However, additional processing based on the complete utterance is required in more complex actions, as reflected by a posterior negativity at the final word when the speech act is in a less-constraining context and a new action sequence is initiated. These findings demonstrate that sentence comprehension in conversational contexts crucially involves recognition of verbal action which begins as soon as it can.

3.1 Introduction

Just like other aspects of social behaviour, conversation involves verbal actions such as requests, greetings and complaints (Schegloff, 2007). The exchange of such speech acts in everyday conversation is the core ecology for language – this is where children acquire language and the great bulk of language usage occurs. The prime task in conversation is to recognize what speech act is being performed and prepare a fitted reply.

Speech act recognition, the topic of this study, is the process of recognizing the action of an utterance in a given context. As discussed in the introduction to the thesis, speech act recognition may be quite direct in some cases, particularly in ritualized expressions where there is a one-to-one correspondence between the form of the utterance and its speech act function (like *Gesundheit!* in response to a sneeze). Illocutionary force indicators (see 1.2.1) can likewise provide clues about the relevant speech act. However, listeners cannot always rely on such morphosyntactic clues. For example, a declarative utterance like *I have a car* could be used to offer somebody help with moving, to indirectly reject an offer for a ride, or to answer a question about commuting. The utterance is thus compatible with multiple speech acts and listeners have to rely on the context to recognize which action is being performed. This problem of underspecification at the action level is pervasive in everyday conversation.

The challenge for participants in regards to speech act recognition in conversation is further enhanced by the very tight time constraints in turn-taking, as discussed in 1.2.2. The most frequent gaps between turns in conversation are only around 200 ms (De Ruiter et al., 2006; Heldner & Edlund, 2010; Stivers et al., 2009). A gap of 200 ms does not leave much time for both recognizing the action in the prior turn and planning a response to it. Findings from word production experiments indicate that it takes people at least 600 ms just to plan a one-word utterance (Indefrey & Levelt, 2004) and even longer for sentences with multiple words (Schnur et al., 2006). These timing facts imply that listeners start planning their responses before the prior speaker has finished speaking, and since the appropriateness of the response crucially depends on recognizing the speech act, they suggest that speech act recognition must be an early process (Levinson, 2013). On this *early speech act recognition account*, recognition of the action is not made at the final stage in the comprehension process, occurring at the last word of incoming utterances, but takes place early on when the turn has only been partially processed. The primary aim of the present study was to test the early speech act recognition account, investigating the time-course of verbal action understanding.

Given that the turn construction doesn't always give us indications about what speech act is coming up, how could early recognition of the action be made? One source of information that can constrain the speech act possibility space is the sequential context, i.e., preceding turns in the conversation. Speech acts do not exist in a vacuum; they are coherently organized into larger action sequences (sequence organization; see Schegloff, 2007). One of the basic action sequences is the adjacency pair (Schegloff, 2007; Schegloff & Sacks, 1973), where the first utterance puts powerful constraints on the following turn. The first part in an adjacency pair "projects a prospective relevance, and not only a retrospective understanding. It makes relevant a limited set of possible second pair parts, and thereby sets some of the terms by which a next turn will be understood" (Schegloff, 2007, p. 16). Prior talk in conversation is thus not merely the background to speech act comprehension, but rather contains a rich structure of action sequences that could proactively funnel possible interpretations of upcoming talk. If this notion is correct, participants in conversation can profit from implicit knowledge of action sequences during speech act comprehension.

The goal of the present experiment was twofold: a) to investigate the time-course of speech act recognition in utterances that are underspecified for the action, and b) examine the effects of sequential context on speech act recognition. The larger theoretical relevance of this investigation is that it addresses the fundamental problem of how listeners map speech act functions onto underspecified utterances – a core cognitive ability that gives language its basic functionality.

3.1.1 Prior research

Several domains of research have addressed speech act comprehension broadly construed. For a short review of the literature, the reader is referred to section 1.3 in the introduction to the thesis. Most relevant for the present investigation, a growing body of research has made use of the excellent temporal resolution of EEG and MEG to investigate pragmatic language comprehension in real time. Studies on non-conversational discourse have shown that spoken or written words are related to the wider discourse context extremely rapidly, from 150 ms after word onset (see, for instance, Van Berkum et al., 1999; Van Berkum, Zwitserlood, et al., 2003). However, research on pragmatic inferencing – which may play a role in speech act recognition – indicates that language processing at the level of pragmatics is not always so fast. For instance, understanding irony involves late inferential processes, reflected by modulations of the P600 component (Regel et al., 2011; Spotorno et al., 2013), and scalar inferences are associated with N400 effects only in some cases (Nieuwland, Ditman, & Kuperberg, 2010; see also Noveck & Posada, 2003). The above studies highlight that pragmatic language comprehension involves both early and late processes. Importantly, they do not address comprehension at the speech act level, i.e., the processing of utterances in conversational contexts, so their relevance for the time-course of speech act recognition is unclear.

Turning to research that comes closer to dialogue, a recent EEG study investigated speech act processing in written words that performed either a requesting or a naming speech act depending on a prior video-taped context sentence (e.g., *What are these called?/What can I get you? – PLANT*) (Egorova, Shtyrov, et al., 2013). The brain responses for the two speech acts diverged as early as 120 ms after the onset of the critical words. A follow-up MEG study (Egorova, Pulvermüller, et al., 2013) reported that the requests engaged comprehension systems in the right hemisphere within 100 ms after word onset, followed by theory of mind activations in the medial prefrontal and temporo-parietal areas from 200 to 300 ms. Naming speech acts, on the other hand, activated brain areas involved in lexico-semantic retrieval from 100 to 150 ms. In a study on visually presented (i.e., written) indirect requests (e.g., *My soup is too cold to eat* in a restaurant context), differences between indirect requests and literal statements were found from the second word onwards, but no ERP differences were present at the final word (Coulson & Lovett, 2010). These EEG and MEG findings provide some supportive evidence for the early speech act recognition account, at least in the visual modality. However, these studies are far removed from conversation and overlook the importance of spoken language input. When a sentence is visually presented, one word at a time, the linguistic signal is artificially spread out over a much longer time period than in spoken language (or natural reading for that matter). This is problematic for investigating the time-course of speech act recognition, as effects of sentence-level factors such as speech act function may not be confined to a single word. Although a few studies have indirectly addressed speech act comprehension by investigating processing of prosody in question-answer dialogues (e.g., Dimitrova et al., 2012; Magne et al., 2005; Wang et al., 2011), to our knowledge there are no studies directly investigating the time-course of speech act processing using auditory stimuli. A critical next step therefore is to shed light on speech act recognition in spoken language, the prime modality of natural conversation.

3.1.2 The present study

Against this background, several important questions need to be addressed. What is the time-course of speech act recognition in the auditory modality? In light of the extraordinarily fast transitions between turns in conversation, how quickly can listeners recognize the speech act in action-underspecified utterances? How does the type of action and how it fits into the action sequence influence this process? We used ERPs to investigate these issues, using spoken dialogues approximating informal everyday conversation.

The experimental paradigm was briefly introduced in sections 1.5 and 2.2 of this thesis, but will be described in more detail below. The paradigm was designed based on the following criteria. First, in order to examine how listeners recognize speech acts in utterances that are underspecified for the action, the critical utterances do not contain morphosyntactic speech act clues such as question words or imperative verbs. Secondly, to get a better understanding of the role of sequential context, we used speech acts that differ in how they fit into the larger action sequence (for details, see below). Third, since passively overhearing a dialogue is quite different from taking an active part in one (see, for instance, Schober & Clark, 1989), we use an action categorization task to mimic the response demands and attention level necessary for everyday interaction. While the absence of a behavioural task is appropriate for ERP studies on passive reading or listening to non-conversational discourse, the crucial task in conversation is comprehending-for-responding. As we have argued above, a critical part of that task is identifying the speech act category of the incoming turn (see also Levinson, 2013).

For convenience, examples of stimuli are presented again in Table 3.1 below. The dialogues contain target utterances (e.g., *I have a credit-card*) that deliver three functionally distinct speech acts (Answer, Pre-offer, Declination) depending on the prior turn. The Answer condition involves a question-answer sequence (*How are you going to pay for the ticket? - I HAVE A CREDIT-CARD*). The Answers serve as the control condition as they should be easiest to comprehend. This assumption is supported by the results of the self-paced reading study in Chapter 2, in which reading times were shortest for Answers on all measures (see also Gisladottir, Chwilla, Schriefers, & Levinson, 2012). The Declination condition consists of an offer, followed by a rejection (*I can lend you money for the ticket. – I HAVE A CREDIT-CARD*). The Pre-offer condition contains a first turn expressing need or desire for something, followed by a prelude to an offer, called a pre-offer (Schegloff, 1988, 2007) in conversation analysis (*I don't have any money to pay for the ticket. – I HAVE A CREDIT-CARD*). To balance control and variety in the format of the critical utterances, we divided the dialogues in two stimulus sets (see and Methods in 3.2); Set 1 contains utterances starting with *I have…* but Set 2 includes other verbs (see Table 3.1 and Methods section below). In none of the conditions can listeners rely on clues in the utterance to recognize the speech act. Instead, it is the sequential context, that is the prior turn, which determines the action. Since the target sentences are identical across conditions, ERP differences between them can be attributed to their speech act function and how listeners arrive at that function. This design allows an investigation of speech act recognition in spoken dialogues that do not contain semantic or pragmatic anomalies.

	Set 1		Set 2	
Condition	**Context**	**Target Utterance**	**Context**	**Target Utterance**
Answer (Control)	*Hoe ga je voor het ticket betalen?*	*Ik heb een creditcard.*	*Waar koop je je shampoo?*	*Ik ga naar de Kruidvat.*
	How are you going to pay for the ticket?	I have a credit-card.	Where do you buy your shampoo?	I go/am going to the Kruidvat [drugstore].
Declination (Context highly-constraining + target utterance ends the sequence)	*Ik kan je wat geld lenen voor het ticket.*	*Ik heb een creditcard.*	*Ik kan wel shampoo voor je meenemen?*	*Ik ga naar de Kruidvat.*
	I can lend you money for the ticket.	I have a credit-card.	I can bring some shampoo for you?	I go/am going to the Kruidvat [drugstore].
Pre-offer (Context less constraining + target utterance starts a new sequence)	*Ik heb geen geld om het ticket te betalen.*	*Ik heb een creditcard.*	*Mijn shampoo is op.*	*Ik ga naar de Kruidvat.*
	I don't have any money to pay for the ticket.	I have a credit-card.	My shampoo is finished.	I go/am going to the Kruidvat [drugstore].

Table 3.1: Examples of stimuli in Dutch and English translations

An important aspect of the design is that while the critical actions – Declinations and Pre-offers – are both relatively indirect, they differ in how they fit into the larger action sequence. Declinations are second parts of adjacency pairs, which entails that the context turn (first part of the pair) should be highly constraining in terms of what type of action can follow. An offer, for instance, sets up a normative expectation for an acceptance or declination. Out of the vast possibility space for speech acts in conversation, the prior turn has narrowed the likely actions down to two. This is not the case for Pre-offers. Although typically responding to a telling of some trouble, pre-offers do not close an adjacency pair but rather initiate a new sequence, a so-called *pre-sequence* (Schegloff, 1988, 2007). Pre-sequences are preliminary to the main course of action, in this case a more direct offer if the conversation were to continue. This is illustrated below (for a similar example from a real conversation, see Schegloff, 2007):

1 A:		I don't have any money to pay for the ticket.
2 B:	Pre-Offer	I have a credit card.
3 A:	Go-ahead	You do?
4 B:	Offer	Want to use it?
7 A:	Acceptance	Yeah.

Understanding Pre-offers may therefore involve forward inferences about upcoming talk, akin to causal consequence inferences in text processing (Magliano et al., 1993). An additional difference between Pre-offers and Declinations is that the context turn is less constraining in Pre-offers. The first turn in the Pre-offer dialogues can be followed by a large number of actions; there is no normative expectation for a Pre-offer. The utterance *I don't have any money* could, for instance, be followed by responses such as condolences (*Oh dear*, *That sucks*), a telling of one's own experience (*Me neither*), or a suggestion (*Why don't you ask somebody for a loan?*); a direct offer or a pre-offer are just two possibilities. Thus by comparing Pre-offers and Declinations we can go beyond the traditional distinction between direct and indirect speech acts and investigate how speech act recognition is modulated by the type of action being performed and how it fits into the sequential context (high- vs. low-constraining context, end of an action sequence vs. a start of a new one).

In Chapter 2 I reported a behavioural study using the same dialogues, in which the target replies were presented visually in self-paced reading (see also Gisladottir et al., 2012). After each dialogue, participants were asked to categorize the action in the target sentences (*I have a credit card*) as doing answering, offering or declining. The categorization accuracy was very high (95.8%), indicating that listeners are very good at identifying the speech act in such action-underspecified sentences. Reading times were longer in Pre-offers than Answers at the first word, while Declinations took longer at the verb and the final word, relative to Answers. However, the differences in reading times were very small and standard deviation large, preventing conclusive interpretations. The ERP method has advantages over self-paced reading in that it is compatible with spoken language input and can indicate with much higher precision the time-course of differences between experimental conditions. Moreover, ERPs allow to determine whether there are quantitative or qualitative processing differences between the actions (see, for instance, Kutas, 1993). Given that Declinations and Pre-offers have different properties, it is possible that they recruit in part qualitatively different comprehension processes. If so, this should be reflected by differences in polarity, morphology (wave shape) and/or scalp distribution of the ERP effects to Declinations vs. Pre-offers.

If the early speech act recognition account is correct and auditory speech act recognition takes place early in the turn when the utterance has only been partially processed, then the three speech acts should differ early on in the target utterance (*I have a credit-card*), for instance at the first word (*I*) or the verb (*have*). Moreover, there should be no ERP differences at the utterance-final word, i.e., *credit-card* (note that this is one word in Dutch, the language of the stimuli); if the action has already been recognized at that point, the processing of the final word should only add to the propositional meaning of the utterance, which for each target is the same for all three conditions and hence no differences should occur at the final word. Our predictions for the experiment were therefore as follows. We expected that both Declinations and Pre-offers, the critical conditions, would elicit ERP effects relative to the control condition (Answers) in an early time-window corresponding to the first word and the verb, and not at the final word. The critical questions are how early the effects appear and whether there are quantitative or qualitative ERP differences between Pre-offers and Declinations, reflecting that the type of action and sequential context influences speech act recognition. Given that Pre-offers have a more complex action sequence structure (the context is less-constraining and the target utterance initiates a new sequence), we hypothesized that they might elicit additional ERP effects relative to Declinations. We did not have specific predictions regarding the ERP components, given that the two visual ERP experiments on speech act recognition discussed above (Coulson & Lovett, 2010; Egorova, Shtyrov, et al., 2013) do not yield a clear picture in terms of which ERP components are involved. However, we speculated that frontal ERP effects would be involved, as both studies (Coulson & Lovett, 2010; Egorova, Shtyrov, et al., 2013) report frontal effects.

3.2 Methods

3.2.1 Participants

Forty-four right-handed speakers of Dutch with no hearing or speech problems and normal or corrected-to-normal vision were recruited from the subject database of the Max Planck Institute for Psycholinguistics in Nijmegen (28 female, 16 male, mean age 20, age range 18-27). The study was approved by the Ethische Commissie Gedragswetenschappelijk Onderzoek at Radboud University

Nijmegen. Participants gave written informed consent according to the Declaration of Helsinki prior to the study and were paid 8 Euro per hour for their participation. EEG data from two participants were removed from analysis due to excessive artifacts.

3.2.2 Construction of materials

The stimuli are auditory versions of the dialogues described in Chapter 2 (see section 2.3.2 and Gisladottir et al., 2012). Eight dialogues (out of 378) were changed for the present experiment due to low accuracy and certainty ratings in Chapter 2. The number of words in the target utterances ranges from three to seven words (median: four words), and average utterance duration is 1175 ms. The sentences were recorded in a soundproof room at 44.1 kHz sampling rate and 16-bit resolution. Four native speakers of Dutch (two male, two female) were instructed to act out the written dialogues as naturally and informally as possible in four different pairings (male1-male2; female1-female2; male1-female1; male2, female2). The partners of each pair took turns in acting context utterances and critical utterances. The context utterances were extracted from those recordings, while the critical utterances were recorded separately from a list (without context) to prevent the prosody of the critical utterance from biasing one condition over another. The overall sound intensity of the recordings was normalized to prevent loudness differences between the items. The stimuli were pseudo-randomized and balanced across three lists, such that participants heard each critical utterance only once. Each list contained 126 dialogues with an equal number of trials across conditions and stimulus sets. Care was taken that the voices of the native speakers appeared as equally as possible in each action within each list.

3.2.3 Procedure

Participants were given written instructions that included one example of each action. They were instructed to pay attention to the underlying meaning of the responses in the dialogues (the target utterances) and answer a comprehension probe: after each dialogue, participants indicated with a mouse click what the second speaker was doing with his response. For this task the options were Answering, Offering and Declining (Dutch antwoorden, aanbieden, weigeren). Since pre-offer is not a colloquial term, the broader concept of offering was chosen. Participants were seated in a comfortable chair in a soundproof room facing a computer monitor. There were 126 experimental trials presented auditorily through loudspeakers, preceded by 18 practice items. On each trial a fixation cross appeared in the middle of the screen which lasted throughout the entire dialogue. Participants were instructed to avoid eye movements and other movements during the presentation of the fixation cross. The context utterance was played 500 ms after the appearance of the fixation cross, followed by a 250 ms pause before the target recording was played. In order to prevent an abrupt start and ending of the sentences, the recordings included a 50 ms buffer before sentence onset and after offset, such that the pause between context and target was in total 350 ms. This pause is similar to average gap durations reported in corpus studies of Dutch, which range from 8 to 380 ms depending on the study (Bosch, Oostdijk, & Boves, 2005; Heldner & Edlund, 2010; Stivers et al., 2009). The fixation cross disappeared 1200 ms after the offset of the target utterance recording. A blank screen was then presented for 1500 ms, until the comprehension probe was presented. This delayed task reduced contamination of the ERPs of interest by movement related EEG activity. Upon answering the comprehension probe (see above) a blank screen appeared for 2000 ms and then the next trial began. The trials were presented in 6 blocks, allowing participants to make eye movements and rest between them. After the EEG recording, participants filled out the Empathy Quotient questionnaire (EQ; see 8.2 in Appendix) (Baron-Cohen & Wheelwright, 2004).

3.2.4 Electrophysiological recordings

The EEG was recorded with 36 active electrodes mounted in a cap (actiCap), referenced to the left mastoid. After the recording, the data were re-referenced off-line to the average of the left and right mastoids. Vertical eye movements were monitored with an electrode placed below the left eye and an electrode in the cap right above the left eye (Fp1). Horizontal eye movements were monitored through two electrodes in the cap placed approximately at the left and right outer canthi (F9 and F10). Bipolar EOGV and EOGH was computed. Electrode impedances were kept below 20 KΩ. EEG and EOG data were amplified with a bandpass filter of .02 to 250 Hz with a 10 second time constant and digitized at a sampling frequency of 500 Hz. Recording and analyses were performed with Brain Vision Analyzer.

3.3 Results

3.3.1 Behavioural analysis and results

Behavioural responses from all participants included in the EEG analysis were analyzed (42 out of 44 participants). Mean accuracy in the action categorization task was very high, 96.6% (SD 5.4%) (see Table 3.2). Participants correctly categorized 98.5% of Answers, 94.8% of Pre-offers and 96.5% of Declinations. The accuracy data were analysed with mixed-effects logistic regression using the lme4 package (D. Bates et al., 2012) in the statistics software R (R Core Team, 2013). Mixed-effects logistic regression is better suited for the analysis of categorical outcome variables (such as question-answer accuracy) than ANOVA and allows the inclusion of participants and items as random factors in a single analysis (Jaeger, 2008). The fixed effects were Action and Set; while the Set factor (see 2.3.2) was not meant to test any hypothesis of theoretical relevance, it was included in the analysis to check for interactions between action and the linguistic form of the sentence. We used the most maximal random effects structure justified by the experimental design and for which convergence was reached (Barr et al., 2013). This included random intercepts by participant and item, as well as by-participant random slopes for Action and Set and by-item random slopes for Action. For model comparison, see Appendix (section 8.2). The full model with Action, Set and the Action $\times$ Set interaction indicated that Declinations were categorized less accurately than Answers (Estimate: -1.57, SE: 0.40, z = -3.94, p < .001) and Pre-offers were categorized less accurately than Answers (Estimate: -1.48, SE: 0.43, z = -3.47, p < .001). However, the comparison between Declinations and Pre-offers was not significant (Estimate: -0.26, SE: 0.27, z = -0.94, p = .34). An Action $\times$ Set interaction reflected that Pre-offers in Set 1 were categorized more accurately than in Set 2 (Estimate: 0.84, SE: 0.38, z = 2.22, p < .05).

	Answer	**Declination**	**Pre-offer**
Overall	98.5% (2%)	96.5% (3.8%)	94.8% (8%)
Set 1	98.0% (3.7%)	96.2% (4.5%)	97.2% (5.2%)
Set 2	99.1% (2.2%)	96.9% (4.9%)	92.4% (11.9%)

Table 3.2: Behavioural results. Mean accuracy (and standard deviation) in the action categorization task for all items (overall) and for each stimulus set.

3.3.2 ERP analysis

The EEG data were averaged relative to the first word onset and the sentence-final word onset of the target utterance. For each time-locking point the EEG data were segmented into epochs of 1200 ms with a 150 ms pre-stimulus baseline. Artifacts were removed by excluding epochs with excessive EEG (>100 μV) or EOG (>75 μV) amplitude. In the final dataset of 42 participants, 17% of trials were rejected due to artifacts. The percentages of rejected trials did not differ across the three experimental conditions at the two time-locking points (Fs < 1.5, ps $> .23$). Only artifact-free trials were included in the averages.

For a separation of ERP effects elicited early in the target sentences and effects elicited at the final word, we defined two broad time-windows: an *early utterance time-window* from 100 to 600 ms after first word onset and a *late utterance time-window* from 100 to 1000 ms after final word onset. These time-windows were chosen for the following reasons. The early utterance time-window covers the duration of the first word and the verb, and therefore captures early speech act recognition effects (the target sentences involve connected speech with subjects and verbs of relatively short duration; see Figure 1). The final word of the target utterances occurs on average 490 ms after utterance onset; however, since the first 100 ms after word onset reflect exogenous, stimulus-bound ERP components which are largely insensitive to cognitive factors (C. M. Brown & Hagoort, 2000; Coulson, 2004), the first 100 ms of the final word should be very similar across experimental conditions. As a consequence, the endpoint of the early utterance time-window was set at 600 ms (roughly 490 ms + 100 ms), and the starting point of both early and late utterance time-windows was set at 100 ms after first and final word onset. Importantly, these two time-windows are not overlapping, as illustrated in Figure 3.1.

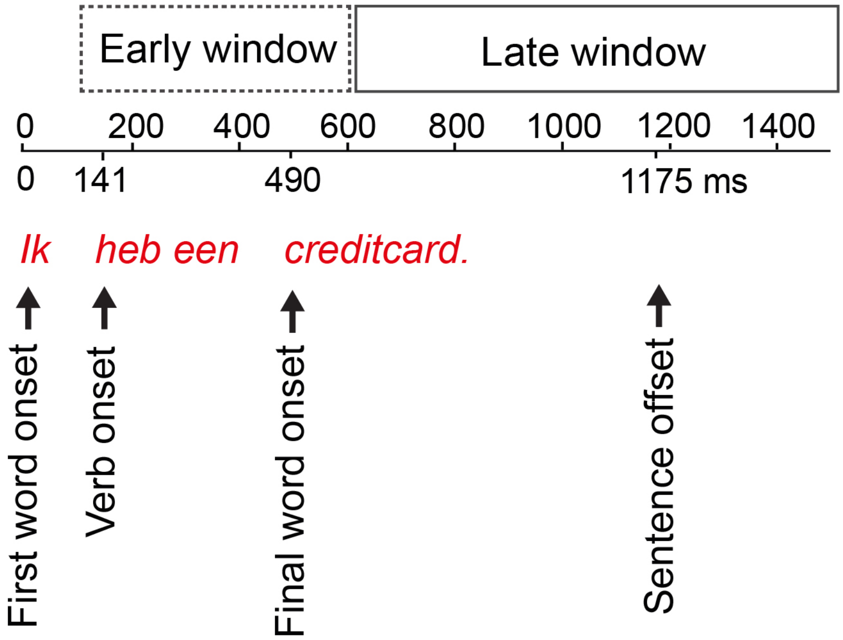

Figure 3.1: Mean onset of the first word, verb and final word in target utterances and a rough timeline for the early and late utterance time-windows. An example sentence (in red) illustrates the onset of the first word, verb and final word, but does not reflect actual word durations.

To further narrow down relevant time-windows, we first performed omnibus repeated measures ANOVAs on mean amplitudes in 100 ms consecutive windows[6] in the early and late utterance time-windows described above, using all electrodes (see Table 3.3 and Table 3.4). Based on these initial omnibus ANOVAs, follow-up analyses were performed in 100 ms consecutive windows at

[6] For a similar approach, where consecutive time-windows of 100 ms were used to get a more fine-grained picture of the time-course of language-relevant ERP effects, see for example Grainger, Lopez, Eddy, Dufau, & Holcomb (2012) and Kwon, Kluender, Kutas, & Polinsky (2013).

the first word from 100 to 600 ms, and at the final word from 100 to 200 ms and 600 to 1000 ms. The lateral ANOVA included the topographical factors AntPost (for Anterior vs. Posterior sites), Hemisphere and Site (i.e., electrode), yielding four regions of four electrodes each (for an illustration of the regions, see Figure 3.2). The medial ANOVA included the topographical factors AntPost and Site, yielding two regions of five sites each. Note that the medial ANOVA included 8 lateral electrodes and only 2 midline electrodes, and is therefore referred to as medial. Since no Action × Set interactions were obtained in the initial omnibus ANOVA (see Table 3.3 and Table 3.4), Set was removed as a factor in these regional analyses. When applicable, the Greenhouse-Geisser correction was applied to correct for violations of the assumption of sphericity; original degrees of freedom are reported. Main effects of Action and relevant interactions in the regional omnibus analysis were followed up with pairwise comparisons (ANOVA), contrasting every action with each one of the others.

Analysis	**Source**	**DF**	**100-200**	**200-300**	**300-400**	**400-500**	**500-600**
Lat	Action × Hem × Site	24, 984	2.77**	3.34**	2.33**	2.48**	2.83**
Mid	Action × Site	6, 246				3.45*	4.63**

Table 3.3: Initial omnibus analyses for the early utterance time-window. *Lat* = lateral sites, *Mid* = midline sites, *Hem* = hemisphere, **p < .05, **p < .01.*

Analysis	Source	DF	100-200	600-700	700-800	800-900	900-1000
Lat	Action × Site	24, 984		2.43*	2.68*	2.51*	2.74*
	Action × Hem	2, 82	3.43*				

Table 3.4: Initial omnibus analyses for the late utterance time-window, for those epochs that showed significant effects. *Lat* = lateral sites, *Hem* = hemisphere, **p* < *.05*.

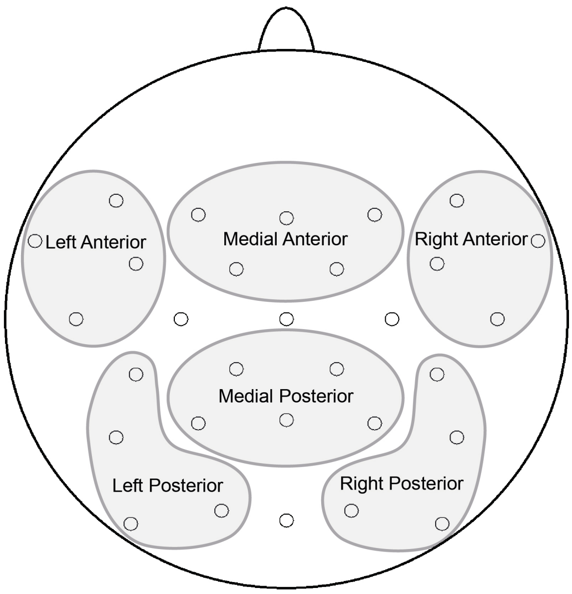

Figure 3.2: Regions used for analyses of EEG data

1.1.1 ERP results

3.3.2.1 Early utterance time-window (first word onset of the target utterance)

Figure 3.3 shows the grand averaged ERPs time-locked to the first word. All conditions elicited the early ERP components characteristic of auditory stimuli, the N1 and P2. The regional omnibus analysis at lateral electrodes (see Table 3.5) yielded Action × AntPost × Hemisphere interactions from 100 to 600 ms (all Fs $\geq$ 3.69, ps $<$ .05) and Action × Hemisphere × Site interactions from 100 to 400 ms (all Fs $\geq$ 2.55, ps $<$ .05). The medial omnibus ANOVA revealed an Action × AntPost interaction from 500 to 600 ms ($F(2, 82) = 4.11$, $p < .05$). Pairwise comparisons between the actions were performed on the basis of these interactions.

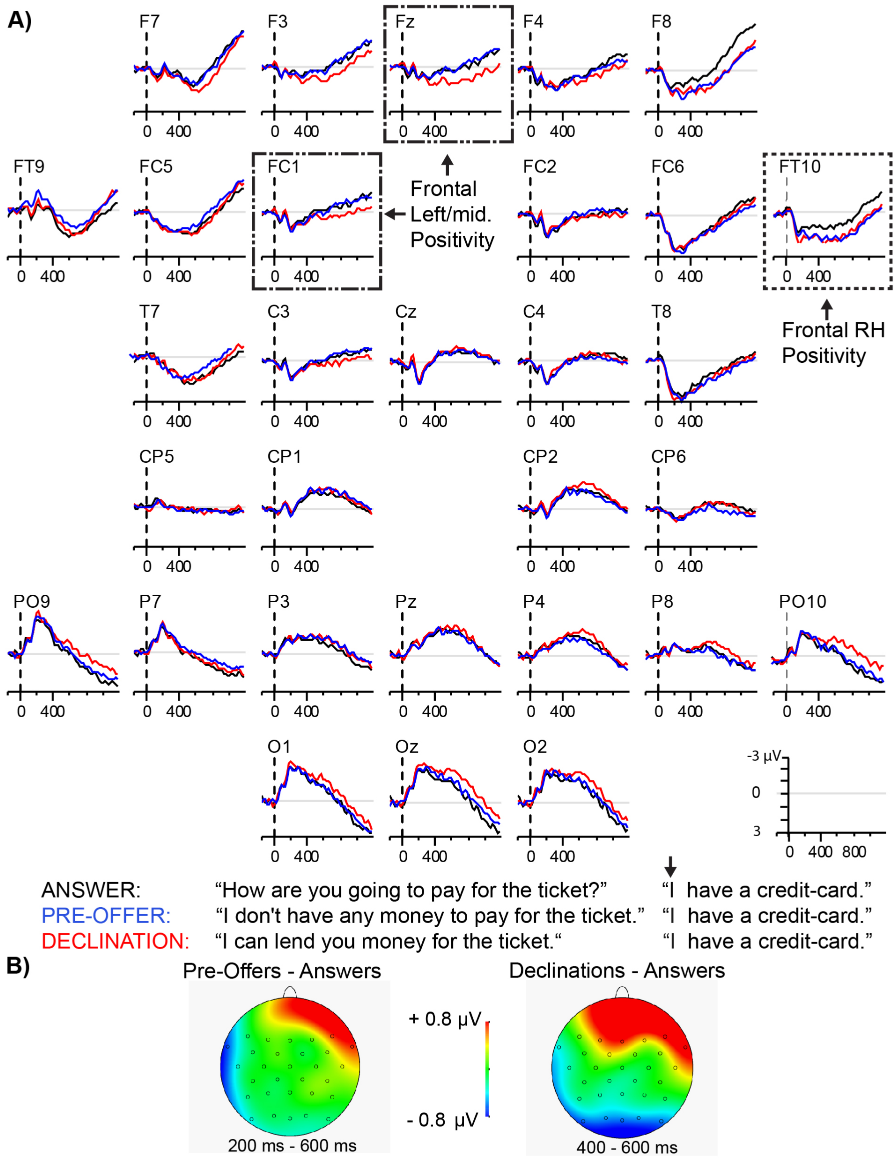

Figure 3.3: Early utterance time-window. A) Grand-averaged ERPs time-locked to the onset of the first word. Representative electrodes showing the relevant effects are highlighted in dashed boxes. B) Scalp distribution of the ERP effects. All waveforms were low-pass filtered (10 Hz) for illustration purposes only. Negativity is plotted upwards.

		Analysis	Source	DF	100-200	200-300	300-400	400-500	500-600
Regional Omnibus		Lat	Action × AntPost × Hem	2, 82	4.14*	7.08**	3.69*	4.08*	5.64**
			Action × Hem × Site	6, 246	3.42*	2.81*	2.55*		
		Med	Action × AntPost	2, 82					4.11*
Pairwise Comparisons	D vs. A	Lat	Action × Hem × Site	3, 123	3.88*	3.63*	3.94*		
			Action × AntPost × Hem	1, 41				6.77*	7.08*
			Action × AntPost × Site	3, 123				2.89*	
		Lat Ant	Action	1, 41					6.70*
			Action × Hem	1, 41				5.40*	4.59*
			Action × Hem × Site	3, 123				3.28*	3.06*
		Right Ant	Action	1, 41				8.29**	
		Med	Action × AntPost	1, 41					6.86*
		Med Ant	Action	1, 41					6.90*
	P vs. A	Lat	Action × Hem	1, 41				6.47*	7.99**
			Action × AntPost × Hem	1, 41	9.63**	15.18**	7.10*	5.46*	10.37**
			Action × AntPost × Hem × Site	3, 123	2.96*			3.40*	2.88*
			Action × Hem × Site	3, 123	7.68**	4.02*	3.21*		
		Lat Ant	Action × Hem	1, 41	4.77*	10.40**		8.56**	11.57**
			Action × Hem × Site	3, 123	6.93**	4.84**	4.08*	3.55*	
		Right Ant	Action	1, 41		7.94**	6.08*	10.61**	6.66*
	P vs. D	Lat	Action × AntPost × Hem	1, 41		5.18*			
		Med	Action × AntPost	1, 41					6.07*

Table 3.5: Regional omnibus analyses and pairwise comparisons for the early utterance time-window. *D* = Declination, *A* = Answer, *P* = Pre-offer, *Lat* = lateral regions, *Med* = medial regions, *Ant* = anterior regions, *Hem* = hemisphere, *AntPost* = Anteriority, *p < .05, **p < .01

Declination versus Answer

The lateral analysis (see Table 3.5) revealed that mean amplitude was more positive-going in Declinations than in Answers at the anterior region in the right hemisphere from 400 to 500 ms (Action; $F(1,41) = 8.29$, $p < .01$), and at bilateral anterior sites from 500 to 600 ms (Action; $F(1,41) = 6.70$, $p < .05$). Mean amplitude was also more positive-going in Declinations than in Answers at the anterior medial region from 500 to 600 ms (Action; $F(1,41) = 6.90$, $p < .05$).

Pre-offer versus Answer

The lateral analysis revealed that mean amplitude in Pre-offers was more positive-going than in Answers at the anterior region in the right hemisphere from 200 to 600 ms (Action; Fs ≥ 6.08, ps $< .05$). The medial ANOVA did not yield any reliable differences across conditions (Fs ≤ 2.34, ps $> .06$).

Pre-offer versus Declination

The lateral ANOVA revealed an Action × AntPost × Hemisphere interaction from 200 to 300 ms ($F(1,41) = 5.18$, $p < .05$), but no reliable differences were obtained in follow-up analyses (Fs ≤ 2.35, ps $> .08$). In the medial analysis, mean amplitude tended to be more positive-going in Declinations than in Pre-offers at the anterior medial region from 500 to 600 ms (Action; $F(1,41) = 3.75$, $p = .06$).

Summary of ERP effects in the early utterance time-window

As predicted, there were early ERP differences between the actions in the early utterance time-window (corresponding to the first word and the verb). In particular, Declinations and Pre-offers both showed a frontal positivity, relative to Answers. In Pre-offers the positivity was restricted to the right hemisphere from 200 to 600 ms. In Declinations the positivity was present at anterior sites in the right hemisphere from 400 to 500 ms and extended to anterior bilateral and medial regions from 500 to 600 ms. The frontal positivity in Declinations was marginally significant relative to Pre-offers at the anterior medial region from 500 to 600 ms ($p = .06$). Differences in scalp distribution of the frontal positivity in Declinations and Pre-offers are illustrated in the topographical maps in Figure 3.4.

3.3.2.2 Late utterance time-window (final word onset of the target utterance)

Figure 3.4 presents the waveforms aligned to the final word onset. The regional omnibus analysis for medial sites (see Table 3.6) yielded an Action $\times$ Site interaction from 100 to 200 ms and 600 to 700 ms (Fs ≥ 2.56, ps $< .05$) and an Action $\times$ AntPost interaction from 900 to 1000 ms ($F(2,82) = 3.57$, $p < .05$). In the lateral analysis an Action $\times$ AntPost interaction was obtained from 600 to 1000 ms (Fs ≥ 3.43, ps $< .05$) in addition to an Action $\times$ Site interaction from 700 to 800 ms ($F(6,246) = 2.65$, $p < .05$). Hence follow-up comparisons between the actions were performed at medial sites from 100 to 200 ms, 600 to 700 ms and 900 to 1000 ms and at lateral sites from 600 to 1000 ms.

Declination versus Answer

The medial ANOVA (see Table 3.6) yielded an Action $\times$ AntPost interaction from 100 to 200 ms ($F(1,41) = 6.32$, $p < .05$), but follow up analyses did not reveal significant differences between the actions ($ps > .12$). In the lateral ANOVA there were no significant differences (Fs ≤ 2.34, $ps > .10$).

Pre-offer versus Answer

The lateral analysis revealed that mean amplitude was more negative-going in Pre-offers than in Answers at bilateral posterior regions from 900 to 1000 ms (Action; $F(1, 41) = 5.11$, $p < .05$). In the analysis for medial sites, mean amplitude was also more negative-going in Pre-offers than in Answers in the medial posterior region from 900 to 1000 ms (Action; $F(1, 41) = 4.97$, $p < .05$).

Pre-offers versus Declinations

Mean amplitude was more negative-going in Pre-offers than in Declinations at bilateral posterior regions from 600 to 1000 ms (Action; Fs ≥ 4.24, $p < .05$). Mean amplitude was also more negative-going in Pre-offers than in Declinations at medial posterior sites from 900 to 1000 ms (Action; $F(1, 41) = 6.31$, $p < .05$).

Summary of ERP effects in the late utterance time-window

As expected, there were no differences between Declinations and Answers in the late utterance time-window. Pre-offers, however, elicited a late, posterior negativity relative to both Answers and Declinations. Relative to Declinations, the negativity was present at posterior bilateral regions (600 to 1000 ms) and the posterior medial region (900 to 1000 ms). Relative to Answers, the negativity was present at posterior bilateral and medial regions from 900 to 1000 ms. The scalp distribution of these effects is illustrated in the topographical maps in Figure 3.4.

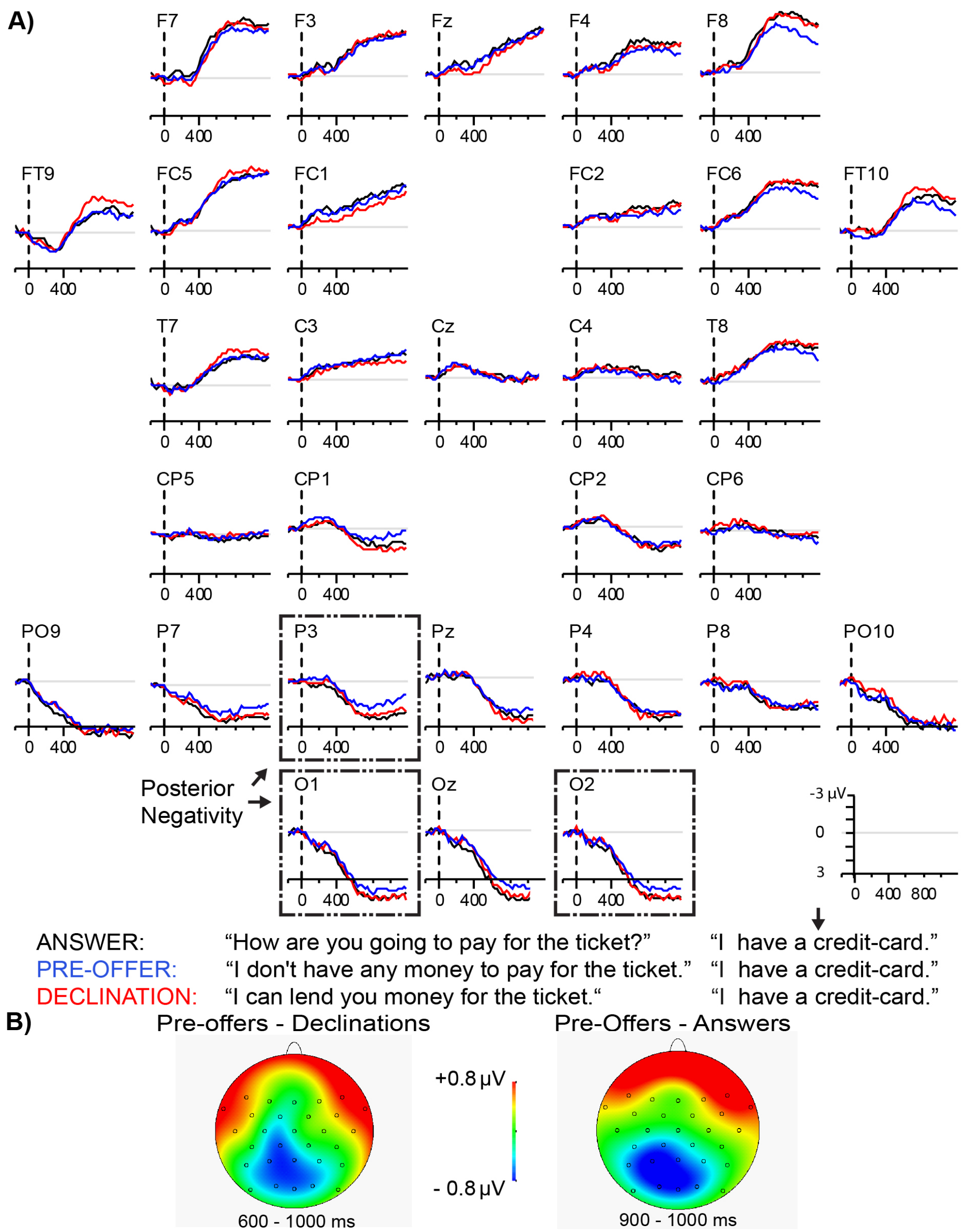

Figure 3.4: Late utterance time-window. A) Grand-averaged ERPs time-locked to the onset of the final word. Representative electrodes showing the relevant effects are highlighted in dashed boxes. B) Scalp distribution of the ERP effects.

		Analysis	Source	DF	100-200	600-700	700-800	800-900	900-1000
Regional Omnibus		Lat	Action × AntPost	2, 82		3.78*	3.43*	4.17*	4.21*
		Lat	Action × Site	6, 246			2.65*		
		Med	Action × Site	8, 328	3.47**	2.56*			
		Med	Action × AntPost	2, 82					3.57*
Pairwise Comparisons	D vs. A	Med	Action × AntPost	1, 41	6.32*				
	P vs. A	Lat	Action × AntPost	1, 41			4.41*	6.79*	8.75**
		Lat Post	Action	1, 41					5.11*
		Med	Action × AntPost	1, 41					5.56*
		Med Post	Action	1, 41					4.97*
	P vs. D	Lat	Action × AntPost	1, 41		7.64**	6.45*	9.04**	8.52**
		Lat Post	Action	1, 41		9.26**	4.24*	4.40*	5.93*
		Med	Action × AntPost	1, 41					4.28*
		Med Post	Action	1, 41					6.31*

Table 3.6: Regional omnibus analyses and pairwise comparisons for the late utterance time-window, for those epochs that showed significant effects. *D* = Declination, *A* = Answer, *P* = Pre-offer, *Lat* = lateral regions, *Med* = medial regions, *Post* = posterior regions, *AntPost* = Anteriority. $^*p < .05$, $^{**}p < .01$.

3.4 Discussion

The aim of this article was to address a fundamental problem in conversation, namely how utterances that are underspecified at the action level can quickly be understood as performing certain speech acts. This study goes beyond prior research by investigating the time-course of speech act recognition in such action-underspecified utterances in spoken dialogues and exploring the effect of sequential context on this process. The target utterances were identical across conditions but differed in the type of speech act performed and how it fits into the larger action sequence (high- vs. low-constraining context, end of an action sequence vs. a start of a new one). The behavioural results indicate that even for action-underspecified utterances participants are very good at identifying the underlying speech act; the mean accuracy in the comprehension task was 96.6%. The ERP results show that speech act recognition begins early in the turn when the utterance has only been partially processed, at least when participants are asked to categorize the action of the utterance. This was reflected by frontal positivities from 200 ms after utterance onset in Pre-offers and from 400 ms in Declinations, relative to the control condition (Answers). At the utterance-final word no ERP effects occurred in Declinations, while Pre-offers elicited a late posterior negativity relative to the other actions. The differences between Pre-offers and Declinations, both of which are relatively indirect speech acts, demonstrate that speech act comprehension is not just modulated by indirectness as traditionally construed, but also by the type of action being performed and the relationships between actions in conversation. Below we describe the results in more detail and discuss their implications.

3.4.1 The time-course of speech act recognition

Conversation is characterized by tight time constraints, allowing listeners limited time between turns (on average only 200 ms) to recognize the speech act and plan an appropriate response (Levinson, 2013). We hypothesized that speech acts are recognized early on in utterances, enabling quick turn transitions. More specifically, we predicted that the critical conditions – Declinations and Pre-offers – would elicit ERP effects relative to the control (Answers) in an early utterance time-window corresponding to the first word and the verb (e.g., *I* *have*), and not at the utterance-final word (*credit-card*). In line with the first aspect of this prediction, both Declinations and Pre-offers elicited ERP effects in the early utterance time-window. In particular, Pre-offers elicited a frontal positivity relative to Answers in the right hemisphere from 200 to 600 ms after utterance onset. A positivity was also observed in Declinations, relative to Answers, at frontal sites in the right hemisphere from 400 to 500 ms after utterance onset, extending to anterior bilateral and medial regions from 500 to 600 ms.

The earlier onset in Pre-offers is consistent with the results from the self-paced reading study in Chapter 2 (see also Gisladottir et al., 2012), in which reading times for the first word were longest in Pre-offers. Future studies are needed to assess the functional significance of this early difference between Declinations and Pre-offers. Studies in the visual modality have reported speech act related ERP effects around 100 ms (Egorova, Pulvermüller, et al., 2013; Egorova, Shtyrov, et al., 2013). The slightly later onset in the present experiment may be due to noisier language input (spoken vs. visual in previous studies (Egorova, Pulvermüller, et al., 2013; Egorova, Shtyrov, et al., 2013)) and greater variability in the target utterances (multi-word sentences vs. one word). It is also possible that the effects observed in the early utterance time-window are early responses to the verb, and not the first word. The dialogues involve connected speech with subjects and verbs of relatively short duration; the duration of the first word in the critical utterances was only 141 ms on average (in 89% of cases it was the short pronoun "I" *ik*; see Figure 3.1). Importantly, however, the early frontal positivities were absent when time-locked to the final word, indicating that the beginning of the utterance, that is the first word and/or the verb, is critical for this aspect of speech act recognition. These early effects of action type show that, at least when an action categorization task is used, listeners do not wait for the full propositional meaning of the sentence, but proceed immediately with speech act recognition based on the partial information available. This is striking given that the critical utterances do not contain any clues regarding what speech act is being performed.

When is speech act recognition achieved, i.e., when has the speech act been successfully recognized? While the ERP method cannot pinpoint the exact moment in time when recognition occurs, the results for the final word are informative about whether additional processing at the end of the utterance is required. In contrast to our prediction that no ERP effects would occur at the utterance-final word (credit-card), a late posterior negativity was observed in Pre-offers from 600 to 1000 ms after final word onset relative to Declinations and from 900 to 1000 ms relative to Answers. This late effect suggests that under certain circumstances listeners do make use of the entire utterance to understand the action. While the final-word effect in Pre-offers is unexpected on an early speech act recognition account, it fits well with our prediction that Pre-offers would elicit additional ERP effects relative to Declinations due to having a more complex action sequence structure. In particular, Pre-offers are less predictable, as the context turn is less constraining (the two turns do not form an adjacency pair). Second, Pre-offers are more predictive, "projecting" further talk in a new action sequence (Schegloff, 1988); understanding a Pre-offer involves knowing that a direct offer would follow if encouraged by the recipient. Pre-offers may thus require additional processing based on the complete utterance because of one or both of these characteristics (for a further discussion of the ERP effects, see below).

The prediction that no ERP effects would be observed at the utterance-final word (*credit-card*) was confirmed for Declinations. As discussed in the introduction, the sequential context is highly constraining in Declinations due to adjacency pair structure; given a proposal (*I can lend you money*), there is a normative expectation for either an acceptance or a declination. The first word of the response (*I*…) helps to narrow down the identity of the unfolding speech act by eliminating obvious acceptances of the form *Oh great, Phew! Thank you* etc. The verb (*I have*…) makes clear that a proto-typical acceptance is not underway, strengthening the likelihood that a declination is involved. By the time the final word is reached, the semantic processing only adds to the propositional meaning, which is the same in all conditions, and hence no ERP differences occur at the final word. We take this to show that in highly constraining contexts – in the present study following the first turn in an adjacency pair – listeners can sidestep the full propositional content of utterances and recognize the speech act before hearing the final word. On the assumption that the action categorization task captures the response demands in everyday conversation, the present results support the early speech act recognition account.

The ERP pattern for Declinations is consistent with the idea that the brain is proactive, constantly generating expectations about upcoming input (Bar, 2009; Kutas, DeLong, & Smith, 2011; Martin J. Pickering & Garrod, 2013; Schacter, Addis, & Buckner, 2007). It has been argued that “we do not interpret our world merely by analysing incoming information, but rather we try to understand it using a proactive link of incoming features to existing, familiar information (e.g., objects, people, situations)” (Bar, 2009, p. 1235). Past experience with speech act sequences during a lifetime of conversation (stored as scripts in memory (Schank & Abelson, 1977)), coupled with minimal information about the utterance, may enable listeners to predict the action in advance of its completion. The implication of the present results is that prediction in language comprehension may not be confined to the level of individual lexical items or their syntactic, semantic or conceptual features (Altmann & Kamide, 1999; Federmeier, 2007; Kutas et al., 2011; Thornhill & Van Petten, 2012), but takes place at the level of speech acts as well. Early action recognition through such predictions could be the key to efficient turn-taking, allowing listeners to plan their reply early and respond within the typical 200 ms window between turns in conversation.

Overall, the current findings on the time-course of speech act recognition contribute to a growing body of ERP research (discussed in section 3.1.1) showing that both early and late processes are involved in pragmatic language comprehension. The early effects of action are in line with the view that "different information types (lexical, syntactic, phonological, pragmatic) are processed in parallel and influence the interpretation process incrementally, that is as soon as the relevant pieces of information are available," (Hagoort, 2008, p. 1055). The results extend reports of early pragmatic effects (from 150-200 ms after word onset in discourse studies) (Van Berkum et al., 1999; Van Berkum, Zwitserlood, et al., 2003) to dialogue contexts. Fast and incremental processing of speech acts challenges classic theories of pragmatics (Grice, 1975; Searle, 1975) which assume that language understanding proceeds by first extracting the propositional or semantic content of the complete utterance and then comparing that to the context, in order to generate additional pragmatic inferences. However, the results for the utterance-final word – in particular, the late negativity in Pre-offers – highlight that pragmatic language comprehension sometimes involves late inferential processes, as research on irony (e.g., Regel et al., 2011; Spotorno et al., 2013) has also shown.

To summarize, we can draw two main conclusions regarding the time-course of speech act recognition. First, even in utterances that are underspecified for the action, speech act recognition begins early in the turn – before the final, critical word has been heard. Second, the time-course of speech act recognition is influenced by how the utterance connects to the larger action sequence. In highly-constraining contexts, when the sequence is coming to a close, additional processing at the utterance-final word is not required. This finding is in line with the early speech act recognition account discussed in the introduction. However, in less-constraining contexts, when a new sequence is initiated, recognition of the speech act involves analysis of the complete utterance, as reflected by a late posterior negativity at the final word.

3.4.2 What do the ERP effects reflect?

There are several parallels between our ERP findings and prior research on speech acts and pragmatic language comprehension more generally which we will briefly discuss. The earliest effect of action was observed at frontal sites in the right hemisphere during the early utterance time-window (e.g., *I have…*), from 200 ms after first word onset in Pre-offers and from 400 ms in Declinations. The timing and anterior scalp distribution of this early positivity differentiates it from other positivities reported in the language domain (see, for instance, Hagoort et al., 1999; Kolk & Chwilla, 2007; Van Petten & Luka, 2012). However, the right hemisphere preponderance of this early frontal positivity is consistent with several studies implicating the right hemisphere in pragmatic language comprehension. A visual half-field lexical-decision study found that the right hemisphere plays a critical role in speech act processing in isolated statements such as *Don't forget to go to the dentist* (Holtgraves, 2012). Similarly, damage to the right hemisphere causes difficulty with understanding indirect speech acts (Foldi, 1987). EEG and fMRI studies in healthy participants have likewise highlighted the role of the right hemisphere in pragmatic language comprehension (Coulson & Wu, 2005) and discourse coherence (Menenti, Petersson, Scheeringa, & Hagoort, 2009). The right hemisphere advantage in pragmatics has been associated with factors such as coarse semantic coding (Beeman, Bowden, & Gernsbacher, 2000), processing of relatively unpredictable semantic relationships (Federmeier, Wlotko, & Meyer, 2008), or forming an integrated representation of ongoing discourse (Menenti et al., 2009). In the

MEG study on speech act comprehension previously discussed (Egorova, Pulvermüller, et al., 2013), several activations were found in the right hemisphere. In particular, an effect was observed between 200 and 300 ms in the right inferior frontal gyrus, suggesting that this area plays a role in "binding information about action and context" (Egorova, Pulvermüller, et al., 2013). At face value, the right hemisphere distribution of the frontal positivity observed in the present study supports the idea that this early effect reflects linguistic and/or discourse-level processes in the right hemisphere that are needed to understand the speech act of the utterance in the given context.

The frontal positivity in Declinations extended to the anterior medial and left hemisphere sites during the early utterance time-window, from 500 to 600 ms after first word onset relative to Answers. This effect is especially prominent at the midline and in the left hemisphere (see waveforms in Figure 3.4) and will therefore be referred to as frontal left/midline positivity. While the frontal left/midline positivity in Declinations was not significant relative to Pre-offers ($p = .06$), the waveforms in Figure 3.4 suggest that it is absent in Pre-offers. This frontal effect bears some resemblance to a frontal positivity reported in a visual ERP study on indirect requests, which was interpreted as reflecting ease of processing in the indirect condition (Coulson & Lovett, 2010). However, that interpretation does not fit the present results, as it counter-intuitively implies that Declinations are the easiest out of the three speech acts to process (whereas we expect Answers to be so). What is it about Declinations that triggers this positivity? In contrast to Pre-offers, the context turn in the Declination dialogues builds up strict expectations about the upcoming action (an acceptance or declining of an offer). In contrast to Answers, the Declination target utterances are relatively indirect (providing just a reason for not needing the offer) and therefore more taxing. We propose that this combination of predictability and indirectness triggers the frontal left/midline positivity: listeners quickly "tune in” to the early signals of the action (expecting either declination or acceptance) and yet have to engage in additional processing because of the indirect response. The nature of this additional processing is unclear. Frontal positivities with a left hemisphere and midline distribution have been linked to discourse structure reanalysis (Dwivedi et al., 2010; Dwivedi, Phillips, Laguë-Beauvais, & Baum, 2006) and update of working memory (MacGregor, Corley, & Donaldson, 2010). The similarity of these positivities to the effect reported here is not obvious, as they are elicited in very different conditions (when a single word is unexpected or pragmatically anomalous). However, an account in terms of discourse structure revision or working memory is compatible with the notion that the beginning of the utterance in Declinations is demanding due to a combination of indirectness and predictability. In particular, the non-canonical beginning of the Declinations (e.g., *I have…*) could trigger revision of the incoming turn in order to make sense of the sentence as a Declination.

An alternative account is that the frontal left/midline positivity reflects socio-emotional factors. As reported in the Introduction, fMRI and MEG studies on indirect speech act comprehension report activations in frontal regions involved in mentalizing, including the medial frontal cortex (Basnakova et al., 2014; Egorova, Pulvermüller, et al., 2013; van Ackeren et al., 2012). Frontal slow waves similar to the frontal left/midline positivity in Declinations have been reported in studies on theory of mind using narratives or story cartoons (Meinhardt, Kühn-Popp, Sommer, & Sodian, 2012; Sabbagh & Taylor, 2000). Moreover, in supplementary analyses of the present data (presented in the Appendix, see 8.2), a correlation was obtained between the frontal left/midline positivity in Declinations and the Empathy Quotient (Baron-Cohen & Wheelwright, 2004), which measures both theory of mind (cognitive empathy) and affective empathy. Although speculative, the frontal left/midline positivity in Declinations may reflect theory of mind processing. Note that frontal theory of mind regions are also activated when participants think about the future (Bar, 2009; C. D. Frith & Frith, 2006), suggesting a link to prediction: these areas "may be concerned with anticipating what a person is going to think and feel and thereby predict what they are going to do" (C. D. Frith & Frith, 2006, p. 532). The combination of predictability and indirectness in Declinations could trigger theory of mind processing at the beginning of the target utterance, resulting in the frontal left/midline positivity. However, further research is clearly required to investigate the two possible accounts for this ERP effect discussed above.

In addition to the frontal left/midline and right hemisphere positivities observed in the early utterance time-window, a late posterior negativity occurred at the final word (*credit-card*) in Pre-offers. This effect was present in Pre-offers from 600 to 1000 ms after final word onset relative to Declinations, and from 900 to 1000 ms relative to Answers. We interpret this ERP effect as a negativity to Pre-offers rather than as a positivity (P600) associated with Answers and Declinations, since it is not plausible that Answers, the control condition, should require more reanalysis, revision, or pragmatic inferencing – processes traditionally associated with late positivities.

One possibility is that the posterior negativity to the final word in Pre-offers reflects uncertainty associated with the subsequent comprehension task, as the behavioural results indicated that accuracy was descriptively lowest for Pre-offers. The difference in accuracy was due to lower accuracy in stimulus Set 2. The target utterances in Set 1 all start with *I have…*, which is a common way to begin an offer, as attested by studies of English and Finnish (Kärkkäinen & Keisanen, 2012). Set 2, on the other hand, is more variable and includes formats such as *I am going…* or *My brother is…*, which may be less common in offering actions and therefore categorized less accurately. If uncertainty in categorization drives the posterior negativity, then the effect should be more pronounced in target sentences in Set 2, for which the task accuracy was lower. However, no differences between Sets were observed in the ERPs, speaking against an uncertainty account for the late ERP effect.

Sustained negativities to sentence-final words have been interpreted as reflecting difficulty in semantic analysis at the message level (see, for instance, Osterhout & Holcomb, 1992), but this "semantic" sentence negativity peaks earlier than in the present experiment (around 400 ms). Sentence-final negativities have also been linked to working memory operations at clause boundaries (clause-ending negativity; Kutas, 1997), in line with studies associating sustained negativities to various linguistic stimuli with increased demands on working memory (see, for instance, King & Kutas, 1995; Kluender & Kutas, 1993; Münte, Schiltz, & Kutas, 1998; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). Negativities associated with working memory have a frontal scalp distribution in most cases and therefore do not match the posterior distribution in the present study (for an exception, see Dwivedi et al., 2006). However, a working memory account for the late negativity in Pre-offers is consistent with the notion that they are more complex than the other two conditions in the study. The context turn in the Pre-offer dialogues is not constraining in terms of what action can follow, and as a consequence the speech act is less predictable. Processing at the final word may be more taxing on working memory in Pre-offers because it requires post-hoc retrieval of the prior turn to figure out what the utterance could mean in that particular context. On this account, a late negativity is not elicited to Declinations, because the prior turn (an offer) has already built up a strong expectation for either an acceptance or a declination, and listeners do not need to compare the utterance to the prior turn in working memory. Whether this working memory account is correct is a topic for further investigation.

3.4.3 Remaining questions

Some questions relating to conceptual and methodological issues remain. The first concerns the task demands. Overhearing a dialogue is different from taking an active part in a conversation, when participants can be held accountable for both the content and timing of their response. Given that the crucial task in conversation involves recognizing the action in the incoming turn and planning a relevant response, we used an action categorization task to ensure that participants paid as much attention to the actions as in natural dialogue. To some degree this task mirrors the situation in everyday conversation, in which listeners have to make an implicit categorization of each utterance in order to prepare a fitted response. However, participants were given labels of the three speech acts in advance and may have used this top-down information to predict the actions based on the contexts. This raises the question whether, and if so to what extent, task-related processes influenced the ERP effects reported in this article. We are currently investigating this in an ERP study. In this follow-up experiment we use a true-false judgment task that probes understanding of the action without providing information about the speech acts in advance. Important for the present purposes, the main ERP effects reported in the current study are replicated under these different task circumstances. On the assumption that using a comprehension task of some kind better captures the response demands in conversation, the generalization of the current results to a different task environment supports the view that the ERP effects are not induced by the action categorization but rather reflect a more natural aspect of speech act comprehension.

Another issue concerns the role of prosody in speech act comprehension. The target utterances were recorded out of their dialogue contexts (from a list) and used in all three conditions to prevent the prosody from biasing any of the three speech acts. One could argue that this neutral prosody, which included falling intonation, may be better suited for some speech acts than others. For instance, in a test recording during development of the materials, the native speakers sometimes used a fall-rise contour (Ward & Hirschberg, 1985) in the Pre-offers. It is therefore possible that a rising intonation is more suitable for the Pre-offers. To what extent the comprehension of the speech acts in the present study was influenced by prosody is not clear. Research on irony has not found an effect of prosody, suggesting that prosody does not necessarily provide crucial cues for utterance interpretation (Regel et al., 2011). Whether this holds for speech act comprehension more generally is an empirical question.

3.5 Conclusions

Three main conclusions can be drawn from the present ERP results. First, the early frontal positivities to the target utterance onset show that – at least when participants are asked to categorize the action – speech act recognition begins very early in the incoming turn, starting already from 200 milliseconds after first word onset when the utterance has only been partially processed. This is the case even though the utterance is relatively indirect and contains no morphosyntactic speech act clues. Second, the ERP findings for the utterance-final word reveal that the time-course of speech act recognition is influenced by how the utterance connects to the larger action sequence. In highly-constraining (adjacency pair) contexts, when the sequence is coming to a close, no late ERP effect to the final word occurs. This suggests that listeners can recognize the speech act before the final word through predictions at the speech act level. Early speech act recognition may well be the key to efficient conversation, allowing listeners to plan their reply early and respond within the 200 ms time frame characteristic for turn-taking. In some cases, however, additional processing based on the complete utterance is required. This is reflected by a posterior negativity at the utterance-final word when the speech act is in a less-constraining context and a new action sequence is initiated. Taken together, the findings show that the time-course of speech act recognition is influenced by the type of action being performed and the sequential context. The present data demonstrate that the speech act dimension is an important aspect of context that should be taken into account in future studies on sentence comprehension in its natural habitat, conversation.

4 Generalization of the ERP correlates of speech act recognition across task environments

This chapter is a modified version of:
Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (In preparation). The influence of speech act type and the sequential context on sentence comprehension: An ERP study.

Abstract
Verbal actions such as asking people out, requesting, and complimenting are the building blocks of conversation. Successful communication relies on the recognition of these actions for appropriate responses. In an ERP experiment, we investigated the time-course of speech act recognition in utterances that are underspecified at the action level. The aim was to determine the robustness of the ERP correlates of speech act recognition reported in Chapter 3 (Gisladottir, Chwilla, & Levinson, 2015). One limitation of the previous study is that participants were asked to categorize the speech acts. Do the early ERP effects – taken as support for an early speech act recognition account – generalize to more natural situations in which overt categorization is not required? In the present study a true/false judgment task was used, which captures the response demands in natural conversation and does not require explicit categorization of the speech acts. Further, filler dialogues were included to reduce strategic processing of the speech acts. Replicating earlier results, Declinations elicited an early frontal positivity to the first word. The frontal positivity is similar to the combined effects of the right fronto-temporal and left/midline frontal positivity in Chapter 3. Also consistent with the previous study, Pre-offers elicited a late

negativity at the utterance-final word. The early frontal positivity in Declinations and the late negativity in Pre-offers are therefore relatively stable across experimental tasks and designs. However, in contrast to the experiment in Chapter 3, no early ERP effects were found in Pre-offers. This indicates that the experimental environment exerts some influence on early action recognition. The present study supports the claim that the early frontal positivity observed to Declinations and the late negativity in Pre-offers are linked to speech act recognition. The results provide further evidence for early recognition of actions in highly-constraining contexts, while highlighting that in some circumstances recognition of action does not begin until at the utterance-final word.

4.1 Introduction

Conversation is an inherently interactive enterprise, characterized by rapid exchange of verbal actions (Schegloff, 1996). Some are routinized and easy to recognize, such as greetings (*hi)* or apologies (*I'm sorry).* Others are more complex. As previously discussed, a simple statement like *I have a car* could be used, for instance, to offer somebody help with moving, to indirectly reject an offer for a ride, or to answer a question. In each case listeners have to go beyond the literal meaning and recognize the speech act of the utterance; was it an offer, rejection, or an answer? Since the response to a turn is critically dependent on what speech act is being performed, recognition of the action is a prerequisite for successful communication.

Not only do participants in conversation need to recognize the speech act of every turn, they also have to plan an appropriate reply and respond in a timely manner to conform to the rules of turn-taking (Sacks et al., 1974). All these processes take place quickly and smoothly, most of the time. As reviewed in the Introduction to this thesis, corpus studies indicate that the gap between turns, i.e., the switch between speakers in conversation, is most frequently only

200 ms and often even shorter (De Ruiter et al., 2006; Heldner & Edlund, 2010; Levinson & Torreira, 2015; Stivers et al., 2009). How conversationalists manage to recognize the speech act of an utterance and respond within 200 ms is a real puzzle, given that it takes people at least between 600 to 1200 ms just to produce a one-word utterance (e.g., naming an object such as an apple; E. Bates et al., 2003; Indefrey & Levelt, 2004; Levelt, 1989; Schnur et al., 2006). Listeners must start planning their own turn well before the other speaker has completed her utterance (Levinson, 2013), which entails that the speech act must have been recognized at an early stage of sentence comprehension. This view was referrred to as the *early speech act recognition account* earlier in this thesis.

The hypothesis that speech acts can be recognized early in utterances is striking, given that turns rarely contain clear speech act clues or so-called illocutionary force indicating devices (Levinson, 1983; Searle, 1969; Searle & Vanderveken, 1985) to aid recognition. As the example above reveals, the form and content of an utterance (e.g., *I have a car*) can be compatible with multiple actions. Given that underspecification at the action level is pervasive in everyday conversation, how is it that listeners recognize speech acts so quickly and efficiently, as the extraordinarily fast transitions between turns suggest? This question taps into a larger issue about how proactive listeners are during language comprehension and whether they make predictions at the level of speech acts. The general aim of this study was to address this by investigating the time-course of speech act recognition, testing the hypothesis that verbal actions are recognized early in utterances based on only partial input.

As reviewed in the Introduction of this thesis, speech act comprehension broadly construed has been investigated in several domains, for instance in relation to reference resolution in requests and questions (e.g., Brown-Schmidt et al., 2008; Hanna & Tanenhaus, 2004), pitch accents in question-answer dialogues (e.g., Magne et al., 2005), and irony (e.g., Regel et al., 2011; Spotorno et al., 2013). The neural substrate of speech act comprehension has been investigated in fMRI and MEG experiments (Basnakova et al., 2014; Egorova, Pulvermüller, et al., 2013; van Ackeren et al., 2012). The time-course of speech act recognition – the topic of this study – has been investigated in ERP studies using visual, non-conversational stimuli (Coulson & Lovett, 2010; Egorova, Shtyrov, & Pulvermüller, 2013). More relevant for research on conversation, the ERP study in Chapter 3 (Gisladottir et al., 2015) investigated speech act recognition in spoken, Dutch dialogues. Since the experiment in this chapter builds on this earlier study, we will first review it again here.

The main goal of Chapter 3 was to test the early speech act recognition account by investigating the time-course of spoken speech act recognition in utterances that are underspecified at the action level. A second aim was to explore how sequential context – i.e., prior turns in conversation – impacts this process. The dialogues used as stimuli in this experiment contained a target turn (e.g., *I have a credit card*) that performed three different speech acts depending on the prior context: an Answer, Declination and Pre-offer (for convenience, example dialogues are repeated in Table 4.1). The rationale was that recognizing the action would be more difficult in Pre-offers and Declinations, which are relatively indirect speech acts, than in Answers (the control condition)[7]. However, while Declinations and Pre-offers are both indirect, they differ in their relationship to prior and upcoming turns in conversation, making it possible to explore the influence of sequential context on speech act recognition.

[7] In the self-paced reading study using the same sentences (Chapter 2 and Gisladottir, Chwilla, Schriefers, & Levinson, 2012), reading times were shortest for Answers on all measures, indicating that Answers are easiest to comprehend.

	Set 1		Set 2	
Condition	**Context**	**Target Utterance**	**Context**	**Target Utterance**
Answer (Control)	*Hoe ga je voor het ticket betalen?*	*Ik heb een creditcard.*	*Waar koop je je shampoo?*	*Ik ga naar de Kruidvat.*
	How are you going to pay for the ticket?	I have a credit-card.	Where do you buy your shampoo?	I go/am going to the Kruidvat [drugstore].
Declination (Context highly-constraining + target utterance ends the sequence)	*Ik kan je wat geld lenen voor het ticket.*	*Ik heb een creditcard.*	*Ik kan wel shampoo voor je meenemen?*	*Ik ga naar de Kruidvat.*
	I can lend you money for the ticket.	I have a credit-card.	I can bring some shampoo for you?	I go/am going to the Kruidvat [drugstore].
Pre-offer (Context less constraining + target utterance starts a new sequence)	*Ik heb geen geld om het ticket te betalen.*	*Ik heb een creditcard.*	*Mijn shampoo is op.*	*Ik ga naar de Kruidvat.*
	I don't have any money to pay for the ticket.	I have a credit-card.	My shampoo is finished.	I go/am going to the Kruidvat [drugstore].

Table 4.1: Examples of stimuli in Dutch and English translations.

As discussed in section 3.1.2, Declinations close an adjacency pair and consequently the context utterance (the first turn of the pair) is highly constraining in terms of what type of action can follow. In contrast, Pre-offers initiate a new action sequence, a pre-sequence, projecting a more direct offer if the conversation were to continue[8]; thus Pre-offers differ from Declinations in that they invite more inferences about upcoming talk and the context utterance

[8] E.g., A: I don't have any money. B: I have a credit card. A: You do? B: Yeah, I can pay for you.

is less constraining (as the dialogues do not form an adjacency pair). It was predicted that both Declinations and Pre-offers would elicit ERP effects relative to Answers in an early time-window corresponding to the first word and the verb, and not at the final word. Such a pattern would support the early speech act recognition account[9]. Secondly, it was hypothesized that Pre-offers might give rise to additional ERP effects due greater complexity (i.e., the context is less constraining context and the target utterance invites inferences about the continuation of the sequence). In line with the first prediction, both Declinations and Pre-offers elicited ERP effects in the early time-window. More specifically, a fronto-temporal positivity was found for the right hemisphere from 200 ms after first-word onset in Pre-offers and from 400 ms in Declinations, extending to anterior medial sites in Declinations only. These early effects suggest that listeners do not wait for the full propositional meaning but proceed with speech act recognition as early as 200 ms after the start of the utterance. No ERP effects occurred at the utterance-final word in Declinations, relative to Answers, providing further support for the early speech act recognition account. Pre-offers, on the other hand, elicited a late, posterior negativity at the utterance-final word (relative to both Declinations and Answers), indicating that additional processing based on the entire utterance is sometimes required. While the late negativity is consistent with the hypothesis that Pre-offers might give rise to additional ERP effects due to a more complex action sequence structure, the occurrence of this effect at the utterance-final word seems surprising under an early speech act recognition account.

One limitation of the ERP study in Chapter 3 relates to task demands.

---

[9] Note that if the action has already been recognized by the time the final word is presented, the processing of the final word should only add to the propositional meaning of the utterance, which for each target is the same for all three conditions and hence no differences should occur at the final word.

Before definitive conclusions can be drawn regarding the time-course and ERP correlates of speech act recognition, it is important to determine the robustness of the ERP results reported in Chapter 3 (and Gisladottir et al., 2015). Participants were instructed to categorize the target turns as doing Answering, Declining and Offering; this task was chosen to approximate the response demands in everyday conversation, in which listeners have to make an implicit categorization of each utterance in order to prepare a fitted response. However, the three speech act categories were known to the participants before the experiment began. The task therefore directed attention to the critical speech acts and a priori restricted possible interpretations of the utterances. It is well-known that task demands and aspects of the experimental design can influence ERPs. In line with this, in the language domain both the N400 and P600 components have been found to be modulated by the type of task (see, for instance, D. J Chwilla et al., 1995; Gunter & Friederici, 1999; Gina R. Kuperberg, 2007). Moreover, early positivities are influenced by experimental factors such as task demands (e.g., Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007). Given these findings, an important question is whether the early frontal positivities or the late negativity reported in Chapter 3 were induced by the categorization task or whether they generalize to a more natural situation in which such categorization is not required.

The aim of the current study was to further investigate the time-course of speech act recognition by determining the robustness of the ERP correlates reported in Chapter 3. Taking into account the task concern raised above, we investigated whether the early positivities to Declinations and Pre-offers and the final-word negativity in Pre-offers are also present when explicit categorization of the target speech act is not required. As previously mentioned at the beginning of the introduction to this chapter, participants in conversation have to juggle several tasks; they need to recognize the speech act of the unfolding turn, plan an appropriate reply, and respond in a timely manner to obey the rules of turn-taking (Sacks et al., 1974). Thus in contrast to passive reading or listening to non-conversational discourse, comprehending-for-responding is a critical aspect of conversation. Simply instructing participants to read and listen for comprehension, without requiring any response, would not capture this important feature of conversation. To better mirror the response demands and attention level necessary in everyday interaction, the current study required a true/false judgment response to a comprehension question. Participants were asked to judge whether a statement about the target utterance was true or false (e.g., *Speaker B gives A the information asked for. – True / Not true*). This task targets understanding of the action without requiring explicit categorization of the speech act.

A second limitation of the experiment in Chapter 3 is that the experimental materials did not include filler dialogues. As a consequence, participants may have used the context turn as a superficial cue to what type of speech act was coming up. ForPa instance, a context utterance performing a proposal of some sort was always followed by a declination, but never by an acceptance. Coupled with the categorization task, the absence of fillers may have rendered the target speech acts predictable. This could have given rise to the early ERP effects reported in Chapter 3. To address this concern, additional filler dialogues were included in the present study, preventing participants from using the context turn as a superficial cue to the target speech act.

We created three types of fillers, one for each condition; these can be referred to as non-answers, non-pre-offers and non-declinations (i.e., acceptances; see Method section). The first turn of the fillers contains the *same* action as the context turn of the corresponding experimental condition (e.g., an offer), while the second turn delivers an action *different* from the target utterance (e.g., an acceptance instead of a declination). As a consequence, the critical speech acts cannot be predicted based on the context turns, in contrast to the prior study. In this new design, an offer is equally likely to be followed by an acceptance (filler) or a declination (experimental condition). Importantly, none of the fillers and the experimental dialogues contain anomalies of any kind. They resemble natural conversations between friends or relatives and span common discourse topics such as working/studying, doing groceries and going out.

The predictions for the experiment were as follows. Under the assumption that the ERP effects reported in the earlier study reflect processes of speech act recognition and are note dependent on the categorization task, we expect to find a similar ERP pattern in the current experiment. More specifically, in the early time-window (corresponding to the first word and the verb) we predict the presence of frontal positivities in both Declinations and Pre-offers, relative to Answers. At the final word, however, we should find a late negativity in Pre-offers, relative to Answers and Declinations, but no ERP effect in Declinations relative to Answers. Such an outcome would confirm the time-course of speech act recognition reported in the earlier study, demonstrating that the ERP correlates of speech act recognition are stable across experimental designs and tasks. An alternative outcome is that some or all of the effects reported in Chapter 3 are not replicated. Depending on the specific pattern of ERP results, this would lead to a reevaluation of the time-course of speech act recognition.

4.2 Methods

4.2.1 Participants

Forty-six native speakers of Dutch participated in the experiment (29 women, 17 men, mean age = 21.2 years, age range 18-27). All participants were right handed, had normal vision and no hearing or speech problems.

4.2.2 Materials

The experimental items are the same spoken dialogues as used in Chapter 3 (see also Gisladottir et al., 2015). The reader is referred to sections 2.3.2 and 3.2.2 for a description of the stimuli. In addition to the experimental items, three types of filler dialogues were added to decrease the predictability of the target speech acts. The first turn of the fillers performs the same action as the context of one of the experimental conditions, while the second turn performs an action that is different from the target utterances (for instance, Non-Answer filler: *What are you bringing to the dinner? – I don't know yet.* Non-Pre-Offer filler: *I don't have any money for the vending maching. – I don't have any either.* Non-Declination filler: *I can buy a ticket for you. – That would be great*). Importantly, the context turns cannot be used as superficial cues to the critical speech acts, as each context sentence can be followed by more than one type of reply (target action or filler action). As an example, an offer (context) can be followed by an acceptance (filler) or a declination (experimental condition).

The stimuli and fillers were pseudo-randomized and balanced across three lists, such that each list contained 126 experimental items and 126 fillers (with an equal number of trials across conditions and filler types). The recording session, which included a short practice list with six dialogues, lasted approximately an hour.

4.2.3 Procedure

Participants were instructed to listen carefully to the dialogues and take the perspective of the speakers. Each trial began with a warning beep and 750 ms later a fixation cross appeared in the middle of the screen. Participants were told to avoid movements, including blinks, during the presentation of the cross (which lasted throughout the entire dialogue). The context utterance was played 500 ms after the appearance of the fixation cross. Between the context recording offset and the start of the target recording there was a 200 ms silence. To prevent an abrupt start and ending of the sentences the recordings included a 50 ms buffer before sentence onsets and after offsets, such that the pause between context and target was in total 300 ms. This pause is similar to average gap durations reported in corpus studies of Dutch, which range from 8 ms to 380 ms depending on the study (Bosch et al., 2005; Heldner & Edlund, 2010; Stivers et al., 2009). The fixation cross stayed on the screen for 1200 ms after the end of the target utterance recording, and was followed by a 1500 ms blank screen interval. A comprehension question then appeared on the screen, e.g., *Speaker B gives A the information asked for* or *Speaker B accepts the suggestion of speaker A*, to which participants responded *True* or *False* (in Dutch: *Waar*, *Niet waar*). To ensure equal probability of questions calling for a true or a false response, each type of experimental items and fillers was presented with two comprehension questions; one which called for a "true" response and another where a "false" response was appropriate. This resulted in six different comprehension questions that targeted understanding of the action. Participants responded by navigating the screen and clicking on the answer with a computer mouse. Upon answering the comprehension probe a blank screen appeared for 4000 ms and then the next trial began. After the EEG recording, participants filled out the Empathy Quotient survey (Baron-Cohen & Wheelwright, 2004) (see Section 8.4 in the Appendix).

4.2.4 EEG recording

The EEG was recorded with 60 active electrodes in a cap (actiCap), referenced to the left mastoid. The data were later re-referenced to the average of the left and right mastoids. Vertical and horizontal eye movements were recorded using four additional electrodes placed above and below the left eye and on the outer canthi. Bipolar EOG was computed. Electrode impedance was kept below 20 KΩ.

EEG and EOG data were filtered by a .02 Hz high pass and 250 Hz low pass filter with a 10 second time constant and sampled with a frequency of 500 Hz.

4.3 Results

4.3.1 Behavioural analysis and results

Accuracy in the comprehension task was analyzed with mixed-effects logistic regression in the statistics software R (R Core Team, 2013), using the lme4 package (D. Bates et al., 2012, p. 4). Data from four subjects were excluded from the behavioural and EEG analysis due to excessive artifacts. Behavioural responses from 42 participants were analyzed. Accuracy in the comprehension task was computed as the percentage of correct responses in each experimental condition. Accuracy on filler items was not analyzed. Overall mean accuracy in the comprehension question was 93.1% (SD 6.7%). Participants correctly answered the question for 97.7% of Answers (SD 2.8%), 95.6% of Declinations (SD 5.3%) and 85.9% of Pre-offers (SD 14.9%). Table 4.2 shows mean accuracy for each set and condition; there seem to be differences in accuracy across sets, particularly in Pre-offers.

	Answer	**Declination**	**Pre-offer**
Overall	97.7% (2.8%)	95.6% (5.3%)	85.9% (14.9%)
Set 1	97.4% (3.5%)	94.6% (6.8%)	92.6% (12.9%)
Set 2	98.0% (3.5%)	96.6% (5.4%)	79.3% (18.6%)

Table 4.2: Mean accuracy in the comprehension task, for all items (overall) and each set.

For the mixed-effects logistic regression analysis, we used the most maximal random effects structure justified by the experimental design and for which convergence was reached (random intercepts by participant and item, as well as by-participant random slopes for Action and Set and by-item random slopes for Action). The fixed effects were Action and Set. The full model with Action, Set and the Action × Set interaction (the best-fitting model, according to model comparison presented in the Appendix; see 8.5) indicated that the accuracy was lower in Pre-offers than in Answers (Estimate: -1.79, SE: 0.32, z = -5.67, p < .001). Accuracy was also lower in Pre-offers than in Declinations (Estimate: 1.40, SE: 0.32, z = 4.32, p < .001), while the comparison between Answers and Declinations was not significant (Estimate: -0.40, SE: 0.32, z = -1.36, p = 0.17). As for interactions between Action and Set, accuracy was higher for Pre-offers in Set 1 than in Set 2 (Estimate: 1.13, SE: 0.26, z = 4.27, p < .001). Note that for Declinations and Answers an opposite pattern was observed, i.e., accuracy was slightly lower in Set 1 than in Set 2 (Answers; Estimate: -1.08, SE: 0.26, z = -4.03, p < .001, Declinations; Estimate: -1.22, SE: 0.27, z = -4.45, p < .001). These behavioural results show that while overall accuracy in the comprehension task is high, there are subtle differences between actions and across sets. Additional analyses of accuracy and the EEG data with low accuracy items (below 93% accuracy) removed are presented in section 4.3.4.

4.3.2 EEG Data Analysis

As in Chapter 3, two time-windows were used for the analyses of the EEG data: an *early utterance time-window* from 100 to 600 ms after first word onset and a *late utterance time-window* from 100 to 1000 ms after final word onset. The time-window at the first word was shorter than the one at the final word to avoid overlapping time-windows. As seen in Figure 4.1, the final word of the target sentences occurs on average 490 ms after utterance onset. Since the first 100 ms after word onset are mainly sensitive to exogenous factors (e.g., modality, intensity) but relatively insensitive to cognitive factors (see e.g., Kutas, Van Petten, & Kluender, 2005), the waveforms for the first 100 ms of the final word should be very similar across the experimental conditions. The endpoint of the early time-window was therefore set at 600 ms (roughly 490 ms + 100 ms), and not at 1000 ms as in the late utterance time-window. The starting point of both early and late time-windows was set at 100 ms after first and final word onset for the same reason, i.e., to exclude exogenous ERP components. These two time-windows allow for a separation of early speech act recognition effects at the beginning of the utterance up to the final word, and effects to the utterance-final word.

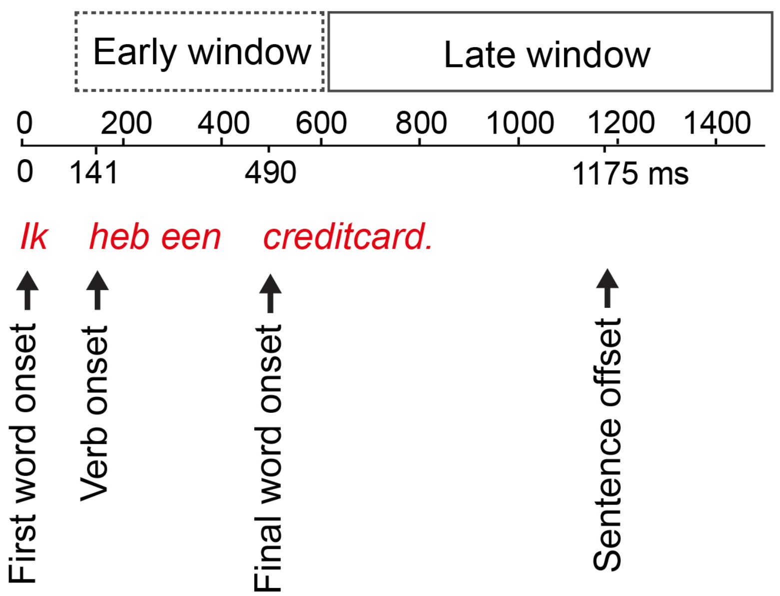

Figure 4.1: Mean onset of the first word, verb and final word in target utterances and a rough timeline for the early and late time-windows. An example sentence (in red) illustrates the onset of the first word, the verb and the final word, but does not reflect word durations.

The EEG data were averaged relative to the first word onset and the sentence-final word onset. For each time-locking point the EEG data were segmented into 1200 ms epochs with a 150 ms pre-stimulus baseline. Artifacts were removed by excluding epochs with excessive EEG (± 100 μV) or EOG (± 75 μV) amplitude. In the final dataset of 42 participants, 14.7% of trials were rejected due to artifacts. The percentages of rejected trials did not differ across the three experimental conditions at the two time-locking points (Fs < 2.21, ps $> .12$).

Statistical analyses were performed on mean amplitudes in the early and late utterance time-windows. We first performed omnibus repeated measures ANOVAs in 100 ms consecutive time-windows, using all 59 electrodes with Action (3), Set (2) and Site (59) as factors, to further narrow down relevant time-windows ased on these initial analyses, follow-up analyses were performed for lateral and midline sites separately in 100 ms consecutive time-windows from 200 to 600 ms at the first word and from 500 to 700 ms and 800 to 1000 ms at the final word. For these analyses a regional approach was used. For the lateral ANOVA the topographical factors were AntPost (for Anterior vs. Posterior sites), Hemisphere and Region, yielding 12 regions of 3 sites each (see Figure 4.2). A similar approach with 3-site regions has been used, for instance, in research on ironic speech acts (Regel et al., 2011). The midline ANOVA analysis included 3 regions (Anterior, Medial, and Posterior) of three sites each; factors used were Action (3), AntPost (i.e., anteriority; 3) and Site (3). Since no Action $\times$ Set interactions were obtained in the initial omnibus ANOVA (see Table 4.3 and Table 4.4), Set was removed as a factor in these regional analyses. Main effects of Action and relevant interactions were followed up with pairwise comparisons. The Greenhouse-Geisser correction was applied to correct for violations of the assumption of sphericity (original degrees of freedom are reported for readability).

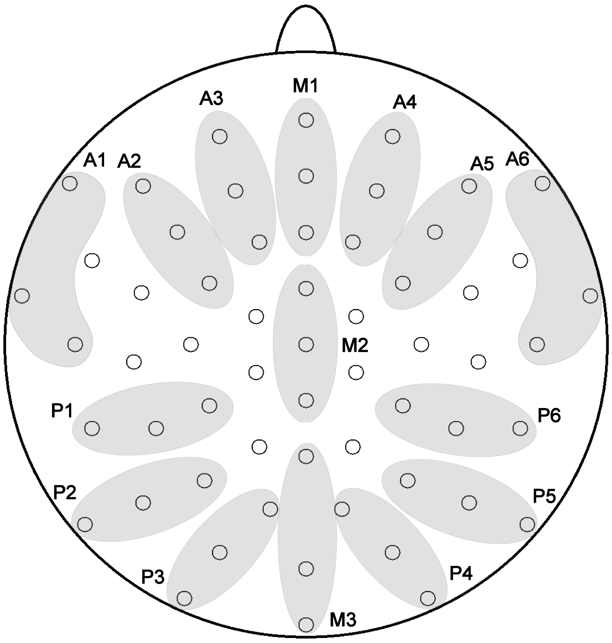

Figure 4.2: Regions of interest used for statistical analyses.

Source	DF	100-200	200-300	300-400	400-500	500-600
Action × Site	116		3.63**	3.64**	3.92**	3.37**

Table 4.3: Initial omnibus ANOVA for the early utterance time-window (all electrodes). **$p<.01$.

Source	DF	100-200	200-300	300-400	400-500	500-600	600-700	700-800	800-900	900-1000
Action x Site	116					1.93*	1.87*		1.89*	1.99*

Table 4.4: Initial omnibus ANOVA for the late utterance time-window (all electrodes). *$p < .05$.

4.3.3 ERP results

4.3.3.1 Early utterance time-window (first word onset of target utterance)

Figure 4.3 shows the grand averaged ERPs time-locked to the first word for a representative subset of electrodes. The overall ANOVA at lateral sites (see Table 4.5) revealed an Action × AntPost interaction from 200 to 600 ms ($Fs > 7.79$, $ps < .01$) and an Action × AntPost × Region interaction from 400 to 600 ms ($Fs > 3.08$, $ps < .05$). The midline ANOVA yielded an Action × AntPost interaction from 200 to 600 ms ($Fs > 5.52$, $ps < .01$) and an Action × AntPost × Site interaction from 200 to 300 ms and 400 to 600 ms ($Fs > 2.32$, $ps < .05$). Based on these interactions, follow-up comparisons between the actions were performed from 200 to 600 ms at both lateral and midline sites.

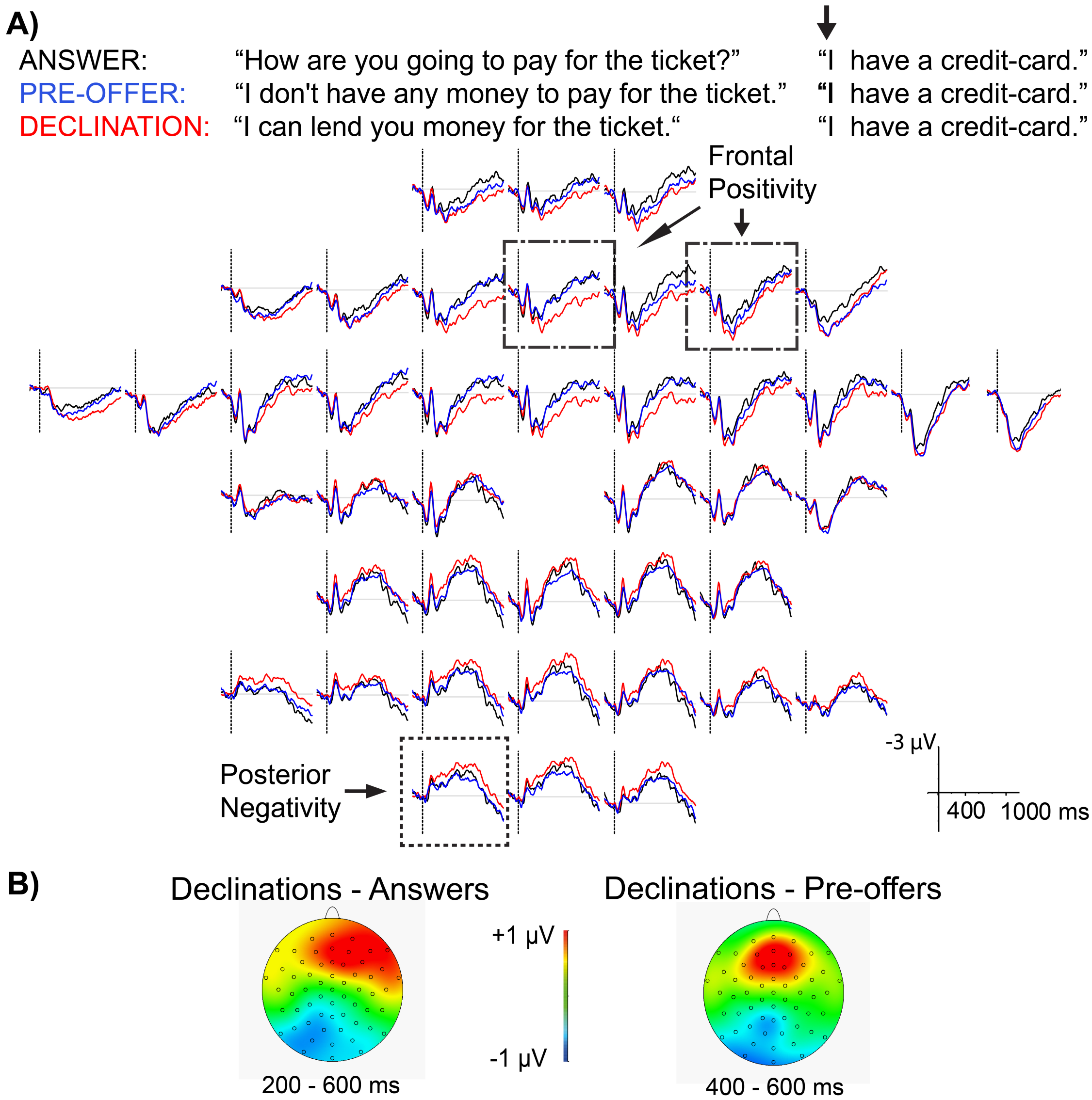

Figure 4.3: Early utterance time-window. A) Grand-averaged ERPs time-locked to the onset of the first word (for a representative subset of electrodes). The relevant effects are highlighted in dashed boxes. B) Scalp distribution of the relevant ERP effects. All waveforms were low-pass filtered (10 Hz) for illustration purposes only. Negativity is plotted upwards.

	Source	DF	200-300	300-400	400-500	500-600
Lateral	Action × AntPost	2	8.59**	8.46**	8.84**	7.79**
Regions	Action × AntPost × Reg	4			4.38**	3.08*
Midline	Action × AntPost	4	5.52**	6.09**	8.28**	6.31**
Regions	Action × AntPost × Site	8	2.32*		3.04*	3.11**

Table 4.5: Regional omnibus analyses at lateral and midline sites for the early utterance time-window (showing only epochs with significant effects). $*p < .05$, $**p < .01$. *AntPost* = Anteriority, *Reg* = Region.

Declination vs. Answer

The lateral ANOVA (see Table 4.6) revealed that mean amplitudes were more positive-going in Declinations than in Answers at anterior regions in the right hemisphere from 200 to 300 ms (Action, $F(1, 41) = 5.20$, $p < .05$). The positivity to Declinations was also present from 400 to 600 ms at anterior bilateral regions (Action; Fs > 5.22, ps $< .05$) and at the anterior midline region (M1) (Action; Fs > 4.73, ps $< .05$). An effect in the opposite direction was observed at posterior sites: mean amplitudes were more negative-going in Declinations than in Answers from 200 to 300 ms at posterior bilateral sites (Action; $F(1, 41) = 6.33$, $p < .05$) and at the posterior midline region (M3) (Action; $F(1,41) = 7.64$, $p < .01$). The negativity to Declinations extended to two posterior regions in the left hemisphere (P2, P3) from 300 to 400 ms (Action, Fs > 7.50, ps $< .01$).

			Source	DF	200-300	300-400	400-500	500-600
Answer vs. Declination	Lateral Regions	All regions	Action × AntPost	1	17.49**	15.04**	18.76**	15.49**
			Action × Hem	1	5.33*			
			Action × Reg	2	4.22*			
			Action × AntPost × Reg	2			6.30*	4.52*
		Anterior	Action	2			7.56**	5.22*
			Action × Hem	2	7.67**			
		Posterior	Action	1	6.33*			
			Action × Reg	2	8.24**	3.86*	6.63**	8.39**
			Action × Reg × Site	4				3.31*
		Ant: RH	Action	1	5.20*			
		Post: P2	Action	1	20.24**	7.57**		
		Post:P3	Action	1	14.99**	7.50**		
		Post: P4	Action	1	7.85**			
	Midline Regions	All regions	Action × AntPost	2	7.72**	9.19**	13.24**	9.11**
			Action × AntPost × Site	4	3.13*	3.74*	5.14**	4.26**
		Anterior	Action	1			7.77**	4.73*
		Posterior	Action	1	7.64**			
Declination vs Pre-offer	Lateral Regions	All regions	Action × AntPost	1	7.42**	5.46*	7.21*	9.32**
			Action × AntPost × ROI	2			6.42**	4.86*
		Anterior	Action × Reg	2			4.75*	
		Posterior	Action × Reg	2	4.06*			5.75*
		Ant: A3	Action	1			4.26*	
		Post: P2	Action	1	4.93*			
		Post: P3	Action	1	6.82*			4.98*
		Post: P4	Action	1	4.45*			
	Midline Regions	All regions	Action × AntPost	2	8.10**	8.54**	14.12**	12.19**
			Action × AntPost × Site	4			3.83*	4.45**
		Anterior	Action	1			8.33**	4.99*
		Posterior	Action	1	6.21*			5.10*

Table 4.6: Pairwise comparisons for the early utterance time-window. *p < .05, **p < .01. *Ant* = anterior regions, *Post* = posterior regions, *AntPost* = anteriority, *Hem* = hemisphere, *Reg* = Region.

Pre-offer vs. Answer

The ANOVA revealed no main effects of Action or relevant interactions at lateral sites (Fs < 3.28, ps $> .08$) nor midline sites (Fs < 2.29, ps $> .09$).

Declination vs. Pre-offer

The lateral analysis revealed that mean amplitude was more positive-going in Declinations than in Pre-offers at the anterior region closest to the midline in the left hemisphere (A3) from 400 to 500 ms (Action; $F(1,41) = 4.26$, $p < .05$) and at the anterior midline region (M1) from 400 to 600 ms (Action; Fs > 4.99, ps $< .05$). In contrast, mean amplitude was more negative-going for Declinations than for Pre-offers in left posterior regions (P2, P3, P4, Action; Fs > 4.45, ps $< .05$) and in the posterior midline region (M3, Action; $F(1,41) = 6.21$, $p < .05$) from 200 to 300 ms. The negativity to Pre-offers was also present in the posterior region closest to the midline in the left hemisphere (P3, Action; $F(1,41) = 4.98$, $p < .05$) and the posterior midline region (M3, Action; $F(1,41) = 5.10$, $p < .05$) from 500 to 600 ms.

Summary of ERP effects in the early utterance time-window

As predicted, Declinations showed a frontal positivity in the early utterance time-window relative to Answers. The frontal positivity was present at anterior sites in the right hemisphere from 200 to 300 ms, extending to anterior bilateral and midline regions from 400 to 600 ms. An additional early effect was observed at posterior sites: a negativity to Declinations, relative to Answers, at posterior bilateral and midline regions from 200 to 300 and at lateral regions in the left hemisphere from 300 to 400 ms. In contrast to our predictions, no ERP effects were present in Pre-offers relative to Answers in the early utterance time-window. The comparison between Pre-offers and Declinations confirmed that the early frontal positivity and posterior negativity observed to Declinations were not present in Pre-offers. Relative to Pre-offers, the frontal positivity was present in Declinations from 400 to 500 ms at an anterior region in the left hemisphere, extending to anterior midline sites from 400 to 600 ms. Similarly, the early negativity in Declinations was reliable relative to Pre-offers roughly at posterior sites in the left hemisphere and at the midline from 200 to 300 ms and 500 to 600 ms after first word onset.

4.3.3.2 Late utterance time-window (final word onset of target utterance)

Figure 4.4 shows the grand averaged ERPs time-locked to the final word for a representative subset of electrodes. The omnibus analysis at lateral regions (see Table 4.7) yielded an Action × Hemisphere interaction from 500 to 700 ms and 800 to 1000 ms (*F*s $>$ 5.33, *p*s $<$ .01) and an Action × AntPost × Hemisphere interaction from 900 to 1000 ms ($F(2,82) = 4.17$, $p < .05$). The midline omnibus revealed no significant main effects of Action or relevant interactions (all *F*s $<$ 2.80, *p*s $>$ .07). Thus follow-up analyses were performed comparing each action with the others from 500 to 700 and 800 to 1000 ms at lateral sites only.

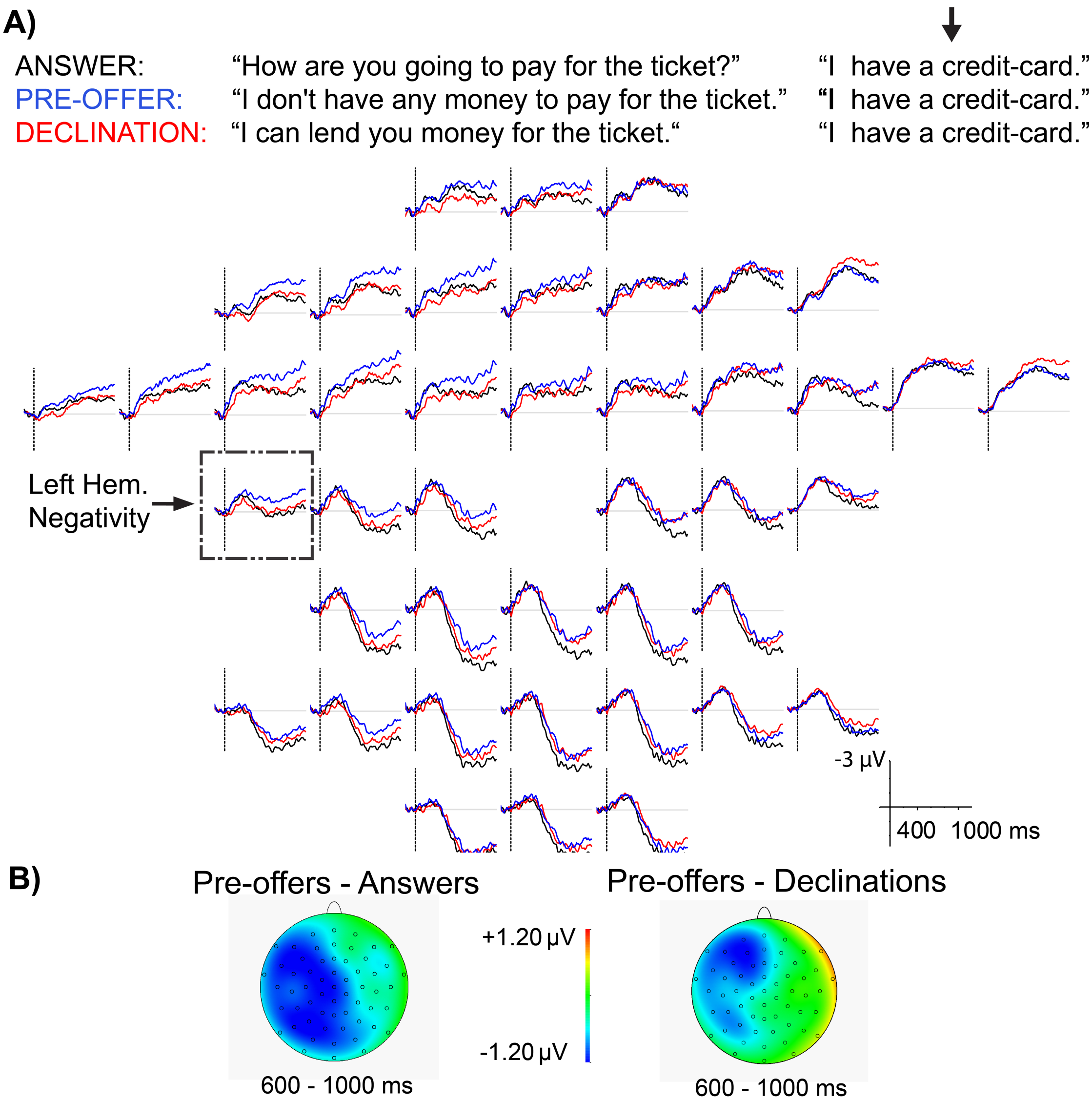

Figure 4.4: Late utterance time-window. A) Grand-averaged ERPs time-locked to the onset of the final word (for a representative subset of electrodes). The left hemisphere negativity is highlighted in dashed boxes. B) Scalp distribution of the left hemisphere negativity.

	Source	DF	500-600	600-700	800-900	900-1000
Lateral Regions	Action × Hem	2	7.53**	5.33**	10.52**	9.07**
	Action × AntPost × Hem	2				4.17*

Table 4.7: Regional omnibus analyses at lateral and midline sites for the late utterance time-window (showing only relevant epochs). *p < .05, **p < .01. *Hem* = hemisphere, *AntPost* = Anteriority.

Declination vs. Answer

The analysis for anterior lateral sites (see Table 4.8) revealed an Action × Hemisphere interaction from 500 to 600 ms ($F(1,41) = 4.15$, $p < .05$) and an Action × Region interaction from 800 to 900 ms ($F(2,82) = 5.22$, $p < .05$), but further analyses did not yield any reliable differences between the conditions (Fs < 3.39, ps > .06).

	Analysis	Source	DF	500-600	600-700	800-900	900-1000
Decl. vs. Answer	**All lateral regions**	Action × AntPost	1	4.43*			
		Action × AntPost × Reg	2			4.91*	
	Ant	Action × Hem	1	4.15*			
		Action × Reg				5.22*	
Pre-offer vs. Answer	**All lateral regions**	Action	1		4.31*		5.06*
		Action × Hem	1		5.00*	15.46**	10.46**
		Action × AntPost × Site	2		4.04*		3.95*
		Action × Hem × Reg	2			3.80*	
		Action × AntPost × Hem	1				4.33*
	Left	Action	1		7.66**	5.73*	10.41**
		Action × AntPost × Site	2		5.43*		
	Right	Action × AntPost × Reg	2				4.02*
	Left Post	Action	1		10.24**		
Declination vs. Pre-offer	**All lateral regions**	Action × Hem	1	14.18 **	12.62**	16.04**	14.43**
		Action × AntPost × Hem	1	6.13*			9.35**
		Action × AntPost × Reg	2				4.56*
		Action × AntPost × Hem × ROI	2		4.14*	4.23*	
		Action × Reg × Site	4			2.84*	
	Left	Action	1		7.28*		5.11*
	Right	Action × Reg	2		3.91*		3.66*
		Action × AntPost × Reg	2		6.46**	5.08*	5.77*
		Action × Reg × Site	4			3.28*	
	Right: Ant	Action × Reg	2		6.44**	4.53*	6.22*
	Right: Post	Action × Reg × Site	4	2.52*	4.39**	4.18**	3.78**

Table 4.8: Pairwise comparisons for the late utterance time-window. *$p < .05$, **$p < .01$. *Ant* = anterior regions, *Post* = posterior regions, *AntPost* = anteriority, *Reg* = region, *Hem* = hemisphere.

Pre-offer vs. Answer

Mean amplitude was more negative-going in Pre-offers than in Answers at lateral regions from 600 to 700 ms and 900 to 1000 ms, reflected by a main effect of Action (Fs > 4.31, ps $< .05$). However, follow-up analyses on interactions with hemisphere (see Table 4.8) revealed that the negativity was more pronounced in the left hemisphere, being present at left posterior sites from 600 to 700 ms (Action; $F(1,41) = 10.24$, $p < .01$) and in the left hemisphere from 800 to 1000 ms (Action; Fs > 5.73, ps $< .05$). An Action × AntPost × Region interaction was present in the right hemisphere from 900 to 1000 ms ($F(2, 82) = 4.02$, $p < .05$), but there were no reliable differences between the actions in follow-up analyses for the right hemisphere (Fs < 3.09, ps $> .07$).

Declination vs. Pre-offer

Pre-offers were more negative-going than Declinations in the left hemisphere from 600 to 700 ms and 900 to 1000 ms (Action; Fs > 5.11, ps $< .05$). Follow-up analyses for the right hemisphere did not reveal any reliable differences between the actions (ps $> .07$).

Summary of ERP effects in the late time-window

As predicted, Pre-offers elicited a negativity relative to both Answers and Declinations in the late utterance time-window. Relative to Answers, the effect was most pronounced in the left hemisphere from 600 to 700 ms and 800 to 1000 ms. Relative to Declinations, the negativity was present in the left hemisphere from 600 to 700 ms and 900 to 1000 ms. Declinations, on the other hand, did not elicit any ERP effects relative to Answers at the utterance-final word, as predicted.

4.3.4 Additional analyses of the utterance-final word

ERP effects at the utterance-final word are unexpected under an early speech act recognition account. To corroborate the results for the late utterance time-window – i.e., the absence of an ERP effect in Declinations and the presence of a late negativity in Pre-offers – additional analyses of the EEG data were performed for the final word.

4.3.4.1 Word length

The final words of the target utterances vary in the number of syllables (range: 1 to 6 syllables; median: 3). One may therefore ask whether a difference in word length across items masked relevant effects at the final word, particularly an N400 effect or later ERP effects (e.g., P600, late negativity) in Declinations. To investigate this we performed supplementary analyses for a subset of the target utterances in which the final words were two syllables (40 items). For completeness, we also report the analysis for Pre-offers. Data from 10 additional participants were excluded because of low number of artifact-free trials per condition ($N = 32$; minimum 10 trials per condition). ANOVAs were conducted for Declinations vs. Answers and Pre-offers vs. Answers at lateral sites (Action, AntPost, Hemisphere, Region, Site) and midline sites (Action, AntPost, Site). To check for a potential N400 effect or later ERP effects, the analyses were carried out in the time-windows from 250 to 500 ms and 600 to 1000 ms.

N400 window (250 – 500 ms)

Declination vs. Answer. The midline ANOVA yielded no significant main effect of Action or relevant interactions. The lateral ANOVA revealed that mean amplitude in Declinations was more negative-going than in Answers at one electrode in the fronto-temporal region in the right hemisphere (A5) (Action $\times$ Site interaction, $F(2,62) = 10.83$, $p < .01$).

Pre-offer vs. Answer. The midline analysis revealed no main effect of Action or relevant interactions. The lateral analysis yielded an Action × AntPost × Hemisphere × ROI interaction ($F(2,62) = 4.31$, $p < .05$), but follow-up analyses revealed no reliable differences.

Late effects (P600, late negativity: 600 to 1000 ms)

Declination vs. Answer. No significant effect of Action or relevant interactions were found in the midline ANOVA. The lateral ANOVA yielded an Action × Site interaction at anterior sites ($F(2,62) = 9.38$, $p < .01$), but t-tests revealed no reliable differences between the conditions.

Pre-offer vs. Answer. The lateral analysis revealed that mean amplitude was more negative-going in Pre-offers than in Answers (Action; $F(1,31) = 5.39$, $p < .05$). The midline ANOVA revealed no reliable differences.

Overall, these additional analyses converge with the results of the main ERP analysis, indicating that difference in word length did not mask an N400 or later ERP effects. In sum, Declinations do not elicit a centro-parietal N400 at the final word, nor do they involve later ERP effects. However, as expected a late negativity is present in Pre-offers relative to Answers in items with two syllables, as was the case in the main analysis of the full set of materials.

4.3.4.2 Accuracy

The behavioural results presented in section 4.3.1 indicate that accuracy is significantly lower in Pre-offers than in the other actions. To follow up on this finding, additional analyses of accuracy were carried out for each item (dialogue). These analyses revealed that the comprehension question was answered with high accuracy for the majority of dialogues; 74% of the dialogues had an overall mean accuracy between 93% and 100%. However, for Pre-offers the accuracy was at, or below, chance level (50%) in eight dialogues (out of 126), whereas this was the case for only two items in Declinations. In the Answer condition the comprehension question was never answered below chance level. The problematic items in the Pre-offers were all in Set 2 and started with *I go/am going* (Dutch “ik ga”), *I am moving* (“ik verhuis”), *I am* (“ik ben”) and *I have to* (“ik moet”).

To clarify whether the late negativity elicited at the final word in Pre-offers is driven by comprehension difficulty mainly in low accuracy items, we conducted additional analyses of the EEG data. Specifically, analyses were carried out for a subset of the target utterances that received very high accuracy in the comprehension task across speech acts (93% to 100% accuracy in each condition; 62 items). The presence of a negativity in the high-accuracy dialogues would indicate that this ERP effect does not reflect task-related comprehension difficulty but rather some other aspect of speech act comprehension in the Pre-offer dialogues. The same analytical approach was used as in section 4.3.4.1 above. For completeness we report the results for both Declinations and Pre-offers and include the N400 time-window. Six participants were removed from this analysis ($n = 36$) due to a low number of artifact-free trials per condition.

N400 window (250 – 500 ms)

Declination vs. Answer. No reliable differences were observed between the conditions in the lateral or midline ANOVA.

Pre-offer vs. Answer. No reliable differences were found in the midline ANOVA. In the lateral ANOVA a marginal main effect of Action was obtained in the left hemisphere ($F(1,35) = 4.07$, $p = .051$), indicating that Pre-offers tended to be more negative than Answers in this time-window.

Late ERP effects (600 – 1000 ms)

Declination vs. Answer. The midline and lateral ANOVAs yielded no relevant main effect of Action or interactions.

Pre-offer vs. Answer. The midline ANOVA yielded a main effect of Action ($F(1,35) = 4.34$, $p < .05$), reflecting that mean amplitude was more negative in Pre-offers across the midline. The lateral analysis revealed that a negativity was also present in the left hemisphere, reflected by a main effect of Action ($F(1,35) = 6.59$, $p < .05$).

The supplementary analyses for high accuracy items are consistent with the overall pattern of results at the final word, reported in section 4.3.3.2. Importantly, they confirm that the utterance-final negativity in Pre-offers does not arise from comprehension difficulty associated with low-accuracy items, as it is also present in items that were answered with very high accuracy.

4.4 Discussion

Conversation is characterized by tight time constraints, allowing listeners limited time between turns to recognize the speech act of the utterance and plan an appropriate response. The general goal of this article was to investigate whether speech acts can be recognized early in utterances, thereby enabling quick turn transitions. The more specific aim was to determine the robustness of the ERP correlates of speech act recognition reported in Chapter 3. The previous study used an action categorization task which a priori restricted possible interpretations of the target speech acts. In the current study we addressed this limitation, using a more natural task that does not require explicit categorization. Note that simply instructing participants to listen for comprehension does not capture the response demands and the attention level necessary for conversational interaction. There is no such thing as passive conversation: in addition to paying attention to the speech act of the unfolding turn, interactants have to plan an appropriate reply and respond in a timely manner. The true/false judgment task used in the present study probes understanding of the action without providing participants with the possible speech act categories in advance. An additional limitation of the study in Chapter 3 is that the absence of filler dialogues made it possible to use the context utterances as superficial cues to the upcoming speech acts. In contrast, the current study used additional filler dialogues to decrease the predictability of the target speech acts.

The main findings of the present study were as follows. Replicating the results of the study resported in Chapter 3, Declinations elicited a frontal positivity in an early time-window roughly corresponding to the first word and the verb. The frontal effect had an earlier onset than in the previous study, starting in the right hemisphere from 200 to 300 ms and extending to anterior bilateral and midline ROIs from 400 to 600 ms. This effect is similar to the combined effects of the right fronto-temporal and left/midline frontal positivity in the study in Chapter 3 (see also Gisladottir et al., 2015). Also consistent with the prior study, no ERP effects were found in Declinations relative to Answers at the final word. This pattern of results provides further empirical support for an early speech act recognition account, indicating that comprehension at the action level takes place early in the utterance before the final word is heard.

A different pattern of results was found in Pre-offers. In contrast to the study reported in Chapter 3, no early ERP effects were found in Pre-offers. However, consistent with prior results, Pre-offers elicited a late negativity roughly from 600 to 700 ms and 800 to 1000 ms after final word onset. These results indicate that early speech act recognition is not always possible.

Moreover, the absence of early effects in Pre-offers shows that the experimental environment does exert some influence on early action recognition. Below we will discuss these findings and their implications in more detail.

4.4.1 Further support for an early speech act recognition account

Despite differences in task demands and experimental design, the pattern of results for Declinations is remarkably similar to that reported in Chapter 3. Speech act recognition in Declinations is reflected by a frontal positivity in the early time-window corresponding to the first word and the verb, while no ERP effects are observed at the final word. The additional statistical analyses for the final word (see section 4.3.4) further confirm the absence of an N400-like effect or a later ERP effect to Declinations. The absence of ERP effects at the utterance-final word in Declinations is important as it indicates that the final word is not needed for recognizing the action, even though it is critical to the propositional meaning of the turn; the utterance-final word provides the key piece of information regarding why the proposal in the prior turn is being declined.

The early frontal positivity has an earlier onset in the present experiment, at 200 ms after utterance onset compared to 400 ms in the previous study. The early onset of the frontal positivity may reflect the influence of the changes in experimental task and/or design. In contrast to the study in Chapter 3, participants were not given labels of the relevant speech act categories in the present experiment, which limited top-down expectations of the target utterances. The predictability of the speech acts was further reduced by including additional filler dialogues. It has previously been reported, based on studies on speech act processing in the visual modality, that speech act comprehension effects have an earlier onset when predictability of the action is reduced (Egorova, Pulvermüller, et al., 2013). In particular, relevant ERP effects in written words performing requesting or naming speech acts had an onset up to 70 ms later in an EEG study using block design (Egorova, Shtyrov, et al., 2013) than in a single-trial MEG study with more variability in the experimental materials (Egorova, Pulvermüller, et al., 2013). The authors argued that in the MEG study showing earlier effects, the presentation of the target speech acts was more challenging and predictability reduced (Egorova, Pulvermüller, et al., 2013). The interpretation that less predictability leads to earlier start of speech act recognition is however somewhat counterintuitive. An alternative explanation is that differences in the onset latency of the frontal positivity may be due to greater measurement sensitivity of detecting possible differences between conditions. The current experiment employed a 64-channel electrode setup instead of the 32-channel setup in the previous study, and as a consequence a slightly different analysis approach was used (e.g., regions with fewer electrodes). The setup and analysis approach used in the present study may be more optimal for capturing the early frontal positivity.

Regardless of which account better explains the early onset of the frontal positivity, the present finding that this effect begins from 200 ms after first word onset is in line with prior ERP research which has shown that spoken words are related to the global discourse context as early as 150 ms after word onset, "well before they have been fully pronounced, and possibly even before they have become acoustically unique" (Van Berkum, 2012, p. 591). The frontal positivity in the current experiment is therefore most likely triggered by the first word of the target utterance, and possibly subsequent words up to the final word as well.

What is it about Declinations that gives rise to this early start of speech act recognition in Declinations? As discussed previously in this thesis, rejections such as those in the Declinations constitute the second part of an adjacency pair. The first part of an adjacency pair is highly constraining in terms of what kind of speech act can follow. In the present experiment, listeners should expect either an acceptance or a declination following the proposal in the context turn (*I can* *lend you money*). The first word of the response (*I…*) can then immediately be related to the speech act context, ruling out obvious acceptances of the form *great*, *thank you* etc. The verb (*I have…*) makes clear that a proto-typical acceptance is not in progress, strengthening the chance that a declination is involved. By the time the final word is reached (*credit-card*), the speech act has already been recognized. The semantic processing of the final word only adds to the propositional meaning, which is the same for all three speech acts, and hence no ERP effects are observed for Declinations at the utterance-final word. On this account, listeners do process the final word of the target utterance, but do not rely on it for recognition of the speech act.

The claim is not that listeners predicted the final word in Declinations, either at the lexical or conceptual level. If this were the case, we should find N400 facilitation (smaller amplitudes) in Declinations at the final word (see, for instance, Federmeier, 2007). The statistical analyses for the N400 window (see 4.3.4) confirm that an N400 effect is not present. Linguistic prediction, either in the "weak" or "strong" sense (Kutas et al., 2011), is not at stake. The current findings show that listeners can get the action of an unfolding utterance in advance of its completion, suggesting that predictions are made at the level of speech acts (or action more generally). Importantly, these predictions cannot be due to a predictable experimental environment, as the presence of filler dialogues in the present study made it impossible to use the context utterances as superficial cues to the target speech act. Rather, we propose that early recognition of the speech act in Declinations reflects the natural tendency to anticipate upcoming talk based on prior knowledge of action sequences. Conversationalists have acquired a lifetime of information about speech act sequences and how they are used in common situations. This information can be stored as scripts (Schank & Abelson, 1977) or schemata (Brewer & Nakamura, 1984) in memory, providing a springboard for predictions that may be necessary for smooth interactions (for the relationship between scripts and predictions, see Bar, 2009). Prediction at the speech act level could be the key to efficient turn-taking; predicting unfolding or upcoming speech acts enables listeners to start planning their responses early, which in turn allows for short gaps between turns in conversation. However, further research is required to shed light on the mechanisms that enable listeners to recognize speech acts based on only partial information.

4.4.2 Early speech act recognition is not always possible

The results for the Declinations indicate that listeners can recognize speech acts early on in the utterance, based on only partial input. This does not hold for all actions, as exemplified by the fact that speech act recognition in Pre-offers does not begin until the final word. What could explain the absence of early ERP effects in this action type? As discussed in Chapter 3, a critical feature of a pre-offer is that it initiates a pre-sequence rather than closing an adjacency pair. As a consequence, the turn immediately prior to a pre-offer (i.e., the context utterance in the present experiment) is much less constraining in regards to what action can follow. The utterance *I don't have any money* could, for instance, be followed by responses such as condolences (*That sucks*), a telling of own experience (*Me neither*), or a suggestion (*Why don't you ask somebody for a loan?*); a direct offer or a pre-offer are just additional possibilities. Since the context turn in the Pre-offer dialogues does not heavily constrain the next speech act, listeners have fewer expectations about the upcoming action and do not immediately proceed with speech act recognition. This is reflected by the absence of the early fronto-temporal positivity in Pre-offers. This account is in line with the finding from research on non-verbal action understanding that predictability is lowest at event boundaries (Zacks, Kurby, Eisenberg, & Haroutunian, 2011); Pre-offers initiate a new action sequence and therefore constitute a boundary between two “events,” which are not in a predictable relationship.

It remains to be explained why an early fronto-temporal positivity was nevertheless observed in the previous study reported in Chapter 3 (and in Gisladottir et al., 2015). In that case participants had more top-down information about the three critical speech act types. First, they were given meta-linguistic labels of the speech acts in advance, as a consequence of using an explicit categorization task. Second, in contrast to the present study, the previous experiment did not contain filler dialogues to direct participants’ attention away from the critical speech acts, thereby making it easier to form expectations about the action of the target utterance. The availability of top-down information about the critical utterances because of these two experimental factors could explain the presence of the early fronto-temporal positivity to Pre-offers in the first study. When such top-down information is limited, as in the current experiment, no strong expectations may build up prior to the Pre-offer being presented so listeners take a wait-and-see approach.

Clearly, the experimental environment exerts some influence on action recognition, in particular on the right hemisphere fronto-temporal positivity reported in Chapter 3. However, the different pattern of results in the two studies is likely reflective of a factor that can also be at play in everyday conversation. Top-down expectations are not only formed based on experimental design or the structure of action sequences, but also based on aspects such as the topic of discussion, the personality of the interlocutor, or the general course of the preceding conversation. It is therefore conceivable that speech act recognition in Pre-offers, and other actions which are not constrained by the prior turn, can under the right circumstances begin early in the utterance. What the current results show is that early recognition of the action in Pre-offers is not possible based on the prior turn alone, when such top-down information is limited.

As for the results at the utterance-final word, the late negativity in Pre-offers converges nicely with the prior study in indicating that the final word is required for comprehension of the action in that speech act type. Although the behavioral results showed lower accuracy in the comprehension task for the Pre-offers than for the other actions, the additional analyses in section 4.3.4 indicate that the late negativity does not arise from general comprehension difficulty (as it is also present in items that were answered with very high accuracy). Rather, it likely reflects that Pre-offers are more complex by nature than the other actions, not only in terms of how constraining the prior turn is, but also in the relationship to upcoming actions in conversation; through their position as a first turn in a pre-sequence, pre-offers invite inferences about what the speaker might do later in the action sequence (i.e., performing a more direct offer).

It is difficult to disentangle whether it is just the lack of constraining context (backward relationship) or the projection of an action (forward relationship) that triggers the late negativity in Pre-offers. Late negativities are rarely reported in language research and occur in situations quite different from naturalistic dialogue. For instance, they have been linked to message-level analysis in sentence-final words preceded by syntactic anomalies (Osterhout & Holcomb, 1992) and to higher processing load in linking unrelated word pairs (C. M. Brown, Hagoort, & Chwilla, 2000). The left hemisphere preponderance of the late negativity in the current experiment may be informative with respect to what it reflects. While the effect was observed at posterior bilateral and medial sites in the study in Chapter 3, it had a clear left hemisphere distribution at both anterior and posterior sites in the present experiment. This is intriguing given that pragmatic processing, including speech act comprehension, has been associated with the right hemisphere (Holtgraves, 2012). The right hemisphere is thought to operate in a bottom-up, post hoc fashion during language comprehension, while the left hemisphere has been argued to be oriented towards predictive language processing and the use of top-down cues (Federmeier, 2007). The left hemisphere's advantage in prediction has mainly been studied in regards to upcoming words and not, for instance, speech acts (for a review, see Federmeier, 2007). At face value, the left hemisphere distribution of the late negativity nevertheless suggests that post-hoc analysis of the dialogues is not at play, and is instead more consistent with the forward explanation of the late negativity, namely the proposal that understanding Pre-offers involves projecting or predicting the continuation of the sequence. Participants in the prior study in Chapter 3 may have had less reason to anticipate what could come next in the Pre-offer dialogues because the task simply required them to categorize the utterance into three known action types, leading to a more bilateral distribution of the late negativity.

The finding that Pre-offers are recognized only very late in the utterance has implications for research on the timing of turn transitions. Speech act recognition in Pre-offers, as reflected by the late negativity, begins around 600 ms after final word onset. Since the duration of the final word in the target utterances was on average 685 ms, recognition of the action begins only shortly before the offset of the final word. After the action has been identified, listeners need to plan an appropriate reply. Response planning requires at least between 600 to 1200 ms, as indicated by studies on word production (see, for instance, E. Bates et al., 2003; Indefrey & Levelt, 2004; Schnur et al., 2006). The picture that emerges is that the gap between the end of a Pre-offer and the beginning of its response should be at least between 500 and 1100 ms long, roughly estimated. This is much longer than the 200 ms average reported for turn transitions (De Ruiter et al., 2006; Heldner & Edlund, 2010; Stivers et al., 2009).

So far, corpus studies have generally ignored sequence organization as a factor that influences the length of gaps between turns, grouping all inter-speaker turn transitions together into one average (Bosch et al., 2005; e.g., Heldner & Edlund, 2010). Similarly, in conversation analysis this issue has only to a limited extent been taken into account (in the context of preferred vs. dispreferred actions; K. H. Kendrick & Torreira, 2015; Sacks et al., 1974; Schegloff, 2007). One testable hypothesis is that initiating actions (including pre-offers) take more time to recognize and are therefore followed by long gaps, while responsive actions (such as declinations and other second parts of adjacency pairs) are easy to recognize and hence followed by short gaps or even overlapping responses (see also K. H. Kendrick, 2012). Future experimental and corpus research could investigate this further to clarify accounts of turn timing and the time-course of speech act recognition in conversation more generally.

4.4.3 Remaining issues

One disadvantage of context-manipulating designs, such as the one used in the present experiment, is that ERP effects to words in the context can spill over into the target interval, leading to artificial, early ERP components (Steinhauer & Drury, 2011). This brings up the question whether the early effect of action in Declinations is caused by ERP differences arising from the context. In addition to the frontal positivity (which began 200 ms after target utterance onset), an early posterior negativity was elicited to Declinations, mainly in the left hemisphere (from 200 to 400 ms relative to Answers, and from 200 to 300 ms and 500 to 600 ms relative to Pre-offers). This effect was not reported in the prior study in Chapter 3. Figure 4.5 shows the early frontal positivity and the early posterior negativity elicited to Declinations in the present study at representative electrodes, using a baseline interval of 150 ms and 500 ms. While the waveforms for the frontal positivity are nicely aligned around target utterance onset (Panel

A), this is not the case for the posterior negativity (Panel B); small differences between the conditions are already visible at 0 ms when the first word is presented, particularly when a 500 ms baseline is used.

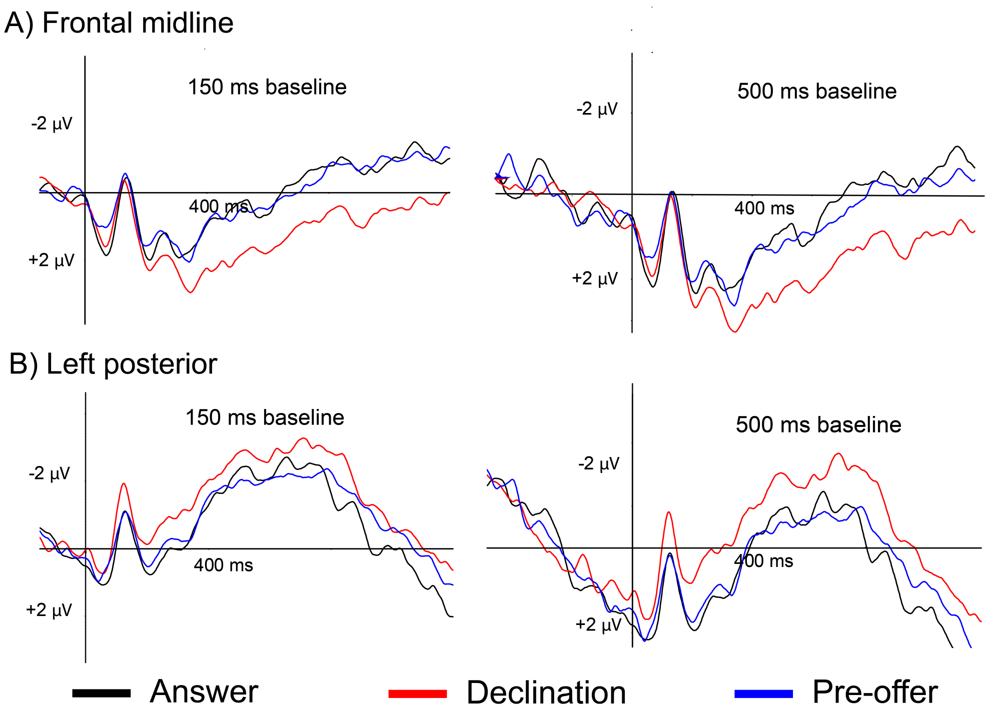

Figure 4.5: Comparison of waveforms for the early frontal positivity (A) and posterior negativity (B) at representative electrodes at the first word, using a baseline of 150 and 500 ms.

To check whether there are reliable differences between Declinations and Answers during the 300-ms gap before the target utterance, we performed additional analyses of the EEG data between the offset of the context turn and the start of the target speech act, using a 150 ms pre-gap baseline. A midline ANOVA comparing Declinations and Answers (factors: Action, AntPost and Site) did not reveal any differences during this time window (-300 to 0 ms before first word onset; $Fs < 1.24$, $ps > .27$). A lateral ANOVA (with Action, AntPost, Hemisphere, Region and Site) revealed an Action×Hemisphere interaction ($F(1,41) = 5.05$, $p < .05$), which was due to an Action×Region interaction in the left hemisphere only ($F(2,82) = 3.85$, $p < .05$). This interaction reflected that mean amplitude was more negative in Declinations at left posterior regions during the gap before target utterance onset (P2 and P3; main effect of Action, $Fs > 4.32$, $ps < .05$). These analyses indicate that the early left posterior negativity in Declinations is present already during the gap, while the frontal positivity is not. When the last 150 ms of the gap were analyzed, i.e., the interval that corresponds to the baseline period in the main EEG analysis (-150 to 0 ms before first word onset), the differences between the conditions were no longer significant ($Fs < 3.24$, $ps > .08$). The analyses for the 300 ms gap nevertheless demonstrate the presence of the posterior negativity prior to the presentation of the target utterance. This effect may therefore reflect speech act comprehension processes that are triggered by the context turn, rather than the target speech act itself. For the sake of brevity, we will leave aside the functional significance of this effect.

Importantly, the additional analyses for the gap interval show that there are no reliable differences between Declinations and Answers at frontal sites, demonstrating that the early frontal positivity is not a spill-over effect from the baseline. This corroborates the interpretation in section 4.4.1 that this early frontal effect is triggered by the first word of the target utterance, and possibly subsequent words up to the final word as well.

Another issue is whether the early or late ERP effects reported in this study may reflect lexical priming from the context. In particular, the first word in the target utterance is repeated from the prior turn in some of the dialogues; a repetition of the first word (usually the pronoun *ik,* 'I') occurs in 54% of

Declinations and 85% of Pre-offers, but not in Answers. Repetition of the verb is less frequent, occurring in 10% of Declinations, 8% of Pre-offers and 3% of Answers. ERPs, in particular the N400, are known to be sensitive to word repetition and other types of lexical priming, such as semantic association (see, for instance, Kutas & Federmeier, 2011). To obtain a measure of semantic relatedness between the context utterance and the target speech act in each condition, Latent Semantic Analysis (LSA) values (Landauer et al., 1998) were computed for the English translation of each dialogue[11]. LSA values have been found to correspond well with the pattern of results in priming studies (Dorothee J. Chwilla & Kolk, 2002; Landauer et al., 1998). We first computed LSA values for the early time-window, comparing the context turn to the target utterance up to the final word. The LSA values for this early part of the target utterance were 0.07 for Answers, 0.31 for Declinations and 0.42 for Pre-offers (the higher the value, the more semantic similarity). An omnibus ANOVA on the LSA values revealed a main effect of Action ($F(2,369) = 172.61$, $p < .01$), and pairwise comparisons confirmed significant differences in LSA values between all actions ($p < .01$). We then computed LSA values for the late time-window, comparing the utterance-final word itself to the preceding words in the dialogue (i.e., the context turn plus the target utterance up to the final word). The LSA values for the final word were 0.13 for Answers, 0.12 for Declinations and 0.13 for Pre-offers; there were no differences in LSA values across actions ($F(2,369) = 0.20$, $p = .81$).

The LSA analyses for the early time-window show that the three action conditions differ in the amount of semantic similarity between the context and the early part of the target utterance. However, we consider it unlikely that

---

[11] The computation was performed using document-to-document mode with "General reading up to 1st year of college" as the semantic space.

semantic priming is a driving factor in the early ERP effects for the following reason. Pre-offers have the highest LSA value and the highest percentage of repetition at the first word as described above (85%), and should therefore show the largest effect of semantic priming in the early time-window. The fact that no early ERP effects were found in Pre-offers relative to Answers (which have the lowest LSA value and almost no repetition) demonstrates that semantic priming is not at play. If the early frontal positivity in Declinations were due to semantic overlap, it would be most pronounced in Pre-offers. As for the late time-window, the LSA analyses show that there are no differences in the semantic relatedness between the utterance-final word and the preceding dialogue across the three actions. This entails that the late negativity to Pre-offers cannot be due to effects of semantic association or lexical repetition. Thus the results of the LSA analyses clearly speak against the idea that the ERP effects reported in this study can be accounted for by a semantic priming account.

4.4.4 Conclusions

This study demonstrates the robustness of two ERP correlates of speech act recognition reported in Chapter 3. First, again an *early frontal positivity* (starting at 200 ms after target utterance onset) was found to Declinations – when the action is in a highly-constraining context and the sequence is coming to a close. Second, as in the first study, a *late negativity* (from 600 ms after final word onset) occurred to Pre-offers – when the context is less-constraining and a new action sequence is being initiated. Based on differences in the timing of the two ERP effects (utterance onset vs. utterance-final word), polarity (positivity vs. negativity) and scalp distribution (anterior sites vs. poterior/left hemisphere), we propose that these effects reflect qualitatively distinct cognitive processes involved in speech act recognition.

While the current experiment provides further empirical support for the early speech act recognition account proposed in Chapter 3, it extends prior findings by demonstrating that the experimental environment has some influence on early action recognition. In particular, in contrast to the results of Chapter 3 no early effects were observed to Pre-offers. The resulting picture of the time-course of speech act recognition is that recognition of the action is made very early in the utterance in highly-constraining action contexts (Declinations), while speech act recognition only begins at the final word when the prior action is less-constraining (Pre-offers). These findings call for revisiting the timing of turn transitions in conversation, as they imply that certain actions are recognized only very late in the utterance and hence should be responded to much later than previously assumed. Most importantly, the findings of Chapter 4 confirm that under the "right" circumstances listeners can get the action of an unfolding utterance in advance of its completion, based on only limited input, that is before the critical final word has been heard. Early speech act recognition in such cases may reflect the natural tendency to anticipate upcoming talk based on prior knowledge of action sequences.

5 Neural oscillations involved in speech act recognition

This chapter is a modified version of:
Gisladottir, R. S., Bögels, S. & Levinson, S. C. (Under review). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialogue.

Abstract
Everyday conversation requires listeners to quickly extract verbal actions, or so-called speech acts, from the underspecified linguistic code and prepare a relevant response within the tight time constraints of turn-taking. The general aim of the present study was to determine the time-course of speech act recognition by investigating oscillatory activity of the electroencephalogram (EEG) during comprehension of spoken dialogue. The more specific aim was to shed light on the role of anticipatory processes in speech act recognition. Participants listened to short, spoken dialogues with target utterances that delivered three distinct speech acts (Answers, Declinations, Pre-offers). The targets were identical across conditions, but differed in the type of speech act performed and how it fit into the larger action sequence. Speech act comprehension was mainly associated with reduced power in the alpha/beta bands just prior to and during the beginning of speech acts in highly constraining action contexts (Declinations). The alpha/beta effect is dependent on the characteristics of the speech act sequences, suggesting that factors such as sequential constraints have an impact on the timing of speech act recognition. Based on the role of alpha and beta desynchronization in anticipatory attention, the present results are taken to indicate that anticipatory processes plays a role in speech act recognition. Anticipatory attention before and during the

beginning of speech acts could be critical for efficient turn-taking, allowing for early speech act recognition based on only partial input.

5.1 Introduction

Having a conversation requires far more than simply retrieving the meaning of individual words and combining them into a syntactic structure. A critical aspect of successful communication is the ability to grasp the function of the utterance in context, i.e., what speech act or verbal action is being performed; for instance, was it an offer, question, rejection, or a compliment? As discussed in section 1.2 of this thesis, speech act recognition is difficult from an individual's cognitive perspective for two reasons. First, it is often the case that an utterance is compatible with multiple speech acts, forcing listeners to rely on the conversational context or other factors to understand the intended action. Second, turn-taking in conversation is characterized by very tight time constraints, giving participants limited time to recognize the speech act and plan a response. Listeners have to quickly extract the speech act from the underspecified linguistic code, prepare a relevant reply and respond within 200 ms, which is the reported average for turn transitions in conversation (De Ruiter et al., 2006; Heldner & Edlund, 2010; Stivers et al., 2009). A crucial question is how listeners are capable of speech act recognition at such an amazing speed.

One proposal put forth in the introduction to this thesis is that speech act recognition takes place early in the incoming turn, allowing listeners more time to plan their response. A factor that could enable such early ascription of the action is the sequential context, i.e., the prior action (Gisladottir et al., 2015; Levinson, 2013). The ERP studies presented in Chapter 3 and 4 suggest that when the prior utterance is highly constraining in terms of what actions can follow, listeners can predict the action of the unfolding turn in advance of the utterance's completion. Early action recognition through anticipation of the action could be critical for efficient turn-taking, enabling listeners to plan their reply early and respond in a timely manner. However, early speech act recognition is not always possible. The ERP results in Chapters 3 and 4 show that when the prior action is not constraining and a new sequence is being initiated, processing based on the complete utterance seems to be required.

The general aim of the present study was to further determine the time-course of speech act recognition by investigating oscillatory EEG activity during comprehension of spoken dialogue. ERPs only reflect a certain part of the event-related EEG signal, namely phase-locked responses which remain after averaging in the time domain (Bastiaansen et al., 2008; Pfurtscheller & Lopes da Silva, 1999). Ocillatory activity, in contrast, is often non-phase locked; oscillations are ongoing, occurring even in the absence of an experimental task, and as a result the phase of the oscillation varies at the time of stimulus presentation (Bastiaansen et al., 2008). The non-phase locked response therefore cancels out in the averaged ERP. By investigating oscillatory dynamics we can get a more complete picture of the time-course of speech act recognition and better understand the cognitive processes underlying successful conversation.

The more specific aim of this study was to shed light on the role of anticipatory processes in speech act recognition. Suppression of EEG oscillations in the alpha (8 to 12 Hz) and beta (13 to 30 Hz) frequency bands have been associated with anticipatory processing, both in relation to motor preparation (see, for instance, Pfurtscheller & Lopes da Silva, 1999) and anticipatory attention, i.e., when attention is oriented towards an upcoming stimulus to facilitate its processing (Bastiaansen et al., 2001; e.g., Bastiaansen & Brunia, 2001; Jones et al., 2010; van Ede et al., 2014). For instance, studies on attention in the visual or somatosensory modalities typically report modulations of alpha and/or beta activity in the interval between some symbolic cue and the anticipated target (such as a tactile event), with the beta activity sometimes extending into the stimulus period (van Ede et al., 2014). Especially relevant for the present study, a recent study on anticipation of turn-endings (using isolated turns from real conversations) reported desynchronization in the low beta range (11 to 18.5 Hz) in turns with predictable compared to non-predictable final words (Magyari, Bastiaansen, De Ruiter, & Levinson, 2014). The beta desynchronization began as early as 1250 ms before the end of the turn and was taken to reflect early anticipation of turn-endings via syntactic, semantic, and temporal processing. Based on the results of these studies, one may predict that the comprehension of relatively predictable speech acts involves a decrease in alpha and/or beta power before and during the beginning of the utterance, reflecting anticipatory attention that facilitates early speech act recognition.

Alternatively, speech act recognition may involve modulations of gamma oscillations (30 to 100 Hz). An increase in gamma band power has been reported in studies on pragmatic phenomena such as irony (Spotorno et al., 2013) and world knowledge violations (Hagoort et al., 2004). More generally, it has been suggested that gamma plays a functional role in normal sentence comprehension, as gamma increase is found in correct, spoken sentences and not in sentences with semantic violations (Hald, Bastiaansen, & Hagoort, 2006). Modulations of gamma oscillations in the present study would indicate that processes not based on anticipatory attention are involved in speech act recognition.

5.1.1 The present study

To our knowledge, no prior studies have investigated oscillatory activity during comprehension of spoken speech acts embedded in conversational contexts. Does oscillatory activity corroborate the time-course of speech act recognition described in Chapter 4? Are anticipatory processes involved in speech act recognition, reflected by modulations in the alpha or beta frequency range? To address these questions we investigated oscillatory brain responses during speech act recognition using the EEG dataset from the experiment in Chapter 4. In contrast to the experiment in Chapter 3, the study in Chapter 4 used a true/false judgment task and additional filler dialogues to reduce strategic processing based on the experimental environment and is therefore better suited to investigate anticipatory processes involved in speech act recognition. The materials are short, spoken dialogues with target utterances that perform three different speech acts – Answers, Declinations, and Pre-offers – depending on the prior turn. Declinations and Pre-offers – the critical conditions – have in common that they are more indirect than Answers, as more inferencing is required to understand the action. However, they differ in the type of action being performed and its relationship to prior and upcoming turns in conversation. More specifically, Declinations occur in highly-constraining contexts, as the context turn (containing an offer or proposal) limits the range of possible responses to only two speech acts – an acceptance or declination. Pre-offers, on the other hand, occur in less-constraining action contexts and initiate a new action sequence, which frequently contains a more direct offer (the reader is referred to Chapters 3 and 4 for details). For convenience, examples of dialogues with the critical speech acts are presented in Table 5.1.

Condition	Context	Target Utterance
Answer	*Hoe ga je voor het ticket betalen?*	*Ik heb een creditcard.*
	How are you going to pay for the ticket?	I have a credit-card.
Declination	*Ik kan je wat geld lenen voor het ticket.*	*Ik heb een creditcard.*
	I can lend you money for the ticket.	I have a credit-card.
Pre-offer	*Ik heb geen geld om het ticket te betalen.*	*Ik heb een creditcard.*
	I don't have any money to pay for the ticket.	I have a credit-card.

Table 5.1: Stimuli in Dutch and English translations

Prior ERP studies using the same dialogues, presented in Chapters 3 and 4, have shown that Declinations and Pre-offers elicit different ERP patterns. In the case of Declinations, frontal positivities are observed as early as 200 ms after first word onset relative to Answers, while no ERP effects are found at the final word. These results show that when the prior utterance is highly constraining in terms of what action can follow, recognition of the action takes place early in the utterance, before the final word. In contrast, Pre-offers mainly elicit a late negativity at the utterance-final word relative to Answers, indicating that processing based on the complete utterance is required when the prior action is less constraining and a new sequence is initiated.

Based on these ERP studies, and the assumption that reduced alpha/beta power reflects anticipatory processes that may play a role in early speech act recognition, we made the following predictions. In the case of Declinations, the highly-constraining context should allow for recognition of the action early in the utterance (*I have a credit-card*). Due to the association of alpha and beta oscillations with anticipatory attention, we hypothesized that such early effects would involve less power in the alpha or beta bands (relative to Answers and Pre-offers) in an early utterance time-window roughly corresponding to the first word or the verb, or during the pre-stimulus interval just before the target utterance. At the final word, however, no effects are expected in Declinations. If the action has already been recognized early in the utterance, the processing of the final word should only add to the propositional meaning of the utterance, which is the same for all three conditions and hence no differences in oscillatory activity should occur at the final word.

As for Pre-offers, recognition of the action should be possible only late in the utterance due to increased complexity; the context is less-constraining and understanding Pre-offers may involve projecting the next action (i.e., that a pre-offer leads to a more direct offer). This should mainly be reflected by oscillatory power differences at the utterance-final word, relative to both Answers and Declinations. As for what frequency band is involved, we speculated on two possibilities. If understanding Pre-offers involves anticipating a more direct offer, then recognition of the action may involve increased anticipatory attention at the utterance-final word, reflected by a decrease in alpha/beta power. Alternatively, if the complexity of Pre-offers mainly calls for increased pragmatic processing, and not anticipatory processes, then we may observe modulations in the gamma frequency range.

5.2 Methods

5.2.1 Participants, materials, procedure, EEG recording

Chapter 4 describes the participant pool, materials, procedure and EEG recording for the dataset used in the current analysis; for details, the reader is referred to sections 4.2.1 through 4.2.4.

5.2.2 Behavioural data analysis

While the behavioural data used in this chapter are the same as those used in Chapter 4, the data were reanalyzed for the current investigation using a smaller subset of participants. More participants were excluded from the time-frequency analysis of the EEG data than in the ERP analysis in Chapter 4 due to a different analysis approach. More specifically, a larger epoch was extracted from the EEG for the time-frequency analysis in order to include the 300 ms gap before the target utterance (capturing anticipatory processes) and to allow for plotting of raw data using pre-stimulus baselines. For consistency between the behavioural and EEG results, the accuracy in the true/false comprehension task was only analyzed for the 37 participants used in the time-frequency analysis. The accuracy data were analysed with mixed-effects logistic regression using the lme4 package (D. Bates et al., 2012) in the statistics software R (R Core Team, 2013).

5.2.3 Time-frequency analysis of oscillatory power

The critical time-windows of interest were from -300 to 600 ms relative to the onset of the first word in the target utterance (*early utterance time-window*) and from 100 to 1000 ms after onset of the final word (*late utterance time-window*). These early and late time-windows are the same as those used in the ERP studies in Chapters 3 and 4, except that the early utterance time-window includes the 300 ms interval (silence) before the target utterance to capture pre-stimulus anticipatory effects. If early speech act comprehension in Declinations does involve anticipatory processes based on the context turn, this may be reflected by power modulations in the alpha and/or beta bands already before target utterance onset in Declinations. Thus the early window spans speech act comprehension just before the target utterance onset up to the presentation of the final word (i.e., including the first word and the verb), while the late time-window captures effects elicited to the final word. Importantly, these time-windows do not overlap.

To reduce boundary effects in the subsequent time-frequency analysis and for plotting of the raw data using pre-window baselines, EEG data were segmented into larger epochs: from -1000 ms to 1000 ms relative to first word onset for the early time-window analysis, and from -600 ms to 1700 ms relative to final word onset for the late time-window analysis. Trials containing artifacts exceeding ±75 μV at eye-monitoring sites and ±100 μV at other sites were excluded[12]. Data from seven participants with less than 22 trials per condition were removed from further analysis, in addition to data from one additional participant with clear heart artifacts (remaining participants N = 37; nr. of trials left for analysis in the final dataset was 3830 or 82%).

Time–frequency representations (TFRs) of power were computed with a sliding time-window approach using the Matlab toolbox FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). For the low-frequency range (2 to 30 hz), power was calculated for each trial with a Hanning taper, using 400 ms time-windows that were advanced in steps of 10 ms and 1 Hz. For the higher frequencies (30 to 90 Hz) a multitaper was used, with 400 ms sliding time-windows, advanced in steps of 10 ms and 2.5 Hz, with frequency smoothing of 5 Hz. Multitapers yield better frequency smoothing which is advantageous for EEG signals above 30 hz ("Time-frequency analysis," 2014). The TFRs of power were averaged over trials for each participant and condition. Since expressing post-stimulus EEG power relative to pre-stimulus EEG power can be problematic (see, for instance, Hu, Xiao, Zhang, Mouraux, & Iannetti, 2014), a baseline correction was not applied.

[12] The artifact rejection was based on baselined data, using a 150 ms baseline prior to the gap between the two utterances, but non-baselined data were used for the time-frequency analysis.

5.2.4 Statistical analysis of TFRs of power

The time-frequency representations of power were submitted to nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) in Fieldtrip (Oostenveld et al., 2011). This approach has the advantage of offering a straightforward way to deal with the multiple comparisons problem. First, two experimental conditions were compared at a time with a dependent-samples t-test for every channel-frequency-time sample in the TFR. Samples that passed a predetermined threshold (p < .05) were selected and clustered based on adjacency in time, space, or frequency. The cluster-level statistics were calculated as the sum of all t-values within the cluster. The Monte Carlo method was then used to determine the significance of the cluster. A null distribution which assumes no differences between conditions is created by randomly assigning participant averages to one of the two conditions. This procedure was repeated 1000 times and cluster-level statistics were computed for each randomization (as above). Finally the observed cluster-level test statistics are compared to the null distribution; the Monte Carlo estimate of the p-value is the proportion of random partitions resulting in a larger test statistic than the observed cluster. The two experimental conditions were then considered significantly different if the p-value was smaller than the critical alpha value, which was set at 0.05 and corrected for a two-tailed test, effectively 0.025.

5.3 Results

5.3.1 Behavioural data

Behavioural responses from all participants included in the time-frequency analysis were analyzed (N = 37). Overall mean accuracy in the comprehension question was 93.3% (SD 6.9%). Participants correctly answered the question for 97.9% of Answers (SD 2.5%), 95.8% of Declinations (SD 5.3%) and 86.1% of

Pre-offers (SD 15.6%). The mixed-effects logistic regression model of the accuracy data included Action as a fixed effect. Stimulus set was not used as a fixed factor in the current analysis, in contrast to previous chapters, for consistency between the behavioural and EEG analysis; the cluster-based permutation test does not allow the straight-forward computation of Action $\times$ Set interactions (since it is designed to compare only two conditions at a time), and since the Set factor is not of theoretical relevance on its own, it was not included in the time-frequency analyses. For the random effect structure of the mixed-effects model we included random intercepts by participant and item, as well as by-participant and by-item random slopes for Action. The model indicated that Answers were categorized more accurately than Pre-offers (Estimate: 2.30 SE: 0.61, $z = 3.76$, $p < .001$). Responses to Declinations were also more accurate than to Pre-offers (Estimate: 1.63 SE: 0.50, $z = 3.27$, $p < .001$). The comparison between Answers and Declinations was not significant (Estimate: -0.51, SE: 0.65, $z = -0.78$, $p = .43$). The same pattern of results was obtained for the same EEG dataset (but with more participants) in Chapter 4.

5.3.2 EEG results

5.3.2.1 Early utterance time-window

Statistical comparison of Declinations and Answers in the early utterance time-window revealed one significant cluster in the theta, alpha and beta frequencies from 2 to 24 Hz, approximately 150 ms before onset of the target utterance until 600 ms into the sentence ($p < .01$). Lower power was observed in Declinations relative to Answers in these frequencies (see Figure 5.1). The difference was largest in the following latency epochs: in the theta range (2 to 8 hz) from -50 to 100 ms; alpha / low beta range (11 to 15 Hz) from -100 ms to 400 ms; and in higher beta frequencies (19 to 22 Hz) from 250 to 450 ms. The effect was widespread over bilateral anterior sites in the theta and alpha/low beta frequencies, with a left hemisphere preponderance at anterior and posterior sites for the higher beta (see Figure 5.1, panel C).

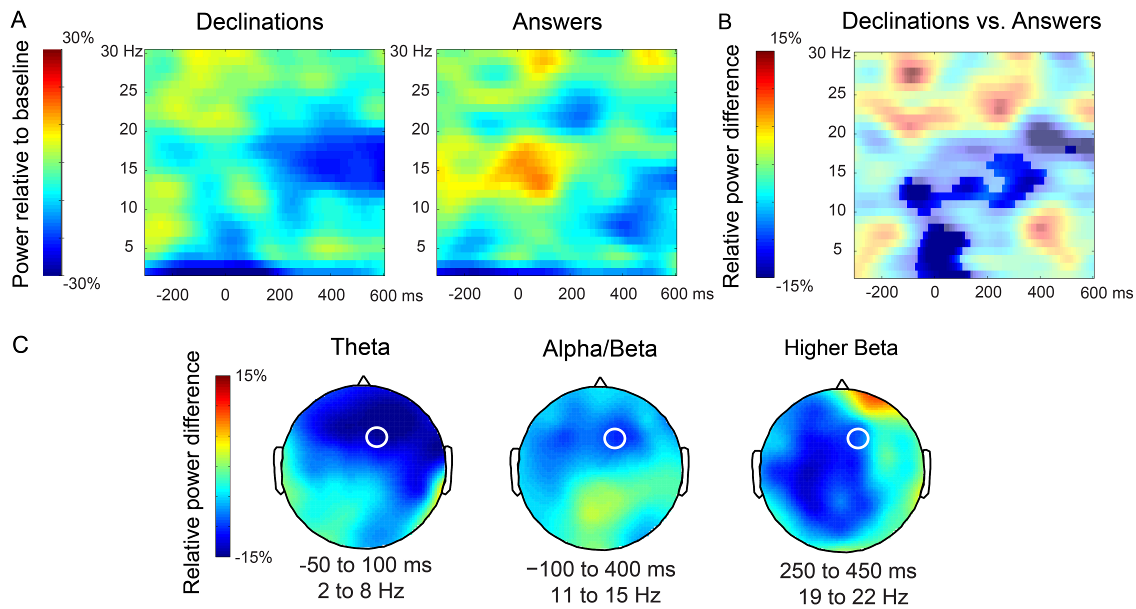

Figure 5.1: Early utterance time-window: Declinations vs. Answers. A) Time-frequency representations of power changes at a representative frontal site for Declinations and Answers separately. The colors indicate power increase/decrease in each condition relative to a pre-stimulus baseline from -650 to -500 ms before target utterance onset (baseline was only used for plotting purposes). For location of the site, see circles in panel C. B) Relative power difference between Declinations and Answers at the representative frontal site in transparent colors, with the significant cluster overlaid in opaque colors. C) Topographic maps illustrating the distribution of the relative power differences between Declinations and Answers in the relevant time-windows and frequencies.

The comparison between Declinations and Pre-offers in the early utterance time-window also yielded one significant cluster approximately -250 ms to 250 ms relative to target sentence onset in the theta and alpha/low beta frequencies, from 2 to 19 Hz (p < .01). Lower power was observed in Declinations relative to Pre-offers in these frequencies (see Figure 5.2). The difference was mainly in the theta band (2 to 8 hz) from -50 to 100 ms and in the alpha / low beta band (8 to 17 hz) from -150 to 50 ms. The effect was most prominent at anterior sites in both frequency bands (see Figure 5.2, panel C).

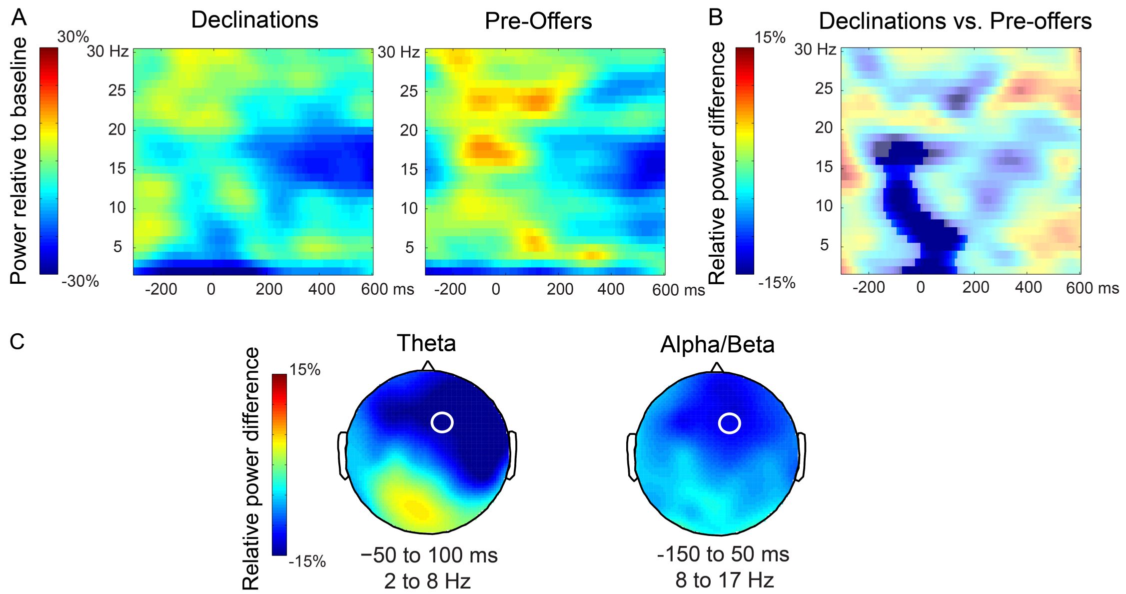

Figure 5.2: Early utterance time-window: Declinations vs. Pre-offers. A) Time-frequency representations of power changes at a representative frontal site for Declinations and Pre-offers separately. The colors indicate power increase/decrease in each condition relative to a pre-stimulus baseline from -650 to -500 ms before target utterance onset (baseline was only used for plotting purposes). For location of the site, see circles in panel C. B) Relative power difference between Declinations and Pre-offers at the representative frontal site in transparent colors with the significant

cluster overlaid in opaque colors. C) Topographic maps illustrating the distribution of the relative power differences between Declinations and Pre-offers in the relevant time-windows and frequencies.

The Pre-offers vs. Answers contrast did not yield any significant clusters in the lower frequencies ($p > .23$). As for the higher frequencies in the gamma band (>30 Hz), no significant clusters were obtained for any of the comparisons ($p > .32$).

5.3.2.2 Late utterance time-window

The statistical analyses did not reveal any significant clusters in the low frequencies ($p > .06$), nor in the higher frequencies ($p > .25$).

5.4 Discussion

Our results show that brain oscillations differentiate speech acts during comprehension of spoken dialogue. Supporting an early speech act recognition account, effects of action were only observed just prior to and during the beginning of the target utterances, while no differences were found between the speech acts at the final word. More specifically, less power was present in the alpha/beta and theta bands in Declinations in this early time-window, relative to both Pre-offers and Answers. Overall, the results show that the time-course of speech act recognition is influenced by the type of action performed and provide evidence that anticipatory processes play a role in the comprehension of speech acts. Below we discuss these findings in more detail.

5.4.1 Alpha/beta power and anticipatory attention to upcoming speech acts

We hypothesized that oscillatory activity would show an early effect of action in Declinations relative to Answers and Pre-offers. This prediction was based on the results from the ERP studies in Chapter 3 and 4 and the assumption that the highly-constraining context in the Declination dialogues allows for early recognition of the action. We also speculated that such early effects would be observed in the alpha or beta bands, which have been associated with anticipatory attention. In line with this, we observed less power in Declinations relative to Answers in the alpha and beta range mainly from -100 to 450 ms after target utterance onset. Less power was also observed in Declinations relative to Pre-offers in the alpha and beta band from -150 to 50 ms. Given that the effects are mainly in the low beta range (also called Beta 1; 13 to 18 Hz (Weiss & Mueller, 2012)), we will focus on the importance of beta oscillations.

The relationship between beta desynchronization and anticipation has been well established, both in the sense of preparation for movement (see, for instance, Pfurtscheller & Lopes da Silva, 1999) and preparation for perception, that is, anticipatory attention (Jones et al., 2010; van Ede, de Lange, Jensen, & Maris, 2011; van Ede et al., 2014). The pre-stimulus onset of the beta effect in the present study (i.e., 150 ms before the presentation of the target utterance), is similar to findings in studies on anticipatory attention reporting beta desynchronization in the interval before an anticipated target, with the beta activity sometimes extending into the stimulus period (see, for instance, van Ede et al., 2014). In the language domain, a link between beta desynchronization and anticipation was recently established in a study using turns from real conversation (Magyari et al., 2014). In contrast to the present experiment, which focuses on the relationship between two turns, Magyari et al. (2014) investigated oscillatory activity in isolated turns that had either a predictable or a non-predictable ending (e.g., predictable: "*eh I live in the same house with four women and with another man", unpredictable: "uh and then, she was again eh alone in eh in the north*"). Predictable turns were accompanied by a power decrease in the low beta band (11 to 18.5 Hz) as early as 1250 ms before the turn ended. As in the present study, the low beta effect was observed in frontal areas, although it was mainly in the left hemisphere. The frontal beta power decrease was localized to areas involved in language as well as anticipation and attention. The authors related the beta power decrease to (non-motor) anticipation of the turn-ending.

Given these prior findings, we take the reduced beta power in the present study to reflect increased anticipatory attention to Declinations. The early onset of the power difference in Declinations relative to Pre-offers (at 150 ms before the start of the target utterance) is in agreement with the assumption that the Declinations are more anticipatable, due to the adjacency pair structure. A characteristic feature of this action sequence type is that the first part of the pair – the context utterance in this study – is highly constraining in terms of what action can follow. The adjacency pair structure may act as a cue to restrict the possibility space for the upcoming speech act, reducing the likely outcomes down to two (acceptance or declination). As a result, the Declinations are relatively anticipatable and engage anticipatory processes just before and during the target utterance. Pre-offers, on the other hand, are initiating actions (i.e., start a new action sequence). As such, they are much less predictable based on the prior context and hence involve anticipatory attention to a lesser degree than Declinations.

However, sequential constraints cannot be the sole factor that determines the beta power effect, in particular the pre-stimulus power difference in Declinations relative to Answers. This is because Answers, the control condition, also form a second part of an adjacency pair, and thus should involve the same degree of constraint as Declinations do, according to the conversation analytic literature (see, for instance, Schegloff, 2007 for a discussion on adjacency pairs). This was not addressed in the ERP studies in Chapter 3 and 4 because these studies investigated the processing of the speech act itself, where indirectness also plays a role, and not the preceding gap: the rationale was that Declinations and Pre-offers are more indirect than Answers and therefore should elicit ERP effects relative to Answers, but critically that the timing of these effects would differ for the actions due to their sequential characteristics. Indirectness only comes into play once the target utterance has started. Thus differences in indirectness (or other features of the utterance) and action sequencing cannot explain the pre-stimulus beta power effect in the Declination dialogues relative to Answers, during the gap between the turns.

One factor that differentiates the Declination and Answer dialogues is what implications are involved in the speech act interaction. While the context in the Answer dialogues (a question) simply calls for information, the first turn in the Declination dialogues contains an offer of assistance, which – if accepted – would involve some additional actions. Moreoever, while the first turn of the Declination dialogues has the potential to be followed by a *face-threatening* action (P. Brown & Levinson, 1988; Goffman, 1955) such as rejection, this is much less the case for the context turns in the Answer dialogues (a question). There is, therefore, "more at stake" in the Declination dialogues. The social significance of a turn could influence processes of speech act comprehension. Studies show that strongly valenced input (e.g., a face with a positive or negative expression) enhances attention and thereby facilitates perception of events (Dolan, 2002; Vuilleumier, 2005; Vuilleumier & Huang, 2009). It is possible that the social implications of the speech acts in the Declination dialogues contributed to increased anticipatory attention just before and during the target utterance in this condition. This could also explain why the beta power difference was longer lasting in Declinations relative to Answers than relative to Pre-offers. Speech acts are not only embedded in sequential contexts, but also interpersonal and social contexts that can influence their interpretation.

The results from the present study suggest that in addition to a high level of sequential constraints, other characteristics of the speech act interaction can lead to increased anticipatory attention, such as the social significance of the response. However, further research is needed to investigate how such characteristics of speech act sequences influence the time-course of speech act recognition.

The relationship between anticipatory attention, as reflected by the reduced alpha/beta power in the present study, and the early speech-act-related ERP effect reported for the same EEG data (in Chapter 4) remains to be addressed. Declinations elicited a frontal positivity with a left and midline distribution starting from 200 ms after target utterance onset. Research on vision has shown an inverse relationship between pre-stimulus alpha and stimulus-evoked ERP components; attention reduces pre-stimulus alpha power but enhances ERP components such as the P1 and N1 (see, for instance, Rajagovindan & Ding, 2011; Sauseng et al., 2005). The pre-stimulus alpha reduction is thought to improve processing of the upcoming stimulus (Rajagovindan & Ding, 2011). However, research in the auditory domain has yielded conflicting results, reporting both inverse and direct relationships between prestimulus alpha power and auditory ERP components (for a discussion, see De Blasio & Barry, 2013). As for the beta band, there is limited research on the association between prestimulus beta and auditory ERP components (but see De Blasio & Barry, 2013). It is therefore not clear what the relationship between alpha/beta power in the present study and the speech-act-related frontal positivity is. However, we speculate that anticipatory attention just before and during the beginning of the Declinations facilitates active processing of the speech act, reflected by the frontal positivity. In other words, the alpha/beta effect does not reflect precise predictions of the target speech act, but rather anticipatory attention that facilitates early recognition.

5.4.2 Theta decrease around target utterance onset

In addition to beta band activity, differences in theta band were also observed for the three speech act conditions; less power was found in the theta range from -50 to 100 ms in Declinations relative to both Answers and Pre-offers. Active processing is usually associated with a power *increase* in the theta band. For instance, theta power increases with more working memory load (Jensen & Tesche, 2002) and top-down cognitive processing (Min & Park, 2010). In the language domain, increased theta power has been associated with the retrieval of lexical-semantic information (Bastiaansen et al., 2008) as well as larger demands on verbal working memory (for a review, see Bastiaansen et al., 2008), for instance during processing of irony (Spotorno et al., 2013). The fact that Declinations are associated with *less* power in the theta band is therefore surprising, given that recognizing the action should be more difficult in Declinations than in Answers. However, a theta power *decrease* has recently been implicated in language processing, more specifically in verb generation (when a noun is presented and the participant's task is to generate a matching verb) (Hermes et al., 2014). Thus, two possible interpretations can be made of the theta power differences in Declinations relative to the other actions around target utterance onset: that *high* theta power in Pre-offers and Answers reflects more working memory or top-down cognitive processing in these conditions, or that *low* theta power in Declinations reflects increased demands on the language processing system in that condition. While the latter account better fits the theoretical assumptions outlined in the introduction to this chapter, more research is needed to understand how theta power relates to speech act comprehension. Moreover, it has to be noted that the effect is short-lived. A good indication of the presence of genuine oscillatory activity – as opposed to power due to ERPs or muscle artifacts – is that it lasts at least one or two periods of the oscillation for low frequencies (Bastiaansen et al., 2008). Given that the theta activity spans only about 150 ms – less than one cycle in this frequency range – it is possible that the theta differences are a reflection of an ERP component[13]. Indeed, the ERP analyses of the same data, reported in Chapter 4, indicate that a posterior negativity was present in Declinations relative to both Pre-offers and Answers during the gap between the context and the target. The robustness and functional significance of the theta effect is therefore unclear.

5.4.3 The time-course of speech act recognition and future questions

Differences in oscillatory activity for the three speech act conditions were only observed for Declinations relative to the other actions in the early utterance time-window, roughly corresponding to the silence before the target utterance, at the first word and the verb. The absence of final-word effects for Declinations converges with the ERP results in Chapter 3 and 4, providing further support for the early speech act recognition account. When the prior turn is highly constraining in terms of what action can follow, listeners seem to recognize the action before the final word. This is the case even though the speech act is relatively indirect and contains no morphosyntactic speech act clues (recall that the target utterances were the same across all conditions and thus underspecified for the action).

In Chapter 3 and 4 I hypothesized that listeners can get the speech act in advance of its completion based on anticipation at the speech act level. The present results, namely low beta power just before and during the beginning of

[13] Note that this argument does not hold for the beta activity, as it is both longer lasting and in a higher frequency band; two cycles of a 15 Hz oscillation in the low beta band last for 130 ms, while the activity was present in Declinations for up to 200 ms relative to Pre-offers and 500 ms relative to Answers.

the target utterance in Declinations, provide more direct evidence for the involvement of anticipatory processes in the recognition of speech acts. Early action recognition in Declinations may reflect the natural tendency to anticipate upcoming talk on the basis of action sequence knowledge and other factors. This proposal is in line with recent models in psycholinguistics and neuroscience which argue that prediction based on past experience plays an important role in cognition (see, for instance, Bar, 2009; Martin J. Pickering & Garrod, 2013; Schacter et al., 2007). The precise mechanism behind such anticipation during speech act comprehension is however still an open question. What, for instance, is the role of action memory, fast and frugal heuristics (Gigerenzer & Goldstein, 1996) or theory of mind (e.g., C. Frith & Frith, 2005) in guiding anticipation? Does our knowledge of speech act sequences consist of abstract sequences, or do we form action scripts with additional information about the situational context to aid prediction (e.g., restaurant scenario)? The relationship between linguistic processing, action knowledge and anticipation is an important topic for future research that aims to understand the efficiency of everyday conversation.

Contrary to our expectations, we did not observe oscillatory effects in Pre-offers at the final word. In comparison to Declinations, Pre-offers are less predictable based on the prior turn and invite more inferences about upcoming talk (i.e., that a more direct offer is underway). Based on the final-word negativity described in Chapter 3 and 4, and oscillatory effects reported in studies on anticipation and linguistic pragmatics, we speculated that this increased complexity in Pre-offers might be reflected in modulations of alpha/beta or gamma oscillations at the final word. The fact that no final-word effects were observed in Pre-offers suggests that phase-locked event-related potentials are more sensitive than non-phase-locked oscillations to the aspect of speech act comprehension reflected by the late negativity. The neural oscillation results are nevertheless revealing, as they do not confirm the speculation put forth in Chapter 4 that understanding Pre-offers involves anticipation of a more direct offer. If the late negativity does reflect the projective nature of Pre-offers, one could expect this to be reflected by reduced alpha/beta power at the utterance final word.

The current results for Pre-offers do, however, converge with the ERP results of the same EEG data (see Chapter 4) in that no effects were found in the early utterance time-window relative to Answers. The lack of early effects in the present study provides convergent evidence for the difference between the time-course of speech act recognition in two relatively indirect speech acts. In Declinations the speech act is recognized early in the utterance, based on only partial information. In contrast, the comprehension of Pre-offers is characterized by a more integrative approach – that is, waiting for the final word before the action can be recognized. The upshot of this is that the exact same utterance (e.g., *I have a credit-card*) is processed in very different ways depending on the speech act it performs and how it fits into the larger action context provided by the prior turn. The speech act dimension is an important aspect of context that should not be ignored in studies of utterance meaning and language use.

5.5 Conclusions

How is it that listeners can recognize speech acts so efficiently, evidenced by the extraordinarily fast transitions between turns in conversation? The primary goal of this study was to address this question by investigating oscillatory activity during comprehension of spoken speech acts. The more specific aim was to shed light on the role of anticipatory processes in speech act recognition. As predicted, speech act comprehension was associated with reduced power in the alpha/beta bands just prior to and during the beginning of speech acts in highly constraining action contexts (Declinations). Based on the association of alpha and beta desynchronization with anticipatory attention, the results are taken to indicate the involvement of anticipatory processes in early speech act recognition. The absence of the alpha/beta effect to actions in less-constraining contexts (Pre-offers) relative to the control condition (Answers) indicates that anticipatory processes are dependent on the characteristics of the speech act interaction, including the sequential constraints between turns in conversation.

6 Summary and implications for future research

Spoken speech act recognition is a core cognitive ability that gives language its basic functionality. However, participants in conversation are faced with a challenge: they must recognize speech acts in utterances that are often radically underspecified for the action level of meaning (Levinson, 2013; Searle, 1975), and do so under tight time constraints to conform to the rules of turn-taking (Sacks et al., 1974). The objective of this thesis was to investigate the time-course of spoken speech act recognition in action-underspecified utterances, testing the hypothesis that verbal actions are recognized early in the utterance. Early speech act recognition could be the key to efficient conversation, enabling quick transitions between speakers. A second goal of this thesis was to explore how the sequential organization of action influences the time-course of speech act recognition. All speech acts are not created equal – they differ in their relationships to prior and upcoming turns in conversation. Going beyond "direct" versus "indirect" speech acts (e.g., Basnakova et al., 2014; H. H. Clark, 1979; Coulson & Lovett, 2010; Searle, 1975; van Ackeren et al., 2012), this thesis investigates how speech act recognition is modulated by the type of action being performed and how it fits into the larger action sequence.

The novelty of this thesis chiefly lies in two aspects. First, this work is based on an interdisciplinary approach that bridges methods from cognitive neuroscience and findings from studies on turn-taking. Such an interdisciplinary perspective raises new, important questions on speech act comprehension and provides methods to address them. Second, in contrast to the majority of EEG research on language (for overviews, see for instance Kolk & Chwilla, 2007; Kutas & Federmeier, 2000; Van Berkum, 2004), this study investigates speech act comprehension in spoken, naturalistic dialogues without any syntactic, semantic or pragmatic anomalies. This research thereby achieves a better balance between ecological validity and experimental control, bringing us closer to understanding language comprehension in its natural habitat – conversation.

In this discussion chapter I will first summarize the main empirical findings of the thesis. I will then discuss their theoretical implications and remaining questions for future research.

6.1 Summary

After an overview of the thesis research questions and methods in Chapter 1, Chapter 2 reports behavioural findings that form a foundation for the remaining empirical chapters. The main aim of Chapter 2 was to investigate how reliably participants can categorize speech acts in utterances that are underspecified for the action level of meaning. A second aim of the behavioural experiment was to obtain a rough estimate of the time-course of speech act recognition, as reflected in self-paced reading times. Chapter 2 introduces an experimental paradigm consisting of short dialogues with target sentences (e.g., *I have a credit card*) that perform three distinct speech acts – Answers, Declinations, Pre-offers – depending on the prior turn. The target sentences are identical across conditions (and hence underspecified for action), but differ in the type of speech act being performed. Declinations and Pre-offers (the critical conditions) have in common that they are more indirect than Answers (the control condition) as they require more inferencing to be understood. However, they differ in their relationship to prior and upcoming turns in conversation. Declinations occur in highly-constraining contexts, being second parts of adjacency pairs (Schegloff, 2007), and are therefore relatively anticipatable. Pre-offers, on the other hand, occur in less-constraining action contexts as they initiate a new action sequence (which frequently contains a more direct offer if the conversation were to continue (Schegloff, 1988, 2007)). Thus by comparing Declinations and Pre-offers it is possible to investigate how speech act recognition is influenced by the type of action being performed and the sequential context (the paradigm is described in more detail in Chapter 3).

In the experiment discussed in Chapter 2, the context sentences were presented auditorily, while the target sentences were presented visually in self-paced reading. After each dialogue participants were asked to categorize the action of the target sentences as doing answering, declining or offering. The results of the action categorization task demonstrate that participants can categorize the speech act in action-underspecified sentences with very high accuracy (95.8%), based on limited context (the prior speech act) and without any prosodic information in the target sentence. The reading time results indicate that speech act recognition is influenced by action type and sequential context, as reflected in different reading time patterns for Declinations and Pre-offers; reliable differences were found at the first word for Pre-offers, relative to Answers, but at the verb and the final word for Declinations. The results suggest that speech act recognition in the visual modality begins relatively early in the target sentence (at the first word or the verb), but also that late processing at the end of the sentence may be required. The experiment in Chapter 2 validates the use of Answers as a control condition in future studies, as they had the shortest reading times on all measures (and should therefore be easiest to comprehend out of the three speech act types). Overall, the results of Chapter 2 demonstrate the feasibility of using the experimental paradigm in research on speech act recognition.

What about speech act recognition in spoken language – the prime modality of everyday conversation? Given the time constraints in spoken conversation, discussed in detail in section 1.2.2 of this thesis, the real puzzle is how speech act recognition proceeds in the auditory modality. Turning to the core of this thesis, Chapter 3 makes use of the excellent temporal resolution of event-related potentials to track speech act recognition in spoken dialogues. How quickly can listeners recognize action-underspecified speech acts in the auditory modality? Keeping in mind that speech acts differ in their relationships to prior and upcoming turns in conversation, how does the sequential context influence the time-course of spoken speech act recognition? Addressing these questions, the experiment in Chapter 3 used spoken versions of the dialogues presented in Chapter 2. The same action categorization task was used to facilitate comparability with the behavioural results. The rationale of the study was that if auditory speech act recognition takes place early in the turn when the utterance has only been partially processed, then there should be ERP differences between the three speech act types early on in the target utterance (*I have a credit-card*), for instance at the first word (*I*) or the verb (*have*). Moreover, there should be no ERP differences at the utterance-final word, i.e., *credit-card*; if the action has already been recognized at that point, the processing of the final word should only add to the propositional meaning of the utterance, which for each target is the same in all three conditions and hence no differences should occur at the final word. In contrast, if speech act recognition takes place at the end of the utterance, based on complete analysis of the turn, there should be ERP differences between the conditions at the final word. Two time-windows of interest were therefore defined for the analysis of ERP data; an *early utterance time-window* roughly corresponding to the first word and the verb in the target utterances, and a *late utterance time-window* corresponding to the final word. Chapter 3 identifies both early and late speech-act-related ERP effects. A *fronto-temporal positivity* in the right hemisphere was elicited to Declinations and Pre-offers during the early time-window relative to Answers, from 200 ms after utterance onset in Pre-offers and from 400 ms in Declinations. Declinations also elicited a *frontal left/midline positivity* from 500 to 600 ms relative to Answers. Finally, Pre-offers elicited a *late, posterior negativity* at the utterance-final word, relative to both Answers (from 900 to 1000 ms) and Declinations (from 600 to 1000 ms). No ERP differences were observed between Declinations and Answers in the late utterance time-window. The early ERP effects indicate that spoken speech act recognition begins early in the turn, when the utterance has only been partially processed. The absence of ERP effects at the utterance-final word in Declinations provides further support for the early speech act recognition account, indicating that when the prior turn is highly constraining in terms of what action can follow (as is the case in the Declination dialogues) recognition of the action can be made before the final word. However, the late negativity to Pre-offers shows that early speech act recognition is not always possible. Additional processing based on the complete utterance is required in more complex actions such as Pre-offers, where the speech act is in a less-constraining context and a new action sequence is initiated. Taken together, Chapter 3 provides partial support for the early speech act recognition account and demonstrates that the time-course of action recognition is influenced by the type of speech act being performed and how it fits into the larger action sequence.

Chapter 4 investigates the robustness of the speech-act-related ERP effects reported in Chapter 3. A disadvantage of the action categorization task used in Chapters 2 and 3 is that it required giving participants labels of the three speech act types in advance (i.e., by forcing them to categorize the utterances as doing answering, offering or declining). Although participants in conversation have to make an implicit categorization of each utterance in order to prepare a fitted reply, the action categorization task directed attention to the critical speech acts and restricted possible interpretations of them. Given that ERP components have been found to be modulated by task demands (see, for instance, D. J Chwilla et al., 1995; Gunter & Friederici, 1999; Gina R. Kuperberg, 2007; Roehm et al., 2007), Chapter 4 investigates whether the ERP effects reported in Chapter 3 were induced by the categorization task or whether they generalize to a more natural situation in which overt categorization is not required. The crucial task in conversation is comprehending-for-responding. Thus simply instructing participants to listen for comprehension, without requiring any response, would not capture this important feature of conversation. To better mirror the response demands and attention level necessary in everyday interaction, Chapter 4 used a true/false judgment task, which probes understanding of the action without requiring explicit categorization of the speech acts. In addition, the ERP experiment reported in Chapter 4 included additional filler dialogues. The absence of fillers in Chapter 3 may have enabled participants to use the context turns as superficial cues to the upcoming speech acts, giving rise to the early ERP effects reported in Chapter 3. The filler dialogues in Chapter 4 reduce strategic processing of the speech acts and render them less predictable.

The experiment in Chapter 4 replicated two of the three ERP effects described in Chapter 3. As in the previous study, Declinations elicited a frontal positivity in the early utterance time-window. This effect was similar to the combined effects of the right fronto-temporal and frontal left/midline positivity in the previous experiment, but had an earlier onset than before (at 200 ms after utterance onset). Also consistent with the prior study, no ERP effects were found in Declinations relative to Answers at the final word. This pattern of results indicates that early speech act recognition in Declinations is robust, even when overt categorization is not required and filler dialogues limit the predictability of the action. As for Pre-offers no early ERP effects were observed, in contrast to the prior study. This indicates that the experimental environment exerts some influence on early speech act recognition, in particular on the right hemisphere fronto-temporal positivity to Pre-offers reported in Chapter 3. However, replicating previous results, Pre-offers elicited a late negativity from 600 ms after the onset of the utterance-final word, relative to Answers and Declinations (although the negativity had a left hemisphere preponderance, in contrast to the bilateral posterior distribution in Chapter 3). The results for Pre-offers was taken to indicate that when participants have limited top-down information about the target speech acts – due to different task demands and additional fillers – listeners do not have strong expectations about the Pre-offers and take a wait-and-see approach. The picture that emerges from Chapter 4 is that speech act recognition involves two robust ERP effects, an early frontal positivity and a late (posterior or left hemisphere) negativity, reflecting qualitatively distinct cognitive processes. In highly-constraining contexts when the action sequence is coming to a close – as is the case in Declinations – speech act recognition begins from 200 ms after utterance onset and is completed before the final, critical word is heard. In contrast, in less constraining contexts when a pre-sequence is initiated – as is the case in Pre-offers – early processing at the action level is less likely, and the full proposition is needed to understand the speech act.

Chapter 5 investigates whether the pattern of results reported in Chapter 4 is supported by converging evidence from time-frequency analyses of the EEG data in the same experiment. Since neural oscillations, as revealed by time-frequency analysis, and ERPs capture different types of brain responses (non-phase locked vs. phase-locked), exploring the oscillatory dynamics of speech act recognition can yield a more complete picture of its time-course. The more specific aim of Chapter 4 was to shed light on the role of anticipatory processes in speech act recognition. Time-frequency representations of power in the EEG data from Chapter 4 were analyzed. Speech act recognition was mainly associated with reduced power in the alpha/beta bands just prior to and during the beginning of Declinations; from -100 to 450 ms after target utterance onset relative to Answers, and from -150 to 50 ms relative to Pre-offers. Given the association between alpha/beta suppression and anticipatory processing (e.g., Bastiaansen & Brunia, 2001; Pfurtscheller & Lopes da Silva, 1999; van Ede et al., 2014), these results are taken to indicate that anticipatory attention plays a role in early recognition of the action in Declinations. The absence of oscillatory effects at the utterance-final word in Declinations converges with the ERP results in Chapters 3 and 4, providing further support for early speech act recognition in Declinations. Finally, no power differences were observed between Pre-offers and Answers, indicating that ERPs (in particular, the late negativity) are more sensitive to speech act recognition in Pre-offers than neural oscillations. Overall, the results substantiate the time-course of speech act recognition as reported in Chapters 3 and 4 and provide evidence that early recognition of the action in Declinations involves anticipatory attention to the speech act already 150 ms before the utterance begins.

6.2 Implications of main findings and future directions

In this section I will discuss the theoretical implications of the main thesis findings and directions for future research.

6.2.1 Speech acts can be recognized early in the utterance

The starting point for this investigation was the observation (discussed in Levinson, 2013) that planning a simple response to an utterance takes at least 600 to 1200 ms (E. Bates et al., 2003; Indefrey & Levelt, 2004; Levelt, 1989; Schnur et al., 2006), while the most frequent gaps between turns in conversation are much shorter – about 200 ms – and gaps of only 0 ms are common (De Ruiter et al., 2006; Heldner & Edlund, 2010; Levinson, 2013; Stivers et al., 2009). These timing facts suggest that listeners begin planning their responses before the prior speaker has finished talking (Levinson, 2013), and since the response to a turn is critically dependent on the speech act being performed, recognition of the action must be made early on in the utterance when the turn has only been partially processed. In line with this early speech act recognition account, the main finding of this thesis is that under certain circumstances listeners can get the action of an unfolding utterance in advance of its completion, before the final, critical word has been heard. More specifically, when the prior turn is highly constraining in terms of what action can follow (as in the Declination dialogues), speech act recognition begins as early as 200 ms after utterance onset and is completed by the time the final word is reached. Early speech act recognition is possible even though the utterance is underspecified for the action and does not contain clues from morphosyntax or prosody to aid recognition.

While the results of Chapter 3 and 4 indicate that early speech act recognition is not always possible – as reflected by the pattern of results for Pre-offers – it is reasonable to believe that early recognition of actions is prevalent in conversation. The sequential environment that seems to engender early speech act recognition is the adjacency pair (Schegloff, 2007); early recognition of the action was reported in Declinations, which form second parts of such pairs. Future research is needed to clarify whether early recognition of the action in Declinations generalizes to other speech acts in the same sequential position. If that is the case, early speech act recognition should be very common, given that the adjacency pair is "the central organizing format for sequences" (Schegloff, 2007, p. 4) and ties together a broad range of actions in everyday talk. Thus although some actions are recognized only at a later stage, early speech act recognition likely explains why turn transitions are on average very short despite long response planning times.

In sections 6.2.1.1 and 6.2.1.2 below I will discuss the implications of early speech act recognition for pragmatic theory and speculate on the cognitive mechanisms behind this ability.

6.2.1.1 Implications for pragmatic theory

The finding that speech acts can be recognized early in the utterance challenges classical theories of linguistic pragmatics (e.g., Grice, 1975; Searle, 1975) which assume that listeners must first process the semantic content of the complete sentence (*sentence meaning*) before getting at what the speaker really means with her utterance (*speaker meaning*). These theories do not explicitly address the time-course of speech act recognition and hence have not received much attention in this thesis. However, they entail that pragmatic inferences or so-called *implicatures* (Grice, 1975; see also Levinson, 2000; D. Sperber & Wilson, 2004) – which speech act recognition can be viewed as a type of – are drawn only at the end of the utterance. The pattern of results in Declinations clearly shows that this is not necessarily the case.

The time-course of speech act recognition has received little consideration in more recent pragmatic frameworks, although it is generally assumed that pragmatic inferencing can take place early in the utterance. In the theory of *presumptive meanings,* Levinson (2000) argues that, based on simple heuristics, default implicatures can be computed "on the fly, given fragments of semantic representation" (Levinson, 2000, p. 168). Levinson's account was specifically developed for default inferences, so-called *generalized conversational implicatures* (Grice, 1975; Levinson, 2000), while the speech act inferences in the present study – if one wants to call them such – are context-dependent and hence not default[14]. Nevertheless, the idea of heuristics guiding pragmatic processing can be extended to speech act recognition. For instance, early recognition of the action in Declinations could be regarded as a conversational implicature that arises via Levinson's M-heuristic (similar to Grice's Maxim of Manner), according

[14] Indeed, the speech act interpretation of the target utterances differs for each condition (Answer, Declination, Pre-offer), depending on the prior action.

to which “what’s said in an abnormal way, isn’t normal” (Levinson, 2000, p. 33). More specifically, the Declinations are effectively merely accounts for why the prior proposal is being rejected and thus form an indirect, marked response. Given the adjacency pair context – with an offer in the preceding turn – and the fact that the unfolding utterance does not perform a prototypical acceptance (it is “abnormal”), the utterance can quickly be understood as performing a rejection. However, such a theory of implicature has to be complemented with an account of action-sequencing, defining the sequential constraints interacting with the heuristics.

The results of this thesis pose similar problems for *relevance theory* (Dan Sperber & Wilson, 1986, 2002), another influential pragmatic framework. Relevance theorists view recognition of illocutionary force as a *higher-level explicature* – a mixture of linguistically decoded material and pragmatically inferred material which requires a higher order of metarepresentational ability than the recognition of the basic proposition (Carston, 2000; D. Sperber & Wilson, 2004). Important for the present purposes, explicatures and even implicatures (which contain only pragmatically inferred material) can be made “before the entire acoustic stimulus has been processed by the linguistic system” (Carston, 2000, p. 6); hearers interpret the utterance until the resulting interpretation meets their expectation of relevance. Under a relevance theory account, speech act recognition may take place early in the turn if the speech act is highly relevant in the context. However, this framework does not capture the fact that the sequential organization of actions is the critical factor defining the relevance of speech acts. How else can we explain that Pre-offers are recognized only at the end of the utterance? Moreover, since the proponents of relevance theory have argued that speech act classification (i.e., recognition) is only necessary in some limited cases (Dan Sperber & Wilson, 1986; see also Bird, 1994), the framework is at odds with the basic assumption made in this thesis that speech act recognition is critical for everyday conversation. Thus even though early recognition of speech acts can be worked into existing pragmatic theories, the results of this thesis show that they need to be complemented with an understanding of the constraints operating between speech acts in conversation.

6.2.1.2 How so fast?

The main goal of this thesis was to investigate the time-course of spoken speech act recognition, but the results inevitably raise the question *how* speech acts can be recognized early in the utterance. The finding that listeners can recognize speech acts in advance of their completion suggests that prediction during language comprehension is not restricted to the level of individual lexical items or their syntactic, semantic or conceptual features (e.g., Altmann & Kamide, 1999; Federmeier, 2007; Kutas et al., 2011; Thornhill & Van Petten, 2012), but takes place at the speech act level as well. As discussed in previous chapters, early speech act recognition via prediction is in line with models of cognition which propose that the brain is "predictive", "proactive", or "prospective" and anticipates upcoming events based on past experience (Bar, 2009; Schacter et al., 2007; see also Van Berkum, 2010). However, since this thesis did not systematically manipulate prediction, further research is needed to elucidate the role of prediction in the recognition of speech acts.

According to one model of predictive processing, we rapidly extract basic features from incoming stimuli and link them via analogy to familiar information in memory in order to build predictions about unfolding or upcoming input (Bar, 2009). In Chapters 3 and 4 I proposed a similar account for early speech act recognition in the Declination dialogues, namely that limited information of the incoming utterance (for instance, the first words), coupled with acquired knowledge of speech act sequences in memory, enables predictions at the speech act level. Events which occur in a non-random order and have a tight temporal relationship are an ideal breeding ground for prediction (Bubic, Von Cramon, & Schubotz, 2010). The adjacency pair is an action sequence with such characteristics and is therefore a good candidate sequence for predictions. On this account, the main ingredients for prediction at the speech act level are memory of action sequences (containing frequency-based information of how actions tend to pattern in sequences and what they commonly look like), linguistic processing, and the ability to link these through analogy.

An alternative or perhaps complementary view is that predictive processing of speech acts involves the use of heuristics – for instance of the pragmatic kind previously described (Levinson, 2000), or the fast and frugal heuristics proposed by Gigerenzer and colleagues (Gigerenzer & Goldstein, 1996; Gigerenzer & Todd, 1999). Fast and frugal heuristics are simple mechanisms of inference that “a mind can actually carry out under limited time and knowledge” (Gigerenzer & Goldstein, 1996, pp. 4–5), without a loss in inferential accuracy (Gigerenzer & Goldstein, 1996). Fast and frugal heuristics consist of searching principles, which guide the search of relevant information, stopping principles, which determine when to stop the search, and finally decision-making principles that are called upon to make a choice based on the search (Gigerenzer & Todd, 1999). The fast and frugal account places less emphasis on the contents of memory representations, giving more weight to the heuristics themselves, and hence may not be compatible with the idea that abstract and complex knowledge of speech act sequences is important for early speech act recognition. It is nevertheless reasonable to assume that simple heuristics of some kind may play a role in early speech act recognition.

A third possibility is that prediction at the speech act level is mediated by theory of mind, the role of which in speech act comprehension is far from clear (as discussed in section 6.3.3 below). However, regardless of whether speech act predictions arise via theory of mind, heuristics, analogy to prior conversational experience or some other mechanism, the different pattern of results for Declinations and Pre-offers suggests that they must be sensitive to the sequential constraints operating between speech acts in conversation. Moroever, these proposed mechanisms are not oblivious to the linguistic input. In Chapter 3 I argued that the first words of the Declinations contribute to early speech act recognition by signalling that a prototypical acceptance is not underway. Based on this information and knowledge of adjacency pairs, listeners can quickly infer that a rejection is being performed. The behavioural results for Pre-offers in Chapter 3 and 4 likewise indicate that, although speech act predictions were not involved, the linguistic form of the utterance can matter for the action interpretation (see 6.3.1 below). Thus while listeners cannot rely on speech act markers (i.e., illocutionary force indicators) in incoming turns, the view adopted in this thesis is that the linguistic format of the action does play a role in speech act recognition by triggering or interacting with other aspects of speech act processing.

The mechanisms behind early speech act recognition, and the involvement of predictive processes, is a topic for future investigation. One avenue of research would be to explore how knowledge of action sequences interacts with bottom-up cues in the utterance to yield speech act predictions. Conversation analysts have noted that noticeably long gaps or hedges (e.g., *uhm, well*) tend to precede dispreferred actions such as declinations to offers, but not preferred actions such as acceptances (Schegloff, 2007; for a discussion, see K. H. Kendrick & Torreira, 2015). One possibility would be to investigate whether long gaps or hedges trigger the prediction that a certain action is coming up in the sequence, as reflected by modulation of the ERP correlates of speech act recognition. Another avenue of investigation is to focus on the nature of verbal action knowledge and its recruitment during language comprehension. Some ideas for future research on this topic are raised in section 6.2.2 below. Research on predictive processing of speech acts could also focus on the role of theory of mind in speech act recognition, which is discussed in section 6.3.3 below.

Another important topic for further investigation is at what level of representation listeners are predicting during speech act comprehension. One issue to consider is the distinction between expectations made regarding upcoming speech acts and expectations made regarding unfolding speech acts. An implicit assumption in this thesis is that these two types of expectations may involve different levels of representation. For instance, it has been argued that the context turn in the Declination dialogues allows listeners to narrow down the speech act possibility space and form the expectation that a certain class of speech acts is appropriate, namely declinations or acceptances. On this account, precise speech act predictions are not made during the context utterance (although it is certainly possible to imagine situations in which such predictions about upcoming speech acts are made). Once the target utterance begins, listeners can use the linguistic input to recognize that a Declination is being performed even though the speech act is not yet completed, essentially "upgrading" their expectation and making a prediction at the speech act level. However, further research is required to better tease apart what levels of representation are involved during these stages of speech act comprehension. For instance, it is possible that expectations made during the context utterances in the Declination dialogues are more specific than this thesis has assumed. An important task for future studies will be to shed light on the nature of the representations involved during predictive processing of speech acts.

6.2.2 The type of action and sequential organization matters

A key finding from this thesis is that sentence comprehension is influenced by the type of action performed and how it fits into the sequential context. This calls for going beyond investigating "direct" versus "indirect" speech act comprehension (e.g., Basnakova et al., 2014; Coulson & Lovett, 2010; van Ackeren et al., 2012) towards a more fine-grained approach that takes into account speech act types and their sequential organization in turn-taking.

The structure of action sequences and their segmentation into smaller units has received considerable attention in research on non-verbal (e.g., manual) action understanding (see, for instance, Avrahami & Kareev, 1994; Baldwin, Andersson, Saffran, & Meyer, 2008; Zacks et al., 2011). It has been argued that sequential cognition – the ability to "extract and utilize the sequential structure of perceptual and motor events in the world in an adaptive and pragmatic manner" (Dominey, Hoen, Blanc, & Lelekov-Boissard, 2003, p. 208) – plays an important role in language processing at other levels than speech acts, such as syntax (Dominey et al., 2003). However, with the exception of this thesis, the sequential structure of verbal action has not been given much consideration in psycholinguistic research on language comprehension.

Abstract knowledge about action sequences can be conceptualized as scripts (Schank & Abelson, 1977) or schemata (Brewer & Nakamura, 1984; Rumelhart, 1980) in memory. Speech act scripts could be built from the observation that certain actions in conversation are usually preceded by some other type of speech act. Knowledge of verbal action sequences might play a critical role in early speech act recognition, particularly in action-underspecified utterances, decreasing the need for higher-level mechanisms such as theory of mind. This proposal opens up a new avenue of research on speech act recognition and language comprehension more generally, as many questions are unanswered with regard to how sequential context and implicit knowledge of speech act organization guides the interpretation of talk. How abstract are such action scripts – do they mainly consist of simple action sequences, such as adjacency pairs or pre-sequences, or do they contain additional script information about the situational context (e.g., restaurant scenario script)? Are speech act scripts stored and accessed in similar ways as non-verbal action knowledge? What is the time-course of such speech act script activation during listening to spoken dialogue? In what kind of situations is knowledge of action sequences not sufficient, triggering other mechanisms such as theory of mind? These questions could be addressed in future research using other methods (fMRI, MEG), new experimental paradigms and localizer tasks that tap into action sequence knowledge.

6.2.3 ERP correlates provide benchmarks for future research

The right hemisphere fronto-temporal positivity that was elicited to both Declinations and Pre-offers when speech act categorization was required (Chapter 3) did not extend to Pre-offers in an experimental environment that engaged participants in a true/false judgment task and included additional filler dialogues (Chapter 4). It is not surprising that task demands do have some impact on the brain signatures of speech act comprehension, since language-relevant ERP components like the N400 and P600 have been found to be modulated by the task (see, for instance, D. J Chwilla et al., 1995; Gunter & Friederici, 1999; Gina R. Kuperberg, 2007). In Chapter 4 I suggested that the absence of the right hemisphere fronto-temporal positivity to Pre-offers in the second ERP experiment was due to limited top-down information about the target speech acts, as a result of additional fillers and a task that did not restrict participants' interpretation of the utterances. If this is correct, the right hemisphere fronto-temporal positivity can nevertheless be regarded as a speech-act-related ERP effect, reflecting early processing of speech acts when listeners have some top-down information of upcoming input.

Most important for the present discussion, this thesis identified two reliable ERP correlates of speech act recognition, which show only subtle differences between studies in terms of latency and scalp distribution. A critical finding from Chapter 4 is that the frontal positivity early in the utterance in Declinations and the late negativity to the final word in Pre-offers are relatively robust across experimental tasks and designs. These EEG signatures can be exploited in follow-up research on speech act recognition. For instance, is speech act recognition – as reflected by these components – an automatic process or is it influenced by factors such as emotional state (see Dorothee J. Chwilla, Virgillito, & Vissers, 2011; Vissers et al., 2010)? Does verbal action understanding differ in young children or clinical populations such as people with autism? Since speech act comprehension likely draws on cognitive mechanisms used for other aspects of linguistic and social processing, the frontal positivity and late negativity are also informative for research in related domains, particularly on pragmatic language comprehension. Thus an important contribution of this thesis is that it provides ERP benchmarks for further studies on speech act recognition and sentence comprehension more generally.

6.3 Remaining questions

In this section I will discuss some remaining questions that are important for a comprehensive understanding of the time-course of speech act comprehension and action understanding more generally.

6.3.1 The influence of action formation

One finding which has not received much discussion in this thesis is that the linguistic form of the target speech act has a subtle influence on results in off-line comprehension tasks. In Chapters 2 and 3, for instance, Pre-offers of the form *I have…* (Set 1) were categorized more accurately than other formats, such as *I go/am going…* (Set 2). Similarly, in Chapter 4 the accuracy in the true-false comprehension task was higher for Pre-offers of the *I have…* format. Importantly, in all three experiments the behavioural results showed an interaction between the action condition and linguistic form (Set), indicating that it is not the case that there is some general feature of the utterance (e.g., frequency of the format within the experiment) which results in higher accuracy across the board, but rather that the form of the target speech act differentially affects the three speech act types. Although the influence of linguistic form was minimal in the EEG data (as reflected by the absence of Action × Set interactions in omnibus analyses in Chapters 3 and 4), the behavioural results are nevertheless a reminder that action formation, i.e., the linguistic form of the utterance, can matter for the action interpretation.

While Schegloff has argued that "sequential features of conversation … overshadow the contribution made by its linguistic form to what an utterance is doing" (Schegloff, 1984, p. 36), other conversation analysts have noted that participants use recurrent linguistic formats to frame actions and that action ascription is partially dependent on these formats (see, for instance, Couper-Kuhlen, 2014). Through analyses of English conversation, Couper-Kuhlen has identified the most common formats for actions that she describes as offers, proposals, requests and suggestions (Couper-Kuhlen, 2014). Although each action type has multiple formats associated with it, and in some cases a format can be used in more than one action, the frequency count indicates that each action has a "preferred" format (e.g., *I can X* is frequent in proposals) (Couper-Kuhlen, 2014). Turning again to the behavioural results for the Pre-offers, which show higher accuracy for Pre-offers starting with *I have…* ("Ik heb"; Set 1), the format *I have X, if you want Y* is a common format for concrete offers in both

English and Finnish (Kärkkäinen & Keisanen, 2012). It is likely that the same is true for Dutch and that utterances starting with *I have…* ("Ik heb") frequently perform offering actions such as offers and pre-offers. Although the *I have…* format is not a discrete speech act marker (i.e., an illocutionary force indicator – see 1.2.1), listeners may learn to associate the linguistic form of the utterance with offering actions through repeated exposure to this format, leading to higher accuracy in the comprehension tasks compared to other formats.

Importantly, the "preferred" formats discussed by Couper-Kuhlen become apparent early in turns. She argues that

> *…pace Schegloff (1984), grammar does tell us something about social action. It provides a basis on which recipients form working hypothesis about what action a co-participant is initiating. And it does this relatively early in the turn, thus enabling recipients (i) to determine an appropriate responsive action and (ii) to implement it in a timely fashion."* (Couper-Kuhlen, 2014, p. 645)

Common action formats could feed into the fast and frugal heuristics (Gigerenzer & Goldstein, 1996) discussed in section 6.2.1.2 above, allowing for early speech act recognition and response planning. While the EEG results for Pre-offers of the *I have…* format do not provide supporting evidence for this, how quickly speech acts are recognized when preferred action formats are involved is still an open question.

Moreover, in addition to the formats described above, some common speech acts in conversation involve routinized expressions, such as *gesundheit* in response to a sneeze. In these cases there is a rare one-to-one mapping between form and speech act function (Levinson, In press, 2013), so the speech act can simply be encoded as a part of the expression's lexical meaning in semantic memory. Lexicalization of speech acts "facilitates speech processing since lexical look-up is a simpler and quicker process than figuring out what the speaker means" (Aijmer, 1996, p. 196). Given that early speech act recognition is possible even when the utterance is underspecified for the action, as the results for Declinations in this thesis reveal, it is reasonable to assume that recognition of routinized speech acts, as well as speech acts with preferred formatting, is made even earlier. This could contribute further to the short turn transitions observed in conversation. However, the time-course of speech act recognition in such "non-underspecified" cases is a topic for further investigation.

6.3.2 Revisiting the timing of turn transitions

The finding that the time-course of speech act recognition is influenced by the type of action and how it fits into the larger action sequence has important implications for research on the timing of turn transitions. In particular, actions that are recognized only at the end of the utterance (such as pre-offers) should be followed by a gap that is considerably longer than the 200 ms reported average (or mode) for turn transitions (keeping in mind that response planning takes at least 600 ms). The sequential organization of action is rarely taken into account in research on turn timings. Corpus studies generally group all inter-speaker turn transitions together into one average (Bosch et al., 2005; e.g., Heldner & Edlund, 2010), and in the field of conversation analysis the type of action has only to a limited extent been considered as a factor that influences the length of gaps between turns (in the context of preferred vs. dispreferred actions; K. H. Kendrick & Torreira, 2015; Sacks et al., 1974; Schegloff, 2007).

As discussed in Chapter 4, one hypothesis that emerges from the current results is that initiating actions (including pre-offers) take more time to recognize and are therefore followed by long gaps, while responsive actions (such as declinations and other second parts of adjacency pairs) are easy to recognize and hence followed by short gaps or even overlapping responses. Further corpus research on this issue would not only clarify accounts of turn timing in conversation but also whether the time-course of speech act recognition as reported in this thesis is corroborated by analyses of gap length in natural conversation.

6.3.3 The role of theory of mind and affective empathy

Chapter 3 reported a correlation between scores on the Empathy Quotient (EQ; Baron-Cohen & Wheelwright, 2004) and the early left/midline frontal positivity observed in Declinations, which reflected that this effect was present only in high empathizers. The EQ measures both affective empathy, the ability to respond emotionally to another person's emotion, and cognitive empathy, which involves inferring the mental state of another person (Baron-Cohen & Wheelwright, 2004). The latter aspect is often referred to as theory of mind, mentalizing or perspective-taking. The empathy correlation was not replicated in Chapter 4, casting doubt on the reliability of this finding. It nevertheless brings up the question what the role of theory of mind and affective empathy is in speech act recognition. Does the frontal positivity in Declinations reflect an affective response, due to the fact that the speech act contains a *face-threatening* rejection (P. Brown & Levinson, 1988; Goffman, 1955), or the engagement of theory of mind? Prior fMRI research indicates that comprehension of indirect speech acts activates both mentalizing and affective empathy areas (Basnakova et al., 2014; van Ackeren et al., 2012), as reviewed in the introduction to this thesis. These two accounts for the frontal positivity are therefore both plausible.

In regards to the possible involvement of affective empathy, some participants in the experiments noted that the Declinations sounded somewhat rude. The rudeness presumably comes about because rejecting a proposal is a dispreferred action. However, recognizing the speech act of the utterance – i.e., that a declination is involved – is a prerequisite for the affective response. Thus even if the frontal positivity in Declinations does reflect emotional processing, a perlocutionary “by-product” of the speech act (see 1.1), it provides important information about the time-course of speech act recognition.

The second account – that the empathy correlation reflects individual variation in the engagement of theory of mind during speech act recognition – suggests that the time-course of speech act recognition is influenced by theory of mind abilities. That listeners differ in how quickly they can extract speech acts from their interlocutors’ utterances is a topic that requires further investigation, one that could be addressed with correlation analyses and MEG experimentation (providing insights into both the time-course and source of brain activity during comprehension of speech acts). The possible association of theory of mind with the frontal positivity also poses the question why this ERP effect and the correlation with empathy were not observed in Pre-offers. Is theory of mind not required for the comprehension of Pre-offers? This taps into a bigger issue of what kind of speech act situation triggers the need for mentalizing during conversation.

The fMRI and MEG research that has reported theory of mind activation during speech act comprehension (see 1.3) used either indirect speech acts (Basnakova et al., 2014; van Ackeren et al., 2012) or visual stimuli that are quite removed from natural conversation (Egorova, Pulvermüller, et al., 2013). It is not clear to what extent mentalizing is involved in other types of speech acts and situations. There is considerable debate within linguistic pragmatics and other fields of research whether intention recognition via theory of mind is necessary for interaction. In contrast to Gricean-inspired pragmatic theories in which reasoning about the mental states of others plays a central role (e.g., Levinson, 1983; Dan Sperber & Wilson, 2002), it has been proposed that intention recognition is not critical for human communication (see, for instance, Gauker, 2001; Gregoromichelaki et al., 2011; M. J. Pickering & Garrod, 2004). A parallel discussion is found within research on non-verbal action understanding, in which the role of theory of mind vis-à-vis the mirror neuron system in action recognition is debated. For instance, based on fMRI studies on manual action comprehension it has been argued that mentalizing regions *can* analyze other people's actions, but only if reflecting on goals, intentions and beliefs is important for the task (de Lange et al., 2008; Thioux, Gazzola, & Keysers, 2008). Whether this is the case for speech act recognition, i.e., that theory of mind activation is to some degree "optional", is an important question for further research. Information about familiar speech acts in semantic memory, top-down knowledge of action sequences, fast and frugal heuristics and general inference mechanisms that do not require theory of mind (see, for instance, Mason & Just, 2011) could obviate the need for mentalizing during speech act recognition. To get a complete picture of the cognitive mechanisms behind fast and efficient speech act recognition it is important to expand the investigation to a variety of actions and formats and, in addition to EEG, use methods such as fMRI and MEG that provide information about the neural substrates involved.

6.3.4 From overhearing to interacting

The current investigation employed an overhearing paradigm in which participants listen to short, pre-designed spoken dialogues. As discussed in the introduction to this thesis, the use of controlled materials reduces noise from confounding factors. Since the dialogues were modeled on natural conversation (based on findings from conversation analysis), the materials nevertheless approximate conversational interaction. This approach therefore offers a good compromise between experimental control and ecological validity. However, overhearing a dialogue is not the same as taking an active part in one. Several studies have shown that understanding differs for addressees and overhearers. For instance, when people are asked to tell another participant how to arrange complex figures in a tangram matching task, those who overhear the matching game are less accurate and slower in arranging the figures than addressees in an interactive situation (Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). This difference is attributed to a lack of *grounding,* that is, overhearers cannot initiate repair in the conversation or use other collaborative means to reach mutual belief (Schober & Clark, 1989; see also Herbert H. Clark & Brennan, 1991). Eye-tracking studies have shown that in interactive dialogue an addressee's initial interpretation of another's utterance is sensitive to the addressees's experience with the speaker, while such partner-specific effects are not found in non-interactive settings (Brown-Schmidt, 2009). Thus listeners may be less likely to engage in sophisticated understanding of speech acts when overhearing dialogue. Although this thesis used comprehension tasks to counter the possible lack of perspective-taking and engagement, participants may have processed the dialogues more superficially or with less speed than in real life, leading to attenuation or delay in ERP components.

Another related issue is how first-hand knowledge and intentions influence speech act recognition. If the experimental dialogues were interactive, the addressees of the target utterances would be the speakers of the context turns. In the case of the Declination dialogues, this entails that the person performing the proposal in the context is already expecting an acceptance or rejection, and does not have to infer this from the adjacency pair structure (i.e., based on her own proposal). More importantly, in the Pre-offer dialogues the speaker of the context turn, which contains an expression of some problem, may already have the intention to elicit an offer of assistance; in this situation the Pre-offer should be readily apparent and possibly recognizable early in the utterance (especially in turns starting with *I have*, given the association of this phrase with offering; see 6.3.1). Thus for both Declinations and Pre-offers it is reasonable to assume that the recognition of the speech act would take place earlier in interactive settings than in non-interactive settings such as the experiments for this thesis. In light of this, the early speech act recognition effect to Declinations, starting from 200 ms after utterance onset, is striking. Moreover, the lack of interactivity could contribute to late recognition of Pre-offers.

Is there a way to investigate the time-course of speech act recognition in a more interactive setting? One possibility for further research is to have the participant interact with a confederate in a situation that is easy to manipulate, for instance in some type of a communication task or a quiz paradigm (Bögels, Magyari, & Levinson, 2014; Van Berkum, 2012). This involves using well controlled, scripted and preferably pre-recorded utterances, which inevitably limits the interaction and is challenging in terms of experimental design (for problems with using confederates in experimental research on dialogue, see Kuhlen & Brennan, 2013). Future studies should aim to address these issues, pushing the limits in research on speech act recognition. However, important strides can be made using an overhearing paradigm, for instance by investigating the automaticity of speech act recognition and exploring individual differences in this ability in healthy people, as well as in children or clinical populations (as discussed in 6.2.3 above).

6.3.5 Relationship to non-verbal action understanding

The previous discussion on the role of theory of mind, prediction and knowledge of action sequences in speech act recognition resonates with similar themes in research on non-verbal action understanding. An important question for future investigation concerns the relationship between the neural substrates of verbal and non-verbal action recognition. Ultimately it may be the case that action recognition in both modalities involves the same systems. Verbal and non-verbal actions are tighly integrated in conversation, often occurring in tandem – as in the case of co-speech gestures (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Willems, Özyürek, & Hagoort, 2007) – or in exchange for each other.

Requesting something can be done with a manual gesture, for example reaching for a glass, or a verbal request. Similarly, raising an eyebrow or leaning forward can function in the same way as “*huh*” or “*what did you say*” in the initiation of repair in conversation (see, for instance, Enfield et al., 2013). There is no reason to believe that the machinery that makes speech act recognition possible is much different from the mechanisms used in the visual modality. Further research on the neural substrate of speech act recognition would provide a more unified understanding of efficient action recognition in both modalities and shed new light on the interaction engine (Levinson, 2006) behind human communication.

6.4 Conclusions

This thesis illustrates how an interdisciplinary approach, combining EEG methodology with findings from conversation analysis and corpus studies, can provide novel insights into the basic cognitive ability underlying conversational interaction – the recognition of speech acts. The findings demonstrate that the time-course of speech act recognition is dependent on the type of action being performed and how it fits into the larger action sequence. Given that the prior turn is highly constraining in terms of what action can follow, speech act recognition begins very early in the incoming utterance and is completed by the time the final word is reached. This finding implies that predictions are made at the level of speech acts, opening up a new arena of research on predictive language comprehension. At a more general level, the results of this thesis show that sentence comprehension is influenced by the give-and-take of the speech act interaction, calling for a more nuanced approach in investigating language comprehension in conversational contexts. Future studies have to take into account the temporal demands and rich structure of action that characterize turn-taking.

7 References

Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. London: Longman.

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

Austin, J. L. (1976). *How to do things with words*. Oxford: Oxford University Press.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition*, *53*(3), 239–261.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Baldwin, D. A., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407.

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Basnakova, J., Weber, K., Petersson, K. M., Van Berkum, J. J. A., & Hagoort, P. (2014). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578.

Bastiaansen, M., Böcker, K. B., Brunia, C. H., de Munck, J. C., & Spekreijse, H. (2001). Event-related desynchronization during anticipatory attention for an upcoming stimulus: A comparative EEG/MEG study. *Clinical Neurophysiology*, *112*(2), 393–403.

Bastiaansen, M., & Brunia, C. H. (2001). Anticipatory attention: An event-related desynchronization approach. *International Journal of Psychophysiology*, *43*(1), 91–107.

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2008). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 31–50). New York: Oxford University Press.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. Retrieved from http://www.cran.r-project.org/web/packages/lme4/index.html

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380.

Beeman, M. J., Bowden, E. M., & Gernsbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. *Brain and Language*, *71*(2), 310–336.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. *European Archives of Psychiatry and Clinical Neuroscience, 87*(1), 527–570.

Bird, G. (1994). Relevance theory and speech acts. In S. Tsohatzidis (Ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives* (pp. 292–311). London: Routledge.

Bögels, S., Magyari, L., & Levinson, S. C. (2014, June). *Cognitive processes involved in turn-taking within Dutch question-answer sequences: An EEG-study*. Presented at the 4th International Conference on Conversation Analysis, Los Angeles, California.

Bosch, L. Ten, Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, *47*(1), 80–86.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.

Brown, C. M., & Hagoort, P. (2000). On the electrophysiology of language comprehension: Implications for the human language system. In M. W. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 213–237). Cambridge, UK: Cambridge University Press.

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language*, *72*(2), 158–190.

Brown, P., & Levinson, S. C. (1988). *Politeness*. Cambridge: Cambridge University Press.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134.

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*(25), 1–15.

Carston, R. (2000). Explicature and semantics. *UCL Working Papers in Linguistics*, *12*, 1–44.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*(3), 274–285.

Chwilla, D. J., & Kolk, H. H. (2002). Three-step priming in lexical decision. *Memory & Cognition*, *30*(2), 217–225.

Chwilla, D. J., Virgillito, D., & Vissers, C. T. W. (2011). The relationship of language and emotion: N400 support for an embodied view of language comprehension. *Journal of Cognitive Neuroscience*, *23*(9), 2400–2414.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology*, *11*, 430–477.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Coulson, S. (2004). Electrophysiology and pragmatic language comprehension. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 187–206). Basingstoke, Hampshire: Palgrave Macmillan.

Coulson, S., & Lovett, C. (2010). Comprehension of nonconventional indirect requests: An event-related brain potential study. *Italian Journal of Linguistics*, *22*(1), 107–124.

Coulson, S., & Wu, Y. C. (2005). Right hemisphere activation of joke-related information: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*(3), 494–506.

Couper-Kuhlen, E. (2014). What does grammar tell us about action? *Pragmatics*, *24*(3), 623–647.

De Blasio, F. M., & Barry, R. J. (2013). Prestimulus alpha and beta determinants of ERP responses in the Go/NoGo task. *International Journal of Psychophysiology*, *89*(1), 9–17.

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology*, *18*(6), 454–457.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535.

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, *30*(11), 3509–3526.

Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. (2012). Less is not more: Neural responses to missing and superfluous accents in context. *Journal of Cognitive Neuroscience*, *24*(12), 2400–2418.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194.

Dominey, P. F., Hoen, M., Blanc, J.-M., & Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and ERP studies. *Brain and Language*, *86*(2), 207–225.

Dwivedi, V. D., Drury, J. E., Molnar, M., Phillips, N. A., Baum, S., & Steinhauer, K. (2010). ERPs reveal sensitivity to hypothetical contexts in spoken discourse. *NeuroReport*, *21*(11), 791–795.

Dwivedi, V. D., Phillips, N. A., Laguë-Beauvais, M., & Baum, S. R. (2006). An electrophysiological study of mood, modal context, and anaphora. *Brain Research*, *1117*(1), 135–153.

Egorova, N., Pulvermüller, F., & Shtyrov, Y. (2013). Neural dynamics of speech act comprehension: An MEG study of naming and requesting. *Brain Topography*, *27*(3), 375–392.

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2013). Early and parallel processing of pragmatic and semantic information in speech acts: Neurophysiological evidence. *Frontiers in Human Neuroscience*, *7*(86), 1–13.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., … Torreira, F. (2013). Huh? What? - A first survey in 21 languages. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 343–380). New York: Cambridge University Press.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Federmeier, K. D., Wlotko, E. W., & Meyer, A. M. (2008). What's "right" in language comprehension: Event-related potentials reveal right hemisphere language capabilities. *Language and Linguistics Compass*, *2*(1), 1–17.

Foldi, N. S. (1987). Appreciation of pragmatic interpretations of indirect commands: Comparison of right and left hemisphere brain-damaged patients. *Brain and Language*, *31*(1), 88–108.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, *15*(17), R644–R645.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Gauker, C. (2001). Situated inference versus conversational implicature. *Nous*, *35*(2), 163–189.

Geluykens, R. (1988). On the myth of rising intonation in polar questions. *Journal of Pragmatics*, *12*(4), 467–485.

Gibbs, R. W. (1979). Contextual effects in understanding indirect requests. *Discourse Processes*, *2*(1), 1–10.

Gibbs, R. W. (1981). Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 431–444.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* Oxford University Press.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLOS ONE*, *10*(3), e0120068.

Gisladottir, R. S., Chwilla, D. J., Schriefers, H., & Levinson, S. C. (2012). Speech act recognition in conversation: Experimental evidence. In N. Miyake & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 1596–1601). Austin, Texas.

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry: Interpersonal and Biological Processes*, *18*(3), 213–231.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face interaction.* Oxford, England: Aldine.

Gordon, D., & Lakoff, G. (1971). Conversational postulates. In D. Adams (Ed.), *Papers from the Seventh Regional Meeting* (pp. 63–83). Chicago: Chicago Linguistic Society.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395.

Graesser, A. C., Swamer, S. S., & Baggett, W. B. (1996). New models of deep comprehension. In B. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 1–32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grainger, J., Lopez, D., Eddy, M., Dufau, S., & Holcomb, P. J. (2012). How word frequency modulates masked repetition priming: An ERP investigation. *Psychophysiology*, *49*(5), 604–616.

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G. J., Cann, R., Meyer-Viol, W., & Healey, P. G. T. (2011). Incrementality and intention-recognition in utterance processing. *Dialogue & Discourse*, *2*(1), 199–233.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, *36*(01), 126–137.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1055–1069.

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In P. Hagoort & C. M. Brown (Eds.), *The neurocognition of language* (pp. 273–316). Oxford: Oxford University Press.

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

Hald, L. A., Bastiaansen, M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90–105.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–1103.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836.

Heritage, J. (2001). Goffman, Garfinkel and conversation analysis. In M. Wetherell, S. Taylor, & S. Yates (Eds.), *Discourse theory and practice: A reader* (pp. 47–56). London: Sage.

Heritage, J. (2012). Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction*, *45*(1), 1–29.

Hermes, D., Miller, K. J., Vansteensel, M. J., Edwards, E., Ferrier, C. H., Bleichner, M. G., … Ramsey, N. F. (2014). Cortical theta wanes for language. *NeuroImage*, *85*, 738–748.

Holtgraves, T. (2008a). Automatic intention recognition in conversation processing. *Journal of Memory and Language*, *58*(3), 627–645.

Holtgraves, T. (2008b). Conversation, speech acts, and memory. *Memory & Cognition*, *36*(2), 361–374.

Holtgraves, T. (2012). The role of the right hemisphere in speech act comprehension. *Brain and Language*, *121*(1), 58–64.

Hu, L., Xiao, P., Zhang, Z. G., Mouraux, A., & Iannetti, G. D. (2014). Single-trial time–frequency analysis of electrocortical signals: Baseline correction and beyond. *NeuroImage*, *84*, 876–887.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–144.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, *15*(8), 1395–1399.

Jones, S. R., Kerr, C. E., Wan, Q., Pritchett, D. L., Hämäläinen, M., & Moore, C. I. (2010). Cued spatial attention drives functionally relevant modulation of the mu rhythm in primary somatosensory cortex. *The Journal of Neuroscience*, *30*(41), 13760–13765.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228.

Kärkkäinen, E., & Keisanen, T. (2012). Linguistic and embodied formats for making (concrete) offers. *Discourse Studies*, *14*(5), 587–611.

Kendrick, K., Brown, P., Dingemanse, M., Floyd, S., Gipper, S., Hayano, K., … Levinson, S. C. (2014, June). *Sequence organization: A universal infrastructure for action.* Presented at the the 4th International Conference on Conversation Analysis, Los Angeles, California.

Kendrick, K. H. (2012, November). *Particles at the transition space*. Presented at the Interactional Foundations of Language Meeting, Schloss Ringberg, Kreuth, Germany.

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 255–289.

King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376–395.

Kissine, M. (2013). *From utterances to speech acts*. Cambridge: Cambridge University Press.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2), 169–195.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*(2), 196–214.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, *100*, 257–261.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572.

Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology*, *34*(4), 383–398.

Kutas, M., & Dale, A. (1997). Electrical and magnetic readings of mental functions. *Cognitive Neuroscience*, 169–242.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford: Oxford University Press.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

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

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

Kutas, M., van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd edition, pp. 659–724). New York: Elsevier.

Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses: Evidence from ERP data. *Language*, *89*(3), 537–585.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259–284.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*(05), 911–920.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–38.

Levinson, S. C. (In press). Speech acts. In Y. Huang (Ed.), *Oxford handbook of pragmatics*. Oxford: Oxford University Press.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Levinson, S. C. (2006). On the human “interaction engine”: Culture, cognition and interaction. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality* (pp. 39–69). Oxford: Berg Publishers.

Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis* (pp. 103–130). Malden, MA: Wiley-Blackwell.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*(731). http://doi.org/doi:10.3389/fpsyg.2015.00731

Luck, S. J. (2005a). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Luck, S. J. (2005b). Ten simple rules for designing ERP experiments. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 17–32). Cambridge, MA: MIT Press.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia, 48*(14), 3982–3992.

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes, 16*(1-2), 35–53.

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of “pop-out” words in spoken French dialogues. *Journal of Cognitive Neuroscience, 17*(5), 740–756.

Magyari, L., Bastiaansen, M., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*(11), 2530–2539.

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

Mason, R. A., & Just, M. A. (2011). Differentiable cortical networks for inferences concerning people’s intentions versus physical causality. *Human Brain Mapping, 32*(2), 313–329.

Meinhardt, J., Kühn-Popp, N., Sommer, M., & Sodian, B. (2012). Distinct neural correlates underlying pretense and false belief reasoning: Evidence from ERPs. *NeuroImage, 63*(2), 623–31.

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience, 21*(12), 2358–2368.

Min, B. K., & Park, H. J. (2010). Task-related modulation of anterior theta and posterior alpha EEG reflects top-down preparation. *BMC Neuroscience, 11*(1), 79.

Motsch, W. (1980). Situational context and illocutionary force. In J. R. Searle, F. Kiefer, & M. Bierwisch (Eds.), *Speech act theory and pragmatics* (pp. 155–168). Dordrecht: D. Reidel Publishing Company.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71–73.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210.

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze? *Psychological Science*, *19*(3), 211–215.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464–496.

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, *110*(11), 1842–1857.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–189.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347.

Purver, M., Gregoromichelaki, E., Meyer-Viol, W., & Cann, R. (2010). Splitting the “I”s and crossing the “you”s: Context, speech acts and grammar. In P. Lupkowski & M. Purver (Eds.), *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th workshop on the semantics and pragmatics of dialogue* (pp. 43–50). Poznan, Poland: Polish Society for Cognitive Science.

Rajagovindan, R., & Ding, M. (2011). From prestimulus alpha oscillation to visual-evoked response: an inverted-U function and its attentional modulation. *Journal of Cognitive Neuroscience*, *23*(6), 1379–1394.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn’t it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259–1274.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *Neuroreport*, *15*(3), 409–412.

Rugg, M. D., & Coles, M. G. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 27–39). Oxford: Oxford University Press.

Rugg, M. D., Schloerscheidt, A. M., & Mark, R. E. (1998). An electrophysiological comparison of two indices of recollection. *Journal of Memory and Language*, *39*(1), 47–69.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. New York: Wiley.

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: An event-related potential study. *Psychological Science*, *11*(1), 46–50.

Sacks, H. (1992). *Lectures on Conversation, Volumes I and II.* (G. Jefferson, Ed.). Oxford: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Sadock, J. M., & Zwicky, A. M. (1985). Speech act distinctions in syntax. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 1, pp. 155–196). Cambridge: Cambridge University Press.

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., … Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, *22*(11), 2917–2926.

Sbisà, M., & Turner, K. (Eds.). (2013). *Pragmatics of speech actions*. Berlin and Boston: De Gruyter Mouton.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Oxford, England: Lawrence Erlbaum.

Schegloff, E. A. (1984). On some questions and ambiguities in conversation. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 28–52). Cambridge: Cambridge University Press.

Schegloff, E. A. (1988). Presequences and indirection: Applying speech act theory to ordinary conversation. *Journal of Pragmatics*, *12*(1), 55–62.

Schegloff, E. A. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, *102*(1), 161–216.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, *8*(4), 289–327.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, *35*(2), 189–213.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, 3: Speech Acts* (pp. 59–82). New York: Academic Press.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, *5*(01), 1–23.

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge, England: Cambridge University Press.

Sidnell, J. (2010). *Conversation analysis: An introduction*. West Sussex, UK: Wiley-Blackwell.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Oxford University Press.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1-2), 3–23.

Sperber, D., & Wilson, D. (2004). Relevance theory. In G. Ward & L. Horn (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford: Blackwell.

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLOS ONE*, *8*(6), e66839.

Steinhauer, K., & Drury, J. E. (2011). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*, 135–162.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587.

Thioux, M., Gazzola, V., & Keysers, C. (2008). Action understanding: How, what and why. *Current Biology*, *18*(10), R431–R434.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392.

Time-frequency analysis using Hanning window, multitapers and wavelets. (2014). Retrieved July 15, 2014, from http://fieldtrip.fcdonders.nl/tutorial/timefrequencyanalysis

Traum, D. R. (1999). Speech acts for dialogue agents. In M. Wooldridge & A. Rao (Eds.), *Foundations of rational agency* (pp. 169–201). Dordrecht: Kluwer.

van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rüschemeyer, S. A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, *24*(11), 2237–2247.

Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras, Jr. & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.

Van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad – Any implications for neuropragmatics? *Italian Journal of Linguistics*, *22*(1), 181–208.

Van Berkum, J. J. A. (2012). The electrophysiology of discourse and conversation. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 589–614). New York: Cambridge University Press.

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

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

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science, 20*(9), 1092–1099.

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

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

Van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82.

van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M., Tesink, C. M., Kos, M., Buitelaar, J. K., & Hagoort, P. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social, Cognitive and Affective Neuroscience*, *7*(2), 173–183.

van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha- and beta-band oscillations. *The Journal of Neuroscience*, *31*(6), 2016–2024.

van Ede, F., Szebényi, S., & Maris, E. (2014). Attentional modulations of somatosensory alpha, beta and gamma oscillations dissociate between anticipation and stimulus processing. *NeuroImage*, *97*, 134–141.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*(4), 380–393.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

Vissers, C. T. W., Virgillito, D., Fitzgerald, D. A., Speckens, A. E., Tendolkar, I., van Oostrom, I., & Chwilla, D. J. (2010). The influence of mood on the processing of syntactic anomalies: Evidence from P600. *Neuropsychologia*, *48*(12), 3521–3531.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594.

Vuilleumier, P., & Huang, Y.-M. (2009). Emotional attention uncovering the mechanisms of affective biases in perception. *Current Directions in Psychological Science*, *18*(3), 148–152.

Walker, T., Drew, P., & Local, J. (2011). Responding indirectly. *Journal of Pragmatics*, *43*(9), 2434–2451.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, *61*(4), 747–776.

Weiss, S., & Mueller, H. M. (2012). "Too many betas do not spoil the broth": The role of beta brain oscillations in language processing. *Frontiers in Psychology*, *3*(201), 1–15.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*(10), 2322–2333.

Wittgenstein, L. (1953). *Philosophical investigations*. (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066.

8 Appendix

8.1 Model comparison for accuracy analysis in Chapter 2

We started by comparing the fit of the full model with Action, Set and the Action $\times$ Set interaction (AIC 1570.2, logLIK -763.1) to a model without the interaction (AIC 1585.3, logLik -772.7); including the interaction significantly improved the fit of the model (loglikelihood difference = 9.56; $\chi 2(2) = 19.11$, $p < 0.001$). We then compared a model with Action and Set as fixed effects (AIC 1585.3, logLik -772.7) to a model without Action (AIC 1582.7, logLik -773.33); this comparison was not significant (loglikelihood difference = 0.67; $\chi 2(2) = 1.34$, $p = 0.51$), indicating that including Action does not improve the model. A comparison of a model with Action and Set as fixed factors (AIC 1585.3, logLik -772.7) to a model with only Action (AIC 1584.1, logLik -773.0) revealed that adding a fixed effect of Set did not significantly improve the model (loglikelihood difference = 0.38; $\chi 2(1) = 0.77$, $p = 0.38$). The best fitting model therefore includes only an Action $\times$ Set interaction, but the more conservative, full model with Action and Set was used for the analysis in Section 2.4.1.

8.2 Empathy Quotient correlation analysis for Chapter 3

Prior fMRI and MEG studies have implicated theory of mind (or mentalizing) and affective empathy in indirect speech act comprehension. For instance, a recent fMRI study found that indirect replies in spoken dialogues (e.g., *Did you like my presentation? – It's hard to give a good presentation*) activate not only typical language regions but also areas for mentalizing and affective empathy (Basnakova et al., 2014). Similarly, an fMRI study on indirect requests (e.g., *It is very hot here* presented with a picture of a closed window) reported activations in theory of mind regions, in addition to activations in areas associated with action planning and motor control (van Ackeren et al., 2012). Theory of mind activations were also reported in an MEG study on speech act processing, in which written words performed either a requesting or a naming speech act depending on a prior video-taped context sentence (e.g., *What are these called?/What can I get you? – PLANT*) (Egorova, Pulvermüller, et al., 2013). Based on this research, we hypothesized that there might be individual differences in the timing or nature of speech act recognition due to possible variation in perspective-taking and social cognition more generally. Such variation can be assessed with the Empathy Quotient (EQ) (Baron-Cohen & Wheelwright, 2004; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). The EQ measures both theory of mind (cognitive empathy; the ability to infer the mental state of another person) and affective empathy (the ability to respond emotionally to another person's emotion) (Baron-Cohen & Wheelwright, 2004). Theory of mind and affective empathy, as measured with the EQ, have previously been found to influence pragmatic language comprehension in ERP research (van den Brink et al., 2012).

To investigate whether theory of mind or affective empathy plays a role in speech act recognition we planned to compute correlations between scores on the Empathy Quotient and the ERP data. More specifically, the goal of the correlation analysis was twofold: to examine whether the observed ERP effects are modulated by empathy, and to check whether individual differences mask other ERP effects that might be present (such as an N400).

8.2.1 Empathy Quotient scores and correlation analysis

All participants filled out the Empathy Quotient questionnaire (EQ) after the EEG recording. EQ scores ranged from 20 to 62 (M = 40.79, SD = 8.46) (highest possible score 80, higher scores indicate more empathy skills). The correlation analysis was performed on mean amplitudes in the early utterance time-window from 100 to 600 ms after first word onset, and in the late utterance time-window from 100 to 600 ms and 600 to 1000 ms after final word onset. The analysis for the early window included the right anterior region and the anterior medial region (to capture modulation of frontal ERP effects), as well as the posterior medial region (to check for effects at posterior sites). The analysis for the final word included the same frontal regions and a posterior region in the left hemisphere (to capture the late negativity). We first averaged the mean amplitude over all sites in each region for each condition. ERP difference scores were then computed for Declinations and Pre-offers by subtracting the mean amplitude for Answers (Declinations - Answers and Pre-offers - Answers, respectively). Two-tailed Pearson's correlations were performed using Bonferroni-adjusted alpha levels, $p = .017$ for the early utterance time-window and p = .008 for the two time-windows at the final word.

EQ scores were positively correlated with the Declination difference score at the anterior medial region in the early utterance time-window, i.e., from 100 to 600 ms after first word onset ($r(40) = .43$, $p = .005$); the more empathy participants have, the bigger the amplitude difference between Answers and Declinations. A scatter plot showing the correlation is provided in Figure 8.1. Participants were split into two groups based on EQ scores, resulting in a Low EQ group (20 participants, EQ 20 to 40) and a High EQ group (22 participants, EQ 41 to 62). Visual inspection of both groups' mean waveforms for Answers and Declinations at the first word (see Figure 8.1) suggests that Declinations separate earlier from Answers in participants with high empathy than in low empathizers. Separate analyses (ANOVAs) for each group in the early utterance time-window revealed that the frontal positivity at the medial region is present from 400 to 600 ms in the High EQ group, indicated by a main effect of Action at the frontal medial region ($Fs \geq 7.06$, $ps < .05$), while no differences were observed in the Low EQ group.

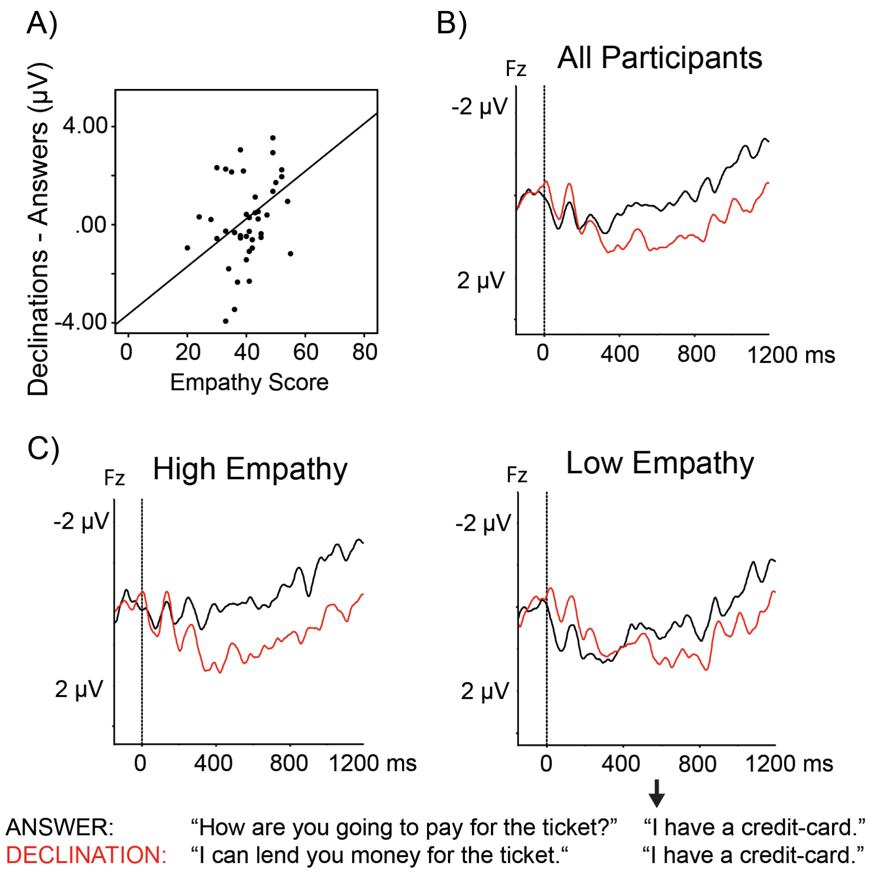

Figure 8.1 Empathy Correlation. A) Scatter plot showing correlation between EQ and Declinations difference score in the early utterance time-window at the frontal medial region. B) ERPs at Fz time-locked to the onset of the first word; for all participants, high empathy participants, and low empathy participants.

The ERP difference scores in Pre-offers were not significantly correlated with EQ at the first word (*ps* $>$ .10). At the final word there were no significant correlations between EQ and ERP difference scores using the corrected alpha level (*ps* $>$ .016; the corrected alpha level was .008 in the late time-windows).

8.2.2 Discussion

A positive correlation was observed between EQ and the Declination difference score at the frontal medial region from 100 to 600 ms after first word onset, indicating that the more empathy participants have, the larger the mean amplitude difference between Answers and Declinations. The follow-up analyses indicated that the frontal medial positivity (referred to as frontal left/midline positivity in Chapter 3) is present from 400 to 600 ms after first word onset in high empathizers only. The correlation suggests that the frontal medial positivity is related to theory of mind and/or affective empathy. Unfortunately it is difficult to disentangle the cognitive and emotional components of empathy as they tend to co-occur (Baron-Cohen & Wheelwright, 2004). Prior fMRI research indicates that comprehension of indirect speech acts activates both mentalizing and affective empathy areas (Basnakova et al., 2014; van Ackeren et al., 2012), as previously discussed.

In Chapter 3 we speculate that the correlation with the Empathy Quotient in Declinations reflects individual differences in theory of mind processing. Declinations may be more taxing on theory of mind than Pre-offers and Answers due to a combination of predictability and indirectness. In contrast to Pre-offers, the context turn in the Declination dialogues builds up strict expectations about the upcoming action (an acceptance or declining of an offer). In contrast to Answers, the Declination target utterances are relatively indirect (providing just a reason for not needing the offer) and therefore more difficult. These factors may trigger theory of mind processing in Declinations, reflected by the frontal medial positivity; listeners quickly "tune in" to the early signals of the action (expecting either declination or acceptance) and yet have to engage in additional processing because of the indirect response. On this account the time-course of speech act recognition is influenced by theory of mind abilities. However, further research is clearly required to investigate the influence of cognitive and emotional empathy on the time-course of speech act recognition.

8.3 Model comparison for accuracy analysis in Chapter 3

We first compared the fit of the full model with Action, Set and the Action x Set interaction (AIC 1394.5, logLIK -675.26) to a model without the interaction (AIC 1397.2, logLik -678.58); including the interaction significantly improved the fit of the model (loglikelihood difference = 3.32; $\chi2(2) = 6.64$, $p < 0.05$). A comparison of a model with Action and Set as fixed effects (AIC 1397.2, logLik -678.58) to a model without Action (AIC 1405.9, logLik -684.97) was significant (loglikelihood difference = 6.39; $\chi2(2) = 12.76$, $p < 0.01$), suggesting that the best fitting model includes Action as a fixed effect. A comparison of a model with Action and Set as fixed factors (AIC 1397.2, logLik -678.58) to a model with only Action (AIC 1396.5, logLik -679.27) revealed that adding a fixed effect of Set did not significantly improve the model (loglikelihood difference = 0.69; $\chi2(1) = 1.38$, $p = 0.24$). The best model therefore includes Action and the Action × Set interaction, but the more conservative, full model with Action, Set and the Action × Set interaction was used for the analysis in section 3.3.1.

8.4 Empathy Quotient correlation analysis for Chapter 4

Chapter 3 (see 8.2 in Appendix) reported a positive correlation between scores on the Empathy Quotient (Baron-Cohen & Wheelwright, 2004) and the early frontal positivity in Declinations; more specifically, the frontal positivity was present in high empathizers from 400 to 600 ms after first word onset while it was absent in low empathizers. The Empathy Quotient measures both cognitive empathy, i.e., theory of mind, as well as affective empathy (Baron-Cohen & Wheelwright, 2004). Both components of empathy have been implicated in the comprehension of indirect speech acts, as investigated by fMRI studies on indirect requests and replies (Basnakova et al., 2013; van Ackeren et al., 2012). To investigate whether individual differences in empathy modulate the ERP correlates of speech act recognition in the current experiment, we carried out correlation analyses on empathy scores from all participants who were included in the main EEG analysis. Empathy scores ranged from 26 to 63 (M = 44.93, SD = 8.80; higher scores indicate more empathy skills). The correlation analysis was performed on averaged mean amplitude in three regions of interest to capture the regions used in the prior study. A Mid-Anterior region included anterior regions at and around the midline (A3, M1 and A4; to capture modulation of the frontal positivity around the midline). A Right-Anterior region included two anterior regions in the right hemisphere (A5, A6; to capture the right hemisphere frontal positivity). Finally, a Posterior region included two left posterior regions (P2 and P3; to check for effects at posterior sites such as the late negativity or an N400). We first averaged the mean amplitude over all sites in each condition for each region. ERP difference scores were then computed for Declinations and Pre-offers by subtracting the mean amplitude for Answers (that is, Declinations - Answers and Pre-offers - Answers). The correlation analysis was performed on mean amplitudes in the early utterance time-window from 100 to 600 ms and in the late utterance time-window from 100 to 600 ms and 600 to 1000 ms. Two-tailed Pearson's correlations were performed using Bonferroni-adjusted alpha levels, p = .017 for the first time-window and p = .008 for the final word.

There were no correlations between the EQ and the difference score in Declinations and Pre-offers in any of the three time-windows (early time-window *ps* $>$ .16; first late time-window (100 to 600 ms) *ps* $>$ .014; second late time-window (600 to 1000 ms) *ps* $>$ .32).

8.4.1 Discussion

The present experiment did not replicate the correlation between Empathy Quotient scores and the frontal positivity in Declinations, reported in Chapter 3. It seems that high empathizers have an advantage when comprehending Declinations in a categorization task (as in Chapter 3), but this advantage disappears under the experimental conditions of the present experiment, as reflected by the finding that the empathy correlation was not replicated. It is likely that these two different outcomes have more to do with the amount of top-down information participants had about the stimuli than the task itself. As discussed in section 4.4.2, participants in the present experiment had less information about the target speech acts before the session started. Moreover, additional fillers directed their attention further away from the critical utterances, limiting top-down expectations. Whether and how empathy interacts with the use of top-down information during speech act recognition is a topic for further investigation.

8.5 Model comparison for accuracy analysis in Chapter 4

We first compared the fit of the full model with Action, Set and the Action $\times$ Set interaction (AIC 2009.91, logLIK -982.96) to a model without the interaction (AIC 2029.90, logLik -994.95). Including the interaction improved the fit of the model (loglikelihood difference = 11.99; $\chi 2(2) = 6.175e\text{-}06$, $p < 0.001$). A comparison of a model with Action and Set as fixed effects (AIC 2029.90, logLik -994.95) to a model with Set only (AIC 2046.59, logLik -1005.29) was significant (loglikelihood difference = 10.34; $\chi 2(2) = 3.227e\text{-}05$, $p < 0.001$), indicating that Action improves the fit of the model. A final comparison of a model with Action and Set (AIC 2029.90, logLik -994.95) to a model with Action only (AIC 2032.45, logLik -997.23) revealed that adding a fixed effect of Set significantly improved the model (loglikelihood difference = 2.28; $\chi 2(1) = 0.03$, $p < .05$). Thus the best fitting model of the accuracy includes Action, Set and the Action × Set interaction.

8.6 Stimuli

Below are the stimuli used for Chapter 3, 4 and 5.

1	Answer	Hoe weet je zoveel over parketvloeren?	Mijn vader is timmerman.
1	Declination	Ik kan je helpen het parket te vervangen.	Mijn vader is timmerman.
1	Pre-offer	Ik zoek iemand die het parket kan vervangen.	Mijn vader is timmerman.
2	Answer	Hoe ben je van de spinnen in je huis afgekomen?	Ik heb insectenspray.
2	Declination	Wil je dat ik met je mee naar huis ga en de spin dood maak?	Ik heb insectenspray.
2	Pre-offer	Er zitten zo veel spinnen in mijn huis.	Ik heb insectenspray.
3	Answer	Hoe ga je de doucheafvoer maken?	Ik heb ontstopper.
3	Declination	Zal ik de doucheafvoer voor je maken?	Ik heb ontstopper.
3	Pre-offer	Ik moet de doucheafvoer maken.	Ik heb ontstopper.
4	Answer	Hoe blijf je eigenlijk zo fit?	Ik ben hardloper.
4	Declination	Zal ik je helpen met de training voor de wedstrijd?	Ik ben hardloper.
4	Pre-offer	Ik weet niet hoe ik moet trainen voor de wedstrijd.	Ik ben hardloper.
5	Answer	Hoe weet je zoveel over Denemarken?	Ik heb een boek over Scandinavië.
5	Declination	Wil je mijn reisgids van Denemarken lenen?	Ik heb een boek over Scandinavië.
5	Pre-offer	Ik ben op zoek naar een reisgids van Denemarken.	Ik heb een boek over Scandinavië.
6	Answer	Wat voor een soort feest ben je aan het plannen?	Ik ga trouwen.

6	Declination	Wil je een keer met me uitgaan?	Ik ga trouwen.
6	Pre-offer	Ik wil ervaring opdoen met het maken van huwelijksreportages.	Ik ga trouwen.
7	Answer	Hoe weet je zoveel over bomen?	Mijn vader is tuinman.
7	Declination	Ik kan je helpen met het planten van de bomen.	Mijn vader is tuinman.
7	Pre-offer	Ik wil bomen planten maar ik weet niet hoe dat moet.	Mijn vader is tuinman.
8	Answer	Wat doe je met het avondeten?	Ik ga iets lekkers koken.
8	Declination	Ik kan een afhaalmaaltijd voor je meebrengen.	Ik ga iets lekkers koken.
8	Pre-offer	Ik wou dat ik een privékok had.	Ik ga iets lekkers koken.
9	Answer	Hoe is het je gelukt om de computer te repareren?	Mijn beste vriend is een computernerd.
9	Declination	Ik kan je een nummer geven van iemand die je computer kan repareren.	Mijn beste vriend is een computernerd.
9	Pre-offer	Ik moet iemand vinden die mijn computer kan repareren.	Mijn beste vriend is een computernerd.
10	Answer	Hoe ga je het slot vinden in het donker?	Ik heb een zaklampje.
10	Declination	Zal ik het licht aandoen zodat je het slot kunt zien?	Ik heb een zaklampje.
10	Pre-offer	Ik kan het slot niet zien in het donker.	Ik heb een zaklampje.
11	Answer	Wat ga je toch met al die bloem doen?	Ik bak vanavond oliebollen.
11	Declination	Wil je met ons mee komen naar het feest?	Ik bak vanavond oliebollen.
11	Pre-offer	Ik zal mijn moeders eten op oudejaarsdag wel missen.	Ik bak vanavond oliebollen.
12	Answer	Waar ga je heen?	Ik moet naar de boekhandel.
12	Declination	Wil je samen met mij lunchen?	Ik moet naar de boekhandel.
12	Pre-offer	Ik heb een nieuwe agenda nodig.	Ik moet naar de boekhandel.
13	Answer	Wat zijn je plannen voor morgen?	Ik ga naar de uni.

13	Declination	Ik kan het boek voor je halen uit de UB.	Ik ga naar de uni.
13	Pre-offer	Ik ben vergeten een boek uit de UB te halen.	Ik ga naar de uni.
14	Answer	Hoe weet je zoveel over klassieke muziek?	Mijn vader is pianist.
14	Declination	Ik kan je aan iemand voorstellen die muziek kan spelen op je receptie.	Mijn vader is pianist.
14	Pre-offer	We zoeken iemand die muziek kan spelen op de receptie.	Mijn vader is pianist.
15	Answer	Waarom ken jij Stockholm zo goed?	Mijn vriendin is Zweeds.
15	Declination	Ik ken Stockholm goed, dus ik kan je wel wat tips geven.	Mijn vriendin is Zweeds.
15	Pre-offer	Ik moet iemand zien te vinden die Stockholm goed kent.	Mijn vriendin is Zweeds.
16	Answer	Waarom weet je zoveel over juridische zaken?	Mijn man is jurist.
16	Declination	Ik kan je juridisch advies geven over het contract.	Mijn man is jurist.
16	Pre-offer	Ik heb juridisch advies nodig over mijn contract.	Mijn man is jurist.
17	Answer	Wat voor zaden heb je voor je tuin?	Ik heb allerlei groentezaden.
17	Declination	Ik kan je zaden van sla geven voor in de tuin.	Ik heb allerlei groentezaden.
17	Pre-offer	Ik zou graag zaden van sla willen voor in de tuin.	Ik heb allerlei groentezaden.
18	Answer	Hoe fiets je met je dochtertje?	Ik heb een fietsstoeltje.
18	Declination	Je kunt onze fietskar lenen voor je dochter.	Ik heb een fietsstoeltje.
18	Pre-offer	Ik moet een fietskar of iets dergelijks lenen.	Ik heb een fietsstoeltje.
19	Answer	Hoe ben je droog gebleven in deze regen?	Ik heb een grote paraplu.
19	Declination	Je mag mijn regenjas wel lenen.	Ik heb een grote paraplu.
19	Pre-offer	Ik had echt mijn regenjas mee moeten nemen vandaag.	Ik heb een grote paraplu.
20	Answer	Waar ga je al deze pizza's laten?	Ik heb een grote vriezer.
20	Declination	Ik kan de pizza's in mijn koelkast leggen voor je.	Ik heb een grote vriezer.
20	Pre-offer	Ik heb geen ruimte in mijn koelkast voor alle pizza's.	Ik heb een grote vriezer.
21	Answer	Waar ga je de slaapbank neerzetten?	Ik heb een grote woonkamer.

21	Declination	We kunnen het feest bij mij houden als je appartement te klein is.	Ik heb een grote woonkamer.
21	Pre-offer	Ik wil een feestje geven, maar mijn appartement is te klein.	Ik heb een grote woonkamer.
22	Answer	Hoe kun je zien wat voor vogels het zijn?	Ik heb een verrekijker.
22	Declination	Ik kan wel wat dichterbij rijden zodat we de vogels beter kunnen zien.	Ik heb een verrekijker.
22	Pre-offer	Ik zou willen dat we de vogels beter konden zien.	Ik heb een verrekijker.
23	Answer	Hoe ga je voor het ticket betalen?	Ik heb een creditcard.
23	Declination	Ik kan je wat geld lenen voor het ticket.	Ik heb een creditcard.
23	Pre-offer	Ik heb geen geld om het ticket te betalen.	Ik heb een creditcard.
24	Answer	Hoe ga je kamperen in de zomer?	Ik heb een tent.
24	Declination	Je mag onze caravan lenen als je wilt.	Ik heb een tent.
24	Pre-offer	Ik moet kampeerspullen kopen.	Ik heb een tent.
25	Answer	Wat gebruik je om het schilderij mee op te hangen?	Ik heb een boormachine.
25	Declination	Zal ik iets brengen om het schilderij mee op te hangen?	Ik heb een boormachine.
25	Pre-offer	Ik moet iets vinden om het schilderij mee op te hangen.	Ik heb een boormachine.
26	Answer	Hoe heb je deze krachtoefeningen geleerd?	Mijn zus is fitnesstrainer.
26	Declination	Ik kan je wat krachtoefeningen laten zien.	Mijn zus is fitnesstrainer.
26	Pre-offer	Ik moet wat goede krachtoefeningen leren.	Mijn zus is fitnesstrainer.
27	Answer	Hoe zorg je dat alles droog blijft op je tuinfeest?	Ik heb een partytent.
27	Declination	Ik kan je wel een waterdichte parasol lenen voor je tuinfeest.	Ik heb een partytent.
27	Pre-offer	Ik wil een tuinfeest geven maar het gaat misschien regenen.	Ik heb een partytent.
28	Answer	Waar ga je het vuur mee aansteken?	Ik heb lucifers.
28	Declination	Zal ik een grote aansteker meebrengen voor de barbecue?	Ik heb lucifers.
28	Pre-offer	Ik heb mijn aansteker thuis laten liggen.	Ik heb lucifers.
29	Answer	Hoe ben je van je allergie afgekomen?	Ik heb goede hooikoortstabletten.
29	Declination	We kunnen wel binnen eten als je allergisch bent?	Ik heb goede hooikoortstabletten.

29	Pre-offer	Ik denk dat ik allergisch voor iets ben.	Ik heb goede hooikoortstabletten.
30	Answer	Hoe ben je van plan je spullen te gaan verhuizen?	Ik heb een auto.
30	Declination	Ik kan je wel helpen verhuizen met mijn bakfiets.	Ik heb een auto.
30	Pre-offer	Ik moet uitzoeken hoe ik mijn spullen kan verhuizen.	Ik heb een auto.
31	Answer	Wat doe je allemaal naast je studie?	Ik werk momenteel als barman.
31	Declination	Heb je zin om mee te gaan naar het theater vanavond?	Ik werk momenteel als barman.
31	Pre-offer	Mijn moeder zoekt mensen die kunnen bedienen op haar feest.	Ik werk momenteel als barman.
32	Answer	Wat voor een fiets gebruik je voor de fietstocht?	Ik heb een mountainbike.
32	Declination	Wil je mijn fiets lenen voor de fietstocht?	Ik heb een mountainbike.
32	Pre-offer	Ik zoek een sportieve fiets voor de fietstocht.	Ik heb een mountainbike.
33	Answer	Wat doe je in je thee?	Ik heb honing.
33	Declination	Ik zal wat suiker voor je thee halen.	Ik heb honing.
33	Pre-offer	Deze kop thee is te bitter.	Ik heb honing.
34	Answer	Wat is je bestemming in de Verenigde Staten?	Ik ga naar New York.
34	Declination	Je kan me in mei komen bezoeken in Berlijn.	Ik ga naar New York.
34	Pre-offer	De camera die ik wil hebben is veel goedkoper in de Verenigde Staten.	Ik ga naar New York.
35	Answer	Waarom ben je eigenlijk zo geïnteresseerd in auto's?	Mijn broer is automonteur.
35	Declination	Misschien kan ik je auto repareren?	Mijn broer is automonteur.
35	Pre-offer	Ik moet mijn auto laten repareren.	Mijn broer is automonteur.
36	Answer	Waar wil je zometeen lunchen?	Ik ga naar de kantine.
36	Declination	Zal ik wat te eten voor je kopen?	Ik ga naar de kantine.
36	Pre-offer	Ik ben vergeten om drinken te kopen.	Ik ga naar de kantine.
37	Answer	Hoeveel contant geld is er nog over?	Ik heb honderd euro.
37	Declination	Ik kan je geld lenen als je dat nodig hebt.	Ik heb honderd euro.
37	Pre-offer	Ik ben vergeten geld te pinnen.	Ik heb honderd euro.
38	Answer	Wat wil je doen vanmiddag?	Ik ga basketballen.
38	Declination	Wil je mee naar de bioscoop vanavond?	Ik ga basketballen.
38	Pre-offer	Ik moet echt sporten vandaag.	Ik ga basketballen.
39	Answer	Wat wil je vanavond doen?	Ik ga naar de bioscoop.

39	Declination	Heb je zin om met ons naar het feest te gaan?	Ik ga naar de bioscoop.
39	Pre-offer	Ik wil vanavond een film kijken of zoiets.	Ik ga naar de bioscoop.
40	Answer	Hoe ben je op internet gekomen?	Ik heb een smartphone.
40	Declination	Ik kan het adres voor je vinden op internet.	Ik heb een smartphone.
40	Pre-offer	Ik moet ergens internettoegang vinden.	Ik heb een smartphone.
41	Answer	Wat voor drank is er vanavond?	Ik heb een krat bier.
41	Declination	Ik kan wel wat drinken halen voor vanavond?	Ik heb een krat bier.
41	Pre-offer	Ik ben helemaal vergeten drank te halen.	Ik heb een krat bier.
42	Answer	Wat zijn je plannen voor het komende weekend?	Ik ga naar een popfestival.
42	Declination	Ik kan een kaartje voor je regelen voor het concert van komend weekend.	Ik ga naar een popfestival.
42	Pre-offer	Ik heb zin om volgend weekend naar een concert te gaan.	Ik ga naar een popfestival.
43	Answer	Wat heb je besloten over de nieuwe woning?	Ik verhuis naar Lent.
43	Declination	We kunnen het appartement in het centrum van Nijmegen wel delen.	Ik verhuis naar Lent.
43	Pre-offer	Ik werk in Nijmegen, maar ik ken niemand in de buurt.	Ik verhuis naar Lent.
44	Answer	Wat ben je van plan in juli?	Ik ga dan op vakantie.
44	Declination	Kom je op mijn feest in juli?	Ik ga dan op vakantie.
44	Pre-offer	Ik zoek een kamer voor een paar weken in juli.	Ik ga dan op vakantie.
45	Answer	Waarom heb je nog een Labrador genomen?	Ik hou van honden.
45	Declination	Zal ik de Labrador in zijn hok stoppen zodat hij je niet irriteert?	Ik hou van honden.
45	Pre-offer	Ik moet iemand vinden die voor mijn Labrador kan zorgen.	Ik hou van honden.
46	Answer	Waar wil je nu naar toe?	Ik ga naar de Coop.
46	Declination	Wil je mee naar het park?	Ik ga naar de Coop.
46	Pre-offer	Ik heb zin in cola.	Ik ga naar de Coop.
47	Answer	Hoeveel golfclubs heb je nu?	Ik heb een complete golfset.
47	Declination	Je kunt een golfclub van mij lenen.	Ik heb een complete golfset.
47	Pre-offer	Ik zal wat golfclubs moeten lenen op de baan.	Ik heb een complete golfset.
48	Answer	Hoe ga je de website maken?	Ik kan programmeren.
48	Declination	Ik kan de website voor je maken.	Ik kan programmeren.

48	Pre-offer	Ik moet iemand vinden om de website te maken.	Ik kan programmeren.
49	Answer	Hoe ga je van je slechte adem afkomen?	Ik heb kauwgom.
49	Declination	Probeer deze pepermuntjes eens.	Ik heb kauwgom.
49	Pre-offer	Ik stink echt uit mijn mond.	Ik heb kauwgom.
50	Answer	Hoe ben jij zo goed geworden in het analyseren van data?	Ik heb een goed statistiekboek.
50	Declination	Heb je mischien hulp nodig met je data-analyse?	Ik heb een goed statistiekboek.
50	Pre-offer	Ik ben helaas niet goed in data-analyse.	Ik heb een goed statistiekboek.
51	Answer	Wat doe je tegen de pijn in je schouders?	Ik heb sterke pijnstillers.
51	Declination	Ik kan je wel een massage geven als je schouders zo zeer doen?	Ik heb sterke pijnstillers.
51	Pre-offer	Mijn schouders doen erg zeer, ik kan wel een massage of zoiets gebruiken.	Ik heb sterke pijnstillers.
52	Answer	Hoe is het je gelukt om zoveel vis te vangen?	Ik heb een goede hengel.
52	Declination	Als je graag wil vissen kan ik je visspullen lenen.	Ik heb een goede hengel.
52	Pre-offer	Ik wil graag vissen maar ik heb geen visspullen.	Ik heb een goede hengel.
53	Answer	Waarop slaapt je broer vanavond?	Ik heb een oude slaapbank.
53	Declination	Je mag mijn logeerbed wel lenen.	Ik heb een oude slaapbank.
53	Pre-offer	Ik zoek een goedkoop bed.	Ik heb een oude slaapbank.
54	Answer	Welke vreemde taal spreek je naast Engels?	Ik kan een beetje Duits.
54	Declination	Ik kan de brief uit München voor je lezen.	Ik kan een beetje Duits.
54	Pre-offer	Ik begrijp niks van deze brief uit München.	Ik kan een beetje Duits.
55	Answer	Wat ga je dit weekend doen?	Ik heb zaterdag een feestje.
55	Declination	Wil je dit weekend met ons uitgaan?	Ik heb zaterdag een feestje.
55	Pre-offer	Ik wil graag uitgaan maar de clubs hier zijn zo saai.	Ik heb zaterdag een feestje.
56	Answer	Hoe is het je gelukt om de Duitse vertaling te maken?	Mijn vriendin is Duits.
56	Declination	Ik kan een Duitser voor je vinden die de vertaling kan maken.	Mijn vriendin is Duits.
56	Pre-offer	Ik moet een Duitser zien te vinden die de vertaling voor me kan maken.	Mijn vriendin is Duits.
57	Answer	Hoe ga je de tijden van de wedstrijd bijhouden?	Ik heb een stopwatch.

57	Declination	Ik kan de tijden van de wedstrijd bijhouden.	Ik heb een stopwatch.
57	Pre-offer	Ik moet de tijden van de wedstrijd bijhouden.	Ik heb een stopwatch.
58	Answer	Hoe weet je hoeveel houtskool je moet gebruiken?	Ik heb al vaak gebarbecued.
58	Declination	Ik kan je helpen met het aansteken van de houtskool.	Ik heb al vaak gebarbecued.
58	Pre-offer	Ik moet uitvinden hoeveel houtskool ik moet gebruiken.	Ik heb al vaak gebarbecued.
59	Answer	Hoe ga je je muziek laten horen op het feest?	Ik heb goede boxen.
59	Declination	Je mag mijn koptelefoon wel lenen, die klinkt geweldig.	Ik heb goede boxen.
59	Pre-offer	Ik kan de muziek alleen op mijn computer afspelen.	Ik heb goede boxen.
60	Answer	Welke teamsport voer je uit?	Ik zit in een voetbalteam.
60	Declination	Je kunt meedoen in ons basketballteam?	Ik zit in een voetbalteam.
60	Pre-offer	Ik wil graag meedoen met een teamsport.	Ik zit in een voetbalteam.
61	Answer	Hoe ga je voor het diner betalen?	Ik heb contant geld.
61	Declination	Ik kan het diner voor je betalen.	Ik heb contant geld.
61	Pre-offer	Ik ben mijn portemonnee vergeten.	Ik heb contant geld.
62	Answer	Hoe is je rijexamen gegaan?	Ik ben geslaagd.
62	Declination	Ik kan je wel helpen met de voorbreiding op je rijexamen.	Ik ben geslaagd.
62	Pre-offer	Ik zoek iemand die zijn rijexamen heeft gehaald.	Ik ben geslaagd.
63	Answer	Hoe kan het dat je zoveel weet over Japan?	Ik heb een vriend uit Tokyo.
63	Declination	Ik kan je voorstellen aan iemand die Japans met je kan oefenen.	Ik heb een vriend uit Tokyo.
63	Pre-offer	Ik moet een Japanner interviewen.	Ik heb een vriend uit Tokyo.
64	Answer	Hoe ben je op het idee gekomen om in een kapsalon te werken?	Mijn moeder is haarstylist.
64	Declination	Ik kan korting voor je regelen bij mijn kapsalon.	Mijn moeder is haarstylist.
64	Pre-offer	Ik moet een goede kapsalon vinden.	Mijn moeder is haarstylist.
65	Answer	Wat eet je vanavond?	Ik maak pizza.
65	Declination	Heb je zin om vanavond friet te eten?	Ik maak pizza.
65	Pre-offer	Ik wil vanavond iets lekkers eten.	Ik maak pizza.
66	Answer	Waar koop je je shampoo?	Ik ga naar de Kruidvat.

66	Declination	Ik kan wel shampoo voor je meenemen?	Ik ga naar de Kruidvat.
66	Pre-offer	Mijn shampoo is op.	Ik ga naar de Kruidvat.
67	Answer	Hoe houd jij de boekhouding toch zo goed op orde?	Mijn broer is accountant.
67	Declination	Ik kan je helpen met de boekhouding.	Mijn broer is accountant.
67	Pre-offer	Ik heb wat hulp nodig met mijn boekhouding.	Mijn broer is accountant.
68	Answer	Waarom ben je je nu aan het omkleden?	Ik ga vanavond naar een concert.
68	Declination	Wil je bij me komen eten?	Ik ga vanavond naar een concert.
68	Pre-offer	Ik wil vandaag iets leuks gaan doen.	Ik ga vanavond naar een concert.
69	Answer	Hoe oefen je met zingen?	Ik zit in een gezellig koor.
69	Declination	Je kunt meezingen met onze sanggroep.	Ik zit in een gezellig koor.
69	Pre-offer	Ik wil weer met anderen gaan zingen.	Ik zit in een gezellig koor.
70	Answer	Wat heb je besloten over de Vierdaagse?	Ik ben al ingeschreven.
70	Declination	Ik kan je de informatie over de Vierdaagse toesturen.	Ik ben al ingeschreven.
70	Pre-offer	Ik hoop dat iemand met me mee wil lopen in de Vierdaagse.	Ik ben al ingeschreven.
71	Answer	Wat doe jij vandaag met deze prachtige zon?	Ik ga naar het strand.
71	Declination	Wil je met ons mee gaan zonnen in het park?	Ik ga naar het strand.
71	Pre-offer	Ik wil zonnen nu het zo'n prachtig weer is.	Ik ga naar het strand.
72	Answer	Wat wil je drinken?	Ik neem een grote Spa Blauw.
72	Declination	Je mag mijn glas water hebben.	Ik neem een grote Spa Blauw.
72	Pre-offer	Ik ben vergeten om water te vragen.	Ik neem een grote Spa Blauw.
73	Answer	Waarom heb je zoveel gehakt gekocht?	Ik maak vanavond lasagne.
73	Declination	Zal ik een pizza bestellen en langsbrengen?	Ik maak vanavond lasagne.
73	Pre-offer	Ik heb zin in echt goed Italiaans eten.	Ik maak vanavond lasagne.
74	Answer	Waar wil je naar toe in België?	Ik ga naar Brussel.
74	Declination	Ik kan wat goede chocola voor je meenemen uit België.	Ik ga naar Brussel.
74	Pre-offer	Ik zou graag wat echte Belgische chocola willen kopen.	Ik ga naar Brussel.
75	Answer	Hoe kreeg je die Hollandaisesaus zo perfect?	Ik heb een goed recept.

75	Declination	Zal ik je helpen om een perfecte Hollandaise saus te maken?	Ik heb een goed recept.
75	Pre-offer	Ik weet niet hoe ik perfecte Hollandaisesaus moet maken.	Ik heb een goed recept.
76	Answer	Wat voor melk heb je om in de koffie te doen?	Ik heb sojamelk.
76	Declination	Hier is wat melk voor in je koffie.	Ik heb sojamelk.
76	Pre-offer	Ik kan geen gewone melk drinken in mijn koffie.	Ik heb sojamelk.
77	Answer	Waar ga je de lasagne in doen?	Ik heb een grote ovenschaal.
77	Declination	Ik kan je een bak lenen voor de lasagne.	Ik heb een grote ovenschaal.
77	Pre-offer	We hebben geen bak voor de lasagne.	Ik heb een grote ovenschaal.
78	Answer	Waar ga je al je spullen in stoppen?	Ik heb een grote koffer.
78	Declination	Als je spullen niet in je koffer passen, dan kan ik je er een lenen.	Ik heb een grote koffer.
78	Pre-offer	Ik weet niet of alle spullen in mijn koffer passen.	Ik heb een grote koffer.
79	Answer	Hoe ken je al deze fitnessoefeningen?	Ik heb een fitness DVD.
79	Declination	Je mag mijn boek met fitnessoefeningen wel lenen.	Ik heb een fitness DVD.
79	Pre-offer	Ik moet eens wat nieuwe fitnessoefeningen uitproberen.	Ik heb een fitness DVD.
80	Answer	Wat ben je vandaag van plan te doen?	Ik ga nu shoppen.
80	Declination	Wil je langskomen om een film te kijken?	Ik ga nu shoppen.
80	Pre-offer	Ik wil echt graag de nieuwe kledingwinkel zien.	Ik ga nu shoppen.
81	Answer	Hoe ga je alles naar huis dragen?	Ik heb een grote fietstas.
81	Declination	Laat mij wat voor je naar huis dragen.	Ik heb een grote fietstas.
81	Pre-offer	Ik kan niet alles naar huis dragen.	Ik heb een grote fietstas.
82	Answer	Hoe kunnen jouw nagels toch altijd zo mooi zijn?	Mijn zus heeft een nagelstudio.
82	Declination	Ik kan je wel een nagelbehandeling geven.	Mijn zus heeft een nagelstudio.
82	Pre-offer	Ik wil echt een nagelbehandeling.	Mijn zus heeft een nagelstudio.
83	Answer	Wat voor toetje eten we vanavond?	Ik heb aardbeienvla.
83	Declination	Zal ik een toetje meenemen voor vanavond?	Ik heb aardbeienvla.

83	Pre-offer	Ik ben vergeten een toetje te kopen voor vanavond.	Ik heb aardbeienvla.
84	Answer	Waar doe je de boodschappen voor het avondeten?	Ik ga naar de markt.
84	Declination	Zal ik zoete aardappelen voor je meenemen?	Ik ga naar de markt.
84	Pre-offer	Ik heb geen zoete aardappelen gevonden in de winkels.	Ik ga naar de markt.
85	Answer	Hoe weet je wat er in de mode is?	Ik werk in een kledingwinkel.
85	Declination	Ik kan je wat tips geven over wat er in de mode is.	Ik werk in een kledingwinkel.
85	Pre-offer	Ik wil iets nieuws kopen, maar ik weet niet wat er in de mode is.	Ik werk in een kledingwinkel.
86	Answer	Hoe ben je toch zo veel afgevallen?	Ik ben begonnen aan een nieuw dieet.
86	Declination	Ik zal je morgen de cake laten proeven.	Ik ben begonnen aan een nieuw dieet.
86	Pre-offer	Ik zoek een methode om af te vallen die echt werkt.	Ik ben begonnen aan een nieuw dieet.
87	Answer	Hoe komt het dat jij zo goed met kinderen om kunt gaan?	Ik heb een klein broertje.
87	Declination	Heb je hulp nodig bij het oppassen op je nichtje?	Ik heb een klein broertje.
87	Pre-offer	Ik moet op mijn nichtje passen, maar ik weet totaal niet wat ik moet doen.	Ik heb een klein broertje.
88	Answer	Waarom heb je het gegrilde vlees niet geprobeerd?	Ik ben vegetariër.
88	Declination	Zal ik wat vlees meenemen om te grillen?	Ik ben vegetariër.
88	Pre-offer	Ik wil meer recepten met groenten leren.	Ik ben vegetariër.
89	Answer	Wat ga je aandoen als het koud wordt?	Ik heb een sjaal.
89	Declination	Als je het koud hebt mag je mijn trui wel lenen.	Ik heb een sjaal.
89	Pre-offer	Ik had mijn trui moeten meebrengen.	Ik heb een sjaal.
90	Answer	Wat doe je in de salade?	Ik heb wat sladressing.
90	Declination	Ik kan olijfolie halen voor de salade.	Ik heb wat sladressing.
90	Pre-offer	Er zit niet zoveel smaak aan mijn salade.	Ik heb wat sladressing.
91	Answer	Wat voor snacks ben je aan het eten?	Ik heb nootjes.
91	Declination	Zal ik wat chips voor je kopen onderweg?	Ik heb nootjes.
91	Pre-offer	Ik wou dat ik wat snacks had meegenomen.	Ik heb nootjes.

92	Answer	Welke Japanse gerechten kun je bereiden?	Ik maak zelf sushi.
92	Declination	Zal ik eten afhalen bij het Japanse restaurant?	Ik maak zelf sushi.
92	Pre-offer	Helaas is er hier geen Japans restaurant.	Ik maak zelf sushi.
93	Answer	Waarom wil je geen wijn of bier?	Ik ben zwanger.
93	Declination	Wil je een glas wijn of een biertje?	Ik ben zwanger.
93	Pre-offer	Ik wil me gaan bezighouden met babyfoto's.	Ik ben zwanger.
94	Answer	Wat gebruik je om de soep te pureren?	Ik heb een staafmixer.
94	Declination	Ik kan je mijn blender wel lenen om de soep te pureren.	Ik heb een staafmixer.
94	Pre-offer	Ik moet een apparaat vinden om de soep te pureren.	Ik heb een staafmixer.
95	Answer	Wat doe je deze zomer in je vakantie?	Ik ga naar Rome.
95	Declination	Ik kan wat goede kaas voor je meenemen uit Italië.	Ik ga naar Rome.
95	Pre-offer	Ik zoek Italiaanse kaas maar het is lastig te krijgen.	Ik ga naar Rome.
96	Answer	Hoe ga je in Wenen het hotel vinden?	Ik heb een navigatiesysteem.
96	Declination	Ik kan je een kaart van Wenen geven.	Ik heb een navigatiesysteem.
96	Pre-offer	Ik ben bang om in Wenen te verdwalen.	Ik heb een navigatiesysteem.
97	Answer	Hoe ben je van plan op het feestje te komen in deze regen?	Ik ga met de auto.
97	Declination	Ik kan je wel even met de fiets naar het feestje brengen.	Ik ga met de auto.
97	Pre-offer	Ik heb geen zin om in de regen naar het feestje te fietsen.	Ik ga met de auto.
98	Answer	Hoe ga je voor iedereen een zitplaats regelen?	Ik heb meerdere klapstoelen.
98	Declination	Wil je dat ik een paar grote zitkussens meebreng?	Ik heb meerdere klapstoelen.
98	Pre-offer	Ik denk niet dat er voldoende zitplaatsen zijn.	Ik heb meerdere klapstoelen.
99	Answer	Waarom ben je niet met de fiets?	Ik woon vlakbij.
99	Declination	Zal ik je een lift naar huis geven?	Ik woon vlakbij.
99	Pre-offer	Ik ben te moe om naar huis te fietsen.	Ik woon vlakbij.
100	Answer	Wat ben je van plan aankomende zaterdag te doen?	Ik ga naar de IKEA.

100	Declination	Wil je mee met de fietstocht zaterdag?	Ik ga naar de IKEA.
100	Pre-offer	Ik moet een paar dingen voor mijn huis kopen.	Ik ga naar de IKEA.
101	Answer	Wat is je volgende bestemming in Amerika?	Ik ga deze maand naar Boston.
101	Declination	Ik zal wat Amerikaans snoep voor je meebrengen.	Ik ga deze maand naar Boston.
101	Pre-offer	Ik mis echt het Amerikaanse snoep.	Ik ga deze maand naar Boston.
102	Answer	Wat voor lekkers is er voor het feestje?	Ik heb nog wat chocola.
102	Declination	Je mag wel een stuk van mijn cake hebben.	Ik heb nog wat chocola.
102	Pre-offer	Ik wou dat we iets zoets hadden gekocht.	Ik heb nog wat chocola.
103	Answer	Wat ben je aan het drinken?	Ik heb nog wat wijn.
103	Declination	Ik zal wat bier voor je halen.	Ik heb nog wat wijn.
103	Pre-offer	Ik heb zin in iets te drinken.	Ik heb nog wat wijn.
104	Answer	Waar wil je een hamburger kopen?	Ik ga naar McDonald's.
104	Declination	Ik kan wel een hamburger voor je maken.	Ik ga naar McDonald's.
104	Pre-offer	Ik heb echt zin in een hamburger.	Ik ga naar McDonald's.
105	Answer	Wat voor frisdrank is er?	Ik heb cola.
105	Declination	Zal ik frisdrank voor je kopen?	Ik heb cola.
105	Pre-offer	Ik had frisdrank moeten halen.	Ik heb cola.
106	Answer	Wat zit er in je tas?	Ik heb zonnebrandcrème.
106	Declination	Zal ik de parasol voor je opzetten?	Ik heb zonnebrandcrème.
106	Pre-offer	Ik voel dat mijn rug begint te verbranden.	Ik heb zonnebrandcrème.
107	Answer	Waar doe jij je boodschapppen?	Ik ga naar de Albert Heijn.
107	Declination	Wil je de film samen met ons kijken?	Ik ga naar de Albert Heijn.
107	Pre-offer	Ik heb een paar dingen voor het avondeten nodig.	Ik ga naar de Albert Heijn.
108	Answer	Wat voor eten is er in huis?	Ik heb nog koekjes.
108	Declination	Ik kan speculaas meebrengen.	Ik heb nog koekjes.
108	Pre-offer	Ik heb zin in iets zoets.	Ik heb nog koekjes.
109	Answer	Hoe ga je je band plakken?	Ik heb een reparatiesetje.
109	Declination	Ik kan iets brengen om je band te plakken.	Ik heb een reparatiesetje.
109	Pre-offer	Ik moet mijn band plakken.	Ik heb een reparatiesetje.
110	Answer	Welk beleg is er voor op de broodjes?	Ik heb salami.
110	Declination	Ik kan rauwe ham gaan halen voor op je broodje.	Ik heb salami.
110	Pre-offer	Er is geen beleg voor op mijn broodje.	Ik heb salami.
111	Answer	Hoe wil je thuiskomen zonder je fiets?	Ik ga met de taxi.
111	Declination	Je mag mijn fiets lenen als je wilt.	Ik ga met de taxi.

111	Pre-offer	Ik wil echt niet naar huis fietsen.	Ik ga met de taxi.
112	Answer	Hoe ga je vannacht slapen zonder verwarming?	Ik heb een extra dekbed.
112	Declination	Je kunt mijn slaapzak morgen lenen voor je vriend.	Ik heb een extra dekbed.
112	Pre-offer	Ik had het koud vannacht want de verwarming was kapot.	Ik heb een extra dekbed.
113	Answer	Hoe komt het dat het hier zo lekker fris ruikt?	Ik heb mijn kamer gisteren schoongemaakt.
113	Declination	Ik kan morgen met wat reinigingsmiddelen langskomen.	Ik heb mijn kamer gisteren schoongemaakt.
113	Pre-offer	Mijn huis is zo'n puinhoop, daar kunnen we niet heen.	Ik heb mijn kamer gisteren schoongemaakt.
114	Answer	Wat voor sla is er voor in de salade?	Ik heb rucola.
114	Declination	Zal ik wat sla voor de salade meebrengen?	Ik heb rucola.
114	Pre-offer	Ik ben vergeten sla mee te brengen voor mijn salade.	Ik heb rucola.
115	Answer	Wat is er te drinken?	Ik heb koffie.
115	Declination	Zal ik een kop thee voor je maken?	Ik heb koffie.
115	Pre-offer	Ik kan nu wel wat warms gebruiken.	Ik heb koffie.
116	Answer	Wat is er te eten bij je barbecue?	Ik heb een hoop vlees.
116	Declination	Zal ik wat worstjes meenemen naar je barbecue?	Ik heb een hoop vlees.
116	Pre-offer	Ik ben vergeten iets voor de barbecue te kopen.	Ik heb een hoop vlees.
117	Answer	Waar ga je de barbecue neerzetten?	Ik heb een grote tuin.
117	Declination	Als je niet genoeg ruimte hebt kunnen we bij mij barbecueën.	Ik heb een grote tuin.
117	Pre-offer	We hebben geen balkon, dus ik kan niet barbecueën.	Ik heb een grote tuin.
118	Answer	Hoe ben je zo snel hersteld van je blessure?	Mijn vrouw is fysiotherapeute.
118	Declination	Ik kan je enkele goede oefeningen laten zien voor je blessure.	Mijn vrouw is fysiotherapeute.
118	Pre-offer	I moet behandeld worden voor mijn blessure.	Mijn vrouw is fysiotherapeute.
119	Answer	Wat doet je partner voor de kost?	Mijn vriendin is zangeres.
119	Declination	Ik zou wel in je band kunnen zingen op het concert.	Mijn vriendin is zangeres.

119	Pre-offer	We hebben iemand nodig die op het concert kan zingen.	Mijn vriendin is zangeres.
120	Answer	Hoe ga je de bladzijden bij elkaar houden?	Ik heb wat paperclips.
120	Declination	Ik zal mijn nietmachine voor je halen.	Ik heb wat paperclips.
120	Pre-offer	De nietmachine werkt niet erg goed.	Ik heb wat paperclips.
121	Answer	Wanneer neem je vrij deze zomer?	Ik heb volgende week vakantie.
121	Declination	Heb je zin om op kantoor te lunchen na dit weekend?	Ik heb volgende week vakantie.
121	Pre-offer	Ik wil snel een dagtrip naar Amsterdam maken.	Ik heb volgende week vakantie.
122	Answer	Waarom kun je niet naar het feest?	Ik ga vanavond studeren.
122	Declination	Heb je zin om een film te kijken?	Ik ga vanavond studeren.
122	Pre-offer	Ik wil met iemand oefenen voor het examen.	Ik ga vanavond studeren.
123	Answer	Wat is je plan voor vandaag?	Ik ga wandelen.
123	Declination	Wil je met ons mee naar het strand?	Ik ga wandelen.
123	Pre-offer	Ik moet wat meer bewegen.	Ik ga wandelen.
124	Answer	Waar gaat je suz logeren deze winter?	Ik heb een lege kamer.
124	Declination	Je zus kan wel in mijn huis logeren.	Ik heb een lege kamer.
124	Pre-offer	Ik moet een logeeradres vinden voor de winter.	Ik heb een lege kamer.
125	Answer	Waar ben je van plan de fietslampjes te kopen?	Ik ga vanmiddag naar de HEMA.
125	Declination	Zal ik nieuwe fietslampjes voor je kopen?	Ik ga vanmiddag naar de HEMA.
125	Pre-offer	Ik moet echt nieuwe fietslampjes kopen.	Ik ga vanmiddag naar de HEMA.
126	Answer	Wat ga je gebruiken om het handvat te maken?	Ik heb superlijm.
126	Declination	Ik kan iets meebrengen om het handvat te maken.	Ik heb superlijm.
126	Pre-offer	Ik moet iets kopen om het handvat te maken.	Ik heb superlijm.

9 Samenvatting

Taalhandelingen (speech acts) zijn verbale acties zoals verzoeken, plagerijen en complimenten (Schegloff, 1996; Searle, 1975). Deelnemers aan een gesprek moeten deze acties kunnen herkennen om communicatie succesvol te laten verlopen. Het herkennen van taalhandelingen is verre van eenvoudig, aangezien verbale uitingen vaak niet aangeven welke actie er wordt uitgevoerd (Levinson, 2013; Searle, 1975). De uitdaging voor luisteraars wordt nog verder vergroot door de buitengewoon snelle beurtwisselingen in een gesprek (De Ruiter, Mitterer, & Enfield, 2006; Stivers et al., 2009), waardoor er maar beperkte tijd overblijft om de actie te herkennen en een relevante respons te plannen (Levinson, 2013).

De doelstelling van dit proefschrift was het onderzoeken van het tijdsverloop van de herkenning van taalhandelingen in gesproken, natuurlijke dialogen binnen een nieuw kader dat bruggen legt tussen methoden van cognitieve neurowetenschappen en onderzoek naar conversationele interactie. Het hoofddoel was het testen van de hypothese dat taalhandelingen vroegtijdig in verbale uitingen worden herkend, wat snelle en efficiënte beurtwisseling faciliteert. Een tweede doel was het onderzoeken hoe de herkenning van taalhandelingen wordt gemoduleerd door de soort actie die wordt uitgevoerd en hoe dit past in de grotere actiesequentie.

Er is een experimenteel paradigma ontwikkeld dat zich richt op deze problemen. Het omvat gesproken mini-dialogen met doeluitingen (bijv. *Ik heb een creditcard*) die afhankelijk van de voorgaande beurt drie afzonderlijke taalhandelingen uitvoeren – Antwoorden, Afwijzingen, Pre-verzoeken (pre-offer). De doeluitingen zijn identiek in alle condities (en specificeren dus geen actie), maar de contexten verschillen wel in de soort taalhandeling die wordt uitgevoerd. De Antwoordconditie bevat een vraag-antwoord sequentie (*Hoe ga je voor het ticket betalen?* – IK HEB EEN CREDITCARD). De Afwijzingconditie bestaat uit een aanbod, gevolgd door een afwijzing (*Ik kan je wat geld lenen voor het ticket.* – IK HEB EEN CREDITCARD). De Pre-verzoekconditie bevat een aankondiging van een probleem, gevolgd door een aanzet tot een aanbod; dit wordt in conversatie-analyse een pre-verzoek (pre-offer) genoemd (Schegloff, 1988, 2007) (*Ik heb geen geld om het ticket te betalen.* – IK HEB EEN CREDITCARD). Afwijzingen en Pre-verzoeken (de cruciale condities) hebben met elkaar gemeen dat ze indirecter zijn dan Antwoorden (de controleconditie) aangezien ze meer inferenties vereisen om begrepen te worden, maar ze verschillen in hun relatie tot de voorgaande en komende beurten in een gesprek. Afwijzingen komen voor in zeer specifieke contexten, waarbij ze een tweede paardeel (second pair part) zijn van een aangrenzend paar (adjacency pair) (Schegloff, 2007), en zijn daarom relatief anticipeerbaar. Pre-verzoeken, echter, komen voor in minder specifieke contexten aangezien ze een nieuwe actiesequentie initiëren, en daarom minder voorspelbaar zijn.

Na de introductie van het onderzoeksveld en de beschrijving van de methodologie in Hoofdstuk 1, beschrijft Hoofdstuk 2 gedragsbevindingen van een 'self-paced' leestaak en een begripstaak. Het doel van het gedragsexperiment was tweevoudig. Het onderzoeken hoe betrouwbaar deelnemers de taalhandeling kunnen categoriseren van zinnen die geen actie specificeren (begripstaak), en het verkrijgen van een grove schatting van het tijdsverloop van de herkenning van taalhandelingen zoals gereflecteerd wordt in self-paced leestijden. De contextuitingen van de hierboven beschreven mini-dialogen werden auditief gepresenteerd, gevolgd door geschreven versies van de doeluitingen. Deze werden gepresenteerd via een self-paced leestaak. Dat wil zeggen dat proefpersonen de uitingen woord-voor-woord op het scherm kregen en zelf met een drukknop konden bepalen wanneer het volgende woord op het scherm kwam. De resultaten van een actiecategorisatietaak laten zien dat deelnemers acties zeer nauwkeurig kunnen categoriseren (95,8%), gebaseerd op alleen de voorgaande actie. De leestijden suggereren dat de herkenning van taalhandelingen in de visuele modaliteit relatief vroeg in de doeluiting begint, namelijk bij het eerste woord of het werkwoord, maar ook dat late verwerking aan het eind van de zin wellicht vereist is.

Hoofdstuk 3 maakt gebruik van de excellente temporele resolutie van event-related brain potentials (ERPs) om het tijdsverloop van de herkenning van taalhandelingen in gesproken dialogen te volgen. Hoe snel kunnen luisteraars taalhandelingen die geen actie specificeren herkennen in de auditieve modaliteit? Hoe beïnvloedt de sequentiële context dit proces? Het ERP experiment dat wordt gepresenteerd in Hoofdstuk 3 gebruikt de gesproken versies van de dialogen in Hoofdstuk 2 en dezelfde actiecategorisatietaak. Hoofdstuk 3 identificeert drie ERP effecten gerelateerd aan taalhandelingen: een vroege fronto-temporale positiviteit in de rechterhersenhelft vanaf 200 milliseconden na de aanvang van de verbale uiting bij Pre-verzoeken en vanaf 400 millieseconden bij Afwijzingen (relatief tot Antwoorden); een iets latere frontale links/middenlijn positiviteit bij Afwijzingen (relatief tot Antwoorden); en een late, posterieure negativiteit bij het laatste woord van de verbale uiting bij Pre-verzoeken (relatief tot Antwoorden en Afwijzingen). De vroege positiviteiten geven aan dat gesproken herkenning van taalhandelingen vroeg in de gespreksbeurt begint, wanneer de verbale uiting pas deels is verwerkt. Het argument voor vroege herkenning van taalhandelingen wordt versterkt doordat er geen ERP verschillen werden waargenomen tussen Afwijzingen en Antwoorden bij het laatste woord van de verbale uiting; herkenning van de actie lijkt plaats te vinden voor het laatste woord wanneer de voorgaande beurt zeer beperkend is in termen van welke actie er mogelijk kan volgen. Echter, de late negativiteit bij Pre-verzoeken laat zien dat aanvullende verwerking die gebaseerd is op de complete verbale uiting vereist is bij complexere acties zoals Pre-verzoeken. Daarbij bevindt de taalhandeling zich in een minder beperkende context en wordt er een nieuwe actiesequentie geïnitieerd. Kort samengevat, levert Hoofdstuk 3 gedeeltelijke steun voor het argument dat taalhandelingen vroeg worden herkend. Bovendien laten de resultaten zien dat het tijdsverloop van de herkenning wordt beïnvloed door het soort taalhandeling dat wordt uitgevoerd en hoe deze in de grotere actiesequentie past.

Hoofdstuk 4 toetst de robuustheid van de ERP effecten gerelateerd aan taalhandelingen die werden waargenomen in Hoofdstuk 3. Een nadeel van de actiecategorisatietaak die werd gebruikt in Hoofdstuk 2 en 3 is dat het de aandacht op de cruciale taalhandelingen richtte en de mogelijke interpretatie hiervan beperkte. Gegeven dat ERP-componenten gemoduleerd kunnen worden door taakeisen (zie bijvoorbeeld: Chwilla, Brown, & Hagoort, 1995; Gunter & Friederici, 1999; Kuperberg, 2007; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007), onderzoekt Hoofdstuk 4 of de ERP effecten die beschreven werden in Hoofdstuk 3 geïnduceerd werden door de categorisatietaak of dat ze generaliseren naar situaties waarin expliciete categorisatie niet vereist is. Het experiment in Hoofdstuk 4 gebruikte een waar/onwaar beoordelingstaak die begrip op het actie niveau meet zonder deelnemers van tevoren de mogelijke categorieën van taalhandelingen te verstrekken. Aanvullende fillerdialogen werden ook toegevoegd om strategische verwerking van de doeltaalhandelingen te verminderen. Hoofdstuk 4 repliceert twee van de drie ERP effecten die beschreven werden in Hoofdstuk 3, wat de robuustheid van deze effecten ondersteunt. Afwijzingen brachten een vroege frontale positiviteit teweeg vanaf 200 ms na het begin van de verbale uiting, vergelijkbaar met de gecombineerde effecten van de twee frontale positiviteiten in het vorige experiment. Weer werden er geen ERP effecten gevonden bij het laatste woord in Afwijzingen relatief tot Antwoorden, wat aangeeft dat vroege herkenning van taalhandelingen in Afwijzingen robuust is. De late negativiteit bij het laatste woord van de verbale uiting in Preverzoeken werd ook gerepliceerd. Echter, er werden geen vroege ERP effecten waargenomen bij Preverzoeken, in tegenstelling tot de vorige studie. Dit geeft aan dat de experimentele omgeving enigszins invloed uitoefent op de vroege herkenning van taalhandelingen, met name op de fronto-temporale positiviteit in de rechterhersenhelft bij Preverzoeken die werd gerapporteerd in Hoofdstuk 3. Het beeld dat ontstaat, gebaseerd op Hoofdstuk 4, is dat in zeer beperkende contexten, wanneer de actiesequentie aan een einde begint te komen – wat het geval is bij Afwijzingen - de herkenning van taalhandelingen 200 ms na het begin van de verbale uiting begint en wordt afgerond voordat het laatste, cruciale woord wordt gehoord. Echter, in minder beperkende contexten, wanneer een pre-sequentie wordt geïnitieerd – wat het geval is in Pre-verzoeken – is vroege verwerking van acties minder waarschijnlijk, en is de volledige propositie nodig om de taalhandeling te begrijpen.

Hoofdstuk 5 onderzoekt of het patroon van resultaten dat werd beschreven in Hoofdstuk 4 wordt ondersteund door tijdsfrequentie analyses van de EEG data in hetzelfde experiment. ERPs alleen reflecteren een bepaald deel van het EEG signaal dat gereleteerd is aan een gebeurtenis, namelijk activiteit met dezelfde fase, welke achterblijft na het middelen in het tijdsdomein (Bastiaansen, Mazaheri, & Jensen, 2008; Pfurtscheller & Lopes da Silva, 1999). Door het analyseren van oscillerende activiteit met verschillende fasen, beoogt Hoofdstuk 5 een completer beeld van het tijdsverloop van de herkenning van taalhandelingen te geven. Het specifieke doel is om licht te werpen op de rol van anticipatieprocessen in de herkenning van taalhandelingen. De tijdsfrequentierepresentaties in de EEG data van Hoofdstuk 4 werden geanalyseerd. De herkenning van taalhandelingen was hoofdzakelijk gerelateerd aan een verminderde sterkte van de alpha/beta golven net voor en tijdens het begin van Afwijzingen; van -150 tot 50 ms na het begin van de doeluiting relatief tot Pre-verzoeken en van -100 tot 450 ms relatief tot Antwoorden. Gegeven de relatie tussen alpha/beta onderdrukking en anticipeerbare verwerking (bijv., Bastiaansen & Brunia, 2001; Pfurtscheller & Lopes da Silva, 1999; van Ede, Szebényi, & Maris, 2014), blijkt uit deze resultaten dat anticipatieaandacht een rol speelt in de vroege herkenning van taalhandelingen in Afwijzingen. Het gebrek aan oscillerende effecten bij het laatste woord van de verbale uiting bij Afwijzingen sluit aan bij de ERP resultaten in Hoofdstuk 3 en 4, wat verdere steun levert voor de vroege herkenning van taalhandelingen in Afwijzingen. Ten slotte werden er geen sterkteverschillen waargenomen tussen Pre-verzoeken en Antwoorden, wat aangeeft dat ERPs (met name late negativiteit) gevoeliger zijn voor de herkenning van taalhandelingen in Pre-verzoeken dan hersengolven. Over het algemeen onderbouwen de resultaten het tijdsverloop van de herkenning van gesproken taalhandelingen zoals beschreven in Hoofdstuk 3 en 4 en leveren bewijs dat vroege herkenning van de actie in Afwijzingen samengaat met anticipeerbare aandacht voor de taalhandeling al 150 ms voordat de verbale uiting begint.

Hoofdstuk 6 geeft een samenvatting van de belangrijkste empirische bevindingen, bespreekt hun theoretische implicaties en geeft richting voor toekomstig onderzoek. De belangrijkste bevinding van dit proefschrift is dat onder sommige omstandigheden luisteraars de taalhandeling van een lopende verbale uiting kunnen oppikken nog voor deze is afgerond – voordat het laatste, cruciale woord is gehoord. Deze bevinding betwist de klassieke pragmatische theorie (bijv., Grice, 1975; Searle, 1975) die aanneemt dat luisteraars eerst de semantische inhoud van een complete zin moeten verwerken voor ze kunnen begrijpen wat de spreker werkelijk bedoelt met zijn of haar verbale uiting. De resultaten geven ook problemen voor de meer recente pragmatische kaders (bijv., Levinson, 2000; Sperber & Wilson, 1986, 2002), aangezien deze geen rekening houden met de sequentiële beperkingen die werkzaam zijn tussen taalhandelingen in een gesprek. De bevinding dat luisteraars taalhandelingen nog voor hun afronding kunnen herkennen suggereert dat predictie tijdens taalcomprehensie niet beperkt is tot het niveau van individuele lexicale elementen of hun syntactische, semantische of conceptuele kenmerken (bijv., Altmann & Kamide, 1999; Federmeier, 2007; Kutas, Delong, & Smith, 2011; Thornhill & Van Petten, 2012), maar ook plaatsvindt op het niveau van taalhandelingen. Hoofdstuk 6 speculeert kort over de mechanismen die zulke predictie van taalhandelingen mogelijk maken.

Een andere belangrijke bevinding van dit proefschrift is dat het begrip van een uiting wordt beïnvloedt door het soort taalhandeling dat de uiting uitvoert en hoe deze past in de sequentiële context. Dit vraagt om een bredere kijk op de kwestie dan het onderzoek naar begrip van "directe" versus "indirecte" taalhandelingen (bijv., Basnakova, Weber, Petersson, Van Berkum, & Hagoort, 2014; Coulson & Lovett, 2010; van Ackeren, Casasanto, Bekkering, Hagoort, & Rüschemeyer, 2012). Een meer verfijnde aanpak die rekening houdt met het soort taalhandeling en hun sequentiële organisatie in beurtwisseling is gewenst. Hoofdstuk 6 stelt dat abstracte kennis van sequenties van taalhandelingen, geconceptualiseerd als scripts (Schank & Abelson, 1977) of schema's (Brewer & Nakamura, 1984; Rumelhart, 1980) in het geheugen, een cruciale rol zouden kunnen spelen in de vroege herkenning van taalhandelingen.

Een derde belangrijke bevinding is dat de frontale positiviteit vroeg in de verbale uiting bij Afwijzingen en de late negativiteit bij het laatste woord bij Pre-verzoeken relatief robuust zijn in verschillende experimentele taken en designs. Dus deze ERP-effecten leveren een maatstaf voor toekomstig onderzoek naar de herkenning van taalhandelingen en pragmatisch taalbegrip in het algemeen, aangezien het begrijpen van taalhandelingen waarschijnlijk voortkomt uit cognitieve mechanismen die ook gebruikt worden voor andere aspecten van linguïstische en sociale verwerking.

Het proefschrift eindigt met een bespreking van overige vragen die gerelateerd zijn aan het begrijpen van taalhandelingen in een bredere context, zoals de invloed van de linguïstische vorm van de verbale uiting op de herkenning van taalhandelingen; implicaties van de bevindingen van het proefschrift voor onderzoek naar de timing van beurtwisselingen; de rol van theory of mind en affectieve empathie in de herkenning van taalhandelingen; het verschil tussen passief belsuiteren van een gesprek en interactie; en de relatie tussen het verbaal en non-verbaal begrip van acties.

Samenvattend illustreert dit proefschrift hoe een interdisciplinaire aanpak, die EEG methodologie combineert met bevindingen van conversatie-analyse en corpus onderzoek, nieuwe inzichten kan leveren in de herkenning van taalhandelingen – een belangrijke cognitieve capaciteit die ten grondslag ligt aan succesvolle conversationele interactie.

10 Curriculum Vitae

Rósa Signý Gísladóttir was born in 1983 in Reykjavík, Iceland, and attended highschool at Menntaskólinn í Reykjavík (Reykjavík Junior College). She was awarded a Kofi Annan International Scholarship for undergraduate studies at Macalester College in St. Paul, USA, and obtained a BA in Linguistics (magna cum laude) in 2006. She then received a Chevening Scholarship from the British Council to study at King's College London, completing an MSc degree (with distinction) in Computational Linguistics and Formal Grammar in 2008. After a short break from research working as a PR and Marketing Director of Innovation Center Iceland, in 2010 she was awarded an IMPRS for Language Sciences Fellowship by the Max Planck Society to do doctoral research at the Max Planck Institute for Psycholinguistics and Radboud University in Nijmegen, the Netherlands. In addition to doing her PhD research in the Language and Cognition Department, she contributed on a collaborative project on conversational repair. Rósa is currently a research scientist at deCODE genetics in Iceland.

Gisladottir, R. S., Chwilla, D., & Levinson, S. C. (In preparation). The influence of speech act type and the sequential context on sentence comprehension: An ERP study.

Gisladottir, R. S., Bögels, S. & Levinson, S. C. (Under review). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialogue.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., & Enfield, N. J. (in press). Universal principles in the repair of communication problems. *PLOS ONE*.

Gisladottir, R. S., Chwilla, D., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLOS ONE*, *10*(3): e0120068. doi:10.1371/journal.pone.0120068.

Gisladottir, R. S. (2015). Other-initiated repair in Icelandic. *Open Linguistics*, *1*(1), 309-328. doi:10.1515/opli-2015-0004.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., Drew, P., Floyd, S., Gipper, S., Gisladottir, R. S., Hoymann, G., Kendrick, K., Levinson, S. C., Magyari, L., Manrique, E., Rossi, G., San Roque, L., & Torreira, F. (2013). Huh? What? – A first survey in 21 languages. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 343-380). New York: Cambridge University Press.

Gisladottir, R. S., Chwilla, D., Schriefers, H., & Levinson, S. C. (2012). Speech act recognition in conversation: Experimental evidence. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of* *the Cognitive Science Society* (CogSci 2012) (pp. 1596-1601). Austin, TX: Cognitive Science Society.

12 MPI Series in Psycholinguistics

1. *The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing.* Miranda van Turennout
2. *The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography.* Niels O. Schiller
3. *Lexical access in the production of ellipsis and pronouns.* Bernadette M. Schmitt
4. *The open-/closed-class distinction in spoken-word recognition.* Alette Haveman
5. *The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach.* Kay Behnke
6. *Gesture and speech production.* Jan-Peter de Ruiter
7. *Comparative intonational phonology: English and German.* Esther Grabe
8. *Finiteness in adult and child German.* Ingeborg Lasser
9. *Language input for word discovery.* Joost van de Weijer
10. *Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe.* James Essegbey
11. *Producing past and plural inflections.* Dirk Janssen
12. *Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea.* Anna Margetts
13. *From speech to words.* Arie van der Lugt
14. *Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language.* Eva Schultze-Berndt
15. *Interpreting indefinites: An experimental study of children's language comprehension.* Irene Krämer

16. *Language-specific listening: The case of phonetic sequences.* Andrea Weber
17. *Moving eyes and naming objects.* Femke van der Meulen
18. *Analogy in morphology: The selection of linking elements in Dutch compounds.* Andrea Krott
19. *Morphology in speech comprehension.* Kerstin Mauth
20. *Morphological families in the mental lexicon.* Nivja H. de Jong
21. *Fixed expressions and the production of idioms.* Simone A. Sprenger
22. *The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria).* Birgit Hellwig
23. *Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies.* Fermín Moscoso del Prado Martín
24. *Contextual influences on spoken-word processing: An electrophysiological approach.* Daniëlle van den Brink
25. *Perceptual relevance of prevoicing in Dutch.* Petra M. van Alphen
26. *Syllables in speech production: Effects of syllable preparation and syllable frequency.* Joana Cholin
27. *Producing complex spoken numerals for time and space.* Marjolein Meeuwissen
28. *Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction.* Rachèl J. J. K. Kemps
29. *At the same time...: The expression of simultaneity in learner varieties.* Barbara Schmiedtová
30. *A grammar of Jalonke argument structure.* Friederike Lüpke
31. *Agrammatic comprehension: An electrophysiological approach.* Marlies Wassenaar
32. *The structure and use of shape-based noun classes in Miraña (North West Amazon).* Frank Seifart

33. *Prosodically-conditioned detail in the recognition of spoken words.* Anne Pier Salverda
34. *Phonetic and lexical processing in a second language.* Mirjam Broersma
35. *Retrieving semantic and syntactic word properties.* Oliver Müller
36. *Lexically-guided perceptual learning in speech processing.* Frank Eisner
37. *Sensitivity to detailed acoustic information in word recognition.* Keren B. Shatzman
38. *The relationship between spoken word production and comprehension.* Rebecca Özdemir
39. *Disfluency: Interrupting speech and gesture.* Mandana Seyfeddinipur
40. *The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation.* Christiane Dietrich
41. *Cognitive cladistics and the relativity of spatial cognition.* Daniel B.M. Haun
42. *The acquisition of auditory categories.* Martijn Goudbeek
43. *Affix reduction in spoken Dutch.* Mark Pluymaekers
44. *Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence.* Valesca Kooijman
45. *Space and iconicity in German Sign Language (DGS).* Pamela Perniss
46. *On the production of morphologically complex words with special attention to effects of frequency.* Heidrun Bien
47. *Crosslinguistic influence in first and second languages: Convergence in speech and gesture.* Amanda Brown
48. *The acquisition of verb compounding in Mandarin Chinese.* Jidong Chen
49. *Phoneme inventories and patterns of speech sound perception.* Anita Wagner
50. *Lexical processing of morphologically complex words: An information-theoretical perspective.* Victor Kuperman
51. *A grammar of Savosavo, a Papuan language of the Solomon Islands.* Claudia Wegener

52. *Prosodic structure in speech production and perception.* Claudia Kuzla
53. *The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension.* Sarah Schimke
54. *Studies on intonation and information structure in child and adult German.* Laura de Ruiter
55. *Processing the fine temporal structure of spoken words.* Eva Reinisch
56. *Semantics and (ir)regular inflection in morphological processing.* Wieke Tabak
57. *Processing strongly reduced forms in casual speech.* Susanne Brouwer
58. *Ambiguous pronoun resolution in L1 and L2 German and Dutch.* Miriam Ellert
59. *Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging.* Ian FitzPatrick
60. *Processing casual speech in native and non-native language.* Annelie Tuinman
61. *Split intransitivity in Rotokas, a Papuan language of Bougainville.* Stuart Robinson
62. *Evidentiality and intersubjectivity in Yurakaré: An interactional account.* Sonja Gipper
63. *The influence of information structure on language comprehension: A neurocognitive perspective.* Lin Wang
64. *The meaning and use of ideophones in Siwu.* Mark Dingemanse
65. *The role of acoustic detail and context in the comprehension of reduced pronunciation variants.* Marco van de Ven
66. *Speech reduction in spontaneous French and Spanish.* Francisco Torreira
67. *The relevance of early word recognition: Insights from the infant brain.* Caroline Junge

68. *Adjusting to different speakers: Extrinsic normalization in vowel perception.* Matthias J. Sjerps
69. *Structuring language. Contributions to the neurocognition of syntax.* Katrien R. Segaert
70. *Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading.* Birgit Knudsen
71. *Gaze behavior in face-to-face interaction.* Federico Rossano
72. *Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space.* Conny de Vos
73. *Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning.* Attila Andics
74. *Lexical processing of foreign-accented speech: Rapid and flexible adaptation.* Marijt Witteman
75. *The use of deictic versus representational gestures in infancy.* Daniel Puccini
76. *Territories of knowledge in Japanese conversation.* Kaoru Hayano
77. *Family and neighbourhood relations in the mental lexicon: A cross-language perspective.* Kimberley Mulder
78. *Contributions of executive control to individual differences in word production.* Zeshu Shao
79. *Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing.* Patrick van der Zande
80. *High pitches and thick voices: The role of language in space-pitch associations.* Sarah Dolscheid
81. *Seeing what's next: Processing and anticipating language referring to objects.* Joost Rommers
82. *Mental representation and processing of reduced words in casual speech.* Iris Hanique

83. *The many ways listeners adapt to reductions in casual speech.* Katja Poellmann
84. *Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners.* Giuseppina Turco
85. *Morphological processing in younger and older people: Evidence for flexible dual-route access.* Jana Reifegerste
86. *Semantic and syntactic constraints on the production of subject-verb agreement.* Alma Veenstra
87. *The acquisition of morphophonological alternations across languages.* Helen Buckler
88. *The evolutionary dynamics of motion event encoding.* Annemarie Verkerk
89. *Rediscovering a forgotten language*. Jiyoun Choi
90. *The road to native listening: Language-general perception, language-specific input.* Sho Tsuji
91. *Infants' understanding of communication as participants and observers.* Gudmundur Bjarki Thorgrímsson
92. *Information structure in Avatime*. Saskia van Putten
93. *Switch reference in Whitesands*. Jeremy Hammond
94. *Machine learning for gesture recognition from videos*. Binyam Gebrekidan Gebre
95. *Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish*. Beyza Sümer
96. *An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music*. Salomi Savvatia Asaridou
97. *lncrementality and Flexibility in Sentence Production*. Maartje van de Velde
98. *Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture.* Edwin van Leeuwen

99. *The request system in Italian interaction*. Giovanni Rossi
100. *Timing turns in conversation: A temporal preparation account*. Lilla Magyari
101. *Assessing birth language memory in young adoptees*. Wencui Zhou
102. *A social and neurobiological approach to pointing in speech and gesture*. David Peeters
103. *Investigating the genetic basis of reading and language skills*. Alessandro Gialluisi
104. *Conversation electrified: The electrophysiology of spoken speech act recognition*. Rósa Signý Gísladóttir