

Mathias Barthel

mes shoe milk Stehlam  
Barriton Mikrophon floor lamp  
clipboard luchtballon mes  
pill Barriton drums Schlagzeug broccoli Kerze  
karton Pizza schelm Blume board olif plant zip  
pill karton drums clapper sombrero Trak  
lamp Pizza schelm Mond clapper Brombeere zippo  
karton raket Fahrrad skateboard  
lamp Mond raket nietmachine skateboard  
Sparschwein Zwiebel  
papegai Sonnenblume  
papegai Sonnenblume

# Speech Planning in Dialogue

Psycholinguistic Studies  
of the Timing of Turn Taking



# Speech Planning in Dialogue

Psycholinguistic Studies of the Timing of  
Turn Taking

---

Mathias Barthel



# Speech Planning in Dialogue

## Psycholinguistic Studies of the Timing of Turn Taking

Proefschrift ter verkrijging van de graad van doctor aan de Radboud  
Universiteit Nijmegen op gezag van de rector magnificus prof. dr.  
J. H. J. M. van Krieken, volgens besluit van het college van decanen in  
het openbaar te verdedigen op

donderdag 23 januari 2020  
om 14.30 uur precies

door

Mathias Barthel

geboren op 7 juli 1987  
te Eilenburg (Duitsland)

**Promotoren**

Prof. dr. Stephen C. Levinson

Prof. dr. Antje S. Meyer

**Manuscriptcommissie**

Prof. dr. Mirjam T.C. Ernestus

Dr. Chiara Gambi (Cardiff University, Verenigd Koninkrijk)

Prof. dr. Falk Huettig

Prof. dr. Herbert J. Schriefers

Dr. Roel M. Willems

MPI Series in Psycholinguistics № 150

ISBN 978-94-92910-08-0

Cover design by Janine Kittler and Mathias Barthel

Cover photo by Janine Kittler

Printed and bound by Ipskamp Printing, Nijmegen

©2020 Mathias Barthel

This research was supported by the Max Planck Society for the Advancement of Science, Munich, Germany.

The educational component of the doctoral training was provided by the International Max Planck Research School for Language Sciences, a joint graduate school of the Max Planck Institute for Psycholinguistics, the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen.





---

## ACKNOWLEDGEMENTS

---

Writing a thesis is a journey. Just like in an interesting conversation, the turns it might take are unclear as the candidate sets out with a more or less specific goal in mind, yet still unprepared to immediately approach that goal without any meandering. For the candidate, learning to navigate the ups and downs of that journey is the real value of that special phase in life. Learning to deal with unforeseen problems, frustration, curiosity, critique, mistakes, pride, confusion, and an amazement about the sheer complexities of human social interaction are fundamental to the young scientist as a person, and are a pre-requisite for successfully adding a valuable piece to the understanding of any relevant problem studied in the humanities.

I am very grateful for having undertaken that journey and for all its phases along the way. Grateful especially for the encouragement I received to take up higher studies, particularly by Sabine Fiedler and Steven Roodenrys, and for the teachings that convinced me that my interest in the fine coordination in social interaction is of general academic value. In that respect, special thanks goes to Thomas Pechmann, who recommended me going to Nijmegen to pursue the quest for understanding how people manage conversation, and to Nick Enfield, who encouraged me to take up a PhD on the psycholinguistics of interaction. Centrally, my gratitude goes to Steve Levinson, who repeatedly pushed me to simultaneously stick to the specific questions I set out to pursue and at the same time to not lose sight of the grand picture that motivated me to ask these questions in the first place. All along the way, Steve was a role model for me in being a researcher striving for an all-encompassing understanding of human sociality, and our personal encounters helped me to pin down what is important in life for myself. Also, I am very thankful for having been part of the marvelous team of researchers Steve assembled in the Language and Cognition Department. The intense exchange of knowledge, ideas and questions inspired and taught me so much. Special thanks to Sara Bögels, Francisco Torreira, Mark Dingemanse, Connie de Vos,



Giovanni Rossi, Kobin Kendrick, Judith Holler, Simeon Floyd, Emma Valtersson, Paul Hömke, Rosa Gísladóttir, Tyko Dirksmeyer, Elisabeth Norcliffe, Lilla Magyari, Julija Baranova, Lila San Roque, Sean Roberts, and Gabriela Garrido for the tour de force of project discussions and data sessions. Also, the scientific environment of the MPI and its frequent visitors was constantly pushing the quality of my research and academic output. Especially discussing projects, thoughts, and problems with Amie Fairs, J.P. de Ruyter, Agnieszka Konopka, Xaver Koch, Matthias Sjerps, Falk Huettig, Franziska Hartung, Laura Hahn, Ernesto Guerra, Elizabeth Couper-Kuhlen, Felicia Roberts, Herbert Schriefers, Herbert Clark and numerous others inspired and motivated me, even though I fail to name all of them here. Of all these people, I wish to give special thanks to Antje Meyer, my co-promoter and strongest critic in methodological questions, who always urged me to thoroughly check the feasibility of my planned studies before setting out to test them. Also, thanks for all technical and administrative support, especially to Ronald Fisher, Johan Weustink, Maarten van den Heuvel, Tanja Marton, Karin Kastens, and Edith Sjoerdsma. The facilities and aid you provide are the foundation of quality research. Special thanks also to Freya Materne for being the assistant and confederate in a major part of the studies in this thesis. Without her patience and stamina in endless recording, testing and coding sessions, this thesis would not have been possible, and I would certainly not have learned so much about how to successfully organize and coordinate a research project. One more person deserves my special gratitude, since the form and shape of this thesis would certainly be different without him. Thanks to Sebastian Sauppe, my friend, colleague, co-author, and paronym, for sharing an office with me during all those years, providing me with critical feedback and joining me in orienting myself in the maze of data analysis. Thank you for always having an open ear for both scientific and any other matters.

Acknowledgements most often speak of the great experience of having been a graduate student. While it certainly was a pleasure, it was also tremendously difficult at times. Founding a family and handling additional responsibilities in a double career family is no piece of cake for a junior researcher, and carefully setting priorities is just as delicate a task as respecting them in difficult moments during the research project. When I was buried in the complications of the empirical cycle,

forgetting to eat or sleep, it was my spouse and partner Manuela who carefully reminded me of what is really important and why I am doing all that. And at another point in time, when I was losing faith and decided to throw in the towel, it was her again who gave me the necessary confidence to continue, maybe even without intending to do so. All along the way, she was both the most essential supporter and general critic of the work that now found its form in this thesis. Thank you so much for that, for being my companion, and for so much more that we share and for all the adventures that are still lying ahead of us. This thesis is dedicated to you.



---

# CONTENTS

---

1	INTRODUCTORY REMARKS	1
2	THE TIMING OF RESPONSE PLANNING IN DIALOGUE	9
2.1	Introduction	10
2.2	The study	15
2.3	Methods and Materials	19
2.3.1	Participants	19
2.3.2	Apparatus	19
2.3.3	Visual stimuli	19
2.3.4	Auditory stimuli	21
2.3.5	Items and Design	21
2.3.6	Procedure	22
2.4	Results	23
2.4.1	Response timing	24
2.4.2	Eye-movements	26
2.5	Discussion	31
2.6	Conclusion	35
2.7	Supplementary Materials	35
3	PROGRESSION OF SPEECH PLANNING IN OVERLAP	45
3.1	Introduction	46
3.2	Experiment 1	52
3.2.1	Method	52
3.2.2	Results	56
3.2.3	Discussion	60
3.3	Experiment 2	61
3.3.1	Introduction	61
3.3.2	Method	62
3.3.3	Results	63
3.3.4	Discussion	67
3.4	Experiment 3	68
3.4.1	Introduction	68
3.4.2	Method	69

3.4.3	Results	74	
3.4.4	Discussion	79	
3.5	General Discussion	80	
3.6	Supplementary Materials	83	
4	PROCESSING LOAD IN SPEECH PLANNING IN DIALOGUE		97
4.1	Introduction	98	
4.2	Methods and Materials	101	
4.2.1	Participants	101	
4.2.2	Apparatus	102	
4.2.3	Stimuli	102	
4.2.4	Procedure	103	
4.2.5	Data Preprocessing and Analyses	104	
4.3	Results	105	
4.4	Discussion	109	
5	THE TIMING OF NEXT TURN INITIATION		115
5.1	Introduction	116	
5.2	Methods and Materials	119	
5.2.1	Participants	121	
5.2.2	Apparatus	122	
5.2.3	Visual stimuli	122	
5.2.4	Auditory stimuli	123	
5.2.5	Items and Design	124	
5.2.6	Procedure	124	
5.3	Results	126	
5.4	Discussion	131	
6	SUMMARY AND MODELLING OF RESULTS		135
6.1	Summary of Results	135	
6.1.1	Summary Chapter 2	135	
6.1.2	Summary Chapter 3	136	
6.1.3	Summary Chapter 4	138	
6.1.4	Summary Chapter 5	139	
6.2	A Model of Turn Taking	140	
	References	153	

Concise Summary	177
Samenvatting	183
Zusammenfassung	189
Curriculum Vitae	195
MPI Series in Psycholinguistics	197
Publications	207



---

## LIST OF FIGURES

---

Figure 2.1	Example item displays for confederate and participant	16
Figure 2.2	Timing of looks for planning	28
Figure 2.3	Looks for planning time-locked to the end of the incoming turn	36
Figure 3.1	Trial structure in Experiment 1	55
Figure 3.2	Distribution of naming latencies in Experiment 1.	56
Figure 3.3	Lexical decision performance in Experiment 1	57
Figure 3.4	Naming latencies in Experiment 2	65
Figure 3.5	Lexical decision performance in Experiment 2	66
Figure 3.6	Trial structure in naming trials and switch trials.	73
Figure 3.7	Naming latencies in Experiment 3.	75
Figure 3.8	Reaction times and error rates in lexical decisions in Experiment 3. Bars signify 95% confidence intervals.	77
Figure 3.9	Prior and posterior distributions of Relatedness effect on lexical decision latencies in Experiment 3.	79
Figure 3.10	Distribution of naming latencies by SOA condition in Experiment 1.	83
Figure 3.11	Distribution of naming latencies by subject in Experiment 1.	84
Figure 3.12	Distribution of lexical decision latencies by subject in Experiment 1.	84
Figure 3.13	Distribution of naming latencies by subject in Experiment 2.	85
Figure 3.14	Distribution of lexical decision latencies by subject in Experiment 2.	85
Figure 3.15	Distribution of naming latencies by subject in Experiment 3.	86



Figure 3.16	Distribution of lexical decision latencies by subject in Experiment 3.	87
Figure 4.1	Pupil changes by condition	106
Figure 5.1	Intonation contours of used conditions	121
Figure 5.2	Example item displays	123
Figure 5.3	Looks for planning by condition	129
Figure 6.1	Model of Language Processing in Conversation	141

---

## LIST OF TABLES

---

Table 2.1	Example sentences of conditions	17
Table 2.2	Verbal response latencies	24
Table 2.3	Statistics of verbal response latencies	26
Table 2.4	Statistics of eye-movements	29
Table 2.5	Statistics of eye-movements - pairwise comparison 1a	37
Table 2.6	Statistics of eye-movements - pairwise comparison 1b	37
Table 2.7	Statistics of eye-movements - pairwise comparison 2a	37
Table 2.8	Statistics of eye-movements - pairwise comparison 2b	38
Table 2.9	Statistics of eye-movements - pairwise comparison 3a	38
Table 2.10	Statistics of eye-movements - pairwise comparison 3b	38
Table 2.11	Statistics of eye-movements - pairwise comparison 4a	39
Table 2.12	Statistics of eye-movements - pairwise comparison 4b	39
Table 2.13	List of Materials	43
Table 3.1	Critical conditions tested in Experiment 1.	54
Table 3.2	Bayesian linear regression model on button press latencies in Experiment 1.	60
Table 3.3	Bayesian linear regression model on button press latencies in Experiment 3.	78
Table 3.4	Logit mixed effects regression model on error rates in Experiment 1.	88
Table 3.5	Mixed effects regression model on button press latencies in Experiment 1.	89
Table 3.6	Logit mixed effects regression model on error rates in Experiment 2.	90

Table 3.7	Mixed effects regression model on button press latencies in Experiment 2.	90
Table 3.8	Bayesian linear regression model on button press latencies in Experiment 2.	90
Table 3.9	Logit mixed effects regression model on error rates in Experiment 3.	91
Table 3.10	Mixed effects regression model on button press latencies in Experiment 3.	91
Table 3.11	List of Materials	95
Table 4.1	Example sentences of the four conditions	103
Table 4.2	Peak and mean amplitudes, and peak latencies by condition	107
Table 4.3	Linear mixed effects regression models predicting task-evoked pupillary responses	108
Table 5.1	Verbal response latencies	127
Table 5.2	Statistics of verbal response latencies	128
Table 5.3	Statistics of eye-movements	130





# I

---

## INTRODUCTORY REMARKS

---

The ability to take turns lies at the heart of human social interaction, with most humans spending a great part of their waking hours talking to other people (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007). Especially in conversation, interlocutors oscillate between uptake and output, taking the roles of listener and speaker in alternation, seemingly without major effort (Clark, 1996; Sacks, Schegloff, & Jefferson, 1974). While language comprehension and production were mostly studied in isolation in non-interactive settings during the history of psycholinguistic research (Fernández & Cairns, 2017; Harley, 2014; Levelt, 2012), the majority of language use as well as language learning takes place during conversation (Bruner, 1974; Bruner & Watson, 1983; Durkin, 1987), where comprehension and production of speech go hand in hand and, as I will argue in the following chapters, often overlap in time.

Take the following simple exchange of turns as an example, where speaker A asks speaker B the question "Are you planning to have lunch later?" and B responds "Yeah, let's go together.", without leaving even half a second of a gap between the end of the question and the beginning of the answer. Interlocutors exchange turns at talk like these with remarkably accurate timing, avoiding long gaps between turns and long stretches of overlapping talk. In this way, only one speaker will speak for most of the time. While there are small differences between speaker communities, short gaps of less than half a second are most common throughout the world's languages (Stivers et al., 2009). Similarly, cases of overlapping talk are also found in all languages that have been examined, being overall rather infrequent and commonly very short. In most cases when the talk of two speakers does overlap, one of the speakers will fall silent shortly, if necessary by dropping her turn without finishing it in order to mend the conversational situation

(Schegloff, 2000). It is interesting to note here that perfect alignment of turn ends and beginnings is also rare, so that turns most commonly begin with a short gap of about a quarter of a second after the end of the preceding turn, i.e. after the floor is open again for another contribution to the conversation (Heldner & Edlund, 2010). We will return to these observations in Chapter 6.

How does talking time get to be distributed among interlocutors so that the timing of turn taking is ordered to the extent just described? A major result of conversation analytic research has been that interlocutors interactively manage who can take the floor when on the basis of a set of turn allocation rules that were first distilled by Sacks et al. (1974) in their seminal generalization from the observable features of turn taking. In their description of the turn taking system, each speaker has the right to produce one turn-constructive unit at a time, which can be between one word and one sentence in length, and each of these units is followed by a transition-relevance place, at which a transition from one speaker to the next may occur and where a set of rules applies to regulate turn allocation. Sacks et al. (1974) find that this allocation of speaking turns is either overt, i.e. that the next speaker is selected by the current turn, or covert, and in that case dependent on the timing of turn initiation. Notably, in the case of covert turn allocation, the first interlocutor to speak up at a point in time where speaker change becomes relevant gains the rights and obligations to produce the next turn. If no other speaker self-selects for the next turn, the current speaker can, but need not, continue his turn until the next transition-relevance place.

This set of rules of turn management creates a time pressure that pushes next speakers, if they intend to take the floor, to hasten to initiate their turn as quickly as possible when a next turn is covertly allocated. This is obviously the case in multi-party conversations, as other listeners might compete for the next turn, but also in dialogue, since the current speaker might continue her turn if the floor is not taken quickly. Time pressure also applies at transition-relevance places where the next turn is allocated overtly, as the observable fast timing in speaker change is coupled with a rich semiotics of turn timing. If, for example, a question is not answered quickly, the mere silence following the question might be interpreted as meaningful by the speaker asking the question, leading him to assume that the questionee is struggling with the question's presuppositions or that the answer will be dispreferred

(Clayman, 2002; Pomeranz & Heritage, 2012). The questioner might even re-select himself and reformulate the question so that potential problems might be solved, for instance by phrasing it in a way that flips the preference of answers or by downsizing a request (Davidson, 1984), or the questioner might give the dispreferred answer himself (Levinson, 1983). While the time window for unmarked turn-transitions can be expected to not be fixed but vary between transitions (Barthel, 2012), analyses of corpora show that turns that are initiated after more than a 700 ms gap are much more likely to contain dispreferred rather than preferred responses (Kendrick & Torreira, 2014), to initiate repair (Kendrick, 2015), or to be disagreeing with assessments (Pomeranz, Atkinson, & Heritage, 1984; F. Roberts, Francis, & Morgan, 2006). The same 700 ms threshold seems to be relevant for passive listeners' ratings on a responder's willingness to comply with a request (F. Roberts & Francis, 2013; F. Roberts et al., 2006), and such effects of turn timing on the interpretation of turns seem to be stable across languages and independent of the semantics or the understanding of the turn's content (F. Roberts, Margutti, & Takano, 2011). In an EEG study, Bögels, Kendrick, and Levinson (2015) played turns from a telephone corpus representing initiating actions such as requests, offers, proposals, and invitations to participants. These turns were either followed by a fast (300 ms gap) or a slow (1000 ms gap) minimal response that was either preferred (e.g. accepting an offer) or dispreferred (e.g. not accepting an invitation). The authors found that fast dispreferred responses evoke an N400 effect relative to fast preferred responses, showing that, after short gaps, dispreferred responses were less expected than preferred responses. In delayed responses however, no difference in N400 amplitude was observed, showing that participants' expectations about the valence of the answer changed merely on the basis of its timing (see also Bögels, Kendrick, & Levinson, 2019). These findings show that the timing of a turn with respect to the end of the preceding turn is relevant for the interpretation of the turn itself or of the speaker's attitude towards the previous turn or the action pursued with it. This means that in order to avoid being interpreted in an unintended way, next speakers need to manage to initiate their turn in tight coordination with the previous turn's end. Thus, the semiotics of turn timing puts immense time pressure on the language production system to rapidly translate a communicative intention into a linguistic form.



Employing the basic set of turn-taking rules and its consequences for talk in interaction needs to be practiced by children as they grow into competent members of the speaker community they are raised in. Babies have been observed to engage in what has been termed ‘proto-conversation’ already at a few months of age (Bruner, 1975; Tomasello, 1999), producing sounds in bursts that show to carry striking similarities in timing with adult conversation (Hilbrink, Gattis, & Levinson, 2015). But as soon as they grow more competent in using words to communicate, the timing of their contributions slows down due to the increasing computational effort that needs to be handled by the growing child (Casillas, Bobb, & Clark, 2016). While they are not yet able to understand the meaning of much of their linguistic input, one-year old toddlers quickly detect the ends of speaking turns and predict speaker change (Casillas & Frank, 2012, 2017). However, only in later childhood do children display the abilities to converse in adult-like speed themselves. Gradually, their language production system successfully adapts to the demanding tasks of turn-timing, enabling them to start the articulation of their contributions only fractions of a second after the ends of the preceding turns (Hilbrink et al., 2015).

This language production system is a fascinating, sophisticated machinery managing complex jobs at high speed in order to prepare utterances for smooth exchanges of turns like the one in the example mentioned above. For an utterance to be produced, minimally three main steps need to be taken from an intention to speak to the articulation of a turn at talk (Bock, 1995; Bock, Levelt, & Gernsbacher, 1994; Kempen & Hoenkamp, 1987; Levelt, 1989). In a first step, the utterance needs to be planned conceptually, meaning that the speaker needs to conceive suitable linguistic concepts fitting her turn’s message. In a second step, this conceptual structure needs to be transferred into a syntactic structure which has to be filled with lexical entries in a grammatical, linear order. In a third step, these lexical entries need to be fleshed out with phonetic forms, which in turn need to be translated into motor programs that can be executed in order to move the respective muscles to produce speech. The sheer rapidity of the underlying cognitive processes is a prerequisite for the speedy timing of turn taking that can be observed in everyday conversation. Yet, even though these processing stages are computed at impressive speed, the sum of these processes takes much longer than the widely observed

turn transition times, even with very simple utterances (Indefrey, 2011; Indefrey & Levelt, 2004), not to speak of whole sentences (Ferreira, 1991; Griffin & Bock, 2000). It is therefore an intriguing question when response preparation begins and when each of the stages of speech planning is processed. We will investigate these questions in Chapters 2 and 3, testing the hypotheses that response planning begins as soon as possible, and regularly so in overlap with the incoming turn, and that all stages of utterance formulation are run through while the incoming turn is still unfolding.

Assuming that planning begins before the end of the incoming turn, that turn's message would need to be anticipated at some point before its end for the response turn to be planned so as to bear reference to the current turn. Indeed, previous research found language comprehension to be predictive on almost any level of processing, including predictions of syntactic constructions (e.g. Staub & Clifton, 2006), morpho-syntactic information (e.g. Kamide, 2012; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; N. Y. Wicha, Bates, Moreno, & Kutas, 2003; N. Y. Wicha, Moreno, & Kutas, 2004), semantic information (e.g. Altmann & Kamide, 1999, 2007; Borovsky, Elman, & Fernald, 2012; Szewczyk & Schriefers, 2013; M. K. Tanenhaus, Carlson, & Trueswell, 1989), and word form information (DeLong, 2009; DeLong, Urbach, & Kutas, 2005), while the phonological level is still under debate and might only be pre-activated under very constrained conditions (Nieuwland, 2019; Nieuwland et al., 2018, see Kuperberg and Jaeger (2016) for review). In addition to the content of incoming language, comprehenders have been found to also anticipate the number of upcoming words in order to estimate when an utterance will come to completion (Magyari, Bastiaansen, de Ruiter, & Levinson, 2014; Magyari & de Ruiter, 2012). Addressees anticipate the message of an utterance in order to predict the speaker's intention as early as possible (Gisladottir, Bögels, & Levinson, 2018; Gisladottir, Chwilla, & Levinson, 2015). While previous research has shown that speech comprehension and speech production interfere with one another (Boiteau, Malone, Peters, & Almor, 2014; Kemper, Herman, & Lian, 2003; Kubose et al., 2006; Schriefers, Meyer, & Levelt, 1990), anticipatory language processing on several linguistic levels might ease the integration of language input during parallel language planning (Pickering & Garrod, 2004). Anticipation could therefore reduce the known effects of interference

and free processing resources that are needed for response planning in overlap with the incoming turn. In this case, it becomes an interesting question whether speech planning is in fact more demanding during the incoming turn than during the gap between turns. We investigate this question in Chapter 4, testing the hypotheses that (a) planning in overlap with the incoming turn is cognitively more demanding than planning during the gap between turns and that (b) predictability of the incoming turn will reduce processing load in next speakers planning their turn.

Now, assuming a relevant response has been prepared early enough, the task for the next speaker remains to align her turn's articulation accurately with the end of the incoming turn. How does the next speaker know when exactly the floor will be open again? Previous research of turn taking examined corpora of speech exchanges and detected a number of characteristics in the current speaker's behaviour and in the current turn that were observable just before speaker change occurred (or did not occur), including gestural, syntactic, and intonational cues (Beattie, 1981; Beattie, Cutler, & Pearson, 1982; Duncan, 1972, 1974; Duncan & Niederehe, 1974; Gravano & Hirschberg, 2011; Local & Walker, 2012; Ward, 2019). The discovery that these cues co-occur together with speaker changes led to the formulation of the so-called signaling model of turn taking, which assumes that turn allocation is handled primarily by the current speaker, who can decide to display any of the available turn-yielding or turn-keeping cues or not. However, the observations on which this model is based remain correlational and do not shed light on the causal links between a current speaker's behaviour and the timing of a next speaker's turn. In the study presented in Chapter 5 of this thesis, we therefore investigate what sources of information are actually used by next speakers to detect that the current turn ends and that speaker change becomes relevant.

In sum, this dissertation tackles the question how the observed finely coordinated timing of turn taking is possible from the point of view of the next speaker. When do next speakers start to plan their turn? What levels of planning are run through in overlap with the incoming turn? Does planning in overlap lead to increased processing load as compared to planning during the gap between turns? And what sources of information do next speakers use to time the articulation of their own turn? These questions and their answers lead towards the

formulation of a cognitive model of turn taking from the perspective of the next speaker in a conversation, connecting the necessary tasks and mechanisms of response preparation and the timing of articulation. This model will be presented and discussed in Chapter 6.

#### A NOTE TO THE READER

Chapters 2 to 5 have been composed to be standalone papers and are already (being) published. These papers are presented here with minimal adjustment – please excuse occasional repetitions and alternative formulations in these chapters.



---

## THE TIMING OF RESPONSE PLANNING IN DIALOGUE

---

Published as:

Barthel, M., Sauppe, S., Levinson, S. C., and Meyer, A. S. (2016). The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm. *Frontiers in Psychology* (7), page 1858.

### ABSTRACT

In conversation, interlocutors rarely leave long gaps between turns, suggesting that next speakers begin to plan their turns while listening to the previous speaker. The present experiment used analyses of speech onset latencies and eye-movements in a task-oriented dialogue paradigm to investigate when speakers start planning their responses. Adult German participants heard a confederate describe sets of objects in utterances that either ended in a noun (e.g. *Ich habe eine Tür und ein Fahrrad* ('I have a door and a bicycle')) or a verb form (*Ich habe eine Tür und ein Fahrrad besorgt* ('I have gotten a door and a bicycle')), while the presence or absence of the final verb either was or was not predictable from the preceding sentence structure. In response, participants had to name any unnamed objects they could see in their own displays in utterances such as *Ich habe ein Ei* ('I have an egg'). The main question was when participants would start to plan their responses. The results are consistent with the view that speakers begin to plan their turns as soon as sufficient information is available to do so, irrespective of further incoming words.

## 2.1 INTRODUCTION

Most psycholinguistic studies are directed at detailed processes in either comprehension or production, testing single participants in isolation. Yet, interactive language use involves both, not only in rapid succession but also in partial overlap. In conversation, the predominant form of language use, interlocutors fluently engage in switching of roles, taking turns at talking with only about 200 ms between turns on average (de Ruiter, Mitterer, & Enfield, 2006; Heldner & Edlund, 2010; Levinson, 2016; Levinson & Torreira, 2015; Sacks et al., 1974; Stivers et al., 2009). One factor that maintains this pace is that markedly delayed turns carry a special semiotics, presaging disagreement or non-compliance with what was said before (Bögels, Kendrick, & Levinson, 2015; Kendrick & Torreira, 2014; Levinson, 1983; F. Roberts & Francis, 2013; F. Roberts et al., 2011).

Given the known latencies involved in speech production of 600 ms or more for a single word in picture naming tasks (Indefrey & Levelt, 2004; Jescheniak, Schriefers, & Hantsch, 2003; Levelt, 1989; Strijkers & Costa, 2011) and over 1500 ms for simple sentences in scene description tasks (Griffin & Bock, 2000; Schnur, Costa, & Caramazza, 2006), this brief interval between turns will often not allow speakers sufficient time to plan and initiate a response (Griffin, 2003). It therefore seems likely that next speakers prepare their response partly while the incoming turn is still unfolding. A model of turn-taking based on these observations has recently been formulated by Levinson and Torreira (2015). In this model, the listener as next speaker tries to anticipate the action carried out with the incoming turn (e.g. a request) early during the turn and begins to conceptualize and formulate a response as soon as the action becomes clear. Parallel to content planning and formulation, the next speaker (predictively) parses the input for possible points of syntactic closure and other cues to turn completion, while a formulated response may be temporarily held in a buffer. As the incoming turn is about to end, the next speaker prepares the articulators and initiates response. Hence, the model accounts for short gaps between turns by assuming that content planning starts as early as possible, comprehension continues in parallel with response preparation, and articulation can be launched from a prepared formulation when transition becomes relevant. Such parallel processing should be cognitively demanding,

since speaking and listening can interfere with one another and are known to take up processing resources (Boiteau et al., 2014; Kemper et al., 2003; Kubose et al., 2006; Schriefers et al., 1990; Sjerps & Meyer, 2015) and partly run on the same neurological system (Hagoort, Brown, & Osterhout, 1999; Kempen, Olsthoorn, & Sprenger, 2012; Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012a). Thus, speakers face the task of producing a response under time pressure while keeping capacity demands and interference between comprehension and production within reasonable bounds. In their parallel processing model, Pickering and Garrod (2013) propose that fluent turn-transitions are made possible by forward modeling of the incoming speech signal with the help of the addressee's own production system (see also Garrod and Pickering (2015)). In this account, the addressee is taken to covertly imitate the production of the incoming turn based on the input that has already been transmitted and thereby anticipate the content and timing of the incoming turn so as to be able to prepare a response in a timely fashion. Irrespective of whether or not the production system is used to imitate the incoming turn, early anticipation of the incoming turn's message and intended action would be a necessary pre-requisite for early response preparation.

Another task of next speakers is to detect when the incoming turn comes to an end and speaker transition becomes relevant. Sacks et al. (1974) hypothesized that listeners predict the end points of the incoming turns using syntactic and prosodic cues to turn closure (see also Ford & Thompson, 1996). They suggested that the projection of upcoming turn-completion points was essential for the close timing observed in conversation. Using experimental evidence for turn end estimation, de Ruiter et al. (2006) claimed that lexico-syntactic cues are essential for accurate projection of turn completion points, which, in their view, is a necessary prerequisite for response planning (see also Riest, Jorschick, & de Ruiter, 2015). Based on this assumption, de Ruiter et al. (2006) hypothesized that response turns could only be planned when the end point of the incoming turn can be accurately projected, meaning that a response could not be planned without knowing the duration of the rest of the incoming turn (Projection-Dependent Hypothesis). Contrary to this hypothesis, based on their quantitative analysis of conversational speech corpora, Heldner and Edlund (2010) claimed that at least about



40% of turn transitions could be explained without the assumption of turn-end projection.

The alternative to the hypothesized projection-dependent planning is that speakers begin to plan their utterance without knowing precisely when the current turn will end and, if necessary, postpone articulation until they detect a turn-completion point, as described in the model by [Levinson and Torreira \(2015\)](#) (Projection-Independent Hypothesis). On this account, the exact syntactic structure and words of the incoming turn do not need to be predicted for response planning to begin. Instead, merely the turn's message or intentions need to be known or anticipated, using the many contextual cues available from the organization of conversational sequences ([Schegloff, 2007](#)), common ground ([Clark, 1996](#)), or general knowledge about the speaker, the environment, and the world. As soon as speakers can anticipate the interlocutor's intention they can allocate some of their computational resources to their own planning processes ([Gisladottir et al., 2015](#)). Thus, if the interlocutor's message can be recognized or anticipated early during their turn, response planning, i.e. conceptualization and formulation, can begin early as well.

The present study tests the hypotheses that (a) response planning starts as early as the incoming turn's message can be anticipated, and (b) that the onset of response planning depends on an accurate projection of the incoming turn's completion point.

A small number of previous studies have set out to investigate when response planning in dialogue starts and whether a projection of the turn-end is necessary for response planning to begin. Their results are not fully consistent. [Magyari, de Ruiter, and Levinson \(2017\)](#) addressed both of these questions. The study investigated whether participants would start planning a response earlier during a question if the answer could be known early on versus only at the last word of the question. Visual displays were used that contained a tiger and a rabbit, each with or without further objects attached to them. Participants heard a question of the format *Which animal has object X and object Y?*, with the answer being available either already before the beginning of the question (early condition, with only one animal with objects) or only with the last object (late condition, with both animals with objects and only the last object being different between animals). Answers were faster in the early condition than in the late condition, suggesting that

response planning was not delayed until the end of the question. The second question was whether participants anticipated exactly when the question would end so as to be able to time their answer accurately to the end of the question. The lengths of the names of the objects were manipulated so that the length of the question could either be accurately projected (congruent condition, with the last objects of each of the animals having equally long names) or not (incongruent condition, with the last objects of the two animals having names of different lengths). No main effect of congruence was found, giving no support to the hypothesis that an accurate projection of a turn's completion point is necessary to plan a response.

Bögels, Magyari, and Levinson (2015) used EEG measurements to track the time course of comprehension and production processes in a quiz-like situation. Participants heard quiz questions to which the answer could be known either mid-sentence or only at the very end of the question, such as *Which character, also called 007 (critical word), appears in the famous movies?* (early condition) and *Which character from the famous movies is also called 007?* (late condition). At both the early and the late time points, they found significant positive deflections after 500 ms in questions containing the critical word (giving away the question) as compared to the respective questions that did not contain the critical word in that position. In a control experiment in which participants did not have to answer the questions but remember them, this effect was substantially reduced. The authors concluded that speech planning began as soon as all information needed to provide an answer was available.

Boiteau et al. (2014) investigated the cognitive load arising in different phases of a conversation using a dual-task paradigm. Participants continuously tracked a point on the screen with their computer mouse while freely talking to either a confederate or a friend. Tracking performance was worse during speaking than during listening and began to decline already about 250 to 450 ms before the end of a listening-turn. The authors concluded that speakers already began to plan their utterance while still listening to their interlocutor.

Sjerps and Meyer (2015) also investigated cognitive load during the temporal overlap between listening and planning using a dual-task paradigm. Participants continuously tapped their fingers in a predefined order while listening to a recorded description of a row of pictures

and subsequently described a second row of pictures before a time-out signal. Whether the recording referred to the top or bottom row varied randomly from trial to trial, but as soon as the participants heard the first noun, they knew which row was being described and could, in principle, prepare for the description of the other row. Nonetheless, both participants' eye-gaze and tapping performance indicated that planning began quite late, only shortly before, or at the very end of the recorded turn. These results do not support the view that speakers begin to plan their utterance as soon as they have understood the message of the incoming turn. Rather, the authors suggest, response planning began much later, perhaps to avoid interference between listening and planning. However, there are a number of reasons that call for caution when generalizing the observed timing of the relevant processes to everyday conversation. First, as there was no interlocutor present, the validity of generalizing the results to live interaction is unknown. Second, all turns, incoming and response, had the same syntactic structure and length. Consequently, the timing of the ends of incoming turns was highly predictable, and the beginnings of response turns could easily be held in working memory. Third, only forty objects were used in the item displays and they were reused twenty-one times, potentially influencing participants' planning strategies. Finally, even though participants only prioritized planning over listening towards the end of the recorded trial, they may have planned the beginnings of their responses already during the recorded utterance, looking at the target object for only a short period and then returning their gaze for comprehension. As the incoming turns were very long, such early looks may be distributed evenly across the incoming turn and are therefore difficult to detect.

To summarize, the reviewed studies came to different conclusions when investigating when next speakers begin to plan a response and whether they rely on projectable turn-completion points to initiate response planning. About the timing of planning, two possible hypotheses are proposed: Next speakers prioritize planning as soon as they have understood or can anticipate the message of the incoming turn, as put forward by [Bögels, Magyari, and Levinson \(2015\)](#) and incorporated in the model by [Levinson and Torreira \(2015\)](#) (Early Planning Hypothesis), or only when the incoming turn is coming to completion, as postulated by [Sjerps and Meyer \(2015\)](#) (Late Planning Hypothesis).

About the necessity of a precise projection of the incoming turn's completion point for response planning, two hypotheses are proposed: Next speakers depend on a projection of the incoming turn's end as proposed by [de Ruiter et al. \(2006\)](#) (Projection-Dependent Hypothesis) or they can start planning their response without an accurate projection, as modeled by [Levinson and Torreira \(2015\)](#) (Projection-Independent Hypothesis). The experiment described in this paper was designed to evaluate these hypotheses.

## 2.2 THE STUDY

The study presented here made use of a novel task-oriented dialogue paradigm, the list-completion paradigm. A female confederate and a participant jointly completed a task while sitting in separate sound proof booths in front of monitors and talking to one another without visual contact via microphones and headphones. Unbeknownst to the participant, most of the critical utterances of the confederate were pre-recorded prior to the experiment and played back by the confederate at the relevant moments during the experiment. In this way, the participant heard the utterances as being produced live and spontaneously by the confederate, fitting the conversational flow. A similar approach of combining live and pre-recorded playback modes was taken by [Bögels, Barr, Garrod, and Kessler \(2014\)](#).

On their screens, participants saw stimuli with differing numbers of objects (cf. [Figure 2.1](#) for an example). The confederate named the objects on her screen and the participant subsequently named all additional objects displayed on their screen. All speech was audio recorded. Moreover, participants' eye-movements were recorded. It was assumed that participants' gaze would follow the objects that are named by the confederate while comprehending the object names, and would move on to the objects that had to be named while planning the response turn ([Altmann & Kamide, 2007](#); [Griffin, 2001](#); [Griffin & Bock, 2000](#); [Huettig, Rommers, & Meyer, 2011](#); [Just & Carpenter, 1980](#); [M. K. Tanenhaus, Magnuson, Dahan, & Chambers, 2000](#)).

The experiment was conducted in German. The confederate's critical turns appeared in four conditions, differing in syntactic structure. The four conditions formed a  $2 \times 2$  design ([Table 2.1](#)). The first factor

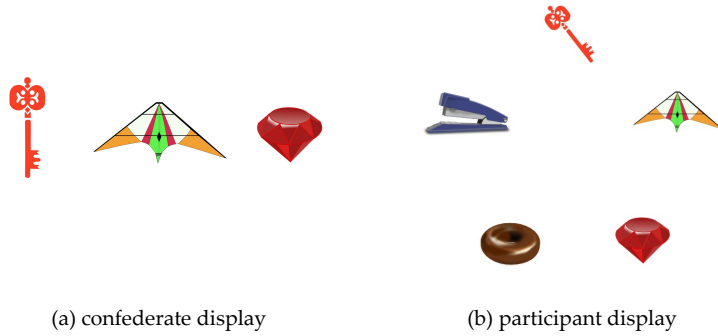


Figure 2.1: Example item displays.

was projectability of the turn ending ( $\pm$ Pend), meaning whether it was projectable or not how a turn would end (either in the last object name of the list or a turn final verb form). -Pend conditions contained the verb form *habe* ('have') in second position. In this position, *habe* was ambiguous as to whether it represents a main verb or an auxiliary requiring another verb form in sentence-final position, in this case *besorgt* ('gotten'). Both meanings of *habe* were used in the experiment. Therefore, sentences in the two -Pend conditions did not allow a precise projection of when they would end. Sentences in the +Pend conditions either contained the main verb *sehe* ('see') or the modal verb *kann* ('can'), which requires another verb form in sentence-final position, in this case *besorgen* ('get'). Therefore, sentences in the two +Pend conditions allowed for a precise projection of their completion point. The second factor was the presence or absence of a sentence-final verb ( $\pm$ Vend). Sentences in -Vend conditions ended right after the object list, whereas sentences in +Vend conditions ended after a sentence-final verb. While the number of objects named by the confederate varied from trial to trial, the last object noun was always preceded by *und* ('and') or, in the case of items with only one object being named, *nur* ('only'), providing a clear lexical cue to the end of the object list.

The timing of the participants' looks for planning and their response latencies were measured. For both measures, the contrasted hypotheses make different predictions. According to the Early Planning Hypothesis, participants should start planning as soon as they recognize the last object of the incoming list and should use the duration of a turn final verb to start planning their response. According to the Late Planning

Condition		Projectable ending or not	
		-Pend	+Pend
Verb position	-Vend	Ich habe einen Schlüssel, einen Lenkdrachen und einen Rubin.	Ich sehe einen Schlüssel, einen Lenkdrachen und einen Rubin.
	+Vend	Ich habe einen Schlüssel, einen Lenkdrachen und einen Rubin besorgt.	Ich kann einen Schlüssel, einen Lenkdrachen und einen Rubin besorgen.

Table 2.1: Example sentences of the four conditions used in the experiment. ‘I have/have gotten/see/can get a key, a kite, and a ruby.’

Hypothesis, however, participants should start planning only when the turn-completion point is reached and would not gain extra planning time in turns with a final verb form.

Eye-movements were analyzed using growth curve modeling (Mirman, 2014), a variety of mixed effects regression that makes use of polynomial time terms as predictors to model differences in fixation likelihoods. Linear, quadratic and cubic time terms were included. The linear time term (Time) models the overall increase in fixations over the time course of a trial. The quadratic time term (Time<sup>2</sup>) models the steepness of the curve, i.e. how “U-shaped” it is. The cubic time term (Time<sup>3</sup>) describes whether fixations increase earlier or later (“S-shaped” curve).

The Early Planning Hypothesis predicts no difference in the moment in time at which participants shift their gaze for planning, measured from the beginning of the turn. In terms of the analyses applied, this is a prediction of null effects of Time<sup>3</sup> × ±Vend. It further predicts a main effect of ±Vend in response latencies, with faster responses after turns with a final verb form (+Vend) than after turns without a final verb form (-Vend), because participants should gain extra planning time at the end of the incoming turn if it ends in a final verb form. The Late Planning Hypothesis predicts participants to shift their gaze for planning later in turns with a turn-final verb form (+Vend) than in turns without (-Vend), which would manifest as an effect of Time<sup>3</sup> × ±Vend. It further predicts a null effect of ±Vend in response latencies because no extra time for planning should be gained in turns with a final verb form.

According to the Projection-Dependent Hypothesis, participants should start planning as soon as they recognize the last object of the incoming list after turns with projectable endings (+Pend), whereas after turns with unprojectable endings (-Pend) they should start planning only upon recognizing whether a turn final verb form follows the object list or not (i.e. only when they can project when exactly the turn will come to an end). According to the Projection-Independent Hypothesis however, participants should in all conditions start planning as soon as they recognized the last object of the incoming list.

Consequently, the Projection-Dependent Hypothesis predicts participants to shift their gaze for planning earlier (measured from the beginning of the turn) in turns with projectable endings (+Pend) than in turns with unprojectable endings (-Pend), which would manifest as an effect of  $\text{Time}^3 \times \pm\text{Pend}$ . It further predicts a main effect of  $\pm\text{Pend}$  in response latencies, with faster responses after projectable turns than after unprojectable turns, since participants could start planning earlier before the end of the turn when its completion point was projectable. The Projection-Independent Hypothesis predicts no difference in the moment in time at which participants shift their gaze for planning, which would manifest as null effects of  $\text{Time}^3 \times \pm\text{Pend}$ . It further predicts a null effect of  $\pm\text{Pend}$  in response latencies.

The timing pattern of response planning as modeled by [Levinson and Torreira \(2015\)](#) results in overlap of comprehension and production processes at the junction of turns, where planning already begins while the incoming turn is not yet complete, as predicted during turn-final verbs in the present study. The studies reviewed above repeatedly found interference effects of incoming speech on response planning ([Bögels, Magyari, & Levinson, 2015](#); [Boiteau et al., 2014](#); [Kemper et al., 2003](#); [Schriefers et al., 1990](#)). Consequently, planning during the turn-final verbs would be hypothesized to be less efficient than planning during silence. This difference in efficiency should manifest as an effect of  $\text{Time}^2 \times \pm\text{Vend}$ , with proportions of looks for planning increasing more slowly in turns with a final verb form than in turns without a final verb form. Furthermore, this difference could be modulated by the projectability of the turn-final verb, since incoming words might be less detrimental to response planning when they can be projected than when they cannot. This influence of projectability of the final verb form should manifest as an effect of  $\text{Time}^2 \times \pm\text{Pend}$  in turns with a

final verb. Both hypotheses about the influence of verb finality and projectability on the efficiency of response planning will be tested in the present study.

## 2.3 METHODS AND MATERIALS

### 2.3.1 *Participants*

Forty-eight German native speakers (30 female) were tested as paid participants at Heinrich-Heine University, Düsseldorf, Germany. All participants reported to have normal or corrected-to-normal vision and normal hearing abilities. Eight participants stated in a questionnaire filled in after the experiment that they noticed the presence of pre-recorded materials. These participants were excluded from the analyses. Two participants were excluded due to technical failures of recording equipment, leaving 38 participants for analysis. Remaining participants had a mean age of 26.3 years ( $SD = 7.6$ ). The experiment was approved by the Ethics Committee of the Faculty of Social Sciences, Radboud University Nijmegen. Informed consent was obtained from all subjects.

### 2.3.2 *Apparatus*

The participant and the confederate were seated in separate cabins about 60 cm away from 21" computer screens. They were unable to see each other and could only communicate via microphones and headphones. The participants' eye-movements were recorded with an SMI RED-m remote eye-tracker (120 Hz sampling rate).

### 2.3.3 *Visual stimuli*

Four-hundred and sixty-eight pictures of objects were used in the experiment. The pictures were sourced online and are under the creative commons license. They were selected to be easy to recognize and name. All pictures, with the exception of twenty pictures used in practice trials, showed inanimate objects.



One-hundred and seventeen pairs of item displays (participant displays and corresponding confederate displays) that showed a differing number of objects drawn from the pool of object pictures were used as visual stimuli (see Figure 2.1 for an example). The participant displays showed between three and five objects. These objects included all objects shown on the corresponding confederate display and zero, one, two, or three further objects. In participant displays that showed three objects, the objects formed an equilateral triangle, when showing four objects, the objects formed a square, when showing five objects, the objects formed an equilateral pentagon. Objects on the displays filled approximately two degrees of visual angle. They had equal distances of about four cm to their neighbors, irrespective of the arrangement they were presented in on the display. That means that to see the individual objects sharply, participants had to move their eyes to focus on them. The most common names of the objects of a display did not start with the same phoneme. Names of objects that were named by the participants had a mid-range frequency. Names of objects that were named by the confederate were sampled from wider frequency ranges (based on German Wortschatz Corpus, [Department of Computer Science, Leipzig University, 2016](#)).

Ninety-six displays were critical test displays, with thirty-two displays each showing three, four, or five objects on the participant display. The confederate displays showed between zero and five objects, so that twenty-four participant displays showed no more objects than the corresponding confederate display, twenty-four participant displays showed one more object, twenty-four participant displays showed two more objects, and twenty-four participant displays showed three more objects.

In the test phase, nine pairs of displays were used as displays for live items (see Auditory stimuli below). Three participant displays in this group of items showed three objects, three showed four objects, and three showed five objects. The confederate displays in this group of items showed between zero and four objects, so that three of the corresponding participant displays showed one more object than the confederate display, three showed two more objects, and three showed three more objects.

The experiment was preceded by a practice phase using twelve display pairs.

### 2.3.4 Auditory stimuli

Sentences accompanying ninety-six of the visual displays were pre-recorded in the same sound protected booth that was used for the experiment, using a unidirectional Sennheiser ME64 microphone attached to a digital flash recorder. Each sentence was recorded in the four conditions exemplified in Table 2.1. When the sentence contained two or more object nouns, the last noun was preceded by *und* ('and'). When it contained only one object noun it was preceded by *nur* ('only'). When it did not contain any object nouns, the object list was replaced by *nichts* ('nothing'), as in *Ich habe nichts (besorgt)*. ('I have (gotten) nothing').

Due to the structures of the sentences, their duration is confounded with the experimental conditions, since the turn-final verb forms in the +Vend conditions are about 600 ms long and there is no word coming after the list of objects in sentences in the -Vend conditions. Therefore, sentence length will be controlled for in the statistical analyses.

The pauses between object nouns were adjusted for the different versions of each sentence with Praat (Boersma & Weenink, 2015) to have random lengths between 400 and 600 ms, imitating the gaps in the original recordings. None of the list contours of the pre-recorded stimuli used in the experiment contained downsteps on non-final items (cf. Selting (2007)) and all sentences ended in a low boundary tone (cf. von Essen (1956)).

Sentences accompanying nine visual displays were not pre-recorded but produced live by the confederate during the experiment (+Live items). The sentences accompanying the twelve practice trials were also produced live. These sentences were produced so as to sound similar to the pre-recorded sentences, using the same verbs and syntactic structures that were used in the pre-recorded sentences. They were included to test for the comparability of participant's response timings after live and pre-recorded stimuli ( $\pm$ Live) so as to validate the assumption that responses after pre-recorded stimuli were given naturally.

### 2.3.5 Items and Design

A participant display in combination with the accompanying sentence constituted an experimental item (see Table 2.13 in Supplementary

Materials for a list of materials). In two thirds of the items in which the confederate named at least one object, the objects were arranged in clockwise order as they were named, starting at the top of the display. In one third of the items, including all +Live items, other arrangements were used, so that the participants had to listen attentively and search for the items mentioned by the participant, rather than scanning the objects in the same order on all trials. Analyses controlled for this order-of-objects variable.

Four lists were constructed, with each sentence and the accompanying display appearing once per list and appearing in a different condition in each of the lists. In each list the same number of items appeared in each condition. Each participant was assigned to one of the lists.

### 2.3.6 Procedure

#### *Familiarization and Instructions*

Participants were invited to the lab to take part in a dialogue experiment. They were the first to enter the lab and told that the other participant of the study would arrive in a few minutes. In the meantime, participants were given a picture booklet containing all pictures used in the experiment and asked to name them. In 1.4% of all cases and in 0.9% of the cases involving pictures to be named by participants, the pictures were not recognized or labeled by participants, and a name was provided by the experimenter. The experimenter recorded participants' responses. The familiarization phase was audio-recorded.

After the familiarization phase, the confederate arrived and was introduced as a second participant. Participant and confederate were informed that they would be seated in separate cabins and talk to each other via headphones and microphones to play the following game. They would see a number of displays on their respective screens, showing things they could get. The confederate was to tell the participant which things she has got already, so that the participant could tell the confederate what *further* objects (s)he could get. Participants were not instructed to use any particular utterance format.

The confederate was instructed to try to remember which objects she had seen and which names she had heard. This served as a cover task to

distract participants from the aim of the study. Participants were told that their eye-movements would be recorded in order to study looking behavior when searching for objects on a screen whose names were heard. After instructions were given, the eye-tracker was calibrated. Calibration was repeated three times during the experiment.

### *Test phase*

Before the beginning of the test phase, participants completed twelve practice trials, where instructions were repeated if necessary. During the test phase, all communication between the participants and the confederate was live, except for ninety-six pre-recorded sentences accompanying the critical displays. The confederate started the presentation of the stimulus displays and the corresponding pre-recorded utterances so as to make them fit naturally into the conversation. Similarly, she produced the sentences accompanying the nine +Live items naturally in the flow of the conversation.

Participants were asked to look at a fixation cross that was presented in the center of the display at the beginning of each trial, which triggered the presentation of the item displays. After a preview of 600 to 1000 ms, the stimulus sentence began. Preview times varied randomly between items.

The experiment took about thirty minutes. After the experiment, participants were asked in a computerized questionnaire whether they had noticed the presence of pre-recorded speech. The entire test session took about seventy minutes, including familiarization, test phase and questionnaire.

## 2.4 RESULTS

Participants' fixation preferences and response latencies were the dependent variables. Statistical analyses are based on linear mixed effects regression models fitted in R (R Core Team, 2014) using the package lme4 (Bates, Maechler, Bolker, & Walker, 2014). The maximal random effects structure justified by design was used for all models (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013). Control variables were not included in the random effects structure. All categorical variables were deviation coded (-0.5 and 0.5). Statistical significance was assessed

with  $F$ -tests with Kenward-Roger approximations of degrees of freedom (Fox & Weisberg, 2011; Halekoh & Hojsgaard, 2014; Kenward & Roger, 1997). We report all data exclusions, all manipulations, and all measures in the study.

#### 2.4.1 Response timing

Response latencies for the 3980 critical turn transitions were measured manually with Praat (Boersma & Weenink, 2015). They were coded as time intervals between the end of the incoming turn and the beginning of the response turn, excluding any non-speech sounds like audible in-breaths. Trials were coded with respect to the verb structure produced by the participants in the critical responses. When participants used the same verbs as in one of the four stimuli conditions (*habe*, *habe besorgt*, *sehe*, *kann besorgen*), trials were coded parallel to the conditions ( $\pm$ Pend;  $\pm$ Vend). All other response structures were coded as ‘other’. Response structure was used as a control variable in the mixed effects regression to control for any differences in response time that are due to the structure of the response turn rather than the structure of the incoming turn. Forty-nine percent of response structures were congruent to the structure of the corresponding confederate turn. Therefore, structural congruency (henceforth  $\pm$ Priming) was included as a control variable in the analyses to control for any priming effects on the dependent variables, since responses repeating the structure of the previous turn might have been produced faster by the participants.

Twenty-four trials were discarded either because participants did not only name the correct objects or due to technical failure. Response latencies ranged from -211 ms to 3132 ms ( $M_{RL} = 806$  ms,  $SD_{RL} = 370$  ms,  $N_{RL} = 3956$ , Table 2.2).

Condition	Format		Mean (SE) in ms
	Pend	Vend	
<i>habe</i>	-	-	842 (11)
<i>habe ... besorgt</i>	-	+	749 (11)
<i>sehe</i>	+	-	867 (12)
<i>kann ... besorgen</i>	+	+	761 (11)

Table 2.2: Response latencies by condition.

For the statistical analyses, thirty-five data-points (1%) were removed from the data set since they were outliers of more than three standard deviations of the mean response latency of the respective subject that produced the data-point.

Turns in conditions with a turn-final verb (+Vend) were longer than corresponding turns in conditions without a turn-final verb (-Vend) due to the presence or absence of a sentence-final verb. Turn length might affect response production processes. Magyari et al. (2017) found participants to answer questions faster the longer the question, irrespective of the content of the question or when the answer could be known. Magyari et al. propose that next speakers' level of preparedness to speak increases as the likelihood that the incoming turn will come to an end increases as the turn unfolds. Therefore, the duration of the critical turns was included as a control variable in the analyses.

To test whether the response latencies after pre-recorded items were the same as after live items, a model was fitted with playback mode as predictor ( $\pm$ Live). The duration of the confederate turns, as well as  $\pm$ Priming, and a binary order-of-objects variable were included as control variables. Playback mode did not influence response latencies ( $\beta = 22$ ,  $SE = 41$ ,  $F(1,15) = 0.30$ ,  $p = 0.58$ ). Hence, data gained with pre-recorded items were regarded as ecologically valid and the following analyses are restricted to these items.

To evaluate the contrasting hypotheses formulated above, Early vs. Late Planning Hypothesis and Projection-Dependent vs. -Independent Hypothesis, a model was fitted to predict response latencies after pre-recorded turns, with  $\pm$ Vend and  $\pm$ Pend as well as their interaction and the duration of the confederate turn as predictors. The syntactic structure of the responses, as well as  $\pm$ Priming were included as control variables. The model revealed a significant main effect of  $\pm$ Vend ( $\beta = 85$ ,  $SE = 15$ ,  $F(1,47) = 37.30$ ,  $p < .001$ ), i.e. participants responded faster after turns that contained a final verb than after turns that did not end in a verb. Projectability did not significantly influence response latencies, nor did the interaction of projectability and verb position, meaning that response latencies were not modulated by the projectability of a turn's ending. Response latencies were significantly shorter with increasing durations of the incoming turns ( $\beta = 17$ ,  $SE = 6.55$ ,  $F(1,77) = 6.91$ ,  $p = .010$ ). This supports the finding by Magyari et al. (2017) that readiness to speak increases with increasing turn length. See Table 2.3 for a model

summary. The analysis was repeated with the duration measured from the end of the confederate's turn to the beginning of the first object noun of the participant's turn (instead of the turn's beginning) as the dependent variable, yielding the same general pattern of results. In sum, the results support the Early Planning Hypothesis and the Projection Independent Hypothesis.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	851.205	36.8	23.121		
Vend	-92.002	14.9	-6.172	42.62(1,46)	***
Pend	23.598	16.5	1.430	2.00(1,60)	n.s.
Vend_structure	-11.954	15.7	-0.760	0.52(1,727)	n.s.
Pend_structure	0.089	16.6	0.005	0.00(1,606)	n.s.
priming	-32.381	12.9	-2.494	5.42(1,461)	*
sentence_dur_cent	-17.151	6.4	-2.642	6.50(1,76)	*
Vend:Pend	16.140	27.4	0.587	0.33(1,33)	n.s.

Table 2.3: Response timing model and *F*-tests. Formula:  $RT \sim 1 + Vend * Pend + Vend\_structure + Pend\_structure + structure.primed + sentence\_duration\_centred + (1 + sentence\_duration\_centred + Pend * Vend | subject) + (1 + Pend * Vend | item)$ . Asterisks indicate significance levels of effects. \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

#### 2.4.2 Eye-movements

In order to explore the time course of participants' comprehension of the confederate's turn and the planning of their own response turn, fixations to the first-mentioned objects in the participants' responses were analyzed. Fixations towards an area of interest covering the first-named objects (target objects) and approximately 0.25 degrees of visual angle around them were categorized as target fixations. Figure 2.2 shows the proportions of target fixations time-locked to the beginning of the last noun in the confederate's utterance. Figure 2.3 shows proportions of looks to target objects time-locked to the offset of the incoming turn.

Participants' eye-movements were analyzed in a time window from 0 ms until 2800 ms, corresponding to the beginning of the last noun in the confederate's turn (0 ms) and the grand mean duration from the time-lock point until the beginning of the first object noun in the participant turn (2800 ms) respectively. Fixations to the target objects were aggregated to empirical logits in 100 ms time bins over

the course of the analysis window by subjects and by items, respectively. This aggregation procedure removes non-independences in the eye movement data that arise from the way how eye movements are planned and executed (Barr, 2008). Where a participant looks at one point in time is highly dependent on where she was looking at the immediately preceding time point, as “[i]t is not physically possible for a participant’s eye gaze to instantaneously travel from one region to another; the gaze must travel through time and space to reach its destination” (Barr, 2008, p. 464). Aggregating all observations from each subject or item for each condition into time bins and applying empirical logit transformation effectively accounts for the problem of non-independent observations. Only trials that included both looks for production and looks for comprehension were analyzed, excluding trials in which the confederate named none or all of the displayed objects. Ninety-two of the remaining trials were discarded due to trackloss, i.e. missing data for a consecutive stretch longer than 500 ms within the time window of analysis. The final dataset included 2124 trials.

Eye-movements were analyzed using quasi-logistic growth curve modeling (Mirman, 2014; Mirman, Dixon, & Magnuson, 2008), a variety of mixed effects regression that makes use of polynomial time terms as predictors to model differences in fixation likelihoods. Linear, quadratic and cubic orthogonal time terms were included as predictors. The linear time term (Time) models the overall increase in fixations over the time course of a trial. The quadratic time term (Time<sup>2</sup>) models the steepness of the curve, i.e. how “U-shaped” it is. The cubic time term (Time<sup>3</sup>) describes whether fixations increase earlier or later (“S-shaped” curve).

Visual inspection of the proportion of fixations indicates that target fixations started to slowly increase about half a second before the onset of the last object noun in the confederate’s turn, probably because the set of candidate objects that needed to be named got smaller as the incoming turn unfolded (see Figure 2.2). The increase of fixations accelerated at about 400 ms after the onset of the last object noun in all conditions, meaning that participants moved their gaze for planning at about the same time for all turns, irrespective of their syntactic structure. From that point in time, it seems that fixations increased and decreased faster in conditions with a sentence-final verb form than



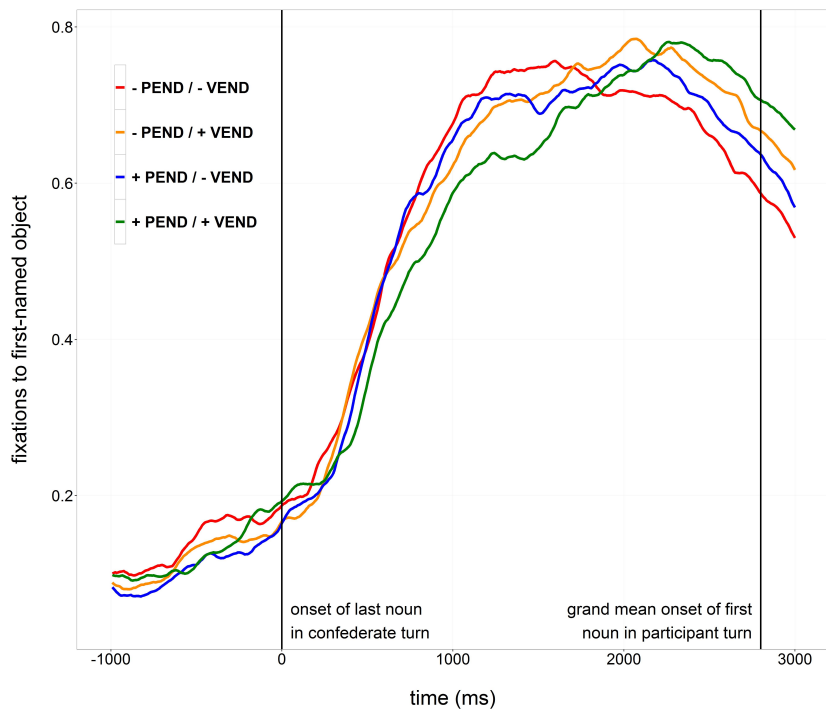


Figure 2.2: Proportions of looks to the object named first by the participant time-locked to the onset of the last object noun of the confederate turn (0 ms).

in items without a sentence-final verb form. In conditions without a sentence-final verb form, fixations appear to develop for the most part in parallel, irrespective of the projectability of the turn's ending. In conditions with a sentence-final verb however, fixations seem to differ from one another dependent on the projectability of the sentence final verb. In the condition with non-projectable sentence-final verbs (-Pend/+Vend), proportions appear to increase and decrease faster than in the condition with projectable sentence-final verbs (+Pend/+Vend).

Two pairs of conditions were compared to test for effects of verb position: trials with a projectable turn ending that contained a final verb form were compared with trials with a projectable turn ending that did not contain a final verb form (+Pend/+Vend vs. +Pend/-Vend, i.e. *kann...besorgen* vs. *sehe*); and trials with an unprojectable turn ending that contained a final verb form were compared with trials with an unprojectable turn ending that did not contain a final verb form (-Pend/+Vend vs. -Pend/-Vend, i.e. *habe...besorgt* vs. *habe*). Similarly, two

pairs of conditions were compared to test for effects of projectability: trials that projectably ended in a turn-final verb were compared with trials that unprojectably ended in a turn-final verb (+Pend/+Vend vs. -Pend/+Vend, i.e. *kann...besorgen* vs. *habe...besorgt*); and trials that projectably ended after the last object noun were compared with trials that unprojectably ended after the last object noun (+Pend/-Vend vs. -Pend/-Vend, i.e. *sehe* vs. *habe*). For each comparison by-subject and a by-item analyses were conducted.

In each test, the interactions of Condition with the cubic time term ( $\text{Time}^3$ ) and the quadratic time term ( $\text{Time}^2$ ) were of most theoretical interest, as they model the hypotheses about latency and speed of the increases of proportions of target looks in the different conditions. The linear time term (Time) itself does not directly relate to the hypotheses, as it only models a linear trend in increases of the proportions of target looks, which is expected to occur in all conditions as the task to name the remaining objects requires participants to look at the target object in all conditions. An interaction effect between Condition and  $\text{Time}^3$  would indicate a difference in the latency of the increase of target fixations between conditions. An interaction effect of Condition and  $\text{Time}^2$  would indicate a difference in the steepness of the increase of target fixations between conditions. Table 2.4 shows an overview of the interactions in question and their statistical significance and Tables 2.5 to 2.12 show summaries of the models and respective *F*-tests.

Comparison	Effect	$\beta$	SE	<i>F</i>	sig.
-Pend/-Vend vs. -Pend/+Vend	$t^2 \times \text{cond.}$	0.52	0.23	$F(1,727)=4.64$	*
+Pend/-Vend vs. +Pend/+Vend	$t^2 \times \text{cond.}$	0.93	0.23	$F(1,735)=15.21$	***
-Pend/-Vend vs. +Pend/-Vend	$t^3 \times \text{cond.}$	-0.06	0.24	$F(1,721)=0.06$	n.s.
-Pend/+Vend vs. +Pend/+Vend	$t^3 \times \text{cond.}$	-0.37	0.28	$F(1,393)=1.55$	n.s.
-Pend/-Vend vs. +Pend/+Vend	$t^2 \times \text{cond.}$	0.32	0.25	$F(1,651)=1.54$	n.s.
-Pend/+Vend vs. +Pend/+Vend	$t^3 \times \text{cond.}$	0.06	0.26	$F(1,554)=0.05$	n.s.
-Pend/+Vend vs. +Pend/+Vend	$t^2 \times \text{cond.}$	0.71	0.21	$F(1,869)=10.89$	***
+Pend/+Vend	$t^3 \times \text{cond.}$	-0.23	0.23	$F(1,843)=1.00$	n.s.

Table 2.4: Eye-movement results of by-subject analysis. Pairwise comparisons of  $\text{Time}^2 \times \text{Condition}$  and  $\text{Time}^3 \times \text{Condition}$  effects in growth curve analyses.  $t^2 = \text{Time}^2$ ,  $t^3 = \text{Time}^3$ . Asterisks indicate significance levels of effects. \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ . By-item analysis yielded similar pattern of results.

Throughout the pairwise comparisons, no interaction effect of Condition  $\times$  Time<sup>3</sup> reached significance, with the single exception of the by-item comparison of +Pend/-Vend trials vs. +Pend/+Vend trials, indicating that the proportions of target looks start to increase at the same point in time in all conditions.

All four comparisons testing for the effects of verb position showed a significant interaction effect of Condition  $\times$  Time<sup>2</sup>, indicating steeper increases and decreases of target fixations in trials without sentence-final verbs as compared to trials with sentence-final verbs, irrespective of whether the turn's endings were projectable or not.

Neither the by-subject, nor the by-item comparison of -Pend/-Vend trials with +Pend/-Vend trials showed an interaction effect of Condition  $\times$  Time<sup>2</sup>, meaning that target fixations increased in the same way in trials without a final verb form, no matter whether the turns' endings were projectable or not. However, both the by-subject and the by-item comparison of -Pend/+Vend trials with +Pend/+Vend trials showed an interaction effect of Condition  $\times$  Time<sup>2</sup> in the direction of target fixations increasing more slowly when the final verb was projectable than when it was not.

Because the finding that participants started gazing at the target object at the same time in all four conditions is based on null effects in the growth curve analyses, breakpoint analyses were conducted for each condition (Baayen, 2008). Breakpoint analysis is based on regression modeling and seeks to identify discontinuities in linear relations, i.e. changes in slope. To identify when participants started to fixate on the target object, a search for breakpoints in target fixations was conducted in a time window between 200 ms after the onset of the last noun in the confederate turn and the grand mean beginning of the participant turn (900 ms) in steps of 100 ms. In the by-subject analyses, breakpoints are located around 400 ms after the onset of the last object noun for all conditions. The by-item analyses yielded a similar pattern of results (cond. 1: 500 ms, cond. 2.: 400 ms, cond. 3.: 400 ms, cond. 4: 300 ms, all conditions together: 400 ms). These results confirm that, irrespective of the incoming turn's structure, participants moved their gaze towards the target object as soon as the last object noun became recognizable, assuming that planning and executing a saccade takes about 200 ms (Allopenna, Magnuson, & Tanenhaus, 1998).

## 2.5 DISCUSSION

This study investigated how speakers coordinate listening and speech planning in a dialogue situation. We contrasted two hypotheses: The Late Planning Hypothesis, as formulated by [Sjerps and Meyer \(2015\)](#), stating that next speakers would start planning their response only at the end of the incoming turn, and the Early Planning Hypothesis, as included in the turn-taking model of [Levinson and Torreira \(2015\)](#), stating that next speakers would start planning as soon as all information that is needed to know what to respond is available. Furthermore, we investigated whether the timing of response planning relies on a projection of the incoming turn's completion point. Again, we contrasted two hypotheses: The Projection-Dependent Hypothesis, as formulated by [de Ruiter et al. \(2006\)](#), stating that next speakers depend on an accurate projection of the incoming turn's completion point to be able to begin planning their response, and the Projection-Independent Hypothesis, as proposed by [Levinson and Torreira \(2015\)](#), stating that planning can begin without an accurate projection of when the incoming turn will end.

To evaluate these hypotheses, an experiment was conducted that made use of the list-completion paradigm, a novel turn-taking paradigm that included two interlocutors, a confederate and a naive participant. The two participants engaged in a cooperative dialogue task that included naming objects on their screens. Which objects participants had to name depended on which objects were named by the confederate. Their conversation was recorded for an analysis of turn transition times and the subject's eye-movements were recorded for analyses of their gazes for comprehension versus gazes for response planning.

Notably, the list-completion paradigm used both live and pre-recorded speech and thereby created a natural dialogue situation that allowed for tight control of critical utterances. The production task was highly naturalistic and resembled a conversational situation, as participants were not restricted to use a limited set of syntactic structures in their responses. The timing of responses was the same for pre-recorded sentences and sentences produced live. The data collected in this study can therefore be regarded as comparable to live situations, especially

with respect to the fact that participants that were analyzed stated that they did not notice the presence of pre-recorded materials.

Participants were found to start planning their responses as soon as they knew which objects they had to name, gazing towards the objects they named in their responses as soon as the last object noun of the incoming turns could be recognized. As a consequence, participants spent more time planning during the incoming turn when it contained a turn-final verb than when it ended with the last object noun, which led to faster responses after turns with a turn-final verb compared to turns without a turn-final verb. These results support the Early Planning Hypothesis over the Late Planning Hypothesis. They are in line with the model by [Levinson and Torreira \(2015\)](#) and with the findings of [Bögels, Magyari, and Levinson \(2015\)](#), who found that when participants had to answer quiz questions, they started planning their responses as soon as the questions could be understood, no matter if that point in time was in the middle or at the end of the question. They are also in line with the findings by [Magyari et al. \(2017\)](#), who found that participants reacted faster to questions about objects on the screen when the answers to the questions could be known longer before the ends of the questions. This advantage of early planning may be an important factor in keeping inter-turn gaps short in conversation.

On the other hand, the results appear to be at odds with the results obtained by [Sjerps and Meyer \(2015\)](#), who found that participants did not start planning until right before or at the end of the incoming turn when taking turns with a computer in naming rows of four pictures. In that study, participants could, in principle, begin to plan their utterance as soon as they had identified the first noun of the incoming turn, but were found to initiate planning only when they heard the final noun. In both the present study and the study by Sjerps and Meyer, the measurement of utterance planning was time-locked to the last noun of the incoming turn. In Sjerps and Meyer's study, utterance planning could have been initiated much earlier but apparently participants opted for a late planning strategy. In contrast, in the present study, planning could not have been initiated any earlier but it could have been initiated later in cases where the incoming turn ended in a verb. However, participants apparently opted for an early planning strategy.

In these two studies, participants were in different communicative situations. While in Sjerps and Meyer's study no interlocutor was

present, in the present study participants interacted with another person in a joint task, which might have encouraged them to plan their utterances as soon as all relevant information was available rather than awaiting the end of the turn. Another difference lies in the structures of the utterances heard and produced. Conceptually and linguistically, the task used by Sjerps and Meyer was undoubtedly easier and more constrained than the task used in the present study. Given the simple nature of the planning task in the study by Sjerps and Meyer, participants could afford to postpone utterance planning until the preceding turn was completed. It seems that if next speakers consider the gain of early planning to be low, they can opt for late planning, as in Sjerps and Meyer's study. If, however, next speakers are under pressure to respond in a timely fashion, as they are in a conversational setting (Sacks et al., 1974), they can opt for early planning, resources permitting. The latter situation is arguably more frequent in everyday conversation, where planning might even start based on an anticipation of the incoming turn's message in order to keep inter-turn gaps short. The onset of planning might therefore depend on the information density at the end of the incoming turn (Jaeger, 2006, 2010), which was much higher in the present study than in the study by Sjerps and Meyer. In the present study, the incoming turn contained task-relevant information either until the last word, when the incoming turn ended in a noun, or until the last but one word, when a turn final verb was present. In the sentences used in the study by Sjerps and Meyer, on the other hand, only the first of four nouns was critical for the task, so that the last nine words of each presented sentence were irrelevant for the participants to follow their instructions.

While participants were found to start planning their responses before the end of the incoming turns, this planning during incoming speech was associated with additional processing costs. This conclusion results from two findings. First, proportions of looks for planning increased faster in turns not containing a sentence-final verb than in turns that ended in a verb. And second, even though response planning was already initiated before a sentence-final verb would be heard, response latencies after verb-final turns were shorter than after turns without a final verb by only a fraction of the length of the sentence-final verb. The reduction of the difference in response latencies might, at least partly, arise from interference of the turn-final material with

response planning, rendering planning less efficient during turn-final verbs than during silence. When planning during the incoming turn, next speakers still need to parse the input and predict or detect the upcoming completion point. When planning in silence, there is no such extra effort, making response planning more efficient.

The projectability of the incoming turn's completion point, which was manipulated by using different verbs in second position (ambiguous *habe* ('have') or unambiguous *sehe* ('see') or *kann* ('can')), did not modulate response latencies, which supports the Projection-Independent Hypothesis over the Projection-Dependent Hypothesis, as predicted from the model by Levinson and Torreira (2015). The results illustrate that response planning can be initiated without an exact projection of further upcoming material or the exact locus of the turn end. However, the conjunction *und* ('and') or *nur* ('only') preceded the final noun in all of the confederate's utterances, giving a cue that the turn would end after either one or two additional words. Thus, coarse projection of the turn-completion point was always possible. Accurate projection of the turn-completion point was found to be unnecessary for response planning.

However, projectability was found to influence looking behaviour when sentences contained turn-final verbs. The influence was in the opposite direction as expected, with the proportion of looks for planning increasing more slowly in turns where a final verb was projectable than in turns in which the final verb was not projectable. This difference in looking behaviour did not lead to a difference in response latencies, however, and therefore cannot be interpreted as a difference in planning difficulty. It could rather be a manifestation of a specific planning strategy, as participants seem to distribute their planning effort more evenly over time when they are presented with turns that projectably allow them to take extra time for planning at the end of the incoming turn. They may do this by planning their response early conceptually, returning their gaze for comprehension, and finally look for planning again to formulate and articulate the target object's name. With such a strategy, next speakers could avoid inefficiencies in planning due to interference of incoming speech and thereby reduce cognitive effort.

Taken together, the results suggest that the timeline of the processes involved in taking turns in a conversation seems to be far from ballistic. Contrary to classical monologic tasks commonly used in psycholin-

guistic studies, conversational situations are more complex and allow for more variability in the succession of the different aspects of language processing, especially regarding the interplay of comprehension and production planning. Cognitive resources seem to be distributed depending on the needs and possibilities of different conversational situations and may well be influenced by interlocutors' decisions and preferences. Since conversation can be regarded as the core ecology of language, this variability deserves more attention in future psycholinguistic research, calling for further studies concerning the psychology of dialogue in order to understand (the limits of) the involved flexibility, which is responsible for the general tendencies in turn-taking behaviour as well as the observable deviations from them.

## 2.6 CONCLUSION

In this experiment, participants started to plan their responses as early as possible. Starting to plan a response during the incoming turn is costly, but leads to efficient timing of turn-taking and might be a key factor to keep gaps between turns short in conversation. Early planning does not depend on accurate projection of the incoming turn's completion point. The results support turn-taking models that include early response planning (Heldner & Edlund, 2010; Levinson, 2012; Levinson & Torreira, 2015; Sacks et al., 1974).

## 2.7 SUPPLEMENTARY MATERIALS

### *Eye-movements time-locked to the end of turn 1*

As can be seen in Figure 2.3, proportions of target looks start to increase earlier in +Vend conditions than in -Vend conditions, namely one second versus about half a second before the offset of the incoming turn, respectively. Proportions of target looks in -Vend conditions seem to develop in parallel. In +Vend items however, proportions seem to differ from one another. In items with a non-projectable sentence-final verb (-Pend/+Vend), the increase of proportions appears to be steeper than in items with a projectable sentence-final verb (+Pend/+Vend).



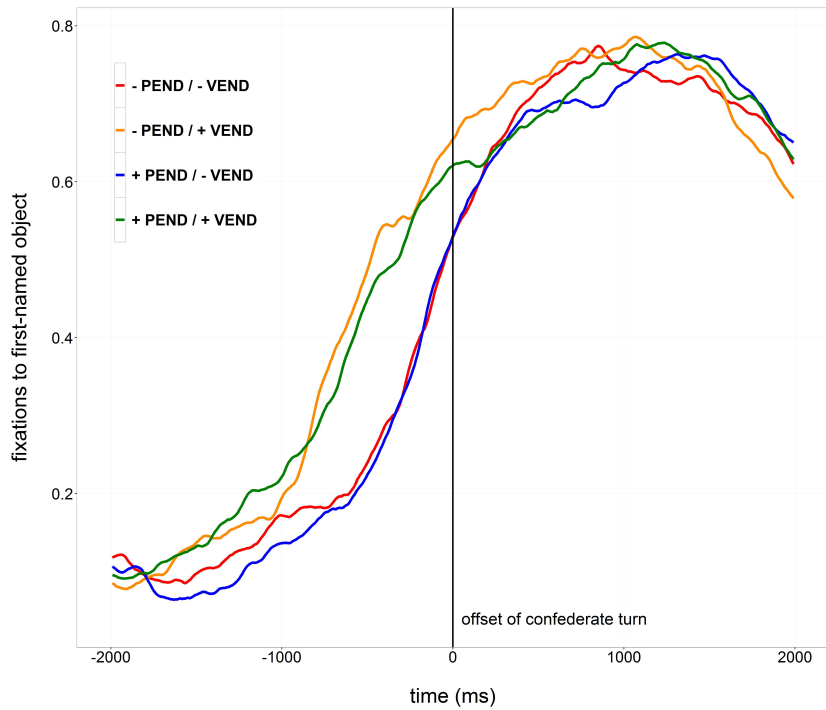


Figure 2.3: Proportions of looks to the target object time-locked to the offset of the confederate turn (0 ms).

### *Eye-movement statistics*

Formula for all comparisons:  $\text{emplogit} \sim 1 + (\text{time} + \text{time}^2 + \text{time}^3) * \text{condition} + (1 + (\text{time} + \text{time}^2 + \text{time}^3) * \text{condition} | \text{subject})$  or:  $\text{emplogit} \sim 1 + (\text{time} + \text{time}^2 + \text{time}^3) * \text{condition} + (1 + (\text{time} + \text{time}^2 + \text{time}^3) * \text{condition} | \text{item})$ , respectively. Asterisks indicate significance levels of effects. \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.461	0.04	9.284		
time	3.289	0.29	11.059	108.21(1,98)	***
time <sup>2</sup>	-2.618	0.22	-11.735	139.36(1,180)	***
time <sup>3</sup>	0.361	0.12	2.971	8.10(1,501)	**
condition	0.054	0.09	0.589	0.05(1,175)	n.s.
time:condition	0.835	0.28	2.901	7.70(1,411)	**
time <sup>2</sup> :cond	0.529	0.23	2.229	4.64(1,727)	*
time <sup>3</sup> :cond	-0.066	0.24	-0.269	0.06(1,721)	n.s.

Table 2.5: Growth curve model and *F*-tests comparing -Pend/-Vend with -Pend/+Vend by-subject.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.546	0.08	6.648		
time	3.532	0.22	15.976	260.33(1,189)	***
time <sup>2</sup>	-2.443	0.17	-14.046	177.90(1,242)	***
time <sup>3</sup>	0.256	0.14	1.720	2.81(1,306)	.
condition	-0.049	0.06	-0.756	1.73(1,506)	n.s.
time:condition	0.614	0.32	1.880	3.30(1,259)	.
time <sup>2</sup> :cond	0.646	0.26	2.467	5.75(1,435)	*
time <sup>3</sup> :cond	0.021	0.19	0.114	0.01(1,868)	n.s.

Table 2.6: Growth curve model and *F*-tests comparing -Pend/-Vend with -Pend/+Vend by-item.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.406	0.05	6.778		
time	3.709	0.31	11.845	129.54(1,89)	***
time <sup>2</sup>	-2.07	0.23	-8.790	63.97(1,137)	***
time <sup>3</sup>	0.274	0.15	1.819	2.77(1,440)	.
condition	-0.046	0.07	-0.666	0.03(1,437)	n.s.
time:condition	0.547	0.27	1.965	3.53(1,383)	.
time <sup>2</sup> :cond	0.932	0.23	4.034	15.21(1,735)	***
time <sup>3</sup> :cond	-0.374	0.28	-1.306	1.55(1,393)	n.s.

Table 2.7: Growth curve model and *F*-tests comparing +Pend/-Vend with +Pend/+Vend by-subject.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.517	0.08	5.813		
time	3.978	0.22	17.332	279.04(1,203)	***
time <sup>2</sup>	-1.928	0.19	-9.753	93.27(1,187)	***
time <sup>3</sup>	0.273	0.12	2.117	4.41(1,483)	*
condition	-0.019	0.06	-0.292	0.03(1,336)	n.s.
time:condition	0.937	0.32	2.857	7.68(1,269)	**
time <sup>2</sup> :cond	0.845	0.30	2.809	7.45(1,350)	**
time <sup>3</sup> :cond	-0.686	0.25	-2.661	6.72(1,465)	**

Table 2.8: Growth curve model and *F*-tests comparing +Pend/-Vend with +Pend/+Vend by-item.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.431	0.04	8.889		
time	3.146	0.32	9.704	72.45(1,90)	***
time <sup>2</sup>	-2.707	0.21	-12.637	153.39(1,197)	***
time <sup>3</sup>	0.422	0.14	2.884	7.65(1,451)	**
condition	-0.003	0.07	-0.047	0.34(1,298)	n.s.
time:condition	0.564	0.24	2.333	5.06(1,683)	*
time <sup>2</sup> :cond	0.328	0.25	1.290	1.54(1,651)	n.s.
time <sup>3</sup> :cond	0.061	0.26	0.234	0.05(1,554)	n.s.

Table 2.9: Growth curve model and *F*-tests comparing -Pend/-Vend with +Pend/-Vend by-subject.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.546	0.08	6.512		
time	3.363	0.21	15.983	260.06(1,205)	***
time <sup>2</sup>	-2.553	0.19	-13.010	156.80(1,186)	***
time <sup>3</sup>	0.427	0.13	3.162	9.65(1,408)	**
condition	-0.053	0.06	-0.809	0.09(1,295)	n.s.
time:condition	0.226	0.34	0.651	0.39(1,260)	n.s.
time <sup>2</sup> :cond	0.390	0.26	1.462	2.01(1,384)	n.s.
time <sup>3</sup> :cond	0.365	0.23	1.565	2.32(1,481)	n.s.

Table 2.10: Growth curve model and *F*-tests comparing -Pend/-Vend with +Pend/-Vend by-item.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.434	0.06	6.738		
time	3.836	0.28	13.544	159.23(1,105)	***
time <sup>2</sup>	-1.991	0.24	-8.199	55.84(1,127)	***
time <sup>3</sup>	0.205	0.14	1.385	2.35(1,444)	n.s.
condition	-0.108	0.07	-1.465	0.55(1,462)	n.s.
time:condition	0.265	0.31	0.846	0.64(1,301)	n.s.
time <sup>2</sup> :cond	0.719	0.21	3.402	10.89(1,869)	**
time <sup>3</sup> :cond	-0.237	0.23	-1.034	1.00(1,843)	n.s.

Table 2.11: Growth curve model and *F*-tests comparing -Pend/+Vend with +Pend/+Vend by-subject.

	Estimate	SE	<i>t</i>	<i>F</i> (Df,Df.res)	sig.
(Intercept)	0.511	0.08	6.145		
time	4.107	0.23	17.601	291.96(1,181)	***
time <sup>2</sup>	-1.826	0.18	-9.954	111.99(1,241)	***
time <sup>3</sup>	0.083	0.14	0.598	0.45(1,391)	n.s.
condition	-0.020	0.07	-0.271	1.02(1,356)	n.s.
time:condition	0.557	0.29	1.906	3.41(1,314)	.
time <sup>2</sup> :cond	0.628	0.28	2.245	4.76(1,398)	*
time <sup>3</sup> :cond	-0.335	0.21	-1.525	2.21(1,556)	n.s.

Table 2.12: Growth curve model and *F*-tests comparing -Pend/+Vend with +Pend/+Vend by-item.

*List of Materials*

Item ID	Confederate Objects	Participant Objects
01	Kabeltrommel, Nagel, Wasserhahn (cable drum, nail, faucet)	-
02	Aktentasche, Fläschchen, LKW (briefcase, baby bottle, truck)	-
03	Cello, Baseball, Cowboystiefel (cello, baseball, cowboy boots)	-
04	Frisbee, Zettel, Angel (frisbee, note, angel)	-
05	Butter, Erde, Wanderschuh (butter, earth, hiking shoe)	-
06	Bildschirm, Tonne, Würfel (screen, cask, dice)	-
07	Gabel, Mikroskop, Papierkorb (fork, microscope, paper basket)	-
08	Schallplatte, Teleskop, Mülleimer, Drucker, Golfball (record, telescope, trash bin, printer, golf ball)	-
09	Öllampe, Weinglas, Vorschlaghammer, Radio, Leinwand (oil lamp, wine glass, suggestion hammer, radio, canvas)	-
10	Pfanne, Tastatur, Wecker (pan, keyboard, alarm clock)	-
11	Burger, Nagellack, Didgeridoo, Safe (burger, nail polish, didgeridoo, safe)	-
12	Honigmelone, Basecap, Zeitung, Feuerzeug (honey melon, basecap, newspaper, lighter)	-
13	Controller, Kaffee, Spitzhacke, Kartoffel (controller, coffee, pickaxe, potato)	-
14	Videokassette, Fernglas, Sessel, Sticksäge (videocassette, binoculars, armchair, jigsaw)	-
15	Arzt Tasche, Laterne, Baseballhandschuh, Golfcart (doctor's bag, lantern, baseball glove, golfcart)	-
16	Birne, Laute, Motorrad, Waage (pear, loud, motorcycle, scales)	-
17	Büroklammer, Stein, Topf, Mikrowelle (paperclip, stone, pot, microwave)	-
18	Teelicht, Zahnrad, Diskokugel, Fussball (tealight, cogwheel, disco ball, football)	-
19	Briefkasten, espressokocher, Käse, Erbsenschote, Tacho (mailbox, espresso maker, cheese, pea pod, speedometer)	-
20	Paprika, Dartscheibe, Energiesparlampe, Kontrabass, Sprühflasche (paprika, dartboard, energy saving lamp, double bass, spray bottle)	-
21	Klemmbrett, Zuckerwatte, Chili, Schuh, Milchtüte (clipboard, cotton candy, chili, shoe, milk carton)	-
22	Haus, Puzzle, Bariton, Mikrophon, Stehlampe (house, puzzle, baritone, microphone, floor lamp)	-
23	Ofen, Pille, Schlagzeug, Sombrero, Kerze (oven, pill, drums, sombrero, candle)	-
24	Schere, Pizza, Klappe, Blume, Zippo (scissors, pizza, clapper board, flower, zippo)	-
25	Tür, Fahrrad (door, bicycle)	Ei (egg)
26	Bombe, Pfeife (bomb, whistle)	Steak (steak)

27	Kasse, Glocke (checkout, bell)	Schwert (sword)
28	Maus, Bleistift (mouse, pencil)	Sonnenbrille (sunglasses)
29	Tasche, Batterie (bag, battery)	Hut (cap)
30	Amphore, Leuchtturm (amphora, lighthouse)	Bier (beer)
31	Telefonhörer, Magnet (telephone receiver, magnet)	Füller (ink pen)
32	Falle, Staubsauger (trap, vacuum cleaner)	Baum (tree)
33	Mond, Brombeere, Traktor (moon, blackberry, tractor)	Skateboard (skateboard)
34	Sparschwein, Waschmaschine, Zwiebel (piggy bank, washing machine, onion)	Sonnenblume (sunflower)
35	Taschenmesser, Banjo, Spitzer (pocketknife, banjo, sharpener)	Buch (book)
36	Steuerrad, Thermometer, Kompass (wheel, thermometer, compass)	Muschel (shell)
37	Notizblock, Erdbeere, Kanister (notepad, strawberry, canister)	Plattenspieler (record player)
38	Feige, Hufeisen, Kegel (fig, horseshoe, cone)	Boot (boat)
39	Satellitenschüssel, Stoppuhr, Bus (satellite dish, stopwatch, bus)	Eichel (acorn)
40	Blasebalg, Drachen, Lupe (bellows, kite, magnifying glass)	Blatt (sheet)
41	Wäscheklammer, Olive, Kleeblatt, Harfe (clothespeg, olive, shamrock, harp)	Stift (pen)
42	Axt, Sanduhr, Papierflieger, Flasche (ax, hourglass, paper plane, bottle)	Maßband (tape measure)
43	Picknickkorb, Radiergummi, Kürbis, Spaten (picnic basket, eraser, pumpkin, spade)	Harke (rake)
44	Truhe, Kaktus, Softeis, Zitrone (chest, cactus, soft ice, lemon)	Trichter (funnel)
45	Turnschuh, Zitronenpresse, Akkuschauber, Flöte (sneaker, lemon squeezer, cordless screwdriver, flute)	Helm (helmet)
46	Fotoapparat, Kettensäge, Polizeiauto, Ananas (camera, chainsaw, police car, pineapple)	Lagerfeuer (campfire)
47	Zaun, E-Gitarre, Bürste, Schachtel (fence, electric guitar, brush, box)	Wollknäuel (ball of wool)
48	Taschenuhr, Wasserkocher, Aktenschrank, Säge (pocket watch, kettle, filing cabinet, saw)	Kassette (cassette)
49	Karton (carton)	Geige, Schirm (violin, screen)
50	Ampel (traffic light)	Brokkoli, Feuerlöscher (broccoli, fire extinguisher)
51	Brille (glasses)	Goldbarren, Rucksack (gold bars, backpack)
52	Cabrio (convertible)	Fernbedienung, Grill (remote control, grill)
53	Fahne (flag)	Zelt, Kastanie (tent, chestnut)
54	Löffel (spoon)	Briefumschlag, Kirsche (envelope, cherry)
55	Nadel (needle)	Kelle, Linial (trowel, ruler)
56	Hose (pants)	Brief, Schatzkiste (letter, treasure chest)

57	Taschenlampe, Weihnachtsbaum (flashlight, christmas tree)	Badehose, Windrad (trunks, wind wheel)
58	Boxhandschuh, Zange (boxing glove, pliers)	Heft, Pyramide (folder, pyramid)
59	Videokamera, Besen (video camera, broom)	Fernseher, Ring (tv, ring)
60	Brot, Schüssel (bread, bowl)	Teekessel, Avocado (tea kettle, avocado)
61	Basketballkorb, Parkbank (basketball basket, park bench)	Formeleinsauto, Gameboy (formula car, gameboy)
62	Klebestreifen, Paket (adhesive tape, package)	Luftballon, Megaphon (balloon, megaphone)
63	Motorboot, Rose (powerboat, rose)	Kleiderbügel, Locher (hanger, punch)
64	Narzisse, Trompete (daffodil, trumpet)	Orange, Jeep (orange, jeep)
65	Gasmaske, Strandkorb, Tomate (gas mask, beach chair, tomato)	Seestern, Wärmflasche (starfish, hot water bottle)
66	Zauberwürfel, Dose, Blumentopf (magic cube, can, flower pot)	Krone, Laptop (crown, laptop)
67	Kleiderständer, Limette, Tasse (clothes rack, lime, cup)	Grillzange, Maiskolben (barbecue tongs, corncob)
68	Schlüssel, Lenkdrachen, Rubin (key, stuntkite, ruby)	Tacker, Donut (tacker, donut)
69	Helicopter, Kerzenständer, Pflaume (helicopter, candlestick, plum)	Schubkarre, Rettungsring (wheelbarrow, lifebelt)
70	Karabinerhaken, Diskette, Lippenstift (snap hook, diskette, lipstick)	Bügeleisen, Radischen (irons, radishes)
71	Rasierer, Sandwich, Beutel (razor, sandwich, bag)	Eishockeyschläger, Kokosnuss (hockey stick, coconut)
72	Vase, Schriftrolle, Gurke (vase, scroll, cucumber)	Eis, Saxophon (ice cream, saxophone)
73	-	Handy, Leiter, Tisch (mobile phone, ladder, table)
74	-	Kran, Pilz, Schneemann (crane, mushroom, snowman)
75	-	Schnuller, Teddybär, Ventilator (pacifier, teddy bear, fan)
76	-	Tennisschläger, Bügelbrett, Croissant (tennis racket, ironing board, croissant)
77	-	Wassermelone, Teekanne, Fliegenklatsche (watermelon, teapot, fly swatter)
78	-	Volleyball, Boomerang, Hemd (volleyball, boomerang, shirt)
79	-	Korb, Spritze, Reissverschluss (basket, syringe, zipper)
80	-	Gürtel, Krug, Bett (belt, pitcher, bed)
81	Flügel (grand piano)	Krawatte, Stethoskop, Rasierapparat (tie, stethoscope, shaver)
82	Schalter (switch)	Salzstreuer, Hydrant, Kaffeebohne (salt spreader, hydrant, coffee bean)
83	Kopfhörer (headphones)	Rohrzange, Gitarre, Taschenrechner (pipe wrench, guitar, calculator)
84	Lampe (lamp)	Schraubenzieher, Sack, Volleyballnetz (screwdriver, bag, volleyball net)
85	Mütze (beanie)	Hotdog, Stöckelschuh, Handfeger (hotdog, high heels, hand brush)
86	Geschenk (gift)	Rad, Stuhl, Apfel (wheel, chair, apple)

87	Koffer (suitcase)	Baseballschläger, Schloss, Piratenschiff (baseball bat, castle, pirate ship)
88	Medaille (medal)	Auto, Kiwi, Handschuh (car, kiwi, glove)
95	Hammer, Schlitten (hammer, sled)	Trommel, Zielscheibe, Banane (drum, target, banana)
96	Filmrolle, Gießkanne (film reel, watering can)	Computer, Uhr, Pflanze (computer, clock, plant)
101	Presslufthammer, Regal (jackhammer, shelf)	Fächer (fan)
102	Pylon, Couch, Handtasche (pylon, couch, handbag)	Hantel (dumbbell)
103	Headset, Muffin, Heißluftballon, Pullover (headset, muffin, hot air ballon, pullover)	Anker (anchor)
104	T-Shirt (t-shirt)	Mutter, Kartenspiel (nut, card game)
105	Etikett, Palme (label, palm)	Reagenzglas, Einrad (test tube, unicycle)
106	Karussell, Hütte, Palette (carousel, hut, palette)	Türklinke, Wasserpistole (doorknob, water gun)
107	-	Tipi, Mixer, Socke (tipi, mixer, sock)
108	Rakete (rocket)	Billardkugel, Messer, Pfirsich (billiard ball, knife, peach)
109	Honigglas, Swimmingpool (honey jar, swimming pool)	Bretzel, Kaffeekanne, Schlauchboot (bretzel, coffee pot, inflatable boat)
p02	Wespe, Elster, Huhn, Schaf (asp, magpie, chicken, sheep)	-
p03	Adler, Tiger, Schmetterling, Delfin, Papagei (eagle, tiger, butterfly, dolphin, parrot)	-
p04	Fisch, Bär (fish, bear)	Weintrauben (grapes)
p05	Himbeere, Giraffe, Angelhaken (raspberry, giraffe, fishhook)	Lautsprecher (speaker)
p06	Eiffelturm, Tischtennisschläger, Raupe, Libelle (eiffel tower, table tennis racket, caterpillar, dragonfly)	Schwan (swan)
p07	Berimbao (berimbao)	Apfelgriebs, Handrechen (apple core, hand rake)
p08	Klarinette, Überwachungskamera (clarinet, surveillance camera)	Schraubenschlüssel, Sicherheitsnadel (wrench, safety pin)
p09	Ladekabel, Verteilerdose, Törtchen (charger cable, junction box, tartlet)	Legosteine, Waffeln (lego, waffles)
p10	-	Ente, Holz, Sushi (duck, wood, sushi)
p11	Getreide (grain)	Dynamit, Klopapier, Straßenlaterne (dynamite, toilet paper, street lamp)
p12	Straße, Fitnessbank (street, fitness bench)	Käfer, Taube, Fledermaus (beetle, pigeon, bat)

Table 2.13: List of materials. Confederate objects were named in the critical turn by the confederate. Participant objects had to be named by the participant.





# 3

---

## PROGRESSION OF SPEECH PLANNING IN OVERLAP

---

Submitted as:

Barthel, M. and Levinson, S. C. (in press). Phonological Planning is Done in Overlap with the Incoming Turn: Evidence from Gaze-contingent Switch Task Performance. *Language, Cognition and Neuroscience*.

### ABSTRACT

To ensure short gaps between turns in conversation, next speakers regularly start planning their utterance in overlap with the incoming turn. Three experiments investigate which stages of utterance planning are executed in overlap. E1 establishes effects of associative and phonological relatedness of pictures and words in a switch-task from picture naming to lexical decision. E2 focuses on effects of phonological relatedness and investigates potential shifts in the time-course of production planning during background speech. E3 required participants to verbally answer questions as a base task. In critical trials, however, participants switched to visual lexical decision just after they began planning their answer. The task-switch was time-locked to participants' gaze for response planning. Results show that word form encoding is done as early as possible and not postponed until the end of the incoming turn. Hence, planning a response during the incoming turn is executed at least until word form activation.

## 3.1 INTRODUCTION

In conversation, interlocutors readily exchange turns of talk, frequently switching from the role of the listener to the role of the speaker without leaving long gaps between turns (Sacks et al., 1974; Stivers et al., 2009). Previous studies consistently find that speech planning takes more time than the average gap between turns in conversation, as it takes speakers at least 600 ms to plan single words (Indefrey, 2011) and about one and a half seconds to prepare a simple sentence (Griffin & Bock, 2000; Schnur et al., 2006). Based on evidence from picture naming studies using the picture-word interference paradigm (e.g. Schriefers et al., 1990; Wilshire, Singh, & Tattersall, 2016), time requirements of the separate levels of the speech production process are estimated to be around 200 ms to activate a mental concept that fits a depicted picture, about 75 ms for the selection of a lemma that matches the concept and represents semantic and syntactic information of a word, and approximately 80 ms to retrieve the phonological code of that word (Indefrey & Levelt, 2004), followed by processes of syllabification and phonetic encoding. Recent models of turn taking postulate that next speakers need to start planning their utterance as early as possible (early-planning hypothesis) and in overlap with the incoming turn (Levinson & Torreira, 2015; Pickering & Garrod, 2013), assuming that the gap between turns would be much longer than regularly observed if next speakers only began to plan their turn in reaction to the end of the incoming turn or even to turn-final cues about the upcoming turn end (Barthel, Meyer, & Levinson, 2017, see ch. 5). Planning the content of a response turn that is contingent upon the incoming turn can only begin when the incoming message is sufficiently clear or can be reliably anticipated. If response planning is executed in overlap with the incoming turn, the respective planning processes might be slowed down due to concurrent speech comprehension. The time pressures of conversation, the most frequently used speech exchange system, might therefore have a great impact on the mechanisms of speech planning.

Experimental studies testing the early-planning hypothesis have indeed shown that planning commonly begins as early as possible during the incoming turn (but see the study by Sjerps and Meyer (2015), and Barthel, Sauppe, Levinson, and Meyer (2016, see ch. 2) for discussion thereof). Barthel et al. (2016, see ch. 2) used a list

completion paradigm with a confederate listing a number of displayed objects and the participant listing the remaining displayed objects. The confederate turns had different syntactic structures, so that they either ended with one of the object names or with a verb form that was redundant for participants to plan their next turn. Eye-movements and voice onset latencies showed that participants started to plan their response turn as early as possible during the incoming turn, even if redundant material predictably followed before the incoming turn's end. [Bögels, Magyari, and Levinson \(2015\)](#) used a confederate who asked participants questions whose answer became clear either in the middle of the question or only at the end of the question (e.g., as in "Which character, also called 007, appears in the famous movies?" (early) vs. "Which character from the famous movies is also called 007?" (late)). Response latencies were shorter when the answer could be deduced in the middle of the question than when it became obvious only at its very end. Additionally, in both early and late questions, 500 ms after the onset of the critical information, the authors recorded a positivity in participants' EEG signal, which was substantially reduced in a control task that did not involve response planning. This positivity was therefore interpreted as an indication of early response planning processes. Consistent findings are reported by [Corps, Crossley, Gambi, and Pickering \(2018\)](#). Manipulating the predictability of an incoming question's end, the authors find that participants answered questions earlier when their end was predictable as compared to unpredictable, suggesting that participants used content prediction to begin to plan their answer in overlap with the incoming question whenever possible.

While these studies show that planning starts in overlap with the incoming turn, they did not investigate which levels of production planning are run through while still listening to incoming speech. Using a post-hoc EEG source localization analysis on the data recorded during their question-answer study, [Bögels, Magyari, and Levinson \(2015\)](#) found activation of the middle frontal and precentral gyri in overlap with the incoming turn and hypothesised this activation to be due to phonological planning in preparation of the answer. However, these brain regions have also been found to be active during memory retrieval ([Rajah, Languay, & Grady, 2011](#); [Raz et al., 2005](#)), which could be responsible for the reported findings instead, since participants needed to retrieve the answers to the posed questions from

long term storage. Alternatively, activation of these brain regions might have been due to ongoing comprehension of the incoming question, which is supposed to result in concurrent activation of related speech production processes (Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985; Pickering & Garrod, 2013). Hence, the question which stages of response planning are run through in overlap with the incoming turn remains unsettled. Speakers need to go through a number of these stages before being prepared to articulate their turn, including at least conceptualization, formation of a syntactic structure, lemma selection, word form retrieval, and phonetic encoding (Indefrey & Levelt, 2004; Levelt, 1989). The turn-taking model by Levinson and Torreira (2015) assumes that all stages of response formulation are run through as early as the action that is intended with the incoming turn can be recognised. The model therefore assumes that all the stages of response formulation regularly occur in overlap with the incoming turn, while articulation is withheld until the incoming turn comes to an end. Whether this is true for all stages of speech planning is an open empirical question.

A major reason to assume that some processing stages might be postponed until the end of the incoming turn is the well established fact that speech production and comprehension compete for processing capacities. Previous studies found that incoming linguistic material interferes with speech production more than non-linguistic material, with interference being most severe on the word form level. Kemper et al. (2003) asked participants open questions to elicit free talk while participants continuously performed different secondary tasks. They found that speech production was more difficult for participants when they had to ignore incoming speech than when they had to ignore noise, as was indicated for example by a higher rate of production errors in the speech condition. Schriefers et al. (1990), using the picture word interference paradigm, compared the effect of auditorily presented distractor words with a noise condition and a condition without distractors (silence) on picture naming performance and found that distracting speech was significantly more detrimental to response latencies than silence or noise. Fargier and Laganaro (2016) tested participants on a dual-task with picture naming as base task 1 and either tone or syllable detection as a go/no-go task 2. Analyzing only no-go trials, they found that naming latencies were longer with syllables

than with tones as concurrent input. Additionally, they found ERP waveform differences between syllables and tones as concurrent input about 400 ms after picture onset, which they interpreted to be caused by increased interference of verbal as compared to non-verbal material with word form encoding processes. Similarly, [Fairs, Bögels, and Meyer \(2018\)](#) found interference on picture naming performance to be larger with a second linguistic task (syllable detection) than with a concurrent non-linguistic task (tone identification).<sup>1</sup> [Klaus, Mädebach, Oppermann, and Jescheniak \(2017\)](#) used a dual-task paradigm asking participants to ignore auditory distractor words and produce subject-verb-object picture descriptions while concurrently performing either a visuospatial or a verbal working memory task. Under verbal but not under visuospatial working memory load, participants' phonological planning scope was reduced to the subject of the sentence, while their abstract lexical planning scope remained unreduced, including the sentence final object. This pattern of results shows that high verbal working memory load interferes with phonological production planning. Taking together these findings, postponing (at least) phonological planning until the end of the incoming turn could therefore be an efficient strategy that might be applied by next speakers to keep the increase in processing costs that come with planning in overlap at a moderate level ([Barthel & Sauppe, 2019](#), see ch. 4). On the other hand, late phonological planning might lead to long gaps between turns that might be undesired because long delays give rise to inferences on the turn's meaning ([Bögels, Kendrick, & Levinson, 2015](#); [Clayman, 2002](#); [Pomeranz & Heritage, 2012](#)) and might commonly be avoided for that reason.

The present study investigates which stages of formulation (lemma selection and word form retrieval) are executed in overlap with incoming speech, mimicking a situation where a participant of a conversation starts to prepare their own turn while still listening to another person speaking. While planning the next turn in overlap with the incoming turn, each level of processing, from conceptual planning to word form retrieval can be hypothesised to add interference of the incoming speech with the respective response planning processes ([Indefrey & Levelt, 2004](#); [Levelt, 1989, 1992](#)). With these processing pressures

---

<sup>1</sup>However, the effect might have been due to differences in acoustic complexity of the tones vs. the syllables used.

standing against the time pressures that are applied by the turn-taking system, there might be a level of processing at which the costs of early response planning match its benefits, the question being where that level is. One hypothesis is that only conceptual planning is done in overlap while formulation is postponed until the end of the incoming turn in order to avoid increased planning effort due to phonological interference. A competing hypothesis is that formulation, including word form retrieval, is done as early as possible and in overlap with the incoming turn in order to keep inter-turn gaps short.

These hypotheses are evaluated here in three experiments making use of a switch task. In Experiment 1, participants were required to name a presented object as fast as possible as a base task. In switch trials (25%), the object disappeared after having been presented for a short amount of time and was replaced by a word that had to be judged to be a real Dutch word or not by giving a button press response. These words were presented either after associated or phonologically related pictures or after unrelated pictures. Words that are associated to pictures are words that come to mind when a particular picture is presented, e.g. *cheese* when a mouse is presented. Phonologically related words on the other hand sound like the presented picture's name, e.g. *mouth* when a mouse is presented. In cases when the respective level of representation (lemma for associative relatedness or word form for phonological relatedness) was activated by the time of the task switch, relatedness of the target picture and the word replacing it should have an effect on participants' lexical decision performance. Assuming a structured mental lexicon consisting of at least three distinct levels of entries, namely concepts, grammatical or semantic entries (lemmas), and word-forms, to produce a word requires selecting the correct word form that belongs to the lemma matching the concept that should be expressed (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). To comprehend a presented word, on the other hand, requires the selection of a concept that belongs to a lemma matching the word form that was presented (Cutler, 2012; Norris, Cutler, McQueen, & Butterfield, 2006). For reading written words, a second word form representation, the orthographic representation, is assumed next to the phonological representation, with the two types of representation being linked in the lexicon, so that an activated orthographic representation leads to activation of the corresponding phonological representation (Coltheart,

Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Ellis & Young, 1988). If, at the time of the task switch, participants activated the lemma corresponding to the picture's name, association of the picture and the word replacing it should lead to associative facilitation (Alario, Segui, & Ferrand, 2000; La Heij, Dirks, & Kramer, 1990; Perea & Rosa, 2002; Plaut, 1995). Similarly, if participants activated the word form of the picture to be named by the time of the task switch, the representations of phonologically related words should be suppressed below their level of resting activation, leading to decreased lexical decision performance in phonologically related words (Levelt, Schriefers, Vorberg, Meyer, & Pechmann, 1991; Pykkänen, Gonnerman, Stringfellow, & Marantz, submitted). As the processes of lemma selection are known to precede the processes of word form retrieval, three different stimulus-onset asynchronies (SOAs) are used in order to target the different levels of production planning (Levelt, 1989; Levelt et al., 1999).

Experiment 2 uses the same materials as Experiment 1 and takes an intermediate step between the monologic setup of Experiment 1 and the dialogic setup of Experiment 3 by adding incoming questions being played to participants as distracting speech which participants were instructed to ignore. In that way, Experiment 2 will allow us to evaluate whether distracting speech as it is commonly used in experimental setups affects the timing of language planning.

In Experiment 3, the same materials were used as in Experiment 2. In Experiment 3 however, participants had to decide based on the question which one out of four displayed pictures they would have to name. In that way, the given task resembled a dialogical situation as participants were required to attend to the presented questions and answer them by naming one of the pictures. The format of these questions was designed to give away the cue to the target picture either during the middle of the question or only at its end. Again, in critical trials (25%), participants had to switch from the picture naming task to the lexical decision task. The relatedness effects of target pictures and words for lexical decision will shed light on the progress of response planning during the incoming turn on the one hand, as compared to at the end of the incoming turn on the other hand. Following the hypothesis that all stages of response planning are run through in overlap with the incoming turn (Levinson & Torreira, 2015) and



consequently activating the respective representations on all levels of the mental lexicon, relatedness of the picture to be named and the word replacing it for lexical decision should have an effect on lexical decision performance both during the incoming question as well as at the end of it. If a relatedness effect was only found at the end of questions, however, when response planning is done in silence, and was absent in the middle of questions, where response planning is done in overlap, that finding would be taken as evidence for delayed response formulation. The filler trials (75%), in which participants have to overtly answer the question by naming the target picture, serve as a replication of the effects of planning in overlap that were described in the previous literature (Barthel et al., 2017, 2016; Bögels, Magyar, & Levinson, 2015, (see ch. 2 and 5)). If the responses are planned as early as possible, naming latencies should be faster in questions that give away the answer early as compared to questions that give away the answer only at their end.

## 3.2 EXPERIMENT 1

### 3.2.1 *Method*

#### *Participants*

Sixty-four Dutch native speakers were recruited as paid participants at Radboud University campus. Data of one participant was lost after recording. All participants reported to have normal or corrected to-normal vision and hearing as well as no speech or language impairments.

#### *Apparatus*

Participants were seated in a sound proof booth approximately 60 cm away from a 21 inch computer screen and a Sennheiser ME64 microphone. They were equipped with a two-precision-buttons response box based on USB-mouse script with 125Hz sampling rate. Stimuli were presented using SMI ExperimentCenter software.

### *Materials and Design*

256 pictures of objects were used in the experiment. The pictures were sourced online and are under the creative commons license. They were selected to be easy to recognize and name. 192 of these pictures served as filler objects in naming trials and were not systematically related to the pictures used in critical trials. The common names for these filler objects cover a broad range of medium frequency counts as extracted from the SUBLEX\_NL corpus (Keuleers, Brysbaert, & New, 2010, mean log frequency per million = 1.95; SD = 1.8) and vary in length between one and five syllables (mean number of syllables = 2.4; SD = 0.95). The remaining 64 pictures served as critical objects in switch trials. The critical objects had very high name agreement, as assessed in a pretest with a different group of 31 participants (mean agreement = 96%, SD = 4%).

256 words were used in the lexical decision task, with half of them being real Dutch words (critical), the other half being pseudowords (filler). Each of the words was either associated with a critical picture or phonologically related to a critical picture's name (Type of Relation: associative/phonological), and would either be presented after the related picture or after another, unrelated picture (Relatedness: unrelated/related). Table 3.1 gives an overview of the tested conditions. Associatively related words were drawn from the Dutch Word Association Database (<http://www.kuleuven.be/semlab/interface/index.php>; see De Deyne and Storms (2008)), and were chosen to be strong associates of the picture name (mean first association strength = 30%, SD = 16%). Phonologically related words had the same syllable length and syllable structure as the related picture name and tended to differ from the picture name in one phoneme towards the end of the word (i.e. in nucleus, coda, or second syllable). Associatively related words were not phonologically related to the respective picture names, with maximally one overlapping segment (mean overlap = 5% of segments, SD = 11%). Phonologically related words were not associated with the pictures. Pseudoword strings were produced by changing one segment of one of the real words.

Eight experimental lists were constructed, with a different word following a given critical picture in each of the lists. Each participant

was tested in one of the lists and assigned to one of three SOA groups (see Section *Procedure* below).





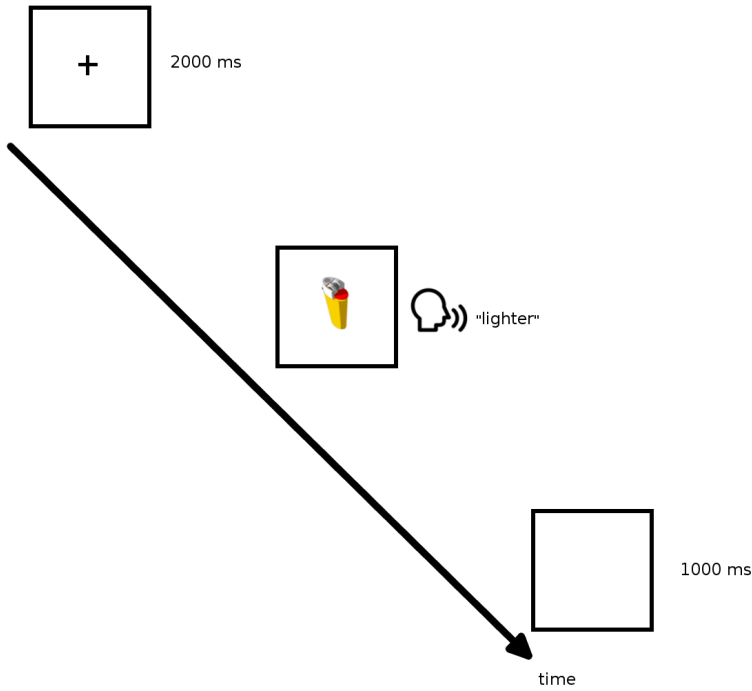
Type of relation	Relatedness	Target (name)	Lexical decision word
phonological	related	 (appel)	ampel (traffic light)
	unrelated	 (zaag)	ampel (traffic light)
association	related	 (appel)	fruit (fruit)
	unrelated	 (zaag)	fruit (fruit)

Table 3.1: Example item showing the four critical conditions tested in Experiment 1. Each condition of an item was tested in a separate list.

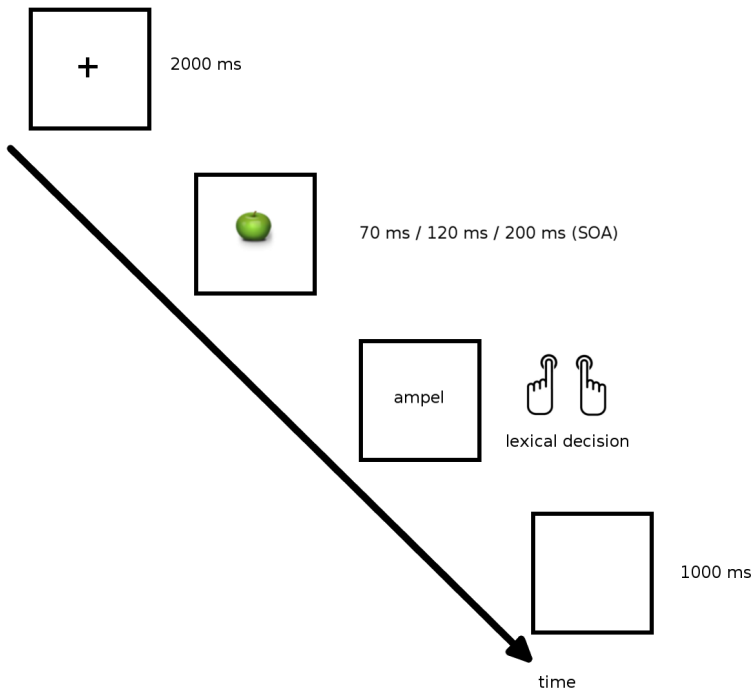
### *Procedure*

Each trial began with a fixation cross in the middle of the screen for 2 seconds, followed by one of the pictures presented at the center of the screen (see Figure 3.1). Participants were instructed to name the picture as fast as possible. The picture disappeared upon voice onset and was replaced by a blank screen for 1 second before the next trial started. In switch trials (25%), the picture was only presented for a short amount of time (SOA) before it was replaced by a letter string. Three SOA conditions of 70 ms, 120 ms, and 200 ms were tested between participants. Participants were instructed to abandon the naming task in case the picture was replaced by a word. In this case, participants were to decide whether the presented word was a real Dutch word or not, and give their response by pressing one of two buttons as fast as possible (with the ‘word’ response lying on the right button). Upon pressing a button, the word disappeared and was replaced by a blank screen for 1 second before the next trial started. Every sixty-four trials, a pause screen was presented, giving participants the chance to take a short break.

The experiment proper was preceded by eight practice trials and followed by a post test in which participants were shown the 64 critical pictures and asked to name them, so as to check whether their responses matched the expected names for the critical pictures. The whole experimental session took about 40 minutes.



(a) naming trial



(b) switch trial

Figure 3.1: Timelines of a naming trial and a switch trial in Experiment 1.

### 3.2.2 Results

Of the 12096 naming trials, 481 trials (3.9%) were regarded as erroneous and consequently discarded, as the voice key was triggered more than four seconds after picture onset. Another 404 trials (3.4%) were discarded because they were outliers of more than 2.5 standard deviations by subject. Remaining trials had a mean naming latency of 1184 ms (SD = 488 ms; CI = <1175 ms, 1193 ms>). Figure 3.2 shows a density plot of the distribution of naming latencies. Figures 3.10 and 3.11 in the Supplementary Materials show density plots of the distributions of naming latencies by SOA condition and by subject.

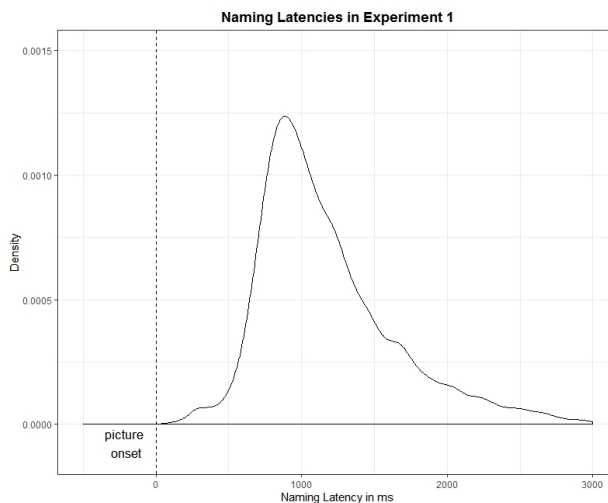


Figure 3.2: Distribution of naming latencies in Experiment 1.

Of the 2016 critical lexical decisions, 111 (5.6%) were discarded since participants did not name the corresponding critical pictures by their standard labels in the post test. Inspecting the distributions of lexical decision latencies for each of the subjects, reaction times by two participants (both in SOA120 condition) were found to behave differently than those of the other subjects in not being uni-modally distributed, possibly hinting at the use of a reaction strategy ignoring the instruction to give a decision as fast as possible. Data from these two subjects were excluded from further analyses.<sup>2</sup> 246 button press responses (13.3%) were erroneous. Notably, almost twice as many errors were produced

<sup>2</sup>Removal of these subjects' data did not change the presented pattern of results, as attested in separate analyses.

with words that were presented after phonologically related pictures (28.2%) as compared to after phonologically unrelated pictures (15.1%, see Figure 3.3).

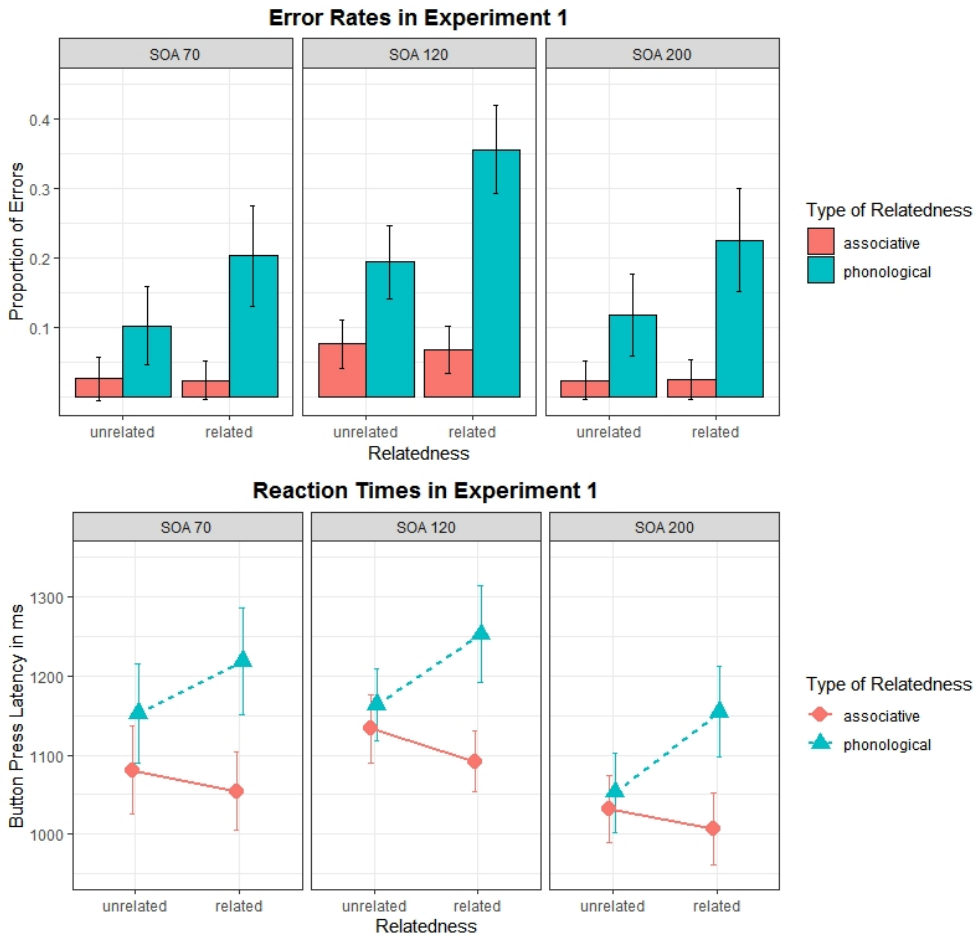


Figure 3.3: Reaction times and error rates in lexical decisions in Experiment 1. Bars represent 95% confidence intervals.

Statistical analyses have been conducted with R (R Core Team, 2019). Mixed effects regression models have been fitted using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and predictors' statistical significance was assessed with  $F$ -tests with Kenward-Roger approximations of degrees of freedom (Fox & Weisberg, 2011; Halekoh & Hojsgaard, 2014; Kenward & Roger, 1997). Bayesian liner models have been fitted using the brms package Bürkner (2017) and

3000 iterations. Bayes factors were calculated using the built in *brms hypothesis*-function.<sup>3</sup> Throughout the study's experiments, the maximal random effects structures justified by design which allowed models to converge were used (Barr, 2013; Barr et al., 2013), with subject and item as random effects. All categorical predictors were deviation coded with the exception of SOA in Experiment 1, which was simple coded with the intercept referring to the grand mean of the three levels of the factor and the effect of the first two levels (SOA70 and SOA120) being compared to the effect of the third level (SOA200).

Error rates were analyzed in a logit mixed effects regression model with SOA, Relatedness and Type of Relation as well as their interactions as predictors (see Table 3.4 in Supplementary Materials). While mean error rates in SOA70 and SOA200 do not differ significantly, error rates in SOA120 are significantly higher than error rates in SOA200 ( $\beta = 0.940$ ,  $SE = 0.407$ ,  $z = 2.307$ ,  $p < .05$ ). This effect, however, does not significantly interact with Relatedness nor with Type of Relation and is hence probably due to differences between the tested populations. The interaction effect between Relatedness and Type of Relation is significant ( $\beta = 1.011$ ,  $SE = 0.467$ ,  $z = 2.164$ ,  $p < .05$ ), indicating that the main effect of Relatedness differs between the phonological and the associative sets of words. To further investigate the effect of Relatedness, corrected post-hoc tests based on estimated marginal means have been calculated using the *emmeans* package (Lenth, 2019). Relatedness was significant in phonologically related words ( $F = 21.269$ ,  $p < .001$ ), but not in associatively related words ( $F = 0.020$ ,  $p = .88$ ), indicating that participants made more errors when words were presented after phonologically related pictures than after non-related pictures and that error rates did not differ between words that were presented after associated pictures versus after non-related pictures.

Erroneous trials were discarded from the following analyses of lexical decision latencies. Further, 39 (2.4%) trials were discarded because their reaction latencies were outliers of more than 2.5 standard deviations by subject. The mean button press latency of the remaining 1560 trials was 1118 ms ( $SD = 300$  ms, see Figure 3.3).

The log-transformed button press latencies of correct trials were analysed in a linear mixed effects regression model with SOA, Relat-

---

<sup>3</sup>A guideline for Bayes factor interpretation can be found in Jeffreys (1961), see Kass and Raftery (1995)

edness and Type of Relation as well as their interactions as predictors (see Table 3.5 in Supplementary Materials). The interaction effect of Relatedness  $\times$  Type of Relation turned out to be highly significant ( $\beta = 0.040$ ,  $SE = 0.008$ ,  $t = 4.82$ ,  $F = 23.196$ ,  $df = 1422$ ,  $p < .001$ ), indicating that the effect of Relatedness goes in opposite directions in the phonological and associative sets of words. To further investigate the effects of Relatedness, corrected post-hoc tests based on estimated marginal means have been calculated. In these tests, Relatedness turns out to significantly affect decision times in both associative words and phonological words with the effect going in opposite directions. While decisions in the associative set were made faster when the words were presented after associated pictures than after unrelated pictures ( $\beta = -0.012$ ,  $SE = 0.005$ ,  $p < .05$ ), decisions in the phonological set were made slower when the words were presented after phonologically related pictures than after non-related pictures ( $\beta = 0.028$ ,  $SE = 0.006$ ,  $p < .001$ ). To test which level of SOA showed the most robust effects of Relatedness, corrected post-hoc tests based on estimated marginal means have been calculated. While none of the effects of association survived the correction for multiple comparisons, the effect was still marginally significant in SOA120 (SOA70:  $\beta = 0.012$ ,  $SE = 0.010$ ,  $p = .250$ ; SOA120:  $\beta = 0.014$ ,  $SE = 0.007$ ,  $p = .061$ ; SOA200:  $\beta = 0.008$ ,  $SE = 0.010$ ,  $p = .395$ ). The effect of phonological relatedness was significant in all three levels of SOA and turned out to be most pronounced in SOA200 (SOA70:  $\beta = -0.022$ ,  $SE = 0.011$ ,  $p = .044$ ; SOA120:  $\beta = -0.024$ ,  $SE = 0.009$ ,  $p = 0.006$ ; SOA200:  $\beta = -0.037$ ,  $SE = 0.011$ ,  $p = .001$ ).

In order to test for the likelihood distribution of the obtained reaction times effects, a Bayesian linear model was used to fit decision latencies, with Relatedness, Type of Relation and SOA as well as their interactions as predictors with default uninformative priors and maximal random effects structures for both subjects and items (see Table 3.2). If 0 lies outside the credible interval, there is sufficient evidence to suggest there is an effect of a particular predictor. As the effect of Relatedness turned out to be decisively affected by the Type of Relatedness ( $\beta = 107$  ms,  $SE = 29$  ms,  $CrI = <50$  ms,  $164$  ms>,  $BF = inf$ ), we conducted two Bayesian inference tests testing the effects of Relatedness separately for the associative and phonological sets of words. The first test revealed decisive evidence for the effect of Relatedness in the associative set of words, with decisions for words being faster when they are presented



after associated pictures than when they are presented after non-related pictures ( $\beta = 31$  ms, SE = 18 ms, CrI =  $\langle 1$  ms , 62 ms $\rangle$ , BF = 23). The second test revealed decisive evidence for the effect of Relatedness in the phonological set of words, with decisions for words being slower when they are presented after phonologically related pictures than when they are presented after non-related pictures ( $\beta = -76$  ms, SE = 22 ms, CrI =  $\langle -113$  ms ,  $-39$  ms $\rangle$ , BF = 1999).

	$\beta$	SE	lower CrI	upper CrI
Intercept	1118.96	27.60	1065.66	1174.18
SOA70	56.83	71.71	-81.29	200.81
SOA120	99.75	64.85	-26.84	228.27
Relatedness	22.29	14.46	-6.91	50.48
Type of Relation	98.12	13.34	71.88	124.35
SOA70 $\times$ Relatedness	-21.18	34.58	-88.54	45.03
SOA120 $\times$ Relatedness	-20.92	31.71	-82.33	41.22
SOA70 $\times$ Type of Relation	17.05	33.35	-47.96	82.72
SOA120 $\times$ Type of Relation	5.25	32.57	-56.45	69.68
Relatedness $\times$ Type of Relation	107.47	28.92	49.66	164.28
SOA70 $\times$ Rel. $\times$ Type of Rel.	-17.98	63.91	-142.18	107.94
SOA120 $\times$ Rel. $\times$ Type of Rel.	3.90	57.31	-107.48	114.01

Table 3.2: Bayesian linear regression model on button press latencies in Experiment 1. For comparison of the presented effects of SOA, SOA200 was used as a baseline. Credible intervals contain 95% area under the posterior likelihood distribution. Model formula = Latency  $\sim$  intercept + SOA \* relatedness \* type.of.relation + (intercept + SOA \* relatedness \* type.of.relation | subject) + (intercept + SOA \* relatedness \* type.of.relation | item).

### 3.2.3 Discussion

Experiment 1 examined the effects of associative relatedness and phonological relatedness of pictures and words on lexical decision performance in a switch task. Participants were instructed to name displayed pictures as fast as possible as a base task. In 25% of trials, the picture was replaced by a word without prior notice after 70 ms, 120 ms, or 200 ms (SOA) and participants had to abandon the naming task and give a lexical decision response instead, evaluating whether the word was a real Dutch word or not. Decisions were faster if words were presented after pictures that were associated with the words than

after pictures that were unrelated to the words, and decisions were slower and yielded more errors if words were presented after a picture whose name was phonologically related to the word as compared to when they were presented after an unrelated picture, with this effect of phonological inhibition being most pronounced at an SOA of 200 ms. The effect of associative facilitation was weaker in absolute terms than the effect of phonological inhibition and only showed in participants' reaction times but not in their error rates. One possible reason for the effects of associative relatedness being weaker than the effects of phonological relatedness might be that association strengths between target pictures and words might vary greatly between participants or were generally too low across participants for activation to spread reliably to the lemmas of the lexical decision words while participants prepared to name the picture. Moreover, [Jongman and Meyer \(2017\)](#) found that effects of associative relatedness disappear in situations where task switches are unpredictable. In their picture naming study, associated auditory primes affected naming latencies only when the task was held constant across trials but did not affect latencies when task switches were unpredictable (as was the case in the present study). Nonetheless, since effects of phonological relatedness were observed reliably throughout Experiment 1, semantic processing of the pictures must have taken place by the time of the respective SOA's. Consequently, we will drop the associative condition and focus on phonological relatedness in Experiment 2, where we aim to replicate the results of phonological inhibition obtained in Experiment 1 in the presence of distracting incoming speech. As the target effect was most robust at SOA 200, we will focus on that SOA in the following Experiment.

### 3.3 EXPERIMENT 2

#### 3.3.1 *Introduction*

In Experiment 2 we take an intermediate step towards a dialogic test situation by adding background speech to the switch task used in Experiment 1. While participants dealt with the respective tasks (picture naming as base task; lexical decision as switch task in 25% of trials), they were auditorily presented with one question per trial in order to

test whether the same effects of phonological inhibition can be observed at the same SOA as in Experiment 1 if participants are presented with distracting speech input while attending to the switch task. If so, the same SOA can be used in a question-answer task in Experiment 3. If not, one probable reason this test might fail to replicate the previous results is that the speech production processes involved in the picture naming task get delayed or slowed down by distracting speech. In that case, the SOA to be used in Experiment 3 should be longer than in Experiment 2.

Based on the results obtained in Experiment 1, Experiment 2 focuses on effects of phonological relatedness of picture names and words for lexical decision at an SOA of 200 ms.

### 3.3.2 *Method*

#### *Participants*

Sixteen Dutch native speakers who did not take part in Experiment 1 were recruited as paid participants on Radboud University campus. All participants reported to have normal or corrected-to-normal vision and hearing as well as no speech or language impairments.

#### *Apparatus*

The apparatus was the same as in Experiment 1, except that participants were additionally equipped with closed headphones.

#### *Materials and Design*

The materials used in Experiment 1 were also used in Experiment 2. Additionally, 256 questions that had been pre-recorded by a male speaker were used. Each question asked for one of the pictures used in the experiment. Questions were of the format 'Which object that has property X also has property Y?' Example: 'Which object that grows on a tree is also edible?' The 64 questions that were used in switch trials were also used in a second version with the mentioned properties in a swapped order ('Which object that is edible also grows on a tree?') (Question Type: A/B). The same questions in these two types will also be used in Experiment 3, where the questions are relevant to the task

of participants and give away their answer either early or late. For now, however, participants were instructed to ignore the questions. Questions had a mean length of 3.74 seconds ( $SD = 0.39$  seconds).

Eight experimental lists were constructed, with a different word following a given critical picture in half of the lists, while a question of either type was played. The same words followed a given critical picture in the other half of the lists, while a question of the other type was played. Each participant was tested in one of the lists.

### *Procedure*

Each trial began with a fixation cross in the center of the screen while a question was played. Participants were instructed to completely ignore the questions. In the middle of the question, at the beginning of the phrase stating either the first (Question Type A) or the second property that was mentioned in the question (Question Type B), the picture corresponding to the question would replace the fixation cross and participants were instructed to name the picture as fast as possible. The picture disappeared upon voice onset and was replaced by a blank screen for 1 second before the next trial started. In lexical decision trials, the picture was replaced by a word after being presented for 200 ms (SOA). In these critical trials, participants were instructed to abandon the naming task and instead press one of two buttons indicating whether the word was a real Dutch word or not. Upon pressing a button, the word disappeared and was replaced by a blank screen for 1 second before the next trial started. Every sixty-four trials, a pause screen was presented, giving participants the chance to take a short break.

The experiment proper was preceded by eight practice trials and followed by a post test in which participants were shown the sixty-four critical pictures and asked to name them, so as to check whether their responses matched the expected names for the critical pictures. The whole experimental session lasted about 50 minutes.

### 3.3.3 *Results*

Inspecting the distribution of naming latencies for each subject, naming latencies of one subject were found to differ from those of the other

subjects in being bi-modally distributed, possibly indicating the use of a waiting strategy that diverges from normal production planning. Data of that subject were removed from analyses.<sup>4</sup> Of the remaining 2880 naming trials, 120 trials (4.2%) were regarded as erroneous and consequently discarded, as the voice key was triggered more than four seconds after picture onset. Another 101 trials (3.7%) were discarded because they were outliers of more than 2.5 standard deviations by subject. Remaining trials had a mean naming latency of 1225 ms (SD = 588 ms; CI = <1202, 1247>; Figure 3.4). The log-transformed naming latencies were analysed in a mixed effects model with Question Type as predictor. Naming latencies did not significantly differ between the two levels of Question Type ( $\beta = 0.007$ , SE = 0.006,  $t = 1.22$ ,  $F = 1.492$ ,  $p = .221$ ). An independent t-test comparing naming performance in Experiments 1 and 2 shows naming latencies to be significantly longer in Experiment 2 ( $t = -3.31$ ,  $df = 3578$ ,  $p < .001$ , CI = <17 ms, 65 ms>).

Of the 480 critical lexical decisions, 6 (1.3%) were discarded since participants did not name the corresponding critical pictures by their standard labels in the post test. Another 86 button press responses (18.1%) were erroneous (see Figure 3.5). Notably, more than twice as many errors were produced when words were presented after pictures with related (25.3%) as compared to unrelated names (11%).

Error rates were analyzed in a logit mixed effects regression model with Relatedness and Question Type as well as their interaction as predictors (see Table 3.6 in Supplementary Materials). Relatedness significantly affected error rates, with participants making more errors in related than in unrelated words ( $\beta = 1.404$ , SE = 0.407,  $z = 3.449$ ,  $p < .001$ ). The main effect of Question Type as well as its interaction with Relatedness turned out non-significant.

Erroneous trials were discarded for the following analyses of decision latencies. Moreover, 10 (2.6%) trials were discarded because their reaction latencies were outliers of more than 2.5 standard deviations by subject. The mean button press latency of the remaining 378 correct trials was 1382 ms (SD = 520 ms). See Figure 3.5 for decision latencies and error rates by condition.

Button press latencies of correct trials were analysed in a mixed effects model with Relatedness and Question Type as well as their

---

<sup>4</sup>Removal of this subject's data did not change the presented pattern of results, as attested in separate analyses.

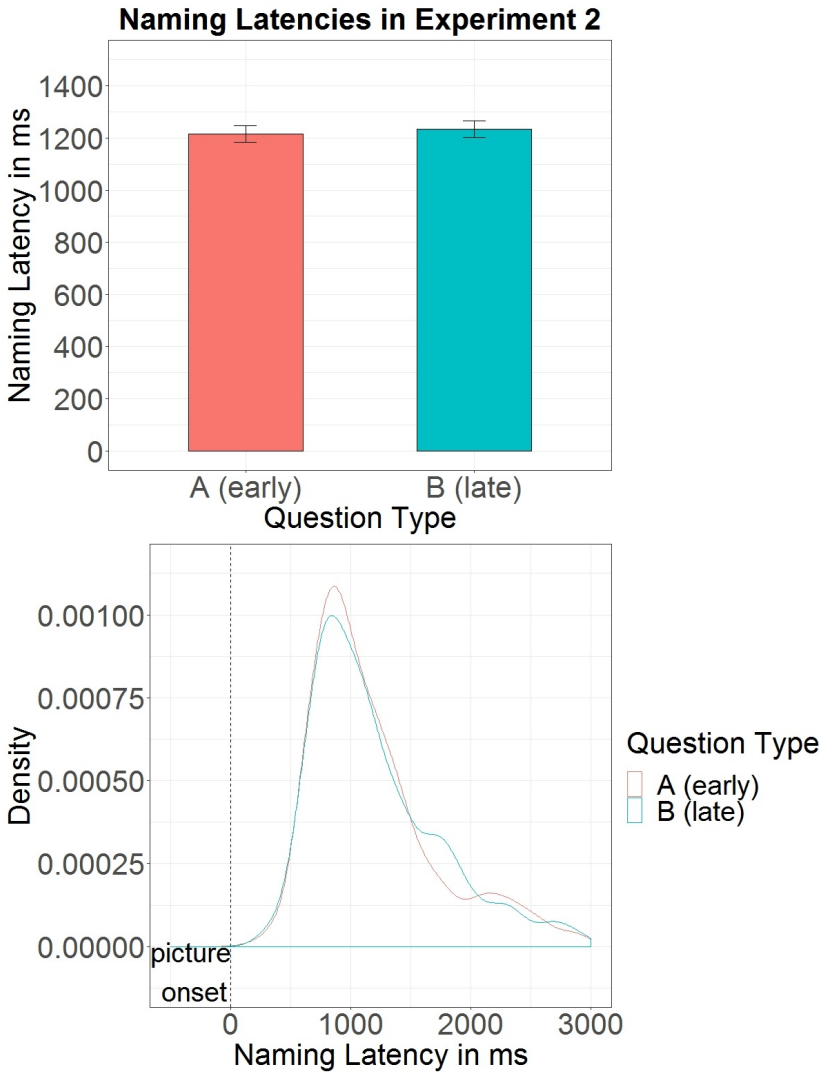


Figure 3.4: Naming latencies in Experiment 2. Bars represent 95% confidence intervals.

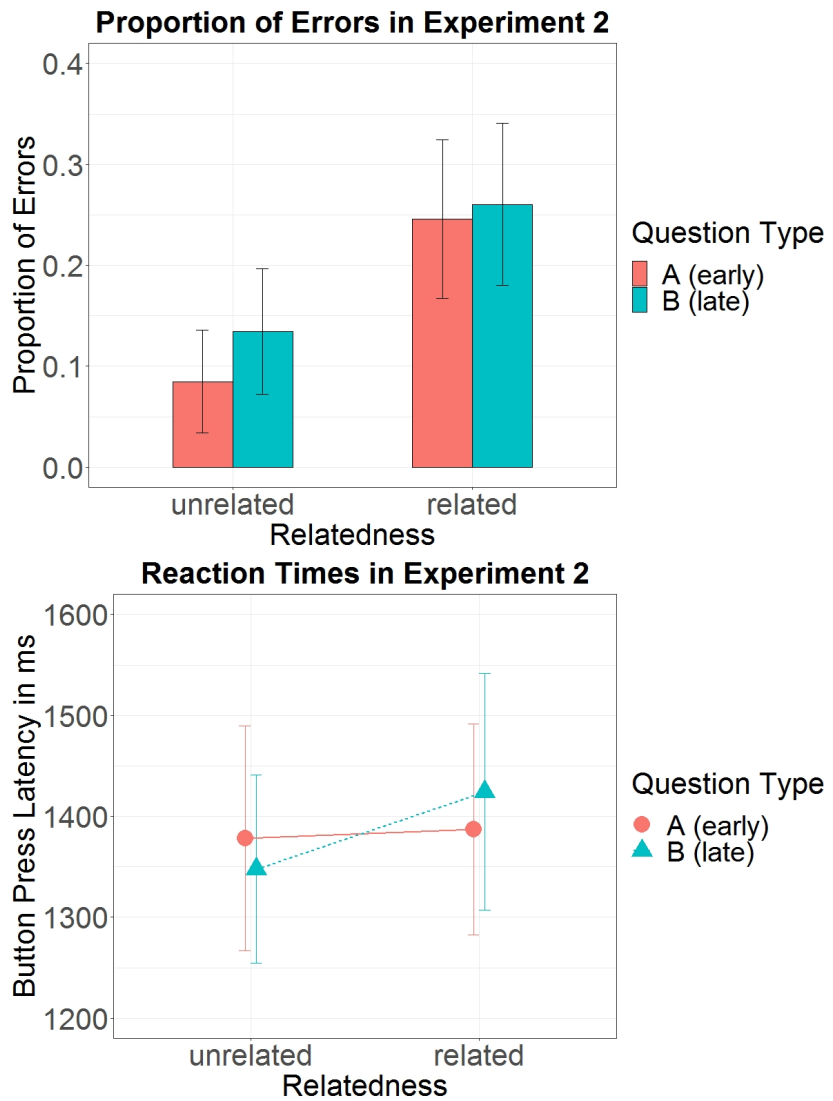


Figure 3.5: Reaction times and error rates in lexical decisions in Experiment 2. Bars represent 95% confidence intervals.

interaction as predictors (see Table 3.7 in Supplementary Materials). The main effect of Relatedness was not significant ( $\beta = 0.015$ ,  $SE = 0.014$ ,  $F = 1.904$ ,  $p = .302$ ), neither was there a significant interaction of Relatedness with Question Type ( $\beta = 0.013$ ,  $SE = 0.019$ ,  $F = 0.388$ ,  $p = .483$ ). The main effect of Question Type was also non-significant ( $\beta = 0.009$ ,  $SE = 0.009$ ,  $F = 0.002$ ,  $p = .345$ ).

To test for the reliability of the attested null results and to get an estimation of the distribution of probability of the observed relatedness effect, a Bayesian linear model was used to fit decision latencies, with Relatedness and Question Type as well as their interaction as predictors and maximal random effects structures for both subjects and items (see Table 3.8 in Supplementary Materials). A normal prior distribution for the expected effect of Relatedness was used, with the mean being the mean Relatedness effect observed in Experiment 1 (74 ms) and the tenfold standard deviation of that previously observed effect (250 ms), so as to make the prior moderately informative. A Bayesian inference test testing for the modeled effect of Relatedness yielded very weak evidence for the effect being higher than zero ( $\beta = 51$  ms,  $SE = 57$  ms,  $CrI = <-40$  ms, 148 ms>,  $BF = 4.68$ ). Similarly, a second test for the modeled interaction effect of Relatedness  $\times$  Question Type yielded very weak evidence for the effect being higher than zero ( $\beta = 71$  ms,  $SE = 82$  ms,  $CrI = <-65$  ms, 203 ms>,  $BF = 4.27$ ).

#### 3.3.4 Discussion

After phonological relatedness of picture names and words for lexical decision led to interference effects in response latencies and error rates in Experiment 1, the present Experiment was designed to replicate these results with distracting background speech. In particular, it was run to test whether the same SOA of 200 ms that led to phonological interference in Experiment 1 can be expected to yield comparable interference effects in the presence of distracting speech or whether response planning gets slightly delayed or slowed down.

Even though participants seemed to ignore the incoming questions, as evidenced by very similar naming latencies in both question types, the effects on lexical decision performance obtained in Experiment 1 could not be fully replicated. While the significant effect of Relatedness



on error rates indicates phonological interference in both early and late questions, Relatedness did not have a reliable effect on reaction times. As the results of the Bayesian analyses have not yielded decisive evidence for the presence or absence of a Relatedness effect, it is possible that a potential effect could not have been detected due to a lack of statistical power. It therefore remains unclear whether participants were planning their verbal responses phonologically during the incoming questions or not. However, as the naming latencies obtained in naming trials were on average 41 ms longer than in Experiment 1, it is likely that incoming speech slowed down the processes of response planning. That means that the results of Experiment 1 (with an SOA of 200 ms) might not fully replicate in the question-answer situation we aim to test in Experiment 3. For that reason, Experiment 3 was designed to use a longer SOA of 300 ms.

### 3.4 EXPERIMENT 3

#### 3.4.1 *Introduction*

Following Experiment 2, in which questions were presented as distracting background speech that had to be ignored by participants, Experiment 3 makes use of a dialogic task in order to investigate whether next speakers plan their utterance phonologically in overlap with the incoming turn. Participants have to attend to auditorily presented questions in order to be able to answer them. The questions ask for one out of four pictures of objects that are presented to participants. They are designed so that they give away their answer either in the middle of the question or only at their end. In that way, speech planning in overlap, which is expected in questions giving away the answer early, can be compared to speech planning in silence, which is expected after questions giving away the answer at their end. Participants' eye-gaze is used as an indicator for the initiation of response planning, assuming that speakers fixate the object they mentally process at a given moment (Barthel et al., 2016; Just & Carpenter, 1980; M. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) (see ch. 2). In a quarter of trials, the task switches after 300 ms of gaze falling on the target object and participants have to give a lexical decision instead of answering the

question. The relatedness of the lexical decision words and the target picture was manipulated in order to investigate whether participants already retrieved the word form of the picture name in overlap with the incoming question or not. If phonological planning was delayed until the end of the incoming turn, relatedness of the lexical decision word and the target picture name should have no effect on lexical decision performance. If, on the other hand, participants planned their answer phonologically already during the incoming question, phonological relatedness of the lexical decision word and the picture name should affect lexical decision performance. If the word form of the picture name was already retrieved by the time of the task switch, activation of the lexical decision word should be inhibited, leading to longer decision latencies and increased error rates.

#### 3.4.2 *Method*

##### *Participants*

Forty-five Dutch native speakers who did not take part in Experiments 1 and 2 were recruited as paid participants on Radboud University campus. All participants reported to have normal or corrected to-normal vision and hearing as well as no speech or language impairments. In thirteen participants tested in a first test session, more than 25% of the critical lexical decision trials were invalid, either due to trackloss of participants' gaze or because participants kept on naming the target picture even though they were instructed to abandon the naming task and switch to lexical decision. Data of these participants were discarded. The other thirty-two participants were tested in a second session on a second experimental list (see Section *Materials and Design* below), with at least one day between the two test sessions. In one of these participants, more than 25% of the critical trials of the second test session were invalid. This participant was replaced.

##### *Apparatus*

The apparatus was the same as in Experiment 2, except that participants' eye-movements were monitored with an SMI RED-m remote eye-tracker (120 Hz sampling rate).

### *Materials and Design*

The same 256 pictures of objects that were used in Experiments 1 and 2 were used as target pictures in Experiment 3, 192 in naming items (75%) and 64 in lexical decision items (25%). In each naming item, four pictures were displayed in the four corners of the screen, with white space between each of the pictures. Each picture was used as target in one naming item and served as a distractor picture in another three naming items. Similarly, in each lexical decision item, four pictures were displayed in the four corners of the screen, with one of the 64 critical pictures as the target picture in one of the display's corners. 192 additional pictures that had not been used in the previous Experiments were used as distractor pictures in the 64 lexical decision items, so that each critical picture would only be displayed once per test session. The position of the target picture on the screen was balanced across the experiment.

256 questions that were used in Experiment 2 were used in Experiment 3, each question asking for one of the target pictures. The questions were of the format "Which object that has property X also has property Y?" One of these properties was uninformative, as all four pictures on the display (target and distractors) carried that property. The other property was informative, since only the target picture carried that property. In lexical decision items, two versions of the respective question were used, with the order of the properties mentioned in the question being swapped between the two versions. The informative property was therefore available either early or late during the question (Question Type: early/late), as illustrated in the following example of a lexical decision trial with the pictures of an apple, a potato, a strawberry, and a broccoli, playing either the question "Which object that grows on a tree is also edible?" (Question Type: early) or the question "Which object that is edible also grows on a tree?" (Question Type: late). Naming items were coupled with only one question, half of the naming items using early questions and the other half using late questions. The questions had a mean length of 3.74 seconds ( $SD = 0.39$  seconds). See Table 3.11 in Supplementary Materials for a list of Materials.

In lexical decision trials (25%), the four pictures were replaced by a word appearing at the position of the target picture. The same words for lexical decision that were used in Experiments 1 and 2 were re-used

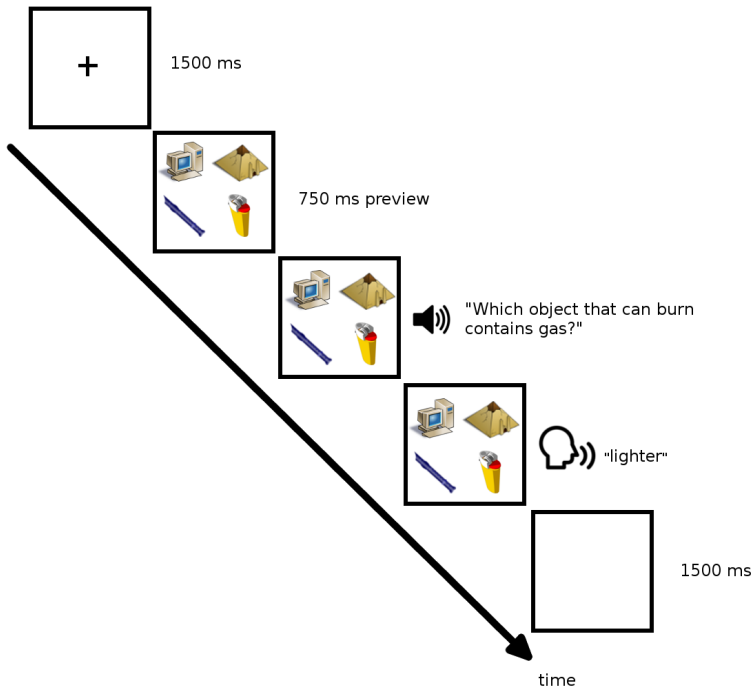
in Experiment 3, with half of the words being real Dutch words, the other half being pseudowords. The words were either presented after target pictures whose name was phonologically related or after pictures whose name was unrelated (Relatedness: unrelated/related).

Eight experimental lists were constructed, with a different word following a given critical picture in half of the lists, while a question of either type was played. The same words followed a given critical picture in the other half of the lists, while a question of the other type was played. Each participant was tested in two of the lists, with at least one day between the two test sessions.

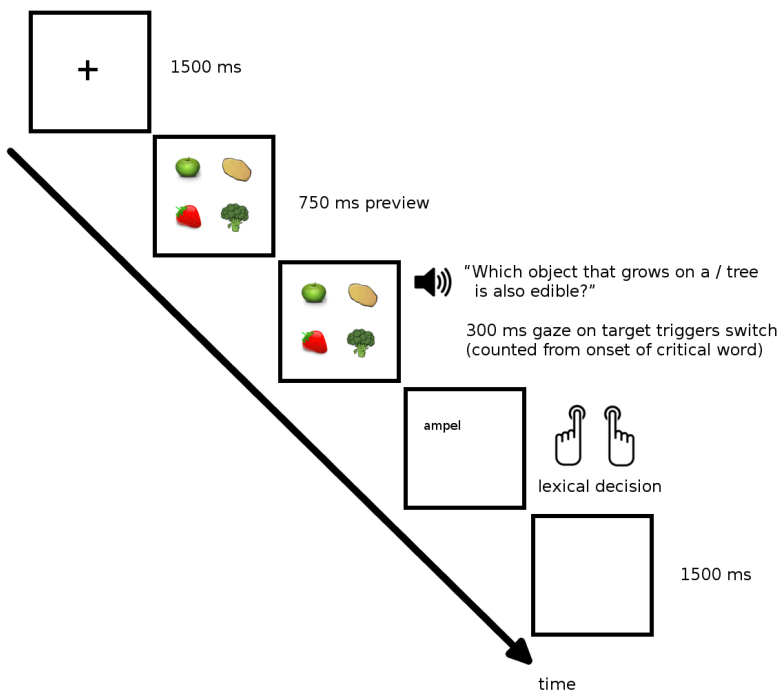
### *Procedure*

Each trial began with a fixation cross for 1.5 seconds to attract participants' gaze to the center of the screen (see Figure 3.6). The fixation cross was replaced by a display showing four pictures of objects in the four corners of the screen. The pictures were approximately  $450 \times 450$  pixels large and occupied about 2.5 degrees of participants' visual angle. 750 ms after the pictures appeared, a question was played. In naming trials (75%), participants had to answer the question as fast as possible by naming one of the displayed objects. Upon voice onset, the pictures would be replaced by a blank screen for 1.5 seconds before the next trial would start. In switch trials (25%), the pictures were replaced by a word that would appear in the position of the target object as soon as the participant's gaze would dwell on the target object for 300 ms (SOA), measured from the onset of the informative part of the particular question in a given trial. That part of the question began on average either after 1.34 seconds (in early questions;  $SD = 0.35$  seconds) or after 2.96 seconds (in late questions;  $SD = 0.44$  seconds). In these switch trials, participants were to abandon the naming task and switch to deciding whether the presented word was a real Dutch word or not and give their response by pressing one of two buttons as fast as possible (with the 'word' response on the right button). Upon button press, the word would be replaced by a blank screen for 1.5 seconds before the next trial would start. Every sixty-four trials, a pause screen was presented, giving participants the chance to take a short break. At the beginning of the experiment, as well as after each of the short breaks, the eye-tracker was calibrated on nine points of the screen.

The experiment proper was preceded by eight practice trials and followed by a post test in which participants were shown the sixty-four critical pictures and asked to name them in order to check whether their responses matched the expected names for the critical pictures. The whole experimental session lasted about 50 minutes.



(a) naming trial



(b) switch trial

Figure 3.6: Timelines of trials in Exp. 3 exemplified with translations of early questions. / = beginning of the informative word in the question. Dutch originals: "Welk object dat kan branden, bevat gas?" in naming trial and "Welk object dat aan een boom groeit is ook eetbaar?" in switch trial.

### 3.4.3 Results

Inspecting the distribution of naming latencies for each subject, naming latencies of one subject were found to differ from those of the other subjects in being bi-modally distributed, possibly indicating the use of a waiting strategy that diverges from normal production planning.<sup>5</sup> 390 naming trials (3.3%) were regarded as erroneous reactions as they triggered the voice key either more than two seconds before the end of the question (when the answer could not yet have been known) or more than four seconds after the end of the question and were consequently discarded. Another 179 trials (1.6%) were discarded because they were outliers of more than 2.5 standard deviations by subject and test session. The remaining 11342 naming trials had a mean naming latency of 919 ms (SD = 848 ms; Figure 3.7), measured from the end of the question. A mixed effects model on log-transformed naming latencies with Question Type as predictor revealed a significant main effect of Question Type, with naming latencies being shorter in early question trials than in late question trials ( $\beta = 0.006$ , SE = 0.002,  $F = 4.258$ ,  $df = 54$ ,  $p < .05$ ).

Of the 1984 critical lexical decision trials, 37 (1.8%) were discarded because participants' name for the respective target objects in the post test did not match the standard label for the object. Moreover, 84 (4.3%) trials were discarded due to trackloss of participants' gaze direction during the trial. 166 (8.9%) critical trials were discarded because participants overtly named at least part of the target picture, contrary to instructions. Of the remaining trials, 354 (20.8%) decisions were erroneous (see Figure 3.8). Error rates in related trials (26.5%) were 81% higher than in unrelated trials (15%).

Error rates were analyzed in a logit mixed effects regression model with Relatedness and Question Type as well as their interaction and Test Session (1/2) as predictors (see Table 3.9 in Supplementary Materials). Participants made marginally significantly less errors in the second as compared to the first test session, indicating a practice effect between test sessions ( $\beta = -0.333$ , SE = 0.177,  $z = -1.876$ ,  $p = .06$ ). While the main effect of Question Type and its interaction with Relatedness are non-significant, the main effect of Relatedness turns out significant ( $\beta = 0.832$ ,

---

<sup>5</sup>Removal of this subject's data did not change the presented pattern of results, as attested in separate analyses.

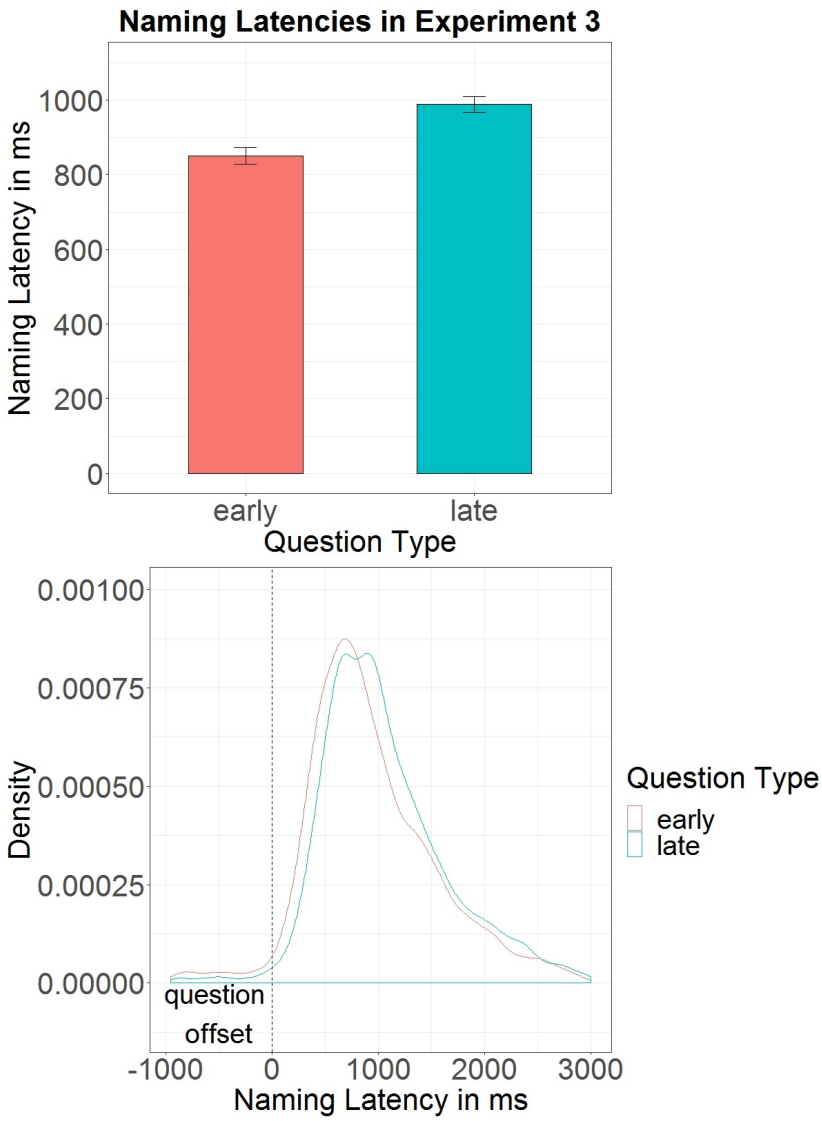


Figure 3.7: Naming latencies in Experiment 3. Bars signify 95% confidence intervals.



SE = 0.135,  $z = 6.134$ ,  $p < .001$ ), with participants making more errors when words for lexical decision are presented after target pictures with related names than when they are presented after target pictures with unrelated names.

Erroneous trials were discarded for the following analyses of decision latencies. Furthermore, 33 (2.4%) trials were discarded because they were outliers of more than 2.5 standard deviations per subject and test session. The mean button press latency (RT) in the remaining 1311 correct lexical decision trials was 1051 ms (SD = 315 ms; Figure 3.8).

Participants triggered the change of display on average after 2417 ms in early questions and after 3888 ms in late questions. Given that the questions had a mean length of 3.74 seconds, displays were generally changed in overlap in early questions and in silence in late questions.

The log-transformed button press latencies were analysed in a mixed effects model with Relatedness, Question Type and Test Session (1/2) as well as the interaction of Relatedness and Question Type as predictors (see Table 3.10 in Supplementary Materials). Test Session significantly affects decision times, with participants taking faster decisions in the second test session than in the first test session, showing a training effect between sessions ( $\beta = -0.068$ , SE = 0.008,  $F = 61.597$ ,  $df = 30$ ,  $p < .001$ ). The main effect of Relatedness turns out significant ( $\beta = 0.028$ , SE = 0.005,  $F = 23.286$ ,  $df = 42$ ,  $p < .001$ ), with decisions being slower when words for lexical decision are presented after target pictures with related names than when they are presented after target pictures with unrelated names. Question Type ( $\beta = 0.001$ , SE = 0.007,  $F = 0.046$ ,  $df = 40$ ,  $p = .832$ ) as well as its interaction with Relatedness ( $\beta = -0.001$ , SE = 0.011,  $F = 0.012$ ,  $df = 39$ ,  $p = .912$ ) turn out non-significant in the model.

In order to test for the likelihood distribution of the obtained effect of Relatedness on reaction times and the evidence for the absence of an interaction effect of Relatedness  $\times$  Question Type, a Bayesian linear model was used to fit decision latencies, with Relatedness and Question Type as well as their interaction and Test Session as predictors and maximal random effects structures for both subjects and items (see Table 3.3). Based on the obtained Relatedness effects in Experiments 1 and 2, we set a normally distributed prior with the mean of the previously observed effects (66 ms) and the tenfold SD of these effects (210 ms), in order to make the prior moderately informative. A Bayesian inference test based on the model revealed substantial evidence for the absence of an

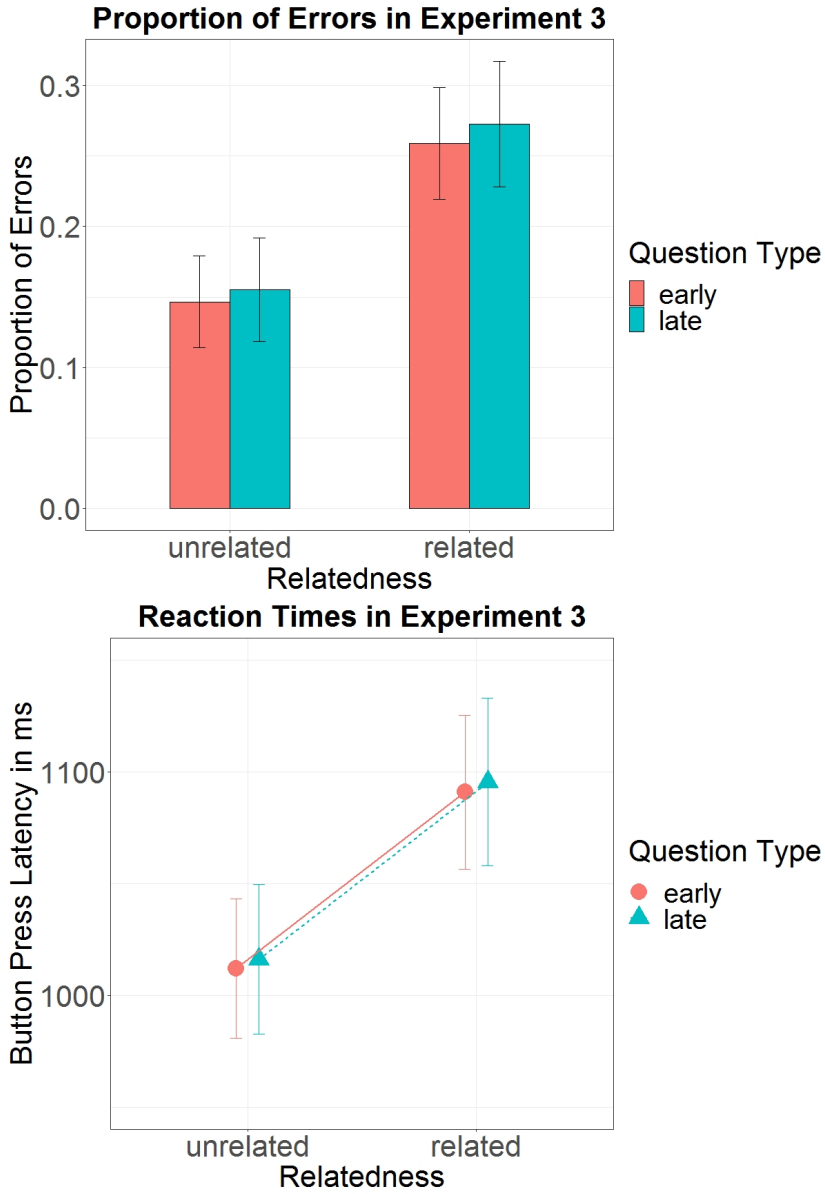


Figure 3.8: Reaction times and error rates in lexical decisions in Experiment 3. Bars signify 95% confidence intervals.

interaction effect of Relatedness  $\times$  Question Type ( $\beta = 1$  ms, SE = 30 ms, CrI =  $\langle -58$  ms, 60 ms  $\rangle$ , BF = 7.15), indicating that the effect of Relatedness does not differ between early and late questions. A second Bayesian inference test yielded decisive evidence for the observed main effect of Relatedness to be higher than zero ( $\beta = 72$  ms, SE = 16 ms, CrI =  $\langle 46$  ms, 98 ms  $\rangle$ , BF = inf; see Figure 3.9), indicating that decision latencies were longer when words were presented after pictures with phonologically related names than when they were presented after unrelated pictures.

	$\beta$	SE	lower CrI	upper CrI
Intercept	1156.30	46.41	1063.81	1250.13
Test Session	-173.97	27.45	-227.47	-118.66
Relatedness	71.81	15.93	40.64	102.63
Question Type	2.87	22.99	-40.91	48.31
Rel. $\times$ Question Type	0.91	30.15	-57.71	60.01

Table 3.3: Bayesian linear regression model on button press latencies in Experiment 3. Credible intervals contain 95% area under the posterior likelihood distribution. Model formula = `brm(1 + Test.Session + Relatedness * Question.Type + (1 + Test.Session + Relatedness * Question.Type | subject) + (1 + Test.Session + Relatedness * Question.Type | item))`.

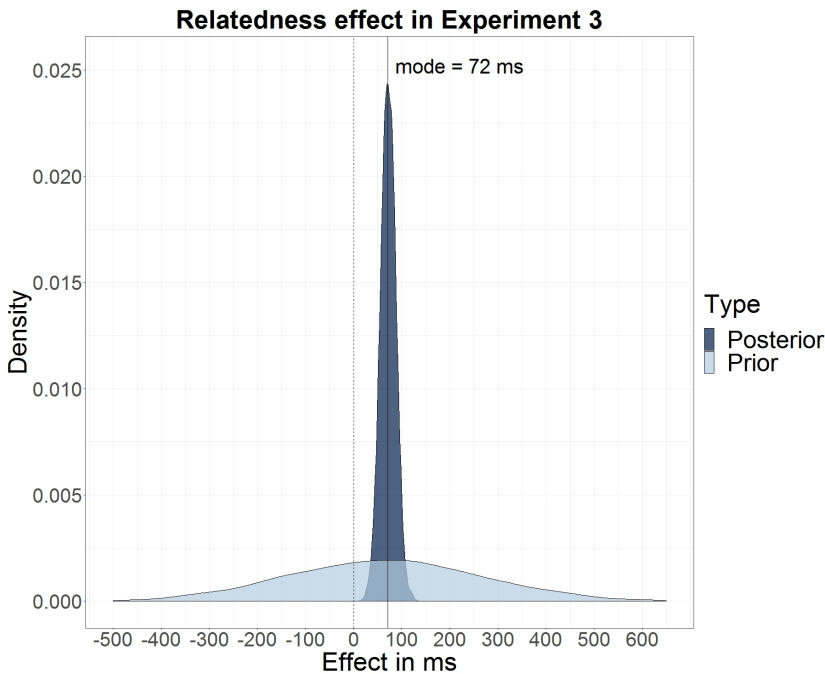


Figure 3.9: Prior and posterior distributions of Relatedness effect on lexical decision latencies in Experiment 3 drawn from Bayesian linear regression model. Prior distribution was informed by the observed effects of Relatedness in Experiments 1 and 2.

#### 3.4.4 Discussion

Experiment 3 tested the time course of speech production planning in overlap with an incoming turn that requires a response. In each incoming turn, participants heard a question they had to answer by naming one of four pictures. These questions either gave away their answer already in the middle of the question or only at their end. Naming latencies were shorter when the answer to the question became clear earlier. This finding is taken to indicate that participants profited from planning in overlap with early questions, replicating previous results (Barthel et al., 2016; Bögels, Magyari, & Levinson, 2015) (see ch. 2).

In a quarter of trials, participants had to attend to a switch task midway through preparing their verbal response and make a lexical decision instead of answering the question. The words for lexical decision appeared shortly after participants' gaze moved towards the target

picture and were either phonologically related to the verbal response in preparation, i.e. the target picture's name, or not. Phonological relatedness led to longer decision latencies and increased error rates. Importantly, the effect of phonological interference was equally strong in questions giving away the answer in the middle of the question and in questions giving away the answer only at their end, indicating that participants were planning their response phonologically as early as possible and already before the incoming question came to an end.

### 3.5 GENERAL DISCUSSION

Previous research into dialogic turn-taking has shown that next speakers regularly start to plan their turn as early as possible and often in overlap with the incoming turn (Barthel et al., 2016; Bögels, Magyari, & Levinson, 2015) (see ch. 2). While the timing of speech planning and turn taking is certainly dependent on speakers' communicative intentions and therefore under some amount of strategic control, the early-planning strategy leads to advantages in turn-timing, as next speakers manage to shorten the gaps between turns when they start planning their response in overlap (Barthel et al., 2017; Corps et al., 2018) (see ch. 5). While a number of studies have shown that next speakers initiate planning in overlap, they did not investigate which steps of response preparation are run through while the current turn is still coming in and potentially interfering with simultaneously running planning processes. Planning in overlap comes at the cost of increased processing load (Barthel & Sauppe, 2019, see ch. 4), which might cause next speakers to postpone certain processing stages until the end of the incoming turn in order to avoid high peaks in processing load.

This study investigated which steps of language planning occur in overlap with the incoming turn, and focused on attesting phonological activation of planned words in the course of the tested experiments. To that end, participants were tested in a switch task combining picture naming and visual lexical decision in a series of three experiments. In Experiment 1, participants had to name pictures as a base task in three quarters of trials and switch to lexical decision instead of naming the picture in one quarter of trials. Effects of associative and phonological relatedness of the pictures lexical decision words on

decision performance indicate that participants prepared their verbal response at least until activating the phonological representation of the picture name until they gave a lexical decision. Experiment 2 was designed as a replication of Experiment 1 while participants were presented with background speech while doing the switch task. In that manner, we investigated whether the effects observed in Experiment 1 can be expected to be replicated in a dialogic test situation using the same SOA's. As relatedness effects could not be fully replicated with background speech, and naming latencies were longer as compared to Experiment 1, response planning might have been slowed down by distracting incoming speech. For that reason, Experiment 3 was designed to use a longer SOA of 300 ms.

In Experiment 3, participants were tested in a responsive test situation in which they had to answer questions by naming one of four pictures. The cue to the answer of the questions was located either early during the question or only towards the end of the question. In critical trials, participants again had to switch from giving a verbal response to the question to making a lexical decision instead. The timing of this switch was tied to the beginning of response planning, which we operationalised as eye-gaze towards the target object triggering the presentation of the lexical decision word (Just & Carpenter, 1980). In line with findings in the previous literature, participants were shown to initiate response planning as early as possible, usually in overlap with the incoming turn (Barthel et al., 2017, 2016; Bögels, Magyari, & Levinson, 2015; Corps et al., 2018) (see ch. 2 and 5). Words for lexical decision were presented after phonologically related target pictures or after unrelated pictures. Comparing the effects of relatedness of the (initially intended) verbal responses with the lexical decision words allowed us to draw inferences about the progress of speech planning at the moment the lexical decision is given. Phonological relatedness led to deteriorated lexical decision performance, showing that the word forms of the picture names had been activated by the time the task switched. Critically, this phonological interference effect was shown to be equally strong in the middle of questions and at their end. These results are taken as evidence that next speakers plan their utterance phonologically as early as possible while the incoming turn is still unfolding.

In conclusion, we have shown that language production planning proceeds right through to word form retrieval even during the incoming speech that is being responded to. While naming latencies in the presented experiments are generally longer than average turn-transition times in conversational settings, the attested effects can be taken as informative about the processes of speech planning in conversation, where context, topic familiarity, predictable sequences of actions and the like speed up turn taking. The presented results support models of the psycholinguistics of dialogue that model response planning as taking place as early as possible (Heldner & Edlund, 2010; Levinson & Torreira, 2015), showing that early planning at least includes the stages of conceptual planning and formulation, even though processing costs in response preparation are higher in overlap with the incoming turn than during the silence between turns (Barthel & Sauppe, 2019, see ch. 4). A recent study by Bögels and Levinson (in prep.), measuring tongue movements using ultrasound visualization, shows that articulatory preparation does not happen as soon as possible but is postponed until articulation becomes immediate. Combining this finding with the present results, the question remains whether early response preparation includes the retrieval and construction of phonetic codes and their translation into motor plans while only the movement of the articulators is postponed or whether phonetic and motor planning stages are postponed and triggered by the incoming turn coming to an end, or possibly by the recognition of turn-final cues (Barthel et al., 2017, see ch. 5). At least up to word form retrieval, the time course of production planning in a dialogue situation seems to be very similar to a monological test situation like the picture naming task used in Experiment 1. However, planning seems to be somewhat slower in overlap with the incoming turn as compared to planning in silence due to increased cognitive load when comprehension and response preparation run in parallel.

## 3.6 SUPPLEMENTARY MATERIALS

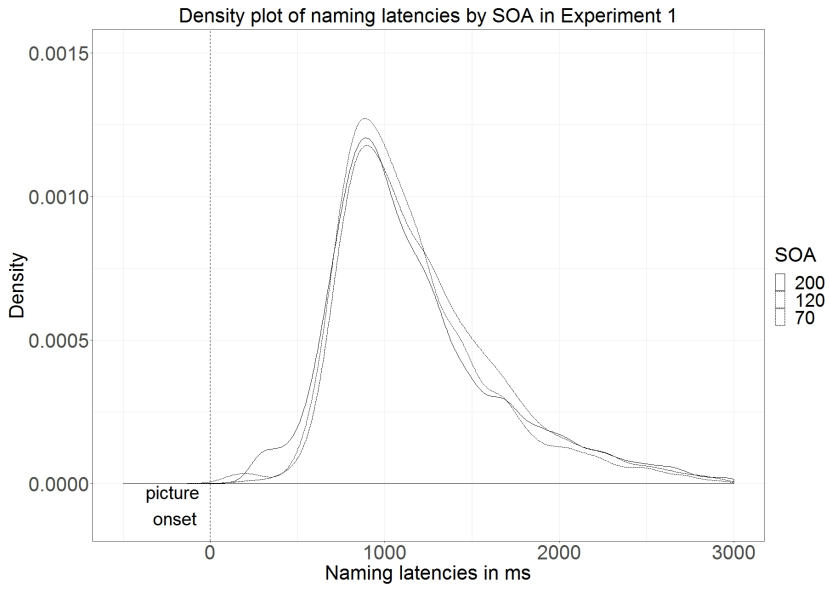
*Figures*

Figure 3.10: Distribution of naming latencies by SOA condition in Experiment 1.



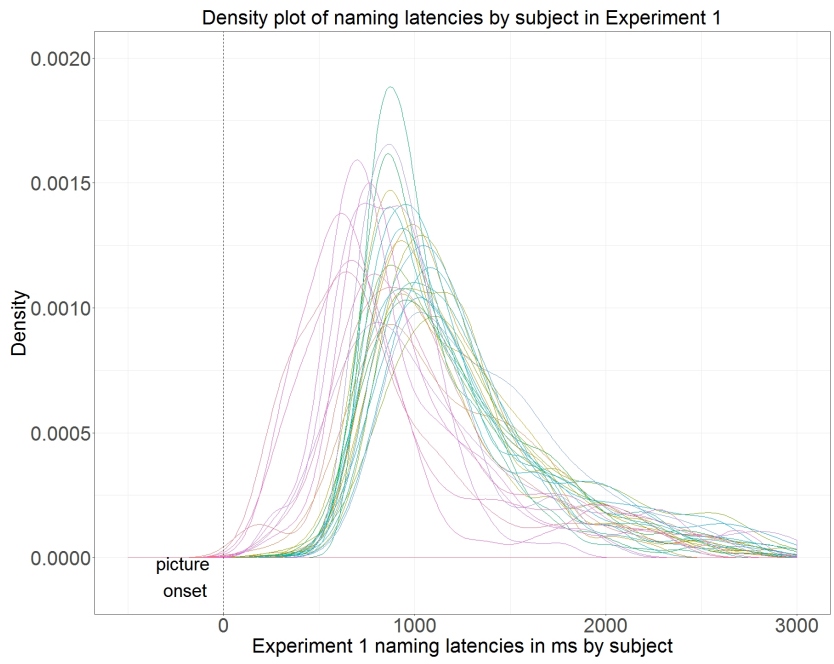


Figure 3.11: Distribution of naming latencies by subject in Experiment 1.

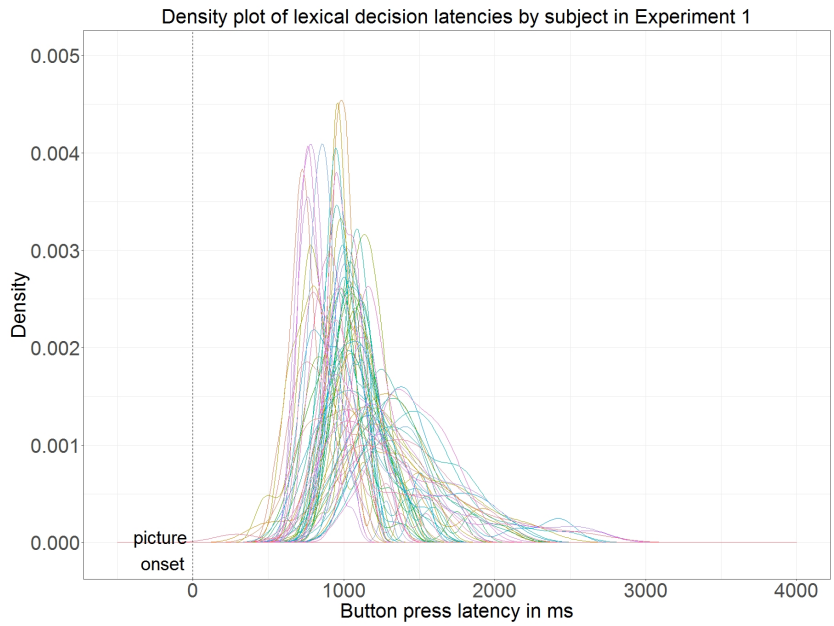


Figure 3.12: Distribution of lexical decision latencies by subject in Experiment 1.

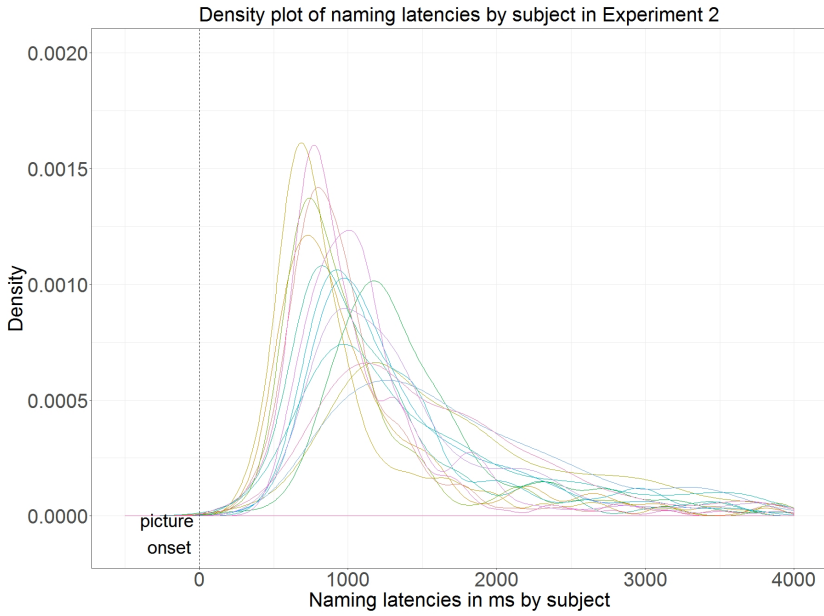


Figure 3.13: Distribution of naming latencies by subject in Experiment 2.

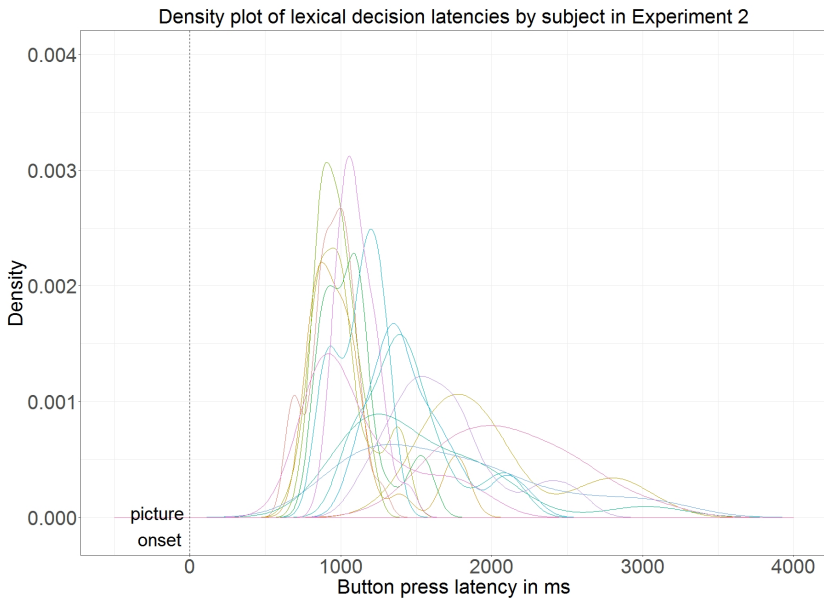


Figure 3.14: Distribution of lexical decision latencies by subject in Experiment 2.

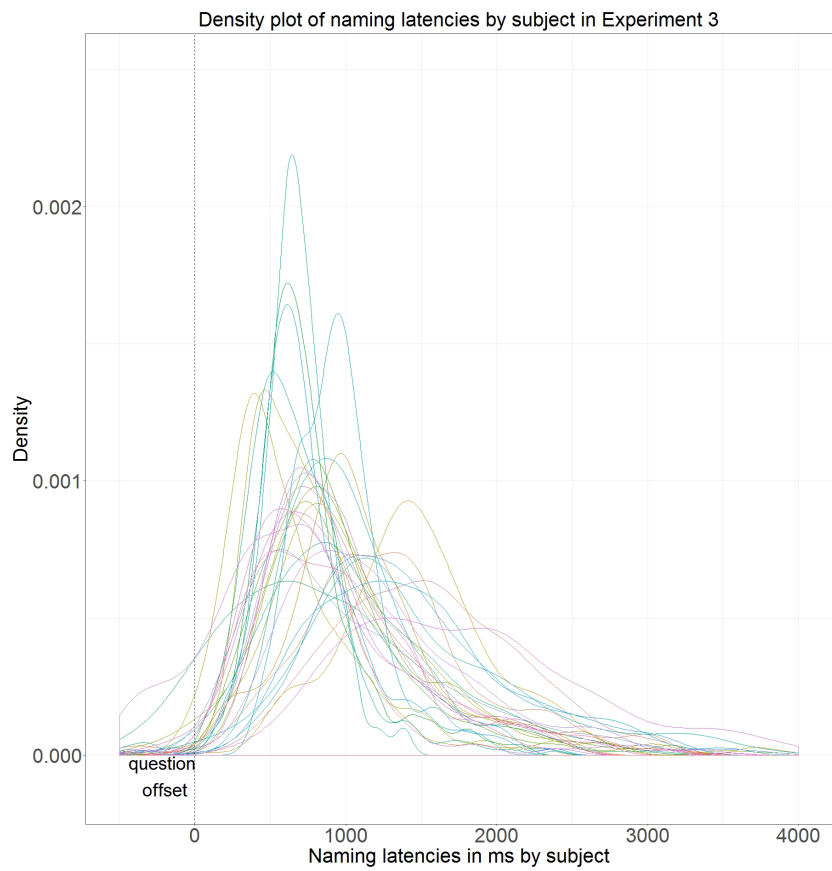


Figure 3.15: Distribution of naming latencies by subject in Experiment 3.

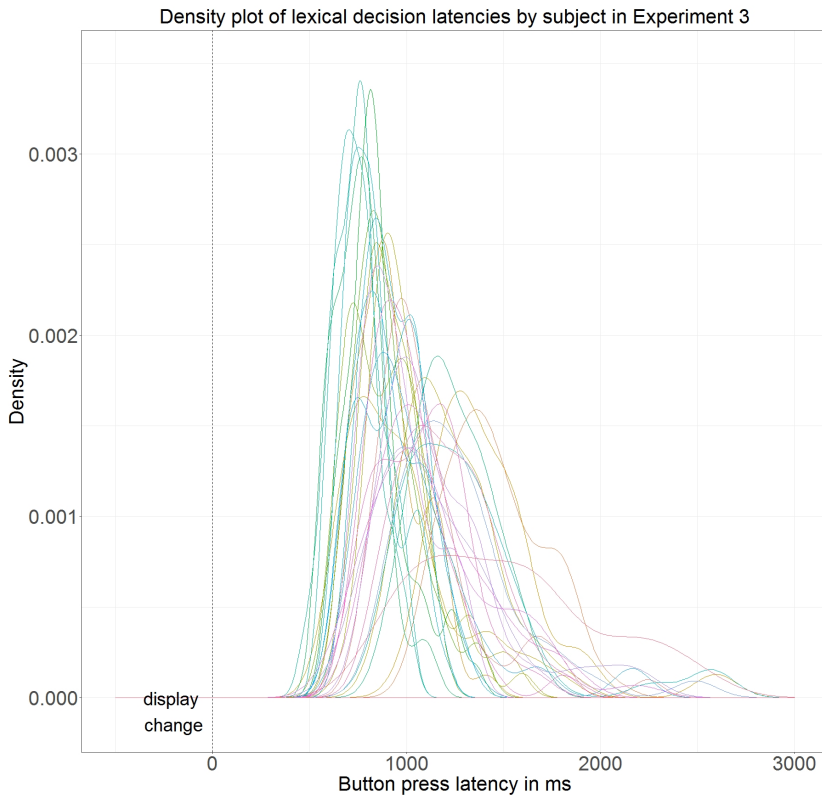


Figure 3.16: Distribution of lexical decision latencies by subject in Experiment 3.

*Tables*

	$\beta$	SE	z	p
Intercept	-2.808	0.208	-13.498	<.001 ***
SOA70	-0.098	0.490	-0.201	.840
SOA120	0.940	0.407	2.307	<.05 *
Relatedness	0.447	0.233	1.913	.055 .
Type of Relation	2.033	0.237	8.568	<.001 ***
SOA70 × Relatedness	-0.103	0.662	-0.156	.875
SOA120 × Relatedness	-0.085	0.517	-0.165	.869
SOA70 × Type of Relation	-0.152	0.664	-0.229	.819
SOA120 × Type of Relation	-0.396	0.521	-0.760	.447
Relatedness × Type of Relation	1.011	0.467	2.164	<.05 *
SOA70 × Rel. × Type of Rel.	0.240	1.325	0.182	.855
SOA120 × Rel. × Type of Rel.	0.334	1.034	0.323	.746

Table 3.4: Logit mixed effects regression model on error rates in Experiment 1. For comparison of the presented effects of SOA, SOA200 was used as a baseline. Model formula = `glmer(Correctness ~ 1 + SOA * Relatedness * Type.of.Relatedness + (1 | subject) + (1 | item))`.

	$\beta$	SE	$t$	$F$	df	$p$
Intercept	3.034	0.009	305.26			
SOA70	0.019	0.026	0.74	1.336	57 (2)	.270
SOA120	0.037	0.023	1.62			
Relatedness	0.008	0.004	1.96	3.854	1422	<.05 *
Type of Relation	0.037	0.004	8.87	78.569	53	<.001 ***
SOA70 $\times$ Relatedness	-0.008	0.010	-0.80	0.448	1408 (2)	.614
SOA120 $\times$ Relatedness	-0.009	0.009	-0.92			
SOA70 $\times$ Type of Relation	0.005	0.011	0.50	0.217	53 (2)	.805
SOA120 $\times$ Type of Relation	0.001	0.009	-0.07			
Rel. $\times$ Type of Rel.	0.040	0.008	4.82	23.196	1422	<.001 ***
SOA70 $\times$ Rel. $\times$ Type of Rel.	-0.010	0.021	-0.49	0.125	1411 (2)	.882
SOA120 $\times$ Rel. $\times$ Type of Rel.	-0.006	0.019	-0.32			

Table 3.5: Mixed effects regression model on log-transformed button press latencies in Experiment 1. For comparison of the presented effects of SOA, SOA200 was used as a baseline. Model formula = `lmer(LogLatency ~ 1 + SOA * Relatedness * Type.of.Relation + (1 + Type.of.Relation | subject) + (1 | item))`.

	$\beta$	SE	$z$	$p$
Intercept	-2.248	0.370	-6.070	<.001 ***
Relatedness	1.404	0.407	3.449	<.001 ***
Question Type	0.408	0.305	1.335	.181
Rel. $\times$ Q.Type	-0.617	0.606	-1.017	.309

Table 3.6: Logit mixed effects regression model on error rates in Experiment 2. Model formula = `glmer(Correctness ~ Position * Relatedness + (1 + Relatedness | subject) + (1 | item))`.

	$\beta$	SE	$t$	$F$	df	$p$
Intercept	3.110	0.030	102.39			
Relatedness	0.015	0.014	1.07	1.904	13	0.302
Question Type	0.009	0.009	0.94	0.002	310	0.345
Rel. $\times$ Q.Type	0.013	0.019	0.70	0.388	317	0.483

Table 3.7: Mixed effects regression model on log-transformed button press latencies in Experiment 2. Model formula = `lmer(LogLatency ~ 1 + Relatedness * Position + (1 + Relatedness | subject) + (1 | item))`.

	$\beta$	SE	lower CrI	upper CrI
Intercept	1361.86	106.77	1151.71	1577.53
Relatedness	51.09	57.59	-59.81	170.08
Question Type	16.98	41.29	-63.83	97.36
Relatedness $\times$ Question Type	70.79	81.56	-91.98	230.67

Table 3.8: Bayesian linear regression model on button press latencies in Experiment 2. Credible intervals contain 95% area under the posterior likelihood distribution. Model formula = `brm(1 + Relatedness * Question.Type + (1 + Relatedness * Question.Type | subject) + (1 + Relatedness * Question.Type | item))`.

	$\beta$	SE	$z$	$p$
Intercept	-1.542	0.202	-7.628	<.001 ***
Question Type	0.111	0.134	0.830	.406
Relatedness	0.832	0.135	6.134	<.001 ***
Test Session	-0.333	0.177	-1.876	.060 .
Question Type $\times$ Relatedness	0.008	0.268	0.031	.975

Table 3.9: Logit mixed effects regression model on error rates in Experiment 3. Model formula = `glmer(Correctness ~ Relatedness * Question.Type + Test.Session + (1 + Test.Session | subject) + (1 | item))`.

	$\beta$	SE	$t$	$F$	df	$p$
Intercept	3.047	0.016	185.21			
Test Session	-0.068	0.008	-7.87	61.597	30	<.001 ***
Relatedness	0.028	0.005	4.87	23.286	42	<.001 ***
Question Type	0.001	0.007	0.21	0.046	40	.832
Rel. $\times$ Q. Type	-0.001	0.011	-0.112	0.012	39	.912

Table 3.10: Mixed effects regression model on log-transformed button press latencies in Experiment 3. Model formula = `lmer(LogLatency ~ 1 + Test.Session + Relatedness * Question.Type + (1 + Test.Session + Relatedness * Question.Type | subject) + (1 + Relatedness * Question.Type + Test.Session | item))`.



## List of Materials

item ID	target object	distractor objects	phon. rel. word	phon. unrel. word	associated word*	non-associated word*	question** (Welk object ... (Which object ...))
01	ballon (balloon)	donut, zakhorloge, lp (donut, pocket watch, record)	balkon (balcony)	mits (given)	lucht (air)	warm (warm)	... dat kan vliegen heeft een ronde vorm? (... that can fly is round shaped?)
02	muts (beanie)	handschoen, stropdas, broek (glove, tie, pants)	mits (given)	balkon (balcony)	warm (warm)	lucht (air)	... dat je op je hoofd zet is een kledingstuk? (... that you put on your head is a garment?)
03	appel (apple)	aardappel, aardbei, broccoli (potato, strawberry, broccoli)	ampel (traffic light)	zaad (seed)	fruit (fruit)	hout (wood)	... dat aan een boom groeit is ook eetbaar? (... that grows on a tree is also edible?)
04	zaag (saw)	schroevendraaier, hamer, schep (screwdriver, hammer, shovel)	zaad (seed)	ampel (traffic light)	hout (wood)	fruit (fruit)	... dat aan een boom groeit is ook eetbaar? (... that grows on a tree is also edible?)
05	auto (car)	fiets, luchtballon, speedboot (bicycle, hot air balloon, speedboat)	aura (aura)	bazuin (trumpet)	rijden (to drive)	krom (crooked)	... met vier wielen wordt gebruikt als middel voor transport? (... with four wheels is used as a means of transport?)
06	banaan (banana)	limoen, ananas, sinaasappel (lime, pineapple, orange)	bazuin (trumpet)	aura (aura)	krom (crooked)	rijden (to drive)	... dat geel is is een exotisch stuk fruit? (... that is yellow is an exotic fruit?)
07	batterij (battery)	ventilator, computerscherm, gloeilamp (fan, computer screen, light bulb)	bakkerij (bakery)	volk (people)	auto (car)	mes (knife)	... dat opgeladen kan worden werkt op elektriciteit? (... that can be charged works with electricity?)
08	vork (fork)	spuit, drillboor, pikhouweel (sprayer, jackhammer, pickaxe)	volk (people)	bakkerij (bakery)	mes (knife)	auto (car)	... dat gebruikt wordt tijdens etenstijd is erg puntig? (... that is used at mealtimes is very spiky?)
09	tomaat (tomato)	aubergine, komkommer, paprika (eggplant, cucumber, pepper)	totaal (totally)	prinses (princes)	rood (red)	papier (paper)	... dat rond is is een lokaale groente? (... that is round shaped is a local vegetable?)
10	printer (printer)	stofzuiger, broodrooster, televisie (vacuum cleaner, toaster, television)	prinses (princes)	totaal (totally)	papier (paper)	rood (red)	... dat gebruikt wordt op kantoor is een elektronisch apparaat? (... that is used in the office is an electronic device?)
11	kiwi (kiwi)	suikerspin, broodje, radijs (cotton candy, bun, radish)	kilo (kilo)	pek (pitch)	groen (green)	hoed (hat)	... dat groen is is ook eetbaar? (... that is green is also edible?)
12	pet (cap)	trui, t-shirt, trouwjurk (sweater, t-shirt, wedding dress)	pek (pitch)	kilo (kilo)	hoed (hat)	groen (green)	... dat beschermt tegen de zon is een kledingstuk? (... that is a piece of garment protects from the sun?)

13	bom (bomb)	pistool, val, kampvuur (gun, trap, campfire)	bof (mumps)	paranoot (brazil nut)	oorlog (war)	regen (rain)	... dat makkelijk kan ontploffen is erg gevaarlijk? (... that can easily explode is very dangerous?)
14	paraplu (umbrella)	zaklamp, barbecue, bergschoen (flashlight, barbecue, mountain shoe)	paranoot (brazil nut)	bof (mumps)	regen (rain)	oorlog (war)	... dat je droog houdt wordt doorgaans buiten gebruikt? (... that keeps you dry is usually used outside?)
15	kaas (cheese)	kokosnoot, ei, framboos (coconut, egg, raspberry)	kaak (jaw)	boel (a lot)	geel (yellow)	lezen (to read)	... dat gemaakt is van melk is ook eetbaar? (... that is made from milk is also edible?)
16	boek (book)	pen, scherm, gum (pen, screen, eraser)	boel (a lot)	kaak (jaw)	lezen (to read)	geel (yellow)	... dat letters bevat wordt vaak gebruikt op school? (... that contains letters is often used at school?)
17	slee (sled)	barkruk, stoel, bankje (barstool, chair, bench)	snee (cut)	schaap (sheep)	sneeuw (snow)	knippen (to cut)	... dat kan glijden is gemaakt om op te zitten? (... that you can slide with is made to sit on?)
18	schaar (scissors)	trechter, bijl, spade (funnel, ax, spade)	schaap (sheep)	snee (cut)	knippen (to cut)	sneeuw (snow)	... dat gebruikt wordt bij de kleermaker is een stuk gereedschap? (... that is used by the tailor is a tool?)
19	cactus (cactus)	palmboom, bloem, plant (palm tree, flower, plant)	campus (campus)	zwaai (wave)	woestijn (desert)	wit (white)	... dat punten heeft is een plant? (... that has spikes is a plant?)
20	zwaan (swan)	giraffe, olifant, arend (giraffe, elephant, eagle)	zwaai (wave)	campus (campus)	wit (whilte)	woestijn (desert)	... dat in het water zwemt is een dier? (... that swims on the water is an animal?)
21	tijger (tiger)	beer, eend, ekster (bear, duck, magpie)	tijdig (timely)	kluit (clod)	strepen (stripes)	geld (money)	... dat in de jungle leeft is een dier? (... that lives in the jungle is an animal?)
22	kluis (safe)	archiefkast, koffer, fles (file cabinet, suitcase, bottle)	kluis (clod)	tijdig (timely)	strepen (stripes)	geld (money)	... dat moeilijk te openen is kan worden dichtgemaakt? (... that is difficult to open can be closed?)
23	riem (belt)	meetlint, filmrol, ijshockeystick (measuring tape, film roll, ice hockey stick)	riet (reeds)	eisen (requirements)	broek (pants)	boom (tree)	... dat dichtgemaakt kan worden is erg lang? (... that can be closed is very long?)
24	eikel (acorn)	teddybeer, piratenschip, wasknijper (teddy bear, pirate ship, clothespin)	eisen (requirements)	riet (reeds)	boom (tree)	broek (pants)	... dat in het bos groeit is bruin van kleur? (... that grows in the woods is of brown colour?)
25	naald (needle)	tandwiel, olievat, sleutel (cogwheel, oil barrel, key)	naakt (naked)	pijn (pain)	draad (thread)	rook (smoke)	... dat erg puntig is is gemaakt van metaal? (... that is very spiky is made of metal?)
26	pijp (pipe)	mand, rugzak, gieter (basket, backpack, watering can)	pijn (pain)	naakt (naked)	rook (smoke)	draad (thread)	... dat in de mond wordt gestopt kan worden gevuld? (... that is put in the mouth can be filled?)
27	sok (sock)	overhemd, zwembroek, sombrero (shirt, swimsuit, sombrero)	som (sum)	maat (measure)	kous (stocking)	nacht (night)	... dat gedragen wordt in je schoenen is een kledingstuk? (... that is worn in shoes is a piece of garment?)
28	maan (moon)	wolk, son, vliegtuig (cloud, son, airplane)	maat (measure)	som (sum)	nacht (night)	kous (stocking)	... dat gezien kan worden in het donker bevindt zich in de lucht? (... that can be seen when it's dark is found at the sky?)
29	kassa (cash desk)	telefoon, rekenmachine, stopwatch (telephone, calculator, stopwatch)	kaste (caste)	duim (thumb)	nacht (store)	vogel (bird)	... dat gebruik wordt om geld te bewaren kan nummers tonen? (... that is used to store money can display numbers?)
30	duif (pigeon)	dolfijn, vleermuis, vis (dolphin, bat, fish)	duim (thumb)	kaste (caste)	vogel (bird)	winkel (store)	... dat leeft in steden is een dier? (... that lives in cities is an animal?)
31	peer (pear)	druiven, watermeloen, pompoen (grapes, watermelon, pumpkin)	pees (tendon)	angel (angel)	boot (apple)	boot (boat)	... dat aan een boom groeit bevat ook zaaden? (... that grows on a tree also contains seeds?)
32	anker (anchor)	horloge, lepel, paperclip (watch, spoon, paper clip)	angel (angel)	pees (tendon)	boot (boat)	appel (apple)	... dat erg zwaar is is gemaakt van metaal? (... that is very heavy is made of metal?)

33	veer (feather)	potlood, label, envelop (pencil, label, envelope)	veen (peat)	deuk (dent)	pluim (plume)	klink (latch)	... dat gebruikt wordt om te vliegen is erg licht? (... that is used to fly is very light?)
34	deur (ENGL)	honing, doosje, schatkist (ENGL)	deuk (ENGL)	veen (ENGL)	klink (ENGL)	pluim (plume)	... dat gebruikt wordt om een kamer te betreden kan geopend worden? (... that is used to enter a room can be opened?)
35	wortel (carrot)	boon, ui, avocado (bean, onion, avocado)	worden (to become)	test (test)	oranje (orange)	kamperen (to camp)	... dat onder de grond groeit is een groente? (... that grows underground is a vegetable?)
36	tent (tent)	schuur, boomhut, hol (barn, treehouse, cave)	test (test)	worden (to become)	kamperen (to camp)	oranje (orange)	... dat ingepakt kan worden kan gebruikt worden om in te slapen? (... that can be packed together can be used to sleep in?)
37	kruiwagen (wheelbarrow)	vrachtwagen, boldekar, winkelwagen (truck, cart, shopping cart)	kruisigen (to crucify)	stoep (sidewalk)	tuin (garden)	zitten (to sit)	... dat slechts een wiel heeft wordt gebruikt om dingen te transporteren? (... that has only one wheel is used to transport things?)
38	stoel (chait)	bank, lamp, kast (couch, lamp, cupboard)	stoep (sidewalk)	kruisigen (to crucify)	zitten (to sit)	tuin (garden)	... dat vier poten heeft is een meubelstuk? (... that has four legs is a piece of furniture?)
39	krant (newspaper)	rol, notitieblaadje, wc-papier (scroll, note paper, toilet paper)	kramp (cramp)	haver (oats)	nieuws (news)	nagel (nail)	... dat iedere dag geprint wordt is gemaakt van papier? (... that is printed every day is made of paper?)
40	hamer (hammer)	tang, nietmachine, zakmes (pliers, stapler, pocket knife)	haver (oats)	kramp (cramp)	nagel (nail)	nieuws (news)	... dat een houten handvat heeft wordt gebruikt als gereedschap? (... that has a wooden handle is used as a tool?)
41	klok (clock)	tennisbal, dartbord, olijf (tennis ball, dart board, olive)	klop (knock)	schelm (rascal)	tijd (tiem)	zee (sea)	... dat wijzers heeft heeft een ronde vorm? (... that has hands is round shaped?)
42	schelp (shell)	diamant, goudstaaf, jadesteen (diamond, gold ingot, jade stone)	schelm (rascal)	klop (knock)	zee (sea)	tijd (time)	... dat parels kan bevatten wordt gebruikt om juwelen van van te maken? (... that can contain a pearl may be used to make jewellery?)
43	spijker (nail)	reageerbuis, liniaal, peper (test tube, ruler, pepper)	spijtig (regrettable)	zwaar (heavy)	hamer (hammer)	ridder (knight)	... dat geplaatst wordt in muren is lang en dun? (... that is put in walls is long and thin?)
44	zwaard (sword)	schroefleutel, cabrio, veiligheidsspeld (spanner, convertible, safety pin)	zwaar (heavy)	spijtig (regrettable)	ridder (knight)	hamer (ENGL)	... dat gebruikt wordt als wapen is gemaakt van metaal? (... that is used as a weapon in made of metal?)
45	rits (zipper)	medaille, kroon, magneet (medal, crown, magnet)	ring (ring)	raken (to touch)	jas (jacket)	maan (moon)	... dat gebruikt wordt om dingen te sluiten is gemaakt van metaal? (... that is used to close things is made of metal?)
46	raket (rocket)	vlieger, vliegtuigje, papegaai (kite, airplane, parrot)	raken (to touch)	ring (ring)	maan (moon)	jas (jacket)	... dat gebruikt wordt door astronauten kan door de lucht vliegen? (... that is used by astronauts can fly through the air?)
47	hoed (hat)	pakketje, cello, pretzel (package, cello, pretzel)	hoef (hoof)	marker (marker)	hoofd (head)	carnaval (carnival)	... dat een kledingstuk is is bruin van kleur? (... that is a piece of garment is of brown colour?)
48	masker (mask)	gasmasker, bril, ice hockey helm (gas mask, glasses, ice hockey helmet)	marker (marker)	hoef (hoof)	hoofd (head)	hoofd (head)	... dat gebruikt wordt om jezelf te verbergen wordt gedragen op het gezicht? (... that is used to disguise oneself is worn on the face?)
49	huis (nouse)	eiffeltoren, atomium, kerk (eiffel tower, atomium, church)	huid (skin)	doop (baptism)	dak (roof)	karton (carton)	... dat gebruikt wordt om in te wonen is een stevig gebouw? (... that is used to live in is a building?)
50	doos (box)	stopcontact, vlieger, toetsenbord (electrical outlet, kite, keyboard)	doop (baptism)	huid (skin)	karton (carton)	dak (roof)	... dat gebruikt wordt om te vullen is rechthoekig? (... that is used as a container has a rectangular shape?)

51	hark (rake)	decoupeerzaag, pijptang, boormachine (jigsaw, pipe wrench, drill)	hard (hard)	kin (chin)	bladeren (leaves)	ei (egg)	... dat gebruikt wordt in de tuin is een stuk gereedschap? (... that is used in the garden is a tool?)
52	kip (chicken)	schaap, rups, muis (sheep, caterpillar, mouse)	kin (chin)	hard (hard)	ei (egg)	bladeren (leaves)	... dat vleugels heeft is een dier? (... that has wings in an animal?)
53	bel (bell)	handhark, satelliet, vishaak (hand rake, satellite, fishing hook)	bek (beak)	kaart (card)	deur (door)	licht (light)	... dat van ver kan worden gehoord is gemaakt van metaal? (... that can be heard very far is made of metal?)
54	kaars (candle)	boomstammen, olielamp, aansteker (tree trunks, oil lamp, lighter)	kaart (card)	bek (beak)	licht (light)	deur (door)	... dat gemaakt is van was kan erg lang branden? (... that can burn is made of wax?)
55	harp (harp)	wandspiegel, houdbet, raam (wall mirror, holding bet, window)	hars (resin)	boog (bow)	muziek (music)	blad (sheet)	... dat snaren heeft heeft een houten geraamte? (... that has chords has a wooden frame?)
56	boom (tree)	zeilboot, tractor, bus (sailboat, tractor, bus)	boog (bow)	hars (resin)	blad (sheet)	muziek (music)	... dat in het bos staat is erg groot? (... that stands in the woods is very tall?)
57	kers (cherry)	perzik, pruim, braam (peach, plum, blackberry)	kerk (church)	held (hero)	taart (pie)	brommer (moped)	... dat rood is van kleur is een stuk fruit? (... that has red colour is a fruit?)
58	helm (helmet)	honkbalhandschoen, kniebeschermer, zonnebril (baseball glove, knee protector, sunglasses)	held (hero)	kerk (church)	brommer (moped)	taart (pie)	... dat gedragen wordt op het hoofd wordt gebruikt voor bescherming? (... that is worn on the head is used for protection?)
59	pan (pan)	zout, blender, koekenpan (salt, blender, frying pan)	pas (pass)	kwart (quarter)	koken (to cook)	verf (paint)	... dat gebruikt wordt om water te verwarmen is handig om eten mee te bereiden? (... that is used to heat water is handy to prepare food?)
60	kwast (brush)	schep, boormachine, kettingzaag (shovel, drill, chainsaw)	kwart (quarter)	pas (pass)	verf (paint)	koken (to cook)	... dat haar heeft haar op een kant is een stuk gereedschap? (... that has hair on one end is a tool?)
61	bezem (broom)	magnetron, strijkijzer, percolator (microwave, iron, percolator)	bezig (busy)	mep (whack)	heks (witch)	scherp (sharp)	... dat gebruikt wordt om schoon te maken wordt gebruikt in het huis? (... that is used to clean is used in the house?)
62	mes (knife)	bord, kom, chopsticks (plate, bowl, chopsticks)	mep (whack)	bezig (busy)	scherp (sharp)	heks (witch)	... dat is gemaakt van metaal is een voorwerp om mee te eten? (... that is made of metal is a piece of dish?)
63	wesp (wasp)	libelle, kever, vlinder (dragonfly beetle and butterfly)	west (west)	kraam (stall)	steek (stab)	water (water)	... dat geel en zwart is is een insect? (... that is yellow and black is an insect?)
64	kraan (crane)	tandenborstel, borstel, spiegel (toothbrush, brush, mirror)	kraam (stall)	west (west)	water (water)	steek (stab)	... dat aangezet kan worden kan worden gevonden in de badkamer? (... that can be turned on can be found in the bathroom?)

Table 3.11: List of materials. Confederate objects were named in the critical turn by the confederate. Participant objects had to be named by the participant. phon. = phonologically; rel. = related; unrel. = unrelated. \* Only used in Experiment 1. \*\* Questions were only used in Experiments 2 and 3. In this Table, only early versions of questions are listed; in late questions, the order of phrases was swapped.



# 4

---

## PROCESSING LOAD IN SPEECH PLANNING IN DIALOGUE

---

Published as:

Barthel, M. and Sauppe, S. (2019). Speech Planning at Turn Transitions in Dialogue is Associated with Increased Processing Load. *Cognitive Science*, 43(7), e12768.

### ABSTRACT

Speech planning is a sophisticated process. In dialogue, it regularly starts in overlap with an incoming turn by a conversation partner. We show that planning spoken responses in overlap with incoming turns is associated with higher processing load than planning in silence. In a dialogic experiment, participants took turns with a confederate describing lists of objects. The confederate's utterances (to which participants responded) were pre-recorded and varied in whether they ended in a verb or an object noun and whether this ending was predictable or not. We found that response planning in overlap with sentence-final verbs evokes larger task-evoked pupillary responses, while end predictability had no effect. This finding indicates that planning in overlap leads to higher processing load for next speakers in dialogue and that next speakers do not proactively modulate the time course of their response planning based on their predictions of turn endings. The turn taking system exerts pressure on the language processing system by pushing speakers to plan in overlap despite the ensuing increase in processing load.

## 4.1 INTRODUCTION

Conversation is the most frequent form of human communication (Levinson, 2006), and taking turns at talk is a well practiced task in which different speakers' contributions usually follow one another with only short gaps in between (Stivers et al., 2009). Planning a verbal response, however, is known to take between about 600 ms for single words (Indefrey, 2011; Strijkers & Costa, 2011) to well more than one second for short sentences (Griffin & Bock, 2000; Myachykov, Scheepers, Garrod, Thompson, & Fedorova, 2013), illustrating that timing a turn at talk in conversation is not a trivial task. To be able to quickly take their turn, next speakers need to start planning their response as early as possible, often in overlap with the incoming turn (Barthel & Levinson, *in press*; Barthel et al., 2017; Bögels, Magyari, & Levinson, 2015; Corps et al., 2018) (see chapters 5 and 3). Barthel et al. (2016, see ch. 2) found that response planning was indeed done as early as the incoming turn's message could be conceived, even if the incoming turn did not end at that point.<sup>1</sup>

Planning the next turn while continuously monitoring the incoming turn for completion, and possibly for content, is a demanding dual task situation. Both language comprehension and planning require allocation of central attention (Hagoort et al., 1999; Kemper et al., 2003; Kubose et al., 2006; Shitova, Roelofs, Coughler, & Schriefers, 2017), and both are known to interfere with concurrent non-linguistic tasks (Boiteau et al., 2014; Roelofs & Piai, 2011; Sjerps & Meyer, 2015). The law of least mental effort proposes that humans try to make decisions and form strategies so as to minimize mental workload in order to achieve an efficient work-benefit ratio (Reichle, Carpenter, & Just, 2000; Zipf, 1949). It is thus a central question whether the language processing system is adapted to this highly frequent task or whether planning in overlap leads to increased processing load in the vicinity of turn transitions, the 'crunch zone' of conversation (S. G. Roberts & Levinson, 2017). Using an auditory picture-word interference paradigm, Schriefers et al. (1990) compared the effects of concurrent noise versus concurrent speech on speech planning and found that naming latencies did not differ between a silent condition and a condition with distracting noise.

---

<sup>1</sup>The study reported here presents pupillometric data from Barthel et al. (2016, see ch. 2), which focused on eye movements.

With distracting words, however, naming latencies increased by 70 ms even when the words were unrelated to the picture names, indicating general interference of speech comprehension with speech planning. As participants were instructed to ignore any incoming speech and as their own utterances were independent of the presented speech input, the measured interference effects are effects of distraction rather than of the processes of integration of speech input, which is the task next speakers face in turn taking. Instead of trying to ignore incoming speech, interlocutors most of the time have to plan their next turn while concurrently listening to the incoming turn. [Fargier and Laganaro \(2016\)](#) studied picture naming performance with either a concurrent syllable or tone detection task and found longer response latencies and differences in ERP components in the syllable condition as compared to the tone condition, indicating increased interference between two concurrent linguistic tasks. [Klaus et al. \(2017\)](#) made use of a dual-task paradigm combining sentence production as task 1 with a concurrent working memory task 2. Participants were instructed to produce subject-verb-object sentences while they had to ignore auditory distractor words that were either phonologically or semantically related to either the subject or the object of the sentence. The concurrently performed working memory task was either visuospatial or verbal in nature. Under visuospatial load, both types of relatedness had effects on both the subject and object of the sentence. The pattern of results was similar under verbal load. Here however, only phonological relatedness to the subject but not to the object affected sentence production performance, showing that verbal load reduced participants' phonological planning scope. These findings make it plausible to assume that next speakers postpone stages of formulation when planning in conversation in order to avoid inefficient processing due to interference. [Barthel and Levinson \(in press, see ch. 3\)](#), however, show that next speakers in a quiz-like situation engage in phonological planning as early as possible and in overlap with the incoming question. To date, evidence on the timing of the different processing stages in conversation is scarce, but the fact that response planning is frequently initiated in overlap with listening to the incoming turn is largely undisputed (but see [Heldner & Edlund, 2010](#)).

The observation that planning in overlap is common can be accounted for in two ways. One account highlights the mechanisms of



turn allocation and the time pressure at turn transitions. According to the simplest systematics of turn taking (Sacks et al., 1974), the first participant that speaks up when a turn transition becomes relevant gains the right to take the next turn. While language production and comprehension are assumed to engage—at least partly—the same cognitive resources (Hagoort & Indefrey, 2014; Kempen et al., 2012; Menenti et al., 2011; Silbert, Honey, Simony, Poeppel, & Hasson, 2014), potentially increased processing load due to parallel processing of the two might be traded for the benefit of early planning, leading to shorter turn transition times (Barthel et al., 2017, 2016, see ch. 2 and 5). The alternative account questions the assumption that the simultaneity of comprehension and production in conversation drastically increases processing load. Previous research shows that participants prefer to use parallel processing over serial processing in dual tasks (Hübner & Lehle, 2007). To investigate the reasons for this tendency, Lehle, Steinhauser, and Hübner (2009) instructed participants explicitly to apply either a parallel or a serial processing strategy when giving parity judgments on two numbers. Lehle et al. found that while a parallel processing strategy increased reaction times and error rates, it decreased processing load, which might be the main reason for preferring parallel over serial processing. Consequently, planning in overlap might not be associated with any significant increase in processing load, especially since turn taking is a highly practiced dual task and cognitive tasks become less demanding with increasing proficiency (Donovan & Radosevich, 1999; Hampton Wray & Weber-Fox, 2013; Neubauer & Fink, 2009; Van Selst, Ruthruff, & Johnston, 1999; Weber-Fox, Davis, & Cuadrado, 2003).

Here, we test whether planning a response while simultaneously comprehending an interlocutor's turn imposes increased processing load on speakers as compared to non-overlapping response planning by analyzing task-evoked pupillary responses from an experiment employing a dialogic paradigm. Changes in pupil diameter in response to task-induced cognitive processes are a reliable indicator of processing load (Beatty, 1982; Beatty & Lucero-Wagoner, 2000; Sirois & Brisson, 2014). The analysis of task-evoked pupillary responses allows studying differences in task demands, i.e. the amount of overall cognitive resources that need to be allocated in order to master a task (Hess & Polt, 1964; Kahneman, 1973; Laeng, Sirois, & Gredebäck, 2012). Most studies using task-evoked pupillary responses to measure processing load

in language processing have focused on comprehension (Engelhardt, Ferreira, & Patsenko, 2010; Just & Carpenter, 1993; Koch & Janse, 2016; Kuchinke, Vo, Hofmann, & Jacobs, 2007; Schmidtke, 2014; Tromp, Haagoort, & Meyer, 2016; Zekveld, Kramer, & Festen, 2010, *inter alia*), and there are only few studies that have investigated language production (Papesh & Goldinger, 2012; Sauppe, 2017). If planning in overlap leads to increased processing load, task-evoked pupillary responses should have larger amplitudes as compared to planning in silence, whereas they are not predicted to differ if overlap does not increase processing load during response planning.

We report a dialogic experiment in which participants took turns with a confederate describing arrays of objects. Participants' pupil diameter was measured as they listened and responded to pre-recorded critical utterances from the confederate. These utterances were designed to on the one hand either allow for response planning in overlap or not and on the other hand to contain either a predictable or a non-predictable ending. In this way, the effects of planning in overlap as compared to planning in silence on task-evoked pupillary responses were tested in the context of predictable and non-predictable overlapping speech input.

## 4.2 METHODS AND MATERIALS

### 4.2.1 *Participants*

Forty-eight German native speakers (mean age = 26.3 years, SD = 7.6 years, 30 female) who reported to have normal hearing and vision participated in the experiment for payment. Eight participants were excluded from the analyses because they reported during a post-test questionnaire that they had noticed the presence of pre-recorded material. Two participants were excluded due to technical failures of recording equipment, leaving 38 participants for analysis. Participants gave informed consent and the experiment was approved by the Ethics Committee of the Faculty of Social Sciences, Radboud University Nijmegen.

#### 4.2.2 *Apparatus*

Participant and confederate were placed in separate sound-proof booths that were equipped with headphones and microphones with which they could communicate with one another. Visual stimuli were presented on a 21" computer screen at a distance of approximately 60 cm. Participants' pupil size was recorded with an SMI RED-m remote eye tracker at 120 Hz sampling rate. Light conditions remained constant across participants.

#### 4.2.3 *Stimuli*

##### 4.2.3.1 *Visual Stimuli*

Coloured pictures of 468 objects were used to generate the visual stimuli. Ninety-six critical stimulus displays showing between three and five objects (32 displays each) were generated. Irrespective of the number of objects shown in an item display, each object filled approximately two degrees of visual angle and was located about four centimeters away from its neighbours, so that participants had to shift their gaze in order to foveally fixate individual objects. Between none and three of the objects had to be named by participants (24 displays each), the remaining objects were named by the confederate (cf. Section 4.2.4).

##### 4.2.3.2 *Auditory Stimuli*

Each of the 96 critical stimulus displays was accompanied by a German sentence in one of four conditions that were pre-recorded by the confederate and crossed according to whether the sentence ended in a verb or not (verb position) and whether it was predictable or not that the sentence would end with or without a final verb (end predictability; see Table 4.1). The presence of a sentence-final verb made planning in overlap possible, since all that participants needed to know to plan their response was which of the displayed objects they would have to name. When a sentence did not end in a verb, it ended in an object noun that was relevant for preparing the response, so that planning could only take place in silence after the turn ended. In predictable sentences, participants could know in advance whether the last word

would be a verb or an object noun, since different verbs in second position (before the list of objects) either required another verb form in sentence-final position (such as the modal verb ‘*can*’) or not (such as the main verb ‘*see*’). In contrast, non-predictable sentences contained ‘*have*’ in second position, which is ambiguous between being a main verb or an auxiliary and consequently either does or does not call for a sentence-final participle. Four pseudo-randomized lists were constructed, so that each item appeared in only one condition per list and the same number of items per condition appeared in each list.

		End predictability	
		unpredictable	predictable
Verb position	not final	Ich habe einen Schlüssel, einen Lenkdrachen und einen Rubin.	Ich sehe einen Schlüssel, einen Lenkdrachen und einen Rubin.
	final	Ich habe einen Schlüssel, einen Lenkdrachen und einen Rubin besorgt.	Ich kann einen Schlüssel, einen Lenkdrachen und einen Rubin besorgen.

Table 4.1: Example sentences of the four conditions used in the experiment. ‘I have/have gotten/see/can get a key, a kite, and a ruby.’

#### 4.2.4 Procedure

Prior to the experiment, participants were shown all objects in a booklet and asked to name them. Participants and the confederate were instructed as follows. In each trial, they would see a number of objects they could get and the confederate should tell the participant what objects she could get, so that the participant could tell the confederate what *further* objects he could get, only listing the objects that had not already been named by the confederate (all objects named by the confederate were also visible on the participant’s display). Participants triggered the beginning of each trial by looking at a fixation cross at the center of the screen. Each trial began with a preview of 600–1000 ms of the stimulus display before the critical sentence was played. The experiment started with twelve practice trials that were of the same structure as experimental trials. The eye-tracker was (re-)calibrated four times at equal intervals during the experiment. The experiment lasted

approximately 30 minutes and was followed by a computerized questionnaire asking participants whether they had noted the presence of pre-recorded material.

#### 4.2.5 *Data Preprocessing and Analyses*

Preprocessing of pupil data and statistical analyses were carried out in R (R Core Team, 2019). Samples recorded with low validity (as indicated by SMI's recording software) and during blinks or saccades were treated as missing values and linearly interpolated separately for each eye. Pupil diameters of both eyes were averaged before time-locking to the offset of the last noun in the confederate turn. For each trial, pupil diameter was baselined by subtracting the mean diameter during a baseline period spanning the 500 ms preceding the offset of the last noun in the confederate turn. Mean task-evoked pupillary response amplitude was calculated for a time window of 3000 ms after the time-lock point and peaks in pupil diameters were identified in this time window (Borchers, 2015).

The data set contained 2736 trials in which both confederate and participant named at least one object. Trials in which participants did not name the correct objects or responded in overlap and trials with more than 30% missing values before interpolation in samples recorded between  $-500$  and  $3000$  ms relative to the offset of the confederate's last noun were excluded from statistical analyses (319 trials). Forty-three additional trials were excluded because their verbal response time was more than 3 SD longer than the participant's mean response time—measured manually in Praat (Boersma & Weenink, 2015) from the offset of the incoming turn to the onset of the first object noun in the participants' turn. Additional items in which sentences were produced live by the confederate (see Barthel et al., 2016, ch. 2) were not considered for analyses (341 trials). On balance, 2377 trials remained for analysis (13.12% of trials were excluded).

Three linear mixed effects regression models were fitted (Bates, Mächler, Bolker, & Walker, 2015) with mean amplitude, peak amplitude, and peak latency as dependent variables. The underlying assumption is that differences in mean and peak amplitude and peak latency relate to differences in processing load and reflect differences in task difficulty

(Beatty & Lucero-Wagoner, 2000). While peak amplitude is a good measure for processing load, accurate peak detection is not straightforward, as the location of peaks is susceptible to noise in the recorded signal (Luck, 2014). Mean amplitude is a more conservative measure for processing load, since it takes into account the whole analysis window and is thus less susceptible to noise. Differences in the latency of peaks between conditions relate to differences in task difficulty, reflecting differences in the time it takes to do the necessary computations in order to give a response. Converging results in these measures is desirable when drawing inferences on cognitive demand on the basis of task-evoked pupillary responses. Verb position and sentence end predictability as well as their interaction were the predictors of interest. Their statistical significance was assessed using *F*-tests with Kenward-Roger approximations of degrees of freedom (Fox & Weisberg, 2011; Halekoh & Hojsgaard, 2014; Kenward & Roger, 1997). The maximal random effects structures as justified by design which allowed models to converge were used (Barr, 2013; Barr et al., 2013). A number of nuisance variables were included in the fixed effects structure of the models (Sassenhagen & Alday, 2016): the duration of the confederate turn, since the pre-recorded sentences differed in complexity; the number of objects to be named by the participant, since task difficulty increases with the number of choices (Hick, 1952); trial number, to account for changes over the course of the experiment; and a binary variable indicating whether the sentence structure of the confederate turn was re-used in the response turn, since processing load might be influenced by structural priming (Pickering & Ferreira, 2008; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012b). The statistical significance of nuisance variables was not assessed. Categorical predictors were deviation coded (−0.5 and 0.5) and continuous predictors were mean centered.

#### 4.3 RESULTS

Average task-evoked pupillary responses are shown in Figure 4.1 and descriptive statistics are presented in Table 4.2.

Linear mixed effects regressions revealed that task-evoked pupillary responses in the verb-final conditions had statistically significantly

higher mean amplitudes, higher peak amplitudes, and greater peak latencies than in non-final conditions. Neither the main effect of predictability, nor its interaction with verb position reached statistical significance in any of the three models (Table 4.3).

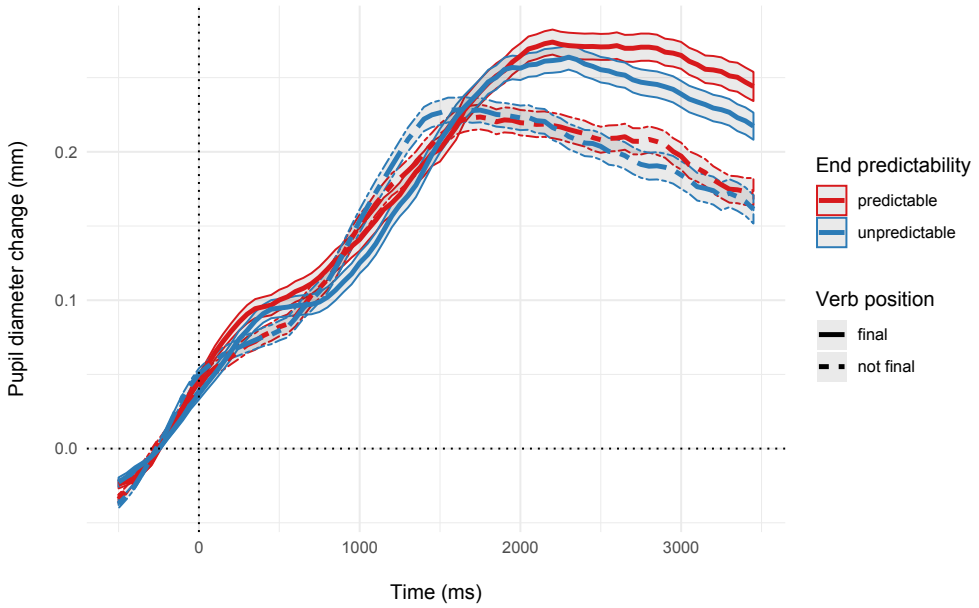


Figure 4.1: Grand average changes in pupil diameter (task-evoked pupillary responses) in mm, time-locked to the offset of the last noun of the confederate's turn (dashed vertical line). Ribbons indicate 95% confidence intervals. The analysis time window ranged from 0–3000 ms. For plotting only, samples were averaged into 50 ms bins within each trial to align time steps across trials before grand averaging.

condition	peak amplitude in mm	mean amplitude in mm	peak latency in ms
no final verb/unpredictable	0.409 (0.231)	0.167 (0.198)	1850 (782)
no final verb/predictable	0.412 (0.222)	0.165 (0.185)	1868 (825)
final verb/unpredictable	0.432 (0.226)	0.179 (0.188)	2030 (756)
final verb/predictable	0.446 (0.241)	0.190 (0.196)	2041 (782)

Table 4.2: Means (standard deviations in parentheses) of peak and mean amplitudes, and peak latencies by condition.



	Mean amplitude (mm)				Peak amplitude (mm)				Peak latency (logarithm of ms)			
	$\hat{\beta}$	$ t $	$F$	$p$	$\hat{\beta}$	$ t $	$F$	$p$	$\hat{\beta}$	$ t $	$F$	$p$
Intercept	0.175	12.366			0.425	21.313			7.397	182.183		
Verb position = final	0.023	2.788	7.847	0.006**	0.034	3.647	13.327	<0.001***	0.114	3.507	12.323	<0.001***
End predictability = pred.	0.006	0.789	0.617	0.438	0.009	1.182	1.393	0.238	-0.012	0.374	0.140	0.708
Verb position $\times$ End pred.	0.009	0.674	0.453	0.501	0.004	0.268	0.072	0.789	0.012	0.185	0.034	0.853
Structural priming = yes	0.009	1.116			0.016	1.754			-0.011	0.317		
Sentence duration (z)	-0.002	0.235			<0.001	0.022			0.081	2.716		
Trial number (z)	-0.015	4.270			-0.016	3.858			-0.060	3.844		
Delta of objects (z)	0.007	0.700			0.013	1.191			0.099	3.329		

Table 4.3: Linear mixed effects regression models predicting mean task-evoked pupillary response amplitude (in mm), peak task-evoked pupillary response amplitude (in mm), and peak task-evoked pupillary response latency (in ms). Statistical significance based on Type II  $F$  tests with Kenward-Roger degrees of freedom (Kenward & Roger, 1997). \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ .

## 4.4 DISCUSSION

We investigated the level of processing load in next speakers in the vicinity of turn transitions in dialogue to answer the question whether planning a turn at talk in overlap with the incoming turn leads to higher processing load than planning it in silence. Task-evoked pupillary responses recorded during a dialogic list-completion task were analyzed, and mean amplitudes and peak amplitudes were found to be higher and peak latencies to be longer when planning was done in overlap than when it was done in silence. While the sentences in conditions that allowed for early planning in overlap were often slightly more complex than the sentences in conditions that did not allow for early planning, the differences in sentence complexity were much greater within than between conditions. Whether a sentence ended in a verb or not influenced pupillary responses beyond the influence of sentence duration, which was included as a nuisance variable to account for the length of a sentence and thereby its complexity. Taken together, the presented results show that planning in overlap is more demanding than planning in silence.

In their analyses of eye-movements from the experiment here, [Barthel et al. \(2016, see ch. 2\)](#) found that participants started to plan their response as early as possible, i.e., as soon as they had identified the last noun of the incoming turn—irrespective of another verb form following before the end of the turn or not. Consequently, participants generally started planning their response in overlap with the incoming turn in verb final conditions and in silence in conditions without a final verb. When planning in overlap, the time gained by starting to plan early was not fully reflected in the reduction of turn-transition times. When participants planned their response in overlap, planning overlapped with turn final verbs which were about 600 ms long. In these cases, however, gaps between turns were shorter by only approximately 100 ms. This means that participants spent considerably more time planning their response when planning started in overlap than when planning was done in silence. The reported pattern of task-evoked pupillary responses sheds light on the cause of this discrepancy: The increase in planning time was due to higher processing load in planning in overlap as compared to planning in silence.

Given that planning in overlap is the norm in conversation, the finding that it is a more demanding strategy as compared to planning in silence shows that the requirements of the systematics of turn taking in conversation (Sacks et al., 1974) receive precedence over the minimization of mental effort. The culturally developed turn taking system exerts pressure on the cognitive mechanisms of language processing, enforcing strategies that raise processing load in order to meet the requirements set by the rules of turn allocation and the semiotics of turn timing. Increased processing load for the sake of finely attuned temporal alignment of turns thus appears to be a cornerstone in the organization of turn allocation: If you want to take a turn at talk, you need to push your language processor in order to speak up before other participants. Trading high processing load for shorter turn transitions is a pre-requisite for the timing of turns to become a meaningful source of information. If the next speaker does not claim her turn in time, she can be interpreted as lacking interest in the conversation, its topic, or her interlocutor, as having trouble understanding the previous turn or parts of it (Kendrick, 2015; Schegloff, Jefferson, & Sacks, 1977), as being unwilling to comply with a request or as preparing to disagree with an assessment (Kendrick & Torreira, 2014; F. Roberts & Francis, 2013; F. Roberts et al., 2011). In that way, turn timing is meaningful in itself, irrespective of the content of the following turn, with a long gap before a turn leading the recipient to expect a dis-preferred response, e.g., a rejection of an invitation (Bögels, Kendrick, & Levinson, 2015). With the timing of turn taking being a source of information that is analyzed by listeners, more information can be inferred from a single unit of talk. This enriches social interaction in conversation but comes at the cost of increased processing load for the individual speaker.

As processing load is high at turn transitions due to time pressures, next speakers might develop strategies to distribute processing load evenly over time when planning their turn. Based on findings that participants in dual tasks can to some degree choose to apply different processing strategies (Hübner & Lehle, 2007; Miller, Ulrich, & Rolke, 2009; Navon & Gopher, 1979; Navon & Miller, 2002; Tombu & Jolicœur, 2005), one conceivable way to avoid high peaks in processing load would be to apply a 'proactive planning' strategy in cases when incoming turns contain highly predictable turn-final words. If predictability of a turn-final word leads to effective changes in response planning,

processing load in sentences with predictable turn ends should be lower than in sentences with unpredictable turn ends. However, none of the analyzed pupillary response measures (peak amplitude, mean amplitude, and peak latency) were significantly affected by predictability, lending no support to the hypothesis that participants applied a proactive planning strategy in order to keep processing load low at turn transitions. We take this as evidence that next speakers did not utilize the predictability of incoming verbal material to adapt the time course of their response planning (cf. also Huettig & Mani, 2016). In order to meet turn timing requirements, next speakers seem to aim to plan their contribution as early and fast as possible, accepting increased processing loads during response planning to avoid risking the consequences of being too slow to take their turn.

By planning their response in overlap with comprehending the incoming turn, participants' behaviour agrees with the general tendency to choose parallel processing over serial processing in dual tasks (Hübner & Lehle, 2007); they do not postpone encoding processes until after a predictable final word. In our experiment, however, the reason for this choice cannot have been reduced processing load, as our analyses of task-evoked pupillary responses show that planning a response in overlap induces *higher* processing load than planning in silence. Instead, participants' motivation was more likely to reduce the length of gap after the incoming turn. Intending to take a well-timed turn, next speakers employed a planning strategy that at the same time took them longer to plan their response and was more demanding as compared to delaying response planning. While it remains possible that the choice of processing strategy is a question of preference of individual speakers (Bögels, Casillas, & Levinson, 2018) or the demands of the dual task situation (Lehle & Hübner, 2009; Reissland & Manzey, 2016), parallel processing appears to be the standard strategy in dialogue.

In sum, the turn taking system requires next speakers to accept higher processing loads induced by planning in overlap in order to be able to respond as fast as possible to an incoming turn so as to avoid the social consequences ensuing from noticeable gaps between turns of talk. In the words of Kahneman (1973), participants in a conversation are forced to trade *efficiency* in terms of processing load for *effectiveness* in terms of short gaps between turns. This means that the turn taking system is not optimized for next speakers' processing, but for overall

effectiveness in social interaction. While putting pressure on cognitive processing in individual speakers, the turn taking system allows for a dense semiotics of turn timing that organizes and enriches social interaction in conversation. In addition to viewing the turn taking system as shaping the evolution of aspects of grammar (Auer, 2005; Ford & Thompson, 2003; S. G. Roberts & Levinson, 2017), the need to meet the timing demands in turn taking might also be shaping the design of the cognitive system. The study presented in this paper shows that examining task-evoked pupillary responses during speech planning is a promising technique to further investigate the mechanisms of speech processing in conversation.





---

## THE TIMING OF NEXT TURN INITIATION

---

Published as:

Barthel, M., Meyer, A. S., and Levinson, S. C. (2017). Next Speakers Plan Their Turn Early and Speak after Turn-Final “Go-Signals.” *Frontiers in Psychology* (8), page 393.

### ABSTRACT

In conversation, turn-taking is usually fluid, with next speakers taking their turn right after the end of the previous turn. Most, but not all, previous studies show that next speakers start to plan their turn early, if possible already during the incoming turn. The present study makes use of the list-completion paradigm (Barthel et al., 2016, see ch. 2), analyzing speech onset latencies and eye-movements of participants in a task-oriented dialogue with a confederate in order to disentangle the contribution of early planning of content and initiation of articulation as a reaction to the upcoming turn-end to the timing of turn-taking. Participants named objects visible on their computer screen in response to utterances that did, or did not, contain lexical and prosodic cues to the end of the incoming turn. In the presence of an early lexical cue, participants showed earlier gaze shifts towards the target objects and responded faster than in its absence, whereas the presence of a late intonational cue only led to faster response times and did not affect the timing of participants' eye movements. The results show that with a combination of eye-movement and turn-transition time measures it is possible to tease apart the effects of early planning and response initiation on turn timing. They are consistent with models of turn-taking that assume that next speakers (a) start planning their response



as soon as the incoming turn's message can be understood and (b) monitor the incoming turn for cues to turn-completion so as to initiate their response when turn-transition becomes relevant.

### 5.1 INTRODUCTION

Taking turns at talk in conversation is an essential feature of human interaction. When talking to one another in everyday encounters, interlocutors efficiently align their turns-of-talk, most of the time leaving only very short gaps of about 200 ms (de Ruiter et al., 2006; Heldner & Edlund, 2010; Levinson, 2016; Sacks et al., 1974; Stivers et al., 2009). How they achieve such rapid timing in turn-taking is still largely unresolved (Levinson, 2012). For such neat alignment of talk, next speakers need to i) start to plan the content of a response to an incoming turn and ii) recognize the incoming turn's point of completion to know when to launch the articulation of their response. Different turn-taking models have been proposed to explain conversational turn management. They vary in the amount of attention they give to the two tasks faced by next speakers. A group of models developed in the 1970s focuses on the transmission of signals about the state of the current turn at talk (Duncan, 1972; Duncan & Fiske, 1977; Duncan & Niederehe, 1974). In their approach, the current turn (or speaker) displays signals for turn continuation or yielding which the next speaker could react to when they are displayed. However, most of these cues, prosodic, syntactic, or gestural in nature, are displayed towards the end of the turn, which is arguably too late to start planning a response and initiate articulation without long gaps due to the latencies involved in speech production (Levinson, 2012). Therefore, more recent models of turn-taking formulated the need for early response planning, i.e. preparing the next turn while the incoming turn is still unfolding (Heldner & Edlund, 2010; Levinson & Torreira, 2015).

In a previous study investigating the timing of planning of the content of a response, Barthel et al. (2016, see ch. 2) came to the conclusion that next speakers begin to plan their response as early as possible, irrespective of how far the current turn's end lies ahead. However, in a dual-task study, Sjerps and Meyer (2015) came to contrary conclusions. In that study, participants tapped their fingers while taking turns with

a pre-recorded voice in naming lines of objects on a screen. On the basis of participants' eye-movements and tapping performance, the authors suggested that planning began only at the very end of the incoming turn. The results of a study by [Bögels, Magyari, and Levinson \(2015\)](#), however, suggested that the participants in their quiz-like experiment started to plan the response to an answer as early as possible, in some cases several words before the end of a question.

Substantial research on turn end detection, suggests that, at least in participants overhearing a conversation, projection of the incoming turn's completion point is influenced by the presence or absence of turn-taking cues ([Beattie et al., 1982](#); [Caspers, 2003](#); [Cutler & Pearson, 1985](#); [Ford, Fox, & Thompson, 1996](#); [Ford & Thompson, 1996](#); [Hjalmarsson, 2011](#); [Kendon, 1967](#); [Schaffer, 1983](#); [Stephens & Beattie, 1986](#); [Walker & Trimboli, 1984](#); [Wesseling & Son, 2005](#)). In particular, a study by [Lammertink, Casillas, Benders, Post, and Fikkert \(2015\)](#) tested toddler and adult participants while observing a conversation without taking part in it. Both toddlers and adults were found to use both syntactic and intonational cues to turn completion in order to anticipate speaker switches, relying more on syntactic than on intonational cues when these were pitted against each other. Another study by [Bögels and Torreira \(2015\)](#) found that listeners who were asked to press a button upon turn completion take advantage of turn-taking cues that are located close to the turn end. In the corpus that was analyzed to serve as a source of stimuli for that experiment, no early cues to when the turn would end were found. While these studies show that some acoustic cues may be helpful to observers of a conversation, they do not shed light on the question whether interlocutors actually do make use of these cues in conversation. What remains to be shown in order to gain further insight into the organization of human interaction is whether these cues are actually used by speakers to keep gaps between turns short.

The present study was designed to disentangle the relative contribution of early planning on the one hand and reaction to the upcoming turn-end on the other hand to the fast timing of turn-taking that is commonly observed in conversation. It makes use of the list-completion paradigm ([Barthel et al., 2016](#), see ch. 2), in which participants listen to sentences of a confederate that contain lists of objects that participants see on a computer screen. The participants' task is to name all objects

that are displayed on the screen and have not been named by the confederate. While participants listen to the incoming utterance and eventually prepare and produce their own turn, their eye-movements are tracked as they move their gaze from the objects they need to comprehend to the objects they need to name themselves. The study's design is based on two assumptions: i) Participants would switch their gaze from confederate objects to participant objects dependent on when they start planning their turn (Griffin & Bock, 2000; Huettig et al., 2011); and ii) Participants would initiate their response only when they are confident that the incoming turn is complete (Sacks et al., 1974).

Conversation analytic work on German investigated the prosodic tools German speakers have at their disposal to indicate turn-finality vs. turn-continuations. In his work on the functions of intonation for turn-taking in German, Gilles (2005) shows that a falling nuclear contour with a low boundary tone is most widely used in German to mark turn-finality in declarative sentences, whereas rises are used to indicate turn-continuation. This way of marking continuation and termination is very prominent in lists like the ones used in the present experiment, such as *I have a key, a kite, a ruby*. Non-final elements are generally produced with rising pitch, whereas the final element is produced with falling pitch, at least in closed lists, i.e. lists with a finite number of items (von Essen, 1956).

To display that the list under construction is a closed list, speakers can (but need not) use downsteps of successive pitch peaks on list items, which means that the rise in pitch in non-final list elements is lower and lower with every successive element (Féry, 1993; Selting, 2007). These downstepped contours require speakers to pre-plan the length of the list in order to plan the size of the pitch steps. Consequently, listeners could use this early cue to project the length of the list before it comes to be complete.

A third cue to the end of a closed list, next to the two intonational cues, can be a conjunction like *and* that often precedes the final item of closed lists and indicates that the turn will end after the following noun phrase, such as in *I have a key, a kite, and a ruby*. Pitch contours, boundary tones, and lexical cues could therefore be monitored by listeners to identify turn-completion points and used to minimize gaps at turn transitions.

To disentangle the contributions of early planning and reaction to the upcoming turn-end, the present study applied two measures, namely

gaze direction as a measure for the timing of planning, and voice onset time as a measure for the latency of launching the response turn. A combination of these two measures can be used to partly disentangle the processes of response preparation and response initiation. Assuming that next speakers aim for short gaps between turns, an earlier start in planning (operationalized as earlier looks for planning in this experiment) should also lead to shorter response latencies (assuming that the head start in planning will not be canceled out by interference of the incoming speech with the planning process). If, however, no difference in planning can be observed in eye-movements, a difference in response latencies should reflect a difference in response initiation. If next speakers can take advantage of any of the cues tested in this experiment to start planning their response early, they should be able to move their gaze for planning earlier and respond faster in turns displaying the cue than in turns without the cue. If however a cue cannot be used to initiate response planning early (e.g. because it was displayed too late in the incoming turn), it could still be useful to detect the end of the incoming turn and to launch the articulation of a response. In that case, the presence of the cue should make no difference to the timing of gaze movements but lead to shorter response latencies compared to the absence of the cue. Early turn-taking cues, including pitch downsteps on non-final items of a list and a lexical cue before the final item of a list, are therefore hypothesized to lead to earlier response planning and consequently shorter gaps between turns. Late cues to turn-completion, however, such as the final boundary tone, were argued to not aid response planning (de Ruiter et al., 2006; Levinson, 2012). Consequently, a turn-final boundary tone can be hypothesized to have no effect on the timing of response planning, but nevertheless it could be useful to detect the turn end and initiate articulation of a response. In that way, it could be used as a “go-signal” for articulation and lead to shorter gaps between turns.

## 5.2 METHODS AND MATERIALS

The present study uses the list-completion paradigm (Barthel et al., 2016, see ch. 2) to investigate the timing of next speakers’ response planning and their orientation towards potential cues to turn comple-

tion. A confederate talks to a participant and plays pre-recorded critical utterances (recorded by the confederate), so that these utterances seem to be produced live in the flow of conversation. The participant and the confederate talk about objects they see on their screens. The confederate names the objects visible on her screen and the participant, seeing the same plus a number of further objects, responds what further objects are visible on his or her screen. It can be assumed that participants' gaze follows the objects that are named by the confederate while comprehending the object names, and moves on to the objects that have to be named during response planning (Altmann & Kamide, 2007; Griffin, 2001; Griffin & Bock, 2000; Huettig et al., 2011; Just & Carpenter, 1980; M. K. Tanenhaus et al., 2000). The experiment was conducted in German and the critical utterances of the confederate appeared in the following conditions, exemplified in (1) to (4).

Sentences in condition (1) (baseline condition) did not contain a lexical cue (like *and*) to mark the final item of the list (-LEX) and ended in a low falling boundary tone (+BT). Non-final list items were produced with high rising intonation and pitch peaks of equal height around 400Hz, i.e. without downsteps of pitch peaks on non-final list items (-DWNS). Sentences in condition (2) (lexical cue condition) were similar to sentences in condition (1), except that the lexical cue *und* ('and') preceded the final list item to mark the item as being the last one of the list (+LEX; if the sentence contained only one item, *nur* 'only' was used instead of 'and'). Sentences in condition (3) (no boundary tone condition) were the same as in condition (1), except that their final intonation contour was manipulated to end in a flat mid tone instead of a low falling boundary tone (-BT). Sentences in condition (4) (downstepped condition) were similar to condition (1), except that non-final list items were produced with consecutive downsteps in pitch peaks in non-final list items (+DWNS). Figure 5.1 shows the difference in intonation contours between sentences in conditions (1), (3), and (4).

- (1) *Ich habe einen Schlüssel, einen Lenkdrachen, einen Rubin.* (L%)  
I have a key, a kite, a ruby.
- (2) *Ich habe einen Schlüssel, einen Lenkdrachen **und** einen Rubin.* (L%)  
I have a key, a kite and a ruby.

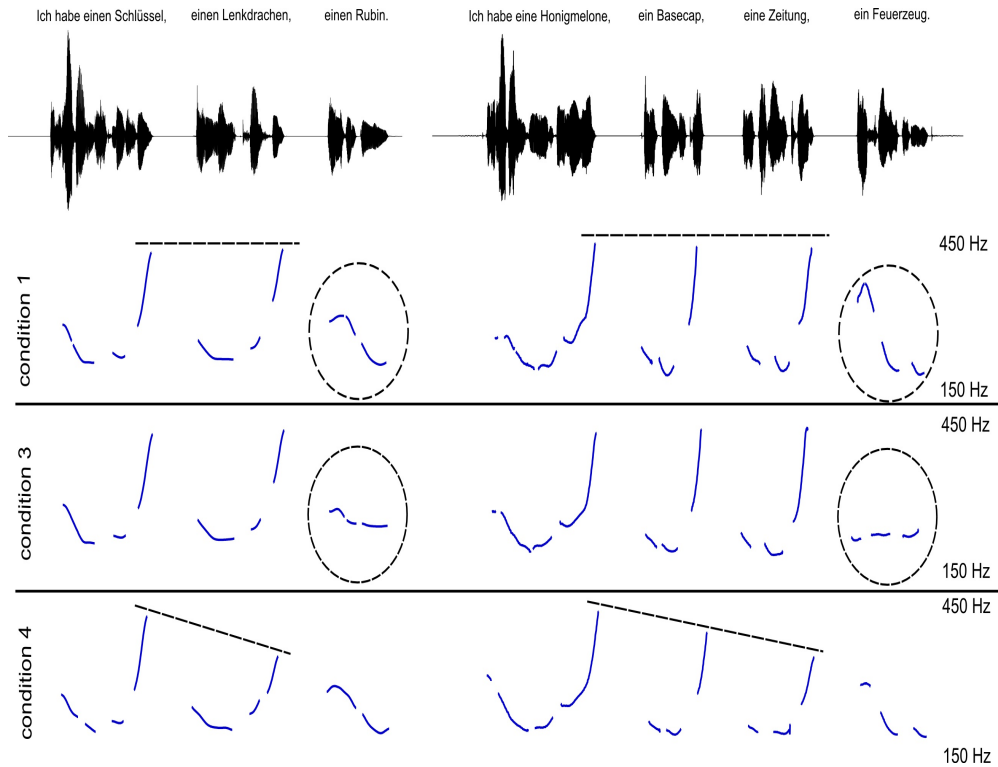


Figure 5.1: Two examples illustrating the intonation contours used in conditions 1 (baseline condition), 3 (no boundary tone condition), and 4 (downstepped condition). Condition 1 (and equally condition 2, not displayed here) contains no downsteps on non-final list items and a low boundary tone at the turn end. By contrast, condition 3 contains no final low boundary tone. Condition 4 contains downsteps and a final low boundary tone.

- (3) *Ich habe einen Schlüssel, einen Lenkdrachen, einen Rubin.* (M%)  
I have a key, a kite, a ruby.
- (4) *Ich habe einen Schlüssel, einen Lenkdrachen, einen Rubin.* (DWNS, L%)  
I have a key, a kite, a ruby.

### 5.2.1 Participants

Thirty-eight German native speakers (mean age = 22.8 years; SD = 2.9) were tested as paid participants at the MPI for Psycholinguistics. All participants reported normal or corrected-to-normal vision and normal hearing abilities. Data of three participants were not considered in the

analyses due to technical failure during recording. Of the remaining participants, 10 answered ‘yes’ to a post-experiment query whether pre-recorded materials were presented to them during the experiment. This factor was included as a binary control variable in the analyses ( $\pm$  recording\_noticed). The experiment was approved by the Ethics Committee of the Faculty of Social Sciences, Radboud University Nijmegen. Written informed consent was obtained from all subjects.

### 5.2.2 *Apparatus*

The participant and the confederate were seated in separate cabins in front of and about 60 cm away from 21 inch computer screens. They were unable to see each other and could only communicate via microphones and headphones. The participants’ eye-movements were recorded with an SMI RED-m remote eye-tracker (120 Hz).

### 5.2.3 *Visual stimuli*

Four-hundred and twenty-four pictures of concrete objects that were used in the study by [Barthel et al. \(2016, see ch. 2\)](#) were used in the experiment. All pictures, with the exception of twenty pictures used in practice trials, showed inanimate objects.

One-hundred and four pairs of displays (participant displays and corresponding confederate displays) that showed a differing number of objects drawn from the pool of object pictures were used as visual stimuli (see [Figure 5.2](#) for an example). The participant displays showed between three and five objects, including all objects shown on the corresponding confederate display plus zero, one, two, or three further objects. In participant displays that showed three objects, the objects formed an equilateral triangle, when showing four objects, the objects formed a square, when showing five objects, the objects formed an equilateral pentagon.

Ninety-two displays were critical test displays, with twenty-eight displays each showing three, twenty-eight showing four, and thirty-six showing five objects on the participant display. The confederate displays showed between zero and five objects, so that twelve participant displays showed no more objects than the corresponding con-

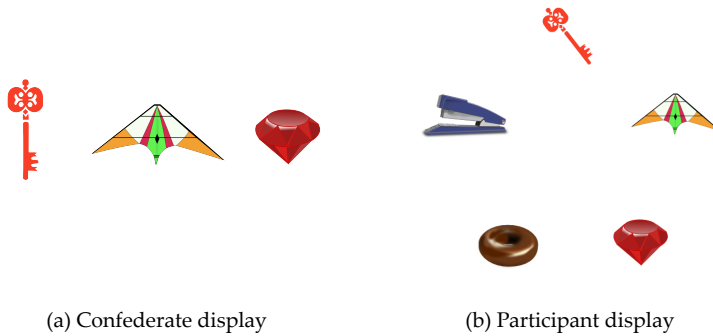


Figure 5.2: Example item displays.

federate display; twenty-eight participant displays showed one more object, twenty-eight participant displays showed two more objects, and twenty-four participant displays showed three more objects. The experiment was preceded by a practice phase using twelve display pairs, with four participant displays each showing three, four, or five objects.

#### 5.2.4 *Auditory stimuli*

Sentences accompanying the visual displays were pre-recorded, using a unidirectional Sennheiser ME64 microphone attached to a digital flash recorder. Each sentence was recorded in the conditions 1, 2, and 4. Sentences in condition 3 (no boundary tone condition) were manipulations of the corresponding sentences in condition 1 (baseline condition). The final low falling boundary tone was flattened to a mid level with Praat (Boersma & Weenink, 2015). Sentences in condition 4 (downstepped condition) were recorded and selected to contain downsteps of pitch peaks on non-final list items, with the first item peaking at about 400 Hz and the penultimate item peaking at about 340 Hz. The more items a list contained, the smaller were the differences in pitch peaks between adjacent list items. Sentences that contained 3 list items were produced with a downstep of 50 to 70 Hz. Sentences that contained 4 list items were produced with two downsteps of 30 to 40 Hz. Sentences that contained 5 list items were produced with three downsteps of 15 to 40 Hz. The pauses between object nouns were manipulated with Praat to have a random length between 400 and 600 ms, equal for the different



versions of each sentence, imitating the average length in the original recordings.

Eight sentences did not contain any object nouns and were used as fillers. In these sentences, the object list was replaced by *nichts* ('nothing'), as in *Ich habe nichts* ('I have nothing'). Sentences accompanying the twelve practice trials were produced live. These sentences were produced to sound similar to the pre-recorded sentences, using the formats that were otherwise used in conditions 1, 2 and 4 in the pre-recorded sentences.

#### 5.2.5 *Items and Design*

A participant display in combination with the accompanying sentence constituted an experimental item. In sixty-one of the items in which the confederate named at least one object, the objects were arranged in clockwise order as they were named, starting at the top of the display. In twenty-three of the items, other arrangements were used, so that the participants had to listen attentively and search for the items mentioned by the participant, rather than scanning the objects in the same order on all trials.

Four lists were constructed, with each sentence and the accompanying display appearing once per list. Since sentences with less than three objects could not appear in condition 4 (downstepped condition), and sentences with less than two objects could not appear in condition 3 (no boundary tone condition), the number of items per condition was not balanced throughout the experiment. In each list, twenty-eight items appeared in condition 1, twenty-eight items in condition 2, sixteen items in condition 3, and twelve items in condition 4. Each participant was assigned to one of the lists.

#### 5.2.6 *Procedure*

##### *Familiarization and Instructions*

The procedure followed [Barthel et al. \(2016, see ch. 2\)](#). Participants were invited to the lab to take part in a dialogue experiment. Upon arrival, they were given a picture booklet containing all pictures used in the experiment and asked to name them. In case a participant

could not recognize or name a picture, a name was provided by the experimenter. The experimenter noted down participants' responses. The familiarization phase was audio-recorded.

After the familiarization phase, the confederate arrived and was introduced as a second participant. Participant and confederate were informed that they would be seated in separate cabins and talk to each other via headphones and microphones to play the following game. They would see a number of displays on their respective screens, showing a number of objects. All objects that were displayed on the confederate display were also displayed on the participant display. The confederate was to tell the participant which things she has got on her display, so that the participant could tell the confederate what *further* objects (s)he has got. Participants were not instructed to use any particular utterance format.

The confederate was instructed to try to remember which objects she had seen and which names she had heard. This served as a cover task to distract participants from the aim of the study. Participants were told that their eye-movements would be recorded in order to study looking behavior when searching for objects on a screen whose names were heard. After instructions were given, the eye-tracker was calibrated and calibration was repeated three times during the experiment.

### *Test phase*

Before the test phase, participants completed twelve practice trials. During the test phase, all communication between the participants and the confederate was live, except for the critical pre-recorded sentences. The confederate started the trials and the corresponding pre-recorded utterances so as to make them fit naturally into the conversation.

Participants were asked to look at a fixation cross that was presented in the center of the display at the beginning of each trial, which triggered the presentation of the item displays. After a preview that varied randomly between items between 600 and 1000 ms, the stimulus sentence began.

After the experiment, participants were asked in a computerized questionnaire whether they had noticed the presence of pre-recorded speech. The answers were used as a control variable ( $\pm$ recording\_noticed).

The experiment took about 25 minutes. The entire test session took about one hour, including familiarization, test, and questionnaire.

### 5.3 RESULTS

Statistical analyses were based on linear mixed effects regression models fitted in R (R Core Team, 2014) using the package lme4 (Bates et al., 2014). Participants' fixation preferences and response latencies were the dependent variables. The maximal random effects structure justified by design was used for all models (Barr, 2013; Barr et al., 2013). Control variables were not included in the random effects structure. All categorical variables were dummy coded (0 and 1). Statistical significance was assessed with F-tests with Kenward-Roger approximations of degrees of freedom (Fox & Weisberg, 2011; Halekoh & Hojsgaard, 2014; Kenward & Roger, 1997).

#### *Response timing*

Response latencies for critical turn transitions were measured manually with Praat (Boersma & Weenink, 2015). They were coded as time intervals between the end of the incoming turn and the beginning of the response turn, excluding any non-speech sounds like audible in-breaths. Participants always named the correct objects that were not named by the confederate. Response latencies ranged from -56 ms (short overlap) to 5113 ms ( $M = 1002$  ms,  $SD = 432$  ms,  $N = 3220$ ). The present latencies are relatively long compared to averages observed in natural conversation, probably due to task demands. They are comparable to the latencies obtained by Barthel et al. (2016, see ch. 2), who used the same paradigm. Table 5.1 shows an overview per condition. For the statistical analyses, thirty-four data points (1%) were removed from the data set since they were outliers of more than three standard deviations of the mean response latency of the respective subject that produced the data-point.

Since confederate turns in the different conditions differ in their average number of objects that are named by the confederate, they are inherently of different average lengths. Because of this difference, the

number	condition			mean (SE)	N
	LEX	BT	DWNS		
1	-	+	-	1010 (12)	988
2	+	+	-	922 (12)	990
3	-	-	-	1077 (14)	560
4	-	+	+	873 (18)	402

Table 5.1: Response latencies by condition. Mean and standard error (SE) in ms.

duration of the critical turns was included as a control variable in the analysis.

To test for the effects of interest, a model was designed to fit response latencies included presence of lexical cue ( $\pm$ LEX), presence of low falling boundary tone ( $\pm$ BT), and presence of downstepped pitch peaks ( $\pm$ DWNS) as predictors and the duration of the confederate turns in seconds, as well as  $\pm$ recording\_noticed as control variables. Response latencies were significantly longer in -LEX items (condition 2) than in the baseline condition ( $\beta = 90$ ,  $SE = 19$ ,  $F(1,35) = 21.04$ ,  $p < .001$ ), i.e. participants responded slower when no lexical cue to the turn end was present. Furthermore, response latencies were significantly longer in -BT items (condition 3) than in the baseline condition ( $\beta = 60$ ,  $SE = 22$ ,  $F(1,34) = 7.39$ ,  $p = .01$ ), i.e. participants responded slower when no final intonational cue to the turn end was present.  $\pm$ DWNS did not significantly influence response latencies, meaning that the apparent difference in the descriptive statistics is merely an artifact of sentence duration.<sup>1</sup> Duration of the confederate turn had a significant effect on response latencies ( $\beta = -49$ ,  $SE = 7$ ,  $F(1,85) = 42.95$ ,  $p < .001$ ), meaning that participants responded faster, the longer the incoming turn, presumably because participants' level of preparedness to speak increases as the likelihood that the incoming turn will come to an end increases in proportion to the likelihood of the unfolding turn coming to an end (cf. Magyari et al. (2017)). Table 5.2 shows a model summary.

<sup>1</sup>A separate model was run to test for the effect of  $\pm$ DWNS in the subset of items that are directly comparable to one another, i.e. that have at least three confederate objects plus at least one additional participant object. The pattern of results is the same as in the full model, showing no effect of  $\pm$ DWNS.

	Estimate	SE	<i>t</i>	<i>F</i>	sig.
(Intercept)	953.482	53.4	17.830		
lexical cue	90.190	19.6	4.602	$F(1,35)=21.041$	***
boundary tone cue	60.344	22.0	2.741	$F(1,34)=7.391$	**
downsteps	8.663	34.4	0.252	$F(1,35)=0.061$	n.s.
sentence_duration	-48.974	7.1	-6.836	$F(1,85)=42.957$	***
recording_noticed	74.624	73.087	1.021	$F(1,32)=0.878$	n.s.

Table 5.2: Response timing model and *F*-tests. Formula:  $RT \sim 1 + LEX + BT + DWNS + recording\_noticed + sentence\_duration\_centered + (1 + LEX + BT + DWNS | subject) + (1 + LEX + BT + DWNS | item)$ . Presences of cues were used as reference levels, so that effects shown are effects of absence of cues. Asterisks indicate significance levels of effects. \*  $p < .05$ ; \*\*  $p = .01$ ; \*\*\*  $p < .001$ .

### *Eye-movements*

In order to investigate the time course of participants' planning of their response to critical confederate turns, fixations to the first-mentioned objects in the participants' responses (target objects) were analyzed. Fixations towards an area of interest covering the target objects and approximately 0.25 degrees of visual angle around them were categorized as target fixations. Figure 5.3 shows proportions of target fixations time-locked to the beginning of the last object noun in the confederate's utterance.

Participants' eye-movements were analyzed in a time window from 0 ms until 2600 ms, corresponding to the beginning of the last noun in the confederate's turn (0 ms) and the grand mean duration from the time-lock point until the beginning of the first object noun in the participant turn (2600 ms) respectively. Fixations to the target object were aggregated to empirical logits in 100 ms time bins over the course of the analysis window by subjects and by items, respectively (Barr, 2008). The empirical logit transformation removes statistical dependencies in the data, which is important to satisfy the assumptions of linear regression. Only trials that included both looks for production and looks for comprehension were analyzed, excluding trials in which the confederate named none or all of the displayed objects. Seventy-eight of the remaining trials were discarded due to trackloss, i.e. missing data for a consecutive stretch longer than 500 ms within the time window of analysis. The final data set included 2442 trials.

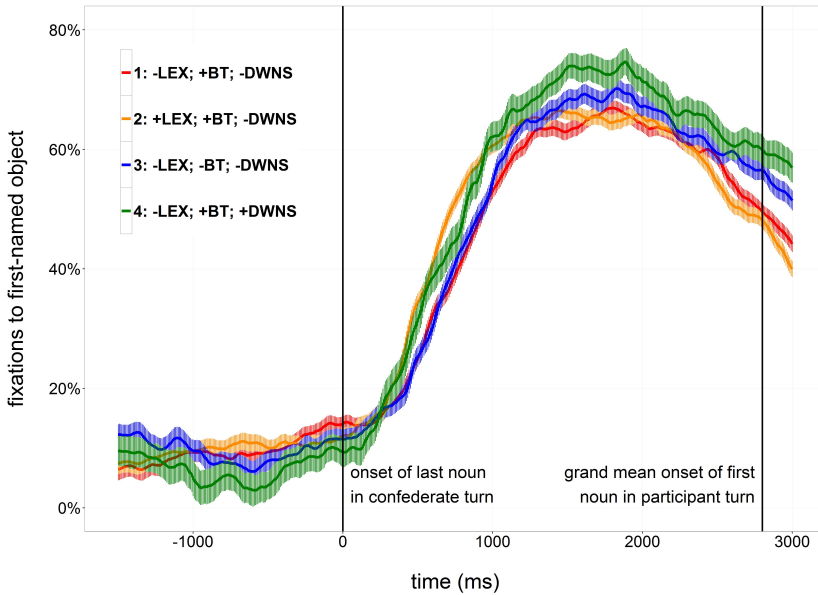


Figure 5.3: Proportions and standard errors of looks to the target object time-locked to the onset of the last object noun of the confederate turn (0 ms).

Eye-movement patterns were analyzed using hierarchical quasi-logistic growth curve modeling (Mirman, 2014). Growth curve analysis is a variety of mixed effects regression that uses orthogonal polynomial time terms as predictors to model differences in curve shapes, in this case differences in growths of fixation likelihoods as expressed by empirical logit transforms. Linear, quadratic, and cubic orthogonal time terms were included as predictors in the model. The linear time term (Time) models the overall increase in fixations over the time course of a trial. The quadratic time term ( $\text{Time}^2$ ) models the steepness of the curve, i.e. how “U-shaped” it is. The cubic time term ( $\text{Time}^3$ ) describes at what point in time fixations increase (“S-shaped” curve). An interaction of the linear time term with a factor of interest would signify a difference in the slope of the increase of proportions over time in one level of the factor versus another level. An interaction of the quadratic time term with a factor of interest would signify a difference in the speed with which proportions increase in one level of the factor versus another level, thereby describing the pointedness of a U-shaped curve. An interaction of the cubic time term with a factor of interest would signify a difference in latency, i.e. a difference in when proportions start to

increase in one level of the factor versus another level. This interaction is most interesting to us, since it models the predictions about when participants shift their gaze towards the target objects in the different conditions. Table 5.3 shows an overview of model summaries and significance levels.

Comparison	Effect	$\beta$	SE	F	sig.
cond. 1 vs. cond. 2 ( $\pm$ LEX)	$t^1 \times \text{cond.}$	3.30	0.29	$F(1,345)=6.51$	*
	$t^2 \times \text{cond.}$	-2.98	0.21	$F(1,740)=4.03$	*
	$t^3 \times \text{cond.}$	-0.51	0.19	$F(1,1116)=6.21$	*
cond. 1 vs. cond. 3 ( $\pm$ BT)	$t^1 \times \text{cond.}$	0.47	0.34	$F(1,270)=1.64$	n.s.
	$t^2 \times \text{cond.}$	-0.52	0.29	$F(1,255)=2.90$	n.s.
	$t^3 \times \text{cond.}$	0.17	0.19	$F(1,924)=0.76$	n.s.
cond. 1 vs. cond. 4 ( $\pm$ DWNS)	$t^1 \times \text{cond.}$	-0.25	0.35	$F(1,161)=0.43$	n.s.
	$t^2 \times \text{cond.}$	-0.24	0.34	$F(1,181)=0.46$	n.s.
	$t^3 \times \text{cond.}$	0.23	0.26	$F(1,273)=0.71$	n.s.

Table 5.3: Eye-movement results of by-subject analysis. Formula =  $\text{emplit} \sim (\text{time1}+\text{time2}+\text{time3}) * \text{condition} + (1 + (\text{time1}+\text{time2}+\text{time3}) * \text{condition} \mid \text{subject/item})$   $t^2 = \text{TIME}^2$ ,  $t^3 = \text{TIME}^3$ .  $\beta$ 's indicate effects of absence of cues. Asterisks indicate significance levels of effects. \*  $p < .05$ . By-item analysis yielded a similar pattern of results.

Visual inspection of the proportions of fixations indicates that proportions of target looks are generally at a low level during the incoming turn, increase suddenly after the onset of the last noun of the confederate turn and start to decrease again after about two seconds (Figure 5.3). In condition 2, which contains a lexical cue to the turn end, the initial increase in proportions of target looks is steeper and takes place earlier than in condition 1, which does not contain this cue. Similarly, the initial increase in proportions in condition 4, containing downsteps of pitch peaks in non-final list items, seems to be slightly steeper than in condition 1, not containing downsteps. Proportions of target looks in condition 3, containing no low falling boundary tone, and in condition 1 do not obviously differ.

Conditions 1 and 2 were compared to test for effects of the lexical cue to turn end ( $\pm$ LEX). Both by-subject and by-item comparisons showed interaction effects of  $\text{Time}^2 \times \text{LEX}$  and  $\text{Time}^3 \times \text{LEX}$  in the direction of earlier and steeper increases in trials with a lexical cue than in trials without a lexical cue.

Conditions 1 and 3 were compared to test for effects of a boundary tone cue ( $\pm$ BT). No interaction of  $\text{Time}^2 \times \text{BT}$  or  $\text{Time}^3 \times \text{BT}$  was found

to be significant, indicating that proportions of target looks were not modulated by the presence or absence of a final low boundary tone.

Conditions 1 and 4 were compared to test for effects of downsteps in pitch peaks on non-final list items ( $\pm$ DWNS). No interaction of  $\text{Time}^2 \times \text{DWNS}$  or  $\text{Time}^3 \times \text{DWNS}$  was found to be significant, indicating that the presence or absence of downstepped pitch peaks on non-final list items had no influence on the growth of proportions of target looks.

#### 5.4 DISCUSSION

The present study set out to investigate whether, in a conversation, next speakers make use of cues to turn ends to temporarily align their next turns to the end of the incoming turn. Three types of cues were tested: a lexical cue that indicated that the turn would end after the following noun phrase, a final boundary tone that prosodically marked the turn as complete, and a pitch contour that allowed for an early estimation of the length of the unfolding turn. To test the use of these different cues in turn-taking, an experiment using the list-completion paradigm was designed, in which a naive subject and a confederate took turns in naming objects (Barthel et al., 2016, see ch. 2). Which objects participants had to name depended on the objects that were named in the critical turns by the confederate. These critical turns either did or did not contain the relevant turn taking cues. The conversation of participant and confederate and the participant's eye-movements were recorded to analyze at what moment participants planned and initiated their response turns.

Participants were found to start planning their turn as soon as they knew which objects they had to name, replicating the results of Barthel et al. (2016, see ch. 2). When the lexical cue *und* ('and') was present before the last item of the list of the incoming turn, participants knew the following list item to be the last item before the turn would be complete. In sentences containing this lexical cue, participants started planning their response earlier than in sentences not containing this cue, showing that they started planning their response as soon as possible. The average length of the lists' final nouns was 670 ms. Dependent on when the turn-final boundary tone becomes recognizable (also indicating the end of the turn) the lexical cue gave participants a head-start in



response planning of at least the length of one syllable. Through this head-start in response planning, participants could respond faster after turns with a lexical cue to the turn end (condition 2) than after turns in the baseline condition.

Contrary to the lexical cue, which was located before the last noun phrase of the turn, the final boundary tone was located right before the end of the turn. It was argued before in the literature that turn-taking cues which are located at the end of a turn could not be used to time the planning of the content of the response (de Ruiter et al., 2006; Levinson, 2012). The present study supports this argument. No difference in the timing of looks for response planning was found between turns that did contain a turn-final prosodic cue and turns that did not. However, participants were found to rely on turn-final cues to minimize the gap between turns. Response times were faster after turns containing a turn-final boundary tone than after turns not containing this cue. This pattern of results suggests that turn-final cues to the turn-end are irrelevant for the timing of response planning but help next speakers to time the initiation of their turn. Consequently, next speakers seem to use turn-final cues as “go-signals” to launch their response when turn transition becomes relevant. The combination of the absence of an effect on the timing of response planning as measured by participants’ gaze movements and the presence of an effect on their response latencies shows that the measure of eye-movements is a good candidate to differentiate between the two processes, response preparation and response initiation.

No evidence was found that next speakers make use of the early downstep prosodic cue to turn length. Participants were not found to use downsteps on pitch peaks in list items to plan their response earlier or respond faster than in turns without a downstepped pitch contour. This early prosodic cue could have been used by participants as much as the lexical cue to the last list item, since the number of list items might have been guessed from the size of the downsteps. However, it is less discrete than the lexical cue, which might be the reason why participants relied on this cue less than on the lexical cue. Both findings on the use of prosodic cues are in line with the conclusions drawn by Bögels and Torreira (2015), who found that participants in their experiments only relied on final intonational turn-taking cues but not

on turn-initial intonational cues when trying to detect turn-completion points.

In conclusion, the results suggest that next speakers plan the content of a response as early as the incoming turn's message becomes recognizable and that turn-final cues can function as "go-signals" to initiate response in a timely fashion. Given that lists are a natural kind of conversational turn that are frequently encountered in everyday situations, the present results can be assumed to be generalizable to casual conversation (Selting, 2007). Turn-final cues can therefore be assumed to be used by speakers to indicate turn-yielding and next speakers can orient to them so as to minimize gaps when taking the floor. The findings show that response turn preparation and the timing of its articulation need to be regarded as separate processes. Response planning depends on (an anticipation of) the incoming turn's message, while response initiation depends on the next speaker's confidence that the incoming turn comes to conclusion and that speaker transition becomes relevant. Consequently, the findings support turn-taking models that include early content planning and the use of turn-final cues as "go-signals" to initiate response (e.g. Levinson & Torreira, 2015; Sacks et al., 1974).



# 6

---

## SUMMARY AND MODELLING OF RESULTS

---

The results of the studies and the previous literature reported and discussed in Chapters 2 to 5 are incorporated into a cognitive model of turn-taking which will be presented in this Chapter. Beforehand, short summaries of the previous chapters shall be provided.

### 6.1 SUMMARY OF RESULTS

#### 6.1.1 *Summary Chapter 2 – The Timing of Response Planning in Dialogue*

The study presented in Chapter 2 made use of a cooperative experimental paradigm in which participants had to listen and respond to turns by a confederate interlocutor in a fairly unrestricted, natural fashion (the list-completion paradigm). While the use of confederates can come with potential drawbacks (Kuhlen & Brennan, 2013), the presented experiment was designed to gain interaction data with high ecological validity while not sacrificing experimental control. To this end, all critical utterances by the confederate were pre-recorded and played to participants at the relevant moments. These utterances were divided into four conditions that were designed to answer two core questions relevant to the mechanisms of turn taking. (1) At what point in time do next speakers start planning their own turn and what is the reference point for the initiation of planning – the earliest possible moment or the end of the incoming turn? (2) Does the end point of the incoming turn need to be projected in order to begin to plan the response turn? The critical confederate utterances were structured so as to, on the one hand, either allow for planning in overlap or not, and, on the other hand, either allow for early projection of the turn end or not. To

explore at what time next speakers start to plan their response turn, we compared eye-movements in early-message conditions that allowed next speakers to plan their turn in overlap and late-message conditions that only allowed for response planning after the incoming turn. Participants moved their gaze for response planning as soon as all necessary information was available to know what a relevant response needed to contain — even in early-message conditions, where all necessary information was available already before the end of the incoming turn. This head start in response planning in the early-message conditions led participants to initiate their response faster than in late-message conditions. These results show that next speakers plan their turn as early as possible, even when the incoming turn is not yet ending at the point in time when its message can be understood. Consequently, the reference point for the initiation of response planning is the earliest possible moment in the incoming turn, rather than the point at which the incoming turn terminates.

To answer the question whether projection of the point in time the incoming turn will come to conclusion is a necessary prerequisite for the initiation of response planning, participants' eye-movements in critical turns that contained projectable turn-end points were compared to those in turns whose end-points were not projectable. Again, participants' gaze moved to the target objects as soon as all necessary information was available, irrespective of the incoming turn-end's projectability. These results show that the end point of the incoming turn does not need to be projected by a next speaker in order to start planning their response.

#### 6.1.2 *Summary Chapter 3 — Progression of Speech Planning in Overlap*

After we established that next speakers start to plan their next turn as early as possible and in overlap with the incoming turn in Chapter 2, we continued to investigate the time-line of utterance planning in overlap in Chapter 3. The results of the study presented in Chapter 2 showed that next speakers prepared their turn at least conceptually in overlap with the incoming turn. In Chapter 3 we asked whether later stages of production planning are executed in overlap as well. The time pressure present in conversational situations would be a good reason to

hypothesize that response planning should indeed proceed to later processing stages during overlap. On the other hand, processes of speech decoding and encoding have been shown to partly rely on the activity of overlapping neural networks in the brain and to interfere with one another. The inefficiency caused by these potential interference effects are potential reasons to delay certain stages of production planning until the end of the incoming turn (see also Chapter 4). Since each planning process that is executed in overlap is prone to an additional interference, there might be a “sweet spot” where the interests of early planning due to time pressure and of delayed planning due to efficiency and ease of planning meet.

The main experiment of the study presented in Chapter 3 combined three paradigms to approach the question which stages of response planning are run through in overlap with the incoming turn. Participants were presented with four pictures while listening to a question they had to answer by naming one of the pictures. Doing this task, participants would look towards the target picture when planning to produce its name. In a quarter of trials, the display changed shortly after participants gazed towards the target picture in order to name it. In that case, the presented pictures disappeared and the target picture would be replaced by a word. In that instance, participants’ task switched from answering the question to making a lexical decision about the presented word, deciding whether it was an actual Dutch word or a non-word. In half of these critical trials, the presented questions made it possible to know and prepare the verbal answers in overlap with the question, while in the other half early planning was impossible since essential information was disclosed only at the questions’ end. Participants’ eye-movements revealed that participants generally started planning their response as soon as possible, replicating the results of the study presented in Chapter 2. The crucial effects of interest concerned participants’ lexical decision performance, however. Words yielded significantly slower and more error-prone decisions when they were presented after pictures with phonologically related names than when they were presented after pictures with unrelated names. Importantly, the size of the interference effects did not differ between trials in which participants were planning their turn in overlap and trials in which they planned their turn in silence after the incoming turn. These findings show that participants were already planning

the phonological form of their response while the incoming turn was still unfolding. Combining this conclusion with the result obtained by Bögels and Levinson (in prep.), that articulatory preparation is held back until the end of the incoming turn, we suggest that the sweet spot of response preparation in overlap with the incoming turn lies somewhere between word form retrieval and articulatory preparation. Whether or not it lies before or after phonetic encoding, syllabification, or the retrieval of articulatory scores is subject to further investigation.

### 6.1.3 *Summary Chapter 4 – Processing Load in Speech Planning in Dialogue*

Chapters 2 and 3 established the facts that next speakers plan their turns at least up until retrieving the relevant word forms in overlap with the incoming turn and that they do so irrespective of an accurate projection of the end point of the current turn. Following from the observation that planning in overlap is common and highly practiced, in Chapter 4 we asked the question whether it is cognitively more demanding than planning in silence. To answer this question, we analysed data collected during the list-completion task study presented in Chapter 2 and operationalized participants' pupillometric responses as an indicator of processing load. We compared pupil dilations of instances of planning in overlap with instances of planning during the gap between turns. We found that increases in pupil diameter were more pronounced and reached their peak later in instances of planning in overlap than in instances of planning in silence. These results indicate that planning in overlap is more effortful and takes longer than planning in silence, which means that, from a processing point of view, planning in overlap is a less efficient strategy; it is, however, nonetheless the norm in conversation. Notably, predictability of the overlapping material at the end of the incoming turn does not modulate processing load while planning in overlap. These findings show that, at turn transitions, the processing system is under time pressure exerted by the turn taking system and that this pressure leads to the fast exchanges of turns that are regularly observed in conversation. This speed of speaker change, in turn, is a pre-requisite for the dense semiotics of turn timing that was discussed in Chapter 1. Fast turn timing in itself and turn timing as a source of meaning lead to increased effectiveness in communication,

which we speculate is the reason why next speakers commonly pursue a planning strategy that entails increased processing load.

#### 6.1.4 *Summary Chapter 5 – The Timing of Next Turn Initiation*

In Chapters 2 to 4 we investigated the timing and mechanisms of planning a turn in a dialogic situation. Assuming a response turn was planned in time, next speakers still face the task to accurately time their turn's articulation in alignment with the incoming turn's ending. In Chapter 5 we therefore investigated what cues next speakers orient to so as to know the timing of the incoming turn's point of completion in order to time the initiation of their response turn. To answer this question, we made use of the list-completion paradigm we described in Chapter 2. Critical utterances by the confederate were presented in four conditions: (1) a baseline condition containing no cues to the timing of the end of the incoming turn; (2) a condition containing a turn final intonational cue that indicated close proximity of turn completion; (3) a condition containing an early intonational cue that allowed for an estimation of the incoming turn's length; and (4) a condition that contained a lexical cue to indicate that the turn would end after the following noun phrase. The lexical cue and the turn-final intonational cue, but not the turn-initial intonational cue, were found to lead to faster response latencies when they were present than when they were not. The lexical cue was also an early cue to the message of the incoming turn, leading participants to start planning their response turn earlier when the cue was present as compared to when it was not present, replicating the results of the study presented in Chapter 2. The presence or absence of any of the two intonational cues did not trigger a similar effect of early response planning. These findings indicate that while next speakers do not seem to rely on turn initial cues to turn length, they use turn final cues to turn completion to time the initiation of the articulation of their own turn. The results illustrate that the mechanisms of speech planning and initiation of articulation are separate in timing, and that a combination of measures of eye-movements and turn-transition times makes it possible to tease them apart.



## 6.2 A MODEL OF TURN TAKING

On the basis of the results presented in Chapters 2 to 5 and in the literature reviewed in these chapters, I formulate a cognitive model of turn taking (Figure 6.1). In this Chapter, the proposed model will be described and will be presented from the perspective of a current listener/potential next speaker in a conversational situation embedded in the current discourse of the interlocutors' interaction. The model is an amplification and modification of the one presented in Levinson and Torreira (2015, Ch. 7) in the light of the experimental findings presented in this thesis. It includes the most important aspects of the cognitive processes that are at work during conversation and proposes a way to implement them so that interlocutors meet the socially grounded rules of turn taking that were first formalized in the sequential production model by Sacks et al. (1974). This seminal model postulated that the current speaker has the right to produce a single turn-constructive unit of variable size, after which a turn transition might occur, with the next speaker who speaks up first at that point gaining the right to the next turn. The model presented here describes the language processing mechanisms that are at work to meet the timing challenge that is born out of the rules of sequential production. While the model by Sacks et al. (1974) describes *what* interlocutors do from a sociological perspective, the present model adds a combination of findings and considerations on *how* they do it from a psycholinguistic perspective.

A cluster of models competing with the sequential production model that offer explanations for the psychological issues of *how* turn taking is achieved by interlocutors are called signalling models (Duncan, 1972, 1974; Duncan & Fiske, 1977; Duncan & Niederehe, 1974). In contrast to the sequential production model, signalling models assume that the transition of turns is handled unilaterally by the current speaker by means of signals that are either present or not present in the current turn at talk. The addressee merely needs to spot these signals and react according to them by either taking the next turn at the moment a turn-yielding signal was given or withhold from taking a turn in the absence of such a signal or in the presence of a turn keeping signal which displays the intention of the current speaker to take yet another turn. This conception of the organization of turn taking does not offer a full picture of the processes governing the timing of talk in conversation,

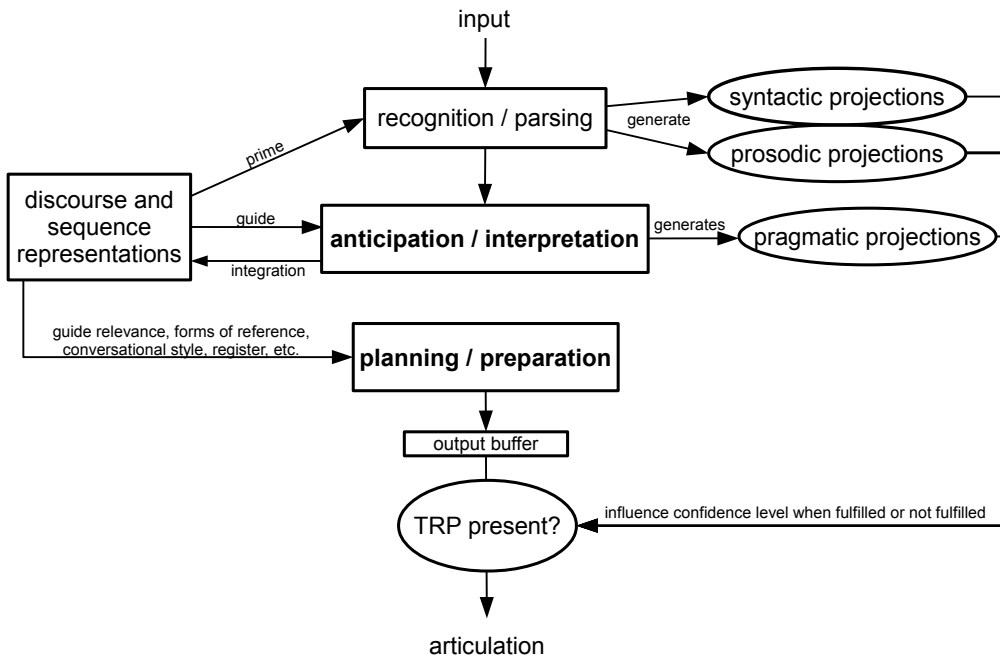


Figure 6.1: Model of Language Processing in Conversation. (TRP = transition relevance place)

and it does not explain the findings presented in this thesis. In the study presented in Chapter 2 we found that next speakers do not wait to plan their turn until they receive a signal of turn completion but rather start planning their response as soon as the incoming message is clear enough to start to formulate a response. This early planning did not depend on the predictability of the incoming turn's end, or, in other words, on any signal of the current speaker yielding their turn. In support of the signalling account, we found in the study presented in Chapter 5 that next speakers do indeed react to cues to turn finality insofar as they launch articulation of their response after having received enough evidence of the incoming turn coming to an end. However, as evidenced by these very findings, these cues to turn finality do not constitute a dichotomous turn yielding signal but are rather taken up as additive evidence of the incoming turn coming to

completion. While the signalling account correctly predicts that cues to the end of the incoming turn are relevant for the timing of turn taking, it is a purely behavioural model remaining agnostic about the processes of preparing the next turn and about their timing, which, as was shown in Chapters 2 and 3, is independent of any cue to the end of the incoming turn as these processes are executed as far as possible not in reaction to but rather in anticipation of the incoming input.

The model proposed here remedies the shortcomings of the signalling models and gives a processing account on how the social rules formulated in the sequential production model are complied with by conversational partners. In this model, each interlocutor carries their own representation of the current discourse, which encompasses all known facts and assumptions about the conversational situation, including knowledge about the interlocutors, their (probable) goals, the type of personal relation between them, and, importantly, what has been said, or rather, what actions have been intended in the course of the current (and previous) conversation(s) (Zwaan & Radvansky, 1998). Consequently, two types of information are held in the discourse model, local information obtained during the current conversation and more global information that was available already before the conversation began. During conversation, these two types of information can influence and update one another while interlocutors continuously aim to coordinate their models of the current discourse (Clark & Brennan, 1991; Kuhlen, Allefeld, Anders, & Haynes, 2015; Kuhlen, Allefeld, & Haynes, 2012; Stephens, Silbert, & Hasson, 2010). Both global and local information contained in the discourse model of a speaker can influence production processes, for instance guiding forms of reference, decisions in register choice, conversational style (e.g. formal vs. casual), grammatical complexity, and precision in enunciation of predictable in contrast to unpredictable words (Bard et al., 2000; Bard & Aylett, 2000; Brennan, 1991; Kuhlen & Brennan, 2010; Mcallister, Potts, Mason, & Marchant, 1994). Amongst the local cues to the current conversational situation, the discourse representations centrally include the position and function of the current turn in the running sequence of turns that are exchanged, especially if they are pursuing a conversational goal that requires a number of turns to be reached (Schegloff, 2007; Stivers, 2012). The next speaker needs to prepare an utterance that is a relevant contribution to the conversation at the particular point in time when

the current turn ends and speaker transition becomes relevant. This means that the next turn needs to be relevant to the conversation in the light of the current turn already being on record and having updated the discourse representations held by the interlocutors. The task of producing a relevant contribution to the ongoing conversation becomes challenging as the next speaker tries to take her turn without leaving a long gap after the incoming turn because of the turn-organizational and semiotic reasons discussed in Chapter 1. The preparation of the next turn needs to be done under remarkable time pressure, since it needs to be a fast and appropriate reaction to the turn that is still unfolding while the response turn is already being prepared.

The next speaker incrementally receives the current turn as input, recognizes the words contained in it and parses the incoming syntactic structure. Morpho-syntactic cues in the incoming turn provide for projections of syntactic structure and slots that need to be filled for the current turn to become syntactically complete (Auer, 2005, see also Chapter 5). The evidence presented in Chapter 5 showed that successful lexico-syntactic projection speeds up turn-transitions. Similarly, unfolding prosodic structures, such as pitch and intensity contours are parsed and their continuations and probable closings are projected while the current turn is unfolding (Bögels & Torreira, 2015; Cutler, 1976; Local & Walker, 2012; Wells & Macfarlane, 1998). As shown in Chapter 5, turn-transitions are shortened in cases where the end of the incoming turn can be accurately projected by either lexico-syntactic or prosodic means. The proposed model assumes these projections to be constantly updated as more information is coming in, allowing for increasingly accurate projections of what it takes for the current turn to become complete.

The output of word recognition and parsing is incrementally fed into an interpretation process that decodes the message and ascribes an intended action to the interlocutor that fits the currently held discourse model. This interpretation process is anticipatory, as the next speaker tries to predict the content of the incoming turn, with predictions being updated as more information becomes available. The predictive and increasingly reliable interpretations of the incoming message are fed into an integration loop, constantly updating the current representations of the discursive context, which in turn influence the processes of input recognition and interpretation, raising the levels of resting

activation of lexical entries (priming) that are related to the current topic of conversation, biasing alternative readings of the incoming turn and aid the processes of interpretation and action ascription.<sup>1</sup> Note here that anticipations of the next interactional move of the interlocutor, and hence also pragmatic projections, can be constructed even before the beginning of the speaking turn and do not depend on the progress of syntactic or prosodic projections. As shown in Chapter 2, a response can be formulated earlier during the incoming turn if the message of the incoming turn can be projected before its end, leading to shorter gaps between turns. Discourse and sequential representations, informed for instance by visual input (e.g. interlocutor picking up objects, displaying facial expressions, averting their gaze, etc.) or background knowledge about the interlocutor or ones own social relation to them can trigger anticipations of the nature of the next speech act, and pre-speech cues such as audible inbreaths (Torreira, Bögels, & Levinson, 2015) or silent gaps between turns (Bögels, Kendrick, & Levinson, 2015) as well as lexical tokens such as discourse particles (Tanaka, 2015) can guide anticipations of the incoming turn's message. In this way, the message the current speaker intends to convey with the current turn can in many cases be anticipated early on (Gisladottir et al., 2018, 2015; Gisladottir, Chwilla, Schriefers, & Levinson, 2012), since the space for possible interpretations is limited by the knowledge of what would constitute a message that would fit the current discourse and sequential position. With these anticipations of the action that the current speaker intends to take with the current turn, the listener forms pragmatic projections of what is needed for such an action to become complete. All projections, lexico-syntactic, prosodic, and pragmatic, are refined and updated with each increment that is processed at the level that formed the projection.

One model of language processing in dialogue proposed by Pickering and Garrod (2007, 2013) and Garrod and Pickering (2015) assumes that the language production system is used to produce projections of these kinds. The results presented in this thesis are challenging for any model that assumes the speech production system to be responsible for such projections. As was shown in the studies presented in Chap-

---

<sup>1</sup>While lexical entries can be primed by the currently held discourse representations, this is not assumed to be the only way in which priming can occur. While this particular sort of priming is central for the challenge of accurate turn timing, priming of lexical entries and morpho-syntactic structures also occurs within the recognition/parsing stage and independent of ensuing interpretation or integration processes.

ters 2, 3, and 5, the production system is occupied with production proper already during the incoming turn, while comprehension (and projection) of the incoming turn are still continuing. Simultaneously producing projections of the incoming turn and speech output for the next speaker's own turn would lead to interference in the speech production system and slow down the processes of either projection or production proper or both. Running both projections and speech planning on the production system should therefore lead to increased processing load. However, in the study presented in Chapter 4 we found that projection did not influence processing load during production planning, calling into question whether the production system is responsible for projections on speech input. On the basis of the results presented in this thesis, it seems more plausible to assume the said projection processes to solely involve the speech comprehension system, as modeled here. As more evidence is coming in and anticipations become more reliable, resources will gradually be shifted away from comprehension towards production of a next turn, while incoming material can be processed more shallowly in order to monitor it for cues to the end of the incoming turn or strong evidence of projection error (Ferreira, Bailey, & Ferraro, 2002; Karimi & Ferreira, 2016; Sanford & Sturt, 2002). As an aside, the claim that lexico-syntactic, prosodic, and pragmatic projections are not generated by the speech production system does not entail that the proposed model stands in opposition to any theory of speech perception that assumes analysis-by-synthesis, such as the motor theory of speech perception (Galantucci et al., 2006; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Stevens, 1960), as the model in its presented form allows for motor simulations to occur in order to decode incoming speech without expecting increased processing load on that level when speech planning and comprehension are executed in parallel, since the model does not assume motor codes for production to be retrieved before the next speaker is certain that turn transition becomes relevant.

Dependent on their own intentions to contribute to or change the current discourse situation, next speakers will plan a turn that is relevant and appropriate to the current discourse and sequential position according to the currently held state of the discourse representation. The model assumes that, due to the time pressure in turn taking, cognitive resources will be drawn from comprehension and used for

production as soon as the discourse and sequence representations have been updated with reliable anticipations of the incoming turn. Planning is pursued incrementally, with the size of increments depending on working memory capacities and reducing with increasing time pressure (Ferreira & Swets, 2002; Griffin, 2001; Konopka, 2012; Korvorst, Roelofs, & Levelt, 2006; Swets, Jacovina, & Gerrig, 2013, 2014; Wagner, Jescheniak, & Schriefers, 2010). As shown by the findings in Chapter 3, all stages of speech planning, including at least conceptual planning, lemma selection and word form retrieval, and possibly down to the preparation of a phonetic plan are run through as early as possible. The output of the speech planning process is sent to an articulatory buffer where it is stored until articulation is launched (Levelt, 1989; Piai, Roelofs, Rommers, Dahlslätt, & Maris, 2015; Postma, 2000). Articulation of the prepared material is initiated only when a transition relevance place is recognized and the floor is open for the next turn to be produced.

The recognition of the presence of a transition relevance place is modeled as being an adjustable threshold of certainty. As the different kinds of projections, syntactic, prosodic, and pragmatic, that are formed along the comprehension and interpretation processes, are fulfilled or at least become increasingly reliable, the certainty of the presence of a transition relevance place increases. How high or low the threshold of certainty is for articulation to be initiated is flexible, and might depend both on personal factors such as the next speaker's mood as well as on inter-personal factors such as the difference in hierarchy of the interlocutors or the nature of the conversation being cooperative or rather competitive, affecting the general speed of conversation. Additionally, the height of the threshold critically depends on the next speaker's intentions. If a current listener's priority is to transmit some piece of information and then end the conversation soon, the certainty threshold will be rather low so that no transition relevance place will be missed, accepting the chance of slight overlap. If, on the other hand, the current listener wants to avoid overlap, e.g. for politeness reasons, the threshold will be rather high, accepting (or in some cases even favouring) possible self-selections of the current speaker for another turn at the next transition relevance place. Accordingly, the threshold can be adjusted during the course of a conversation and usually rests at a level that is not reached by only a single type of projection being fulfilled but only if evidence

for a turn end clusters at moments where the incoming turn comes to syntactic, prosodic, and pragmatic completion (Ford et al., 1996; Ford & Thompson, 1996). Notably, while silence is an important indicator for the presence of a transition relevance place, it is not sufficient in itself without the additional presence of syntactic, prosodic and/or pragmatic closure of the incoming turn (Gravano & Hirschberg, 2011; Yngve, 1970). However, a short silence may be responsible for reaching the certainty threshold at a point it would not have been reached otherwise due to remaining uncertainty, e.g. because pragmatic closure was absent in the presence of syntactic closure or vice versa. In the absence of any other closure, silence would trigger the threshold only after a longer period, leading the conversation to continue after a lapse. As soon as the threshold is reached, articulation of the buffered material will be initiated, presuming that (the beginning of) the next turn has already been successfully prepared to fit the subsequent sequential slot. In case the next speaker is not yet prepared to initiate articulation, fillers might be used to indicate that the presence of a transition relevance place was recognized and the next turn will be taken as soon as the next speaker is ready (Casillas, 2014; Clark & Fox Tree, 2002; Smith & Clark, 1993).

At this point, a significant difference between the proposed model and the model presented in Levinson and Torreira (2015) deserves to be noted. While in the previous model the order of subtasks of the next speaker was (at least implicitly) ordered linearly, with monitoring the input for syntactic completion only following successful action recognition, and monitoring for turn-final cues only in the environment of an imminent syntactic closure, the current model assumes all the respective processes to run in parallel, with syntactic frames and their minimal requirements as well as prosodic developments being projected during parsing even in the absence of any anticipation of the incoming message. In that way, the accurate prediction of the point in time when the current turn will come to completion and the preparation of a relevant next turn run independently of one another, and articulation of the buffered material is initiated only in case of certainty that transition will become immediately relevant. As such, the processes of content or intention apprehension on the one hand and timing estimation on the other hand are assumed to run completely parallel in time (see also Garrod & Pickering, 2015). Consequently, the model assumes, the



next speaker is concurrently predictive in comprehension and action recognition as well as reactive in initiation of articulation.

Where exactly the certainty threshold lies sets the range for expectable response times. The necessary level that needs to be reached for articulation to be launched might be influenced by socio-cultural factors, explaining differences in the timing of turn taking between the sexes (S. G. Roberts, Torreira, & Levinson, 2015), between different speaker communities (Stivers et al., 2009), as well as between different communicative contexts and emotional states of interlocutors (Collins, 2014; Heritage, 1984; Schegloff, 1992). The division of these two sets of processes, dealing with the content of turns on the one hand and their timing on the other hand, explains a major part of the developmental trajectory of turn taking skills in infancy. While detecting the ends of turns is accomplished fairly well already by very young children between one and two years of age (Casillas & Frank, 2013, 2017; Lammertink et al., 2015), planning a next turn in time is challenging, especially when children's utterances begin to contain linguistic content of increasing complexity (Casillas et al., 2016; Hilbrink et al., 2015). As skills in language production and message anticipation improve in later childhood, children learn to prepare their utterances at adult-like speed and have them ready for articulation at the point in time when the incoming turn comes to an end.

The proposed model predicts the most prominent aspects of turn-timing in conversation that have been reported in the literature. Next speakers prepare their turn as early as possible and buffer the output of the language planning process until they are certain enough that transition is relevant, initiating articulation at that time. To launch articulation takes time itself, with minimal speech initiation times around 200 ms (Fry, 1975; Izdebski & Shipp, 1978; Shipp, Izdebski, & Morrissey, 1984) and about 280–340 ms in cases of short words that had to be maintained in overlap and uttered upon a predictable prompt (Jescheniak, Hahne, & Schriefers, 2003), making short gaps the most common type of turn transition. Short overlaps are also common, regularly coming about in cases where the certainty threshold was reached before the actual end of the incoming turn, e.g. when the incoming turn ends in a question tag or in other increments to the core turn. In such cases, the incoming turn contains a point of syntactic, prosodic, and pragmatic completion before its actual end, so that the certainty

of a transition relevance place passes the threshold and articulation of the next turn is released. As the threshold of certainty is modeled to be a variable parameter, turn taking style can vary between different encounters and change during a single conversation. If, for instance, interlocutors' thresholds are very low when they are engaged in a quarrel, more overlap is to be expected, which should still adhere to the observed regularities of turn taking, e.g. occur in the vicinity of transition relevance places. If, on the other hand, the next speaker talks to a socially senior interlocutor and his threshold is set to be rather high, hardly any overlap and longer inter-turn gaps can be expected. Conceivably, the threshold can be influenced by the priorities the speaker sets for any given conversational situation. If understanding the input thoroughly is given priority at a particular moment, for instance when asking for instructions, the threshold will be raised so as to not miss out on any important piece of information. If, on the other hand, passing information, airing an opinion, or presenting knowledge are prioritized, the threshold will be lowered in order to not miss out on a chance to take the next turn at talk.

The model accounts for the observed complexity of turn-timing. It has been shown that a number of processing factors influence the timing of speech production and hence the timing of turn taking (e.g. [Barthel & Sauppe, 2019](#); [Damian & Dumay, 2007](#); [Gleitman, January, Nappa, & Trueswell, 2007](#); [Griffin & Bock, 2000](#); [Jescheniak & Levelt, 1994](#); [Schnur et al., 2006](#), see Chapter 4), and obviously the next turn can only be uttered when it has been prepared. A higher speech rate of the incoming turn, for example, leads to longer turn transition times ([S. G. Roberts et al., 2015](#)), which is predicted by the model. The faster the speech stream of the current turn is coming in, the slower the different types of projections would be to influence the level of certainty that a transition relevance place is coming up. Consequently, the certainty threshold would be reached and articulation would be initiated later when the incoming turn is produced faster. Moreover, the sequential position of a next turn in relation to its prior turn has been shown to influence turn timing, with next turns initiating a new sequence starting after longer gaps than next turns that are responding actions in an already running sequence ([S. G. Roberts et al., 2015](#)). If the next turn is responding to a prior turn which itself was responding to the previous turn, transition times are again shorter, meaning that tran-

sitions become faster as sequences become more projectable. The turn taking model presented here predicts these effects, as richer discourse and sequence representations are more informative for the processes of comprehension, action ascription, and response preparation, leading to more accurate projections on the timing of the incoming turn and faster response planning, which in turn result in shorter turn transition times. The processing model of turn taking presented here also accounts for the findings presented in Chapters 2 to 5, which reveal that speech planning in conversation is done as early and as far as possible during the incoming turn in order to be most flexible at the point of turn transition, even though planning in overlap is more demanding for the cognitive system than planning after the incoming turn. The need to meet the timing demands of the turn taking system force the language processing system to not only accept but virtually seek peaks in processing load at turn transitions. In that way, the setup of the language processing system is well adapted to the timing demands of turn taking, and its design is shaped by both cognitive and social demands.





## REFERENCES

- Alario, F., Segui, J., & Ferrand, L. (2000). Semantic and Associative Priming in Picture Naming. *The Quarterly Journal of Experimental Psychology Section A*, 53(3), 741–764. doi: 10.1080/713755907
- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38, 419–439.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4), 502–518. doi: 10.1016/j.jml.2006.12.004
- Auer, P. (2005). Projection in Interaction and Projection in Grammar. *Text - Interdisciplinary Journal for the Study of Discourse*, 25(1). doi: 10.1515/text.2005.25.1.7
- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, 42(1), 1–22. doi: 10.1006/jmla.1999.2667
- Bard, E. G., & Aylett, M. P. (2000). Accessibility, Duration, and Modeling the Listener in Spoken Dialogue. In *Proceedings of Gotalog 2000, 4th Workshop on the Semantics and Pragmatics of Dialogue* (p. 8). Gotalog, Sweden.
- Barr, D. J. (2008). Analyzing visual world eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. doi: 10.1016/j.jml.2007.09.002
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4(328), 1–2. doi: 10.3389/fpsyg.2013.00328
- 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. doi: 10.1016/j.jml.2012.11.001
- Barthel, M. (2012). *Aspects of the Timing of Turn-Taking* [unpublished MA Thesis]. Leipzig University.
- Barthel, M., & Levinson, S. C. (in press). Phonological planning is done in overlap with the incoming turn: Evidence from gaze-contingent switch task performance. *Language, Cognition and Neuroscience*.
- Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next Speakers Plan Their Turn Early and Speak after Turn-Final “Go-Signals”. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.00393
- Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialogue is associated with increased processing load. *Cognitive Science*, 43(7), e12768. doi: 10.1111/cogs.12768
- Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm. *Frontiers in Psychology*, 7(1858). doi: 10.3389/fpsyg.2016.01858
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, 67(1). doi: 10.18637/jss.v067.i01
- Beattie, G. W. (1981). The regulation of speaker turns in face-to-face conversation: Some implications for conversation in sound-only communication channels. *Semiotica*, 34(1-2), 55–70. doi: 10.1515/semi.1981.34.1-2.55
- Beattie, G. W., Cutler, A., & Pearson, M. (1982). Why is Mrs Thatcher interrupted so often? *Nature*, 300(5894), 744–747. doi: 10.1038/300744a0
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. doi: 10.1037/0033-2909.91.2.276
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook*

- of psychophysiology* (2nd ed., pp. 142–162). Cambridge: Cambridge University Press.
- Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2014). Conversational Interaction in the Scanner: Mentalizing during Language Processing as Revealed by MEG. *Cerebral Cortex*, *25*(9), 1–16. doi: 10.1093/cercor/bhu116
- Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, *109*, 295–310. doi: 10.1016/j.neuropsychologia.2017.12.028
- Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never Say No ... How the Brain Interprets the Pregnant Pause in Conversation. *PLOS ONE*, *10*(12), e0145474. doi: 10.1371/journal.pone.0145474
- Bögels, S., Kendrick, K. H., & Levinson, S. C. (2019). Conversational expectations get revised as response latencies unfold. *Language, Cognition and Neuroscience*, 1–14. doi: 10.1080/23273798.2019.1590609
- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*(12881), 1–11. doi: 10.1038/srep12881
- Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. doi: 10.1016/j.wocn.2015.04.004
- Bock, K. (1995). Sentence Production: From Mind to Mouth. In *Speech, Language, and Communication* (pp. 181–216). Elsevier. doi: 10.1016/B978-012497770-9/50008-X
- Bock, K., Levelt, W., & Gernsbacher, M. A. (1994). Language production: Grammatical encoding. In *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.
- Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer [Computer program]*. Version 5.3.56, retrieved from [www.praat.org](http://www.praat.org).
- Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, *143*(1), 295–311. doi: 10.1037/a0031858
- Borchers, H. W. (2015). *pracma: Practical Numerical Math Functions [R package]*. Version 2.1.4.



- Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436. doi: 10.1016/j.jecp.2012.01.005
- Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-adapted Interaction, 1*(1), 67–86. doi: 10.1007/BF00158952
- Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1). doi: 10.18637/jss.v080.i01
- Bruner, J. S. (1974). From communication to language—a psychological perspective. *Cognition, 3*(3), 255–287. doi: 10.1016/0010-0277(74)90012-2
- Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language, 2*(01). doi: 10.1017/S0305000900000866
- Bruner, J. S., & Watson, R. (1983). *Child's talk: learning to use language* (1st ed ed.). New York: W.W. Norton.
- Casillas, M. (2014). Taking the floor on time: Delay and deferral in children's turn taking. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in Interaction: Studies in honor of Eve V. Clark* (Vol. 12, pp. 101–114). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/tilar.12.09cas
- Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(06), 1310–1337. doi: 10.1017/S0305000915000689
- Casillas, M., & Frank, M. C. (2012). Cues to turn boundary prediction in adults and preschoolers. In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *Proceedings of SemDial 2012 (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 61–69). Paris: Université Paris-Diderot.
- Casillas, M., & Frank, M. C. (2013). The development of predictive processes in children's discourse understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, 299–304.
- Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language, 92*, 234–253. doi: 10.1016/j.jml.2016.06.013

- Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, 31(2), 251–276. doi: 10.1016/S0095-4470(03)00007-X
- 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.
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111. doi: 10.1016/S0010-0277(02)00017-3
- Clayman, S. (2002). Sequence and solidarity. In *Group Cohesion, Trust and Solidarity* (pp. 229–53.). Oxford: Elsevier Science Ltd.
- Collins, R. (2014). *Interaction ritual chains*. Princeton: Princeton University Press.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, 175, 77–95. doi: 10.1016/j.cognition.2018.01.015
- Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, 20(1), 55–60.
- Cutler, A. (2012). *Native listening: language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.
- Cutler, A., & Pearson, M. (1985). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (Ed.), *Intonation in Discourse* (pp. 139–155). London: Croom Helm.
- Damian, M., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, 57(2), 195–209. doi: 10.1016/j.jml.2006.11.001
- Davidson, J. (1984). Subsequent versions of invitations, offers, requests,

- and proposals dealing with potential or actual rejection. In J. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 102–128). Cambridge: Cambridge University Press.
- De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, 40(1), 198–205. doi: 10.3758/BRM.40.1.198
- DeLong, K. A. (2009). *Electrophysiological explorations of linguistic pre-activation and its consequences during online sentence processing* (Unpublished doctoral dissertation). UC San Diego, San Diego, CA.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. doi: 10.1038/nn1504
- Department of Computer Science, Leipzig University. (2016). *Wortschatz Projekt Leipzig University*. Leipzig.
- de Ruiter, J., Mitterer, H., & Enfield, N. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515–535.
- Donovan, J. J., & Radosevich, D. J. (1999). A Meta-Analytic Review of the Distribution of Practice Effect: Now You See It, Now You Don't. *Journal of Applied Psychology*, 84(5), 795–805. doi: 10.1037/0021-9010.84.5.795
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2), 283–292. doi: 10.1037/h0033031
- Duncan, S. (1974). On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, 3(2), 161–180. doi: 10.1017/S0047404500004322
- Duncan, S., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. Hillsdale: Lawrence Erlbaum.
- Duncan, S., & Niederehe, G. (1974). On signalling that it's your turn to speak. *Journal of Experimental Social Psychology*, 10(3), 234–247. doi: 10.1016/0022-1031(74)90070-5
- Durkin, K. (1987). Minds and Language: Social Cognition, Social Interaction and the Acquisition of Language. *Mind & Language*, 2(2), 105–140. doi: 10.1111/j.1468-0017.1987.tb00111.x

- Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. Hove, U.K. ; Hillsdale (USA): L. Erlbaum Associates, Publishers.
- Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*(4), 639–645. doi: 10.1080/17470210903469864
- Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. *Acta Psychologica*, *191*, 131–148. doi: 10.1016/j.actpsy.2018.09.006
- Fargier, R., & Laganaro, M. (2016). Neurophysiological Modulations of Non-Verbal and Verbal Dual-Tasks Interference during Word Planning. *PLOS ONE*, *11*(12), e0168358. doi: 10.1371/journal.pone.0168358
- Fernández, E. M., & Cairns, H. S. (Eds.). (2017). *The Handbook of Psycholinguistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi: 10.1002/9781118829516
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233. doi: 10.1016/0749-596X(91)90004-4
- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. doi: 10.1111/1467-8721.00158
- Ferreira, F., & Swets, B. (2002). How Incremental Is Language Production? Evidence from the Production of Utterances Requiring the Computation of Arithmetic Sums. *Journal of Memory and Language*, *46*(1), 57–84. doi: 10.1006/jmla.2001.2797
- Ford, C. E., Fox, B. A., & Thompson, S. A. (1996). Practices in the construction of Turns: The "TCU" revisited. *Pragmatics*, *6*(3), 427–454.
- Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 134–184). Cambridge: Cambridge University Press.
- Ford, C. E., & Thompson, S. A. (2003). Social Interaction and Grammar. In M. Tomasello (Ed.), *The New Psychology of Language* (Vol. 2). Mahwah: Lawrence Erlbaum.
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd

- ed ed.). Thousand Oaks, CA: SAGE Publications.
- Féry, C. (1993). *German intonational patterns*. Tübingen: Niemeyer.
- Fry, D. (1975). Simple Reaction-Times to Speech and Non-Speech Stimuli. *Cortex*, 11(4), 355–360. doi: 10.1016/S0010-9452(75)80027-X
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377. doi: 10.3758/BF03193857
- Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00751
- Gilles, P. (2005). *Regionale Prosodie im Deutschen. Variabilität in der Intonation von Abschluss und Weiterweisung*. Berlin: de Gruyter.
- Gisladottir, R. S., Bögels, S., & Levinson, S. C. (2018). Oscillatory Brain Responses Reflect Anticipation during Comprehension of Speech Acts in Spoken Dialog. *Frontiers in Human Neuroscience*, 12. doi: 10.3389/fnhum.2018.00034
- 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), 1–24. doi: 10.1371/journal.pone.0120068
- 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.
- Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4), 544–569. doi: 10.1016/j.jml.2007.01.007
- Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3), 601–634. doi: 10.1016/j.csl.2010.10.003
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82, B1–B14.
- Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, 10(3), 603–609. doi: 10.3758/BF03196521

- Griffin, Z. M., & Bock, K. (2000). What the Eyes Say About Speaking. *Psychological Science*, *11*(4), 274–279. doi: 10.1111/1467-9280.00255
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *Neurocognition of Language* (pp. 273–361). Oxford: Oxford University Press.
- Hagoort, P., & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, *37*(1), 347–362. doi: 10.1146/annurev-neuro-071013-013847
- Halekoh, U., & Hojsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbrtest. *Journal of Statistical Software*, *59*(9), 1–30.
- Hampton Wray, A., & Weber-Fox, C. (2013). Specific aspects of cognitive and language proficiency account for variability in neural indices of semantic and syntactic processing in children. *Developmental Cognitive Neuroscience*, *5*, 149–171. doi: 10.1016/j.dcn.2013.03.002
- Harley, T. A. (2014). *The psychology of language: from data to theory* (Fourth edition ed.). Hove, East Sussex: Psychology Press, Taylor & Francis Group.
- Hübner, R., & Lehle, C. (2007). Strategies of flanker coprocessing in single and dual tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 103–123. doi: 10.1037/0096-1523.33.1.103
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. doi: 10.1016/j.wocn.2010.08.002
- Heritage, J. (1984). Conversation analysis. In *Garfinkel and ethnomethodology* (p. 233-292). Cambridge: Polity Press.
- Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190–1192. doi: 10.1126/science.143.3611.1190
- Hick, W. E. (1952). On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. doi: 10.1080/17470215208416600
- Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of

- mother–infant interaction. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01492
- Hjalmarsson, A. (2011). The additive effect of turn-taking cues in human and synthetic voice. *Speech Communication*, 53(1), 23–35. doi: 10.1016/j.specom.2010.08.003
- Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, 31(1), 19–31. doi: 10.1080/23273798.2015.1072223
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, 2. doi: 10.3389/fpsyg.2011.00255
- Indefrey, P., & Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1-2), 101–144. doi: 10.1016/j.cognition.2002.06.001
- Izdebski, K., & Shipp, T. (1978). Minimal Reaction Times for Phonatory Initiation. *Journal of Speech Language and Hearing Research*, 21(4), 638. doi: 10.1044/jshr.2104.638
- Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* (Unpublished doctoral dissertation). Stanford University, Stanford.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. doi: 10.1016/j.cogpsych.2010.02.002
- Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford: Oxford University Press.
- Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: evidence from event-related brain potentials. *Cognitive Brain Research*, 15(3), 261–276. doi: 10.1016/S0926-6410(02)00198-2
- Jescheniak, J. D., & Levelt, W. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824–843. doi: 10.1037/0278-7393.20.4.824
- Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2003). Utterance

- format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 441–454. doi: 10.1037/0096-1523.29.2.441
- Jongman, S. R., & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming? *Acta Psychologica*, 181, 40–50. doi: 10.1016/j.actpsy.2017.10.003
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4), 329–354.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47(2), 310–339. doi: 10.1037/h0078820
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall.
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124(1), 66–71. doi: 10.1016/j.cognition.2012.03.001
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013–1040. doi: 10.1080/17470218.2015.1053951
- Kass, R. E., & Raftery, A. E. (1995, June). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795. Retrieved 2019-12-09, from <http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572> doi: 10.1080/01621459.1995.10476572
- Kempen, G., & Hoenkamp, E. (1987). An Incremental Procedural Grammar for Sentence Formulation. *Cognitive Science*, 11(2), 201–258. doi: 10.1207/s15516709cog1102-5
- Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, 27(3), 345–380. doi: 10.1080/01690965.2010.544583
- Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, 18(2), 181–192. doi: 10.1037/0882-7974.18.2.181



- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63. doi: 10.1016/0001-6918(67)90005-4
- Kendrick, K. H. (2015). The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00250
- Kendrick, K. H., & Torreira, F. (2014). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*, 52(4), 1–35. doi: 10.1080/0163853X.2014.955997
- Kenward, M. G., & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, 53(3), 983–997. doi: 10.2307/2533558
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. doi: 10.3758/BRM.42.3.643
- Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: effects of concurrent verbal and visuospatial working memory load. *Quarterly Journal of Experimental Psychology*, 70(4), 811–831. doi: 10.1080/17470218.2016.1167926
- Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, 139(4), 1618–1636. doi: 10.1121/1.4944032
- Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1), 143–162. doi: 10.1016/j.jml.2011.08.003
- Korvorst, M., Roelofs, A., & Levelt, W. (2006). Incrementality in naming and reading complex numerals: evidence from eyetracking. *Quarterly Journal of Experimental Psychology*, 59(2), 296–311. doi: 10.1080/17470210500151691
- Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*, 20(1), 43–63. doi: 10.1002/acp.1164
- Kuchinke, L., Vo, M., Hofmann, M., & Jacobs, A. (2007). Pupillary responses during lexical decisions vary with word frequency but

- not emotional valence. *International Journal of Psychophysiology*, 65(2), 132–140. doi: 10.1016/j.ijpsycho.2007.04.004
- Kuhlen, A. K., Allefeld, C., Anders, S., & Haynes, J.-D. (2015). Towards a multi-brain perspective on communication in dialogue. In R. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 182–200). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107323667.009
- Kuhlen, A. K., Allefeld, C., & Haynes, J.-D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, 6. doi: 10.3389/fnhum.2012.00266
- Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating Distracted Addressees: How Speakers' Expectations and Addressees' Feedback Influence Storytelling. *Discourse Processes*, 47(7), 567–587. doi: 10.1080/01638530903441339
- 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. doi: 10.3758/s13423-012-0341-8
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. doi: 10.1080/23273798.2015.1102299
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, 7(1), 18–27. doi: 10.1177/1745691611427305
- La Heij, W., Dirx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, 81(4), 511–525. doi: 10.1111/j.2044-8295.1990.tb02376.x
- Lammertink, I., Casillas, M., Benders, T., Post, B., & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00495
- Lehle, C., & Hübner, R. (2009). Strategic capacity sharing between two tasks: evidence from tasks with the same and with different task sets. *Psychological Research Psychologische Forschung*, 73(5), 707–726. doi: 10.1007/s00426-008-0162-6
- Lehle, C., Steinhauser, M., & Hübner, R. (2009). Serial or parallel processing in dual tasks: What is more effortful? *Psychophysiology*, 46(3), 502–509. doi: 10.1111/j.1469-8986.2009.00806.x

- Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means [r package]*.
- Levelt, W. (1989). *Speaking: From Intention to Articulation*. London: MIT Press.
- Levelt, W. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1-3), 1–22. doi: 10.1016/0010-0277(92)90038-J
- Levelt, W. (2012). *A history of psycholinguistics: the pre-Chomskyan era*. Oxford: OUP Oxford. (OCLC: 818851561)
- Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(01), 1–75. doi: 10.1017/S0140525X99001776
- Levelt, W., Schriefers, H., Vorberg, D., Meyer, A. S., & Pechmann, T. (1991). The Time Course of Lexical Access in Speech Production: A Study of Picture Naming. *Psychological Review*, 98(1), 122–142.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (2006). On the Human 'Interaction Engine'. In N. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality - Culture, Cognition and Interaction* (pp. 39–69). Oxford: Berg.
- Levinson, S. C. (2012). Action Formation and Ascription. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 101–130). Chichester, UK: John Wiley & Sons, Ltd.
- Levinson, S. C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, 20(1), 6–14. doi: 10.1016/j.tics.2015.10.010
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6(731), 10–26. doi: 10.3389/fpsyg.2015.00731
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461. doi: 10.1037/h0020279
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. doi: 10.1016/0010-0277(85)90021-6
- Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association*, 42(03), 255–280. doi: 10.1017/S0025100312000187

- Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition ed.). Cambridge, Massachusetts: The MIT Press.
- Magyari, L., Bastiaansen, M. C. 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. doi: 10.1162/jocn-a-00673
- Magyari, L., & de Ruiter, J. P. (2012). Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology*, *3*(376), 1–9. doi: 10.3389/fpsyg.2012.00376
- Magyari, L., de Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. (in press). *Frontiers in Psychology*.
- Mcallister, J., Potts, A., Mason, K., & Marchant, G. (1994). Word Duration in Monologue and Dialogue Speech. *Language and Speech*, *37*(4), 393–405. doi: 10.1177/002383099403700404
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are Women Really More Talkative Than Men? *Science*, *317*(5834), 82–82. doi: 10.1126/science.1139940
- Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared Language: Overlap and Segregation of the Neuronal Infrastructure for Speaking and Listening Revealed by Functional MRI. *Psychological Science*, *22*(9), 1173–1182. doi: 10.1177/0956797611418347
- Miller, J., Ulrich, R., & Rolke, B. (2009). On the optimality of serial and parallel processing in the psychological refractory period paradigm: Effects of the distribution of stimulus onset asynchronies. *Cognitive Psychology*, *58*(3), 273–310. doi: 10.1016/j.cogpsych.2006.08.003
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton: CRC Press/Taylor & Francis Group.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. doi: 10.1016/j.jml.2007.11.006
- Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1601–1619. doi: 10.1080/

- 17470218.2012.754910
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214–255. doi: 10.1037/0033-295X.86.3.214
- Navon, D., & Miller, J. (2002). Queuing or Sharing? A Critical Evaluation of the Single-Bottleneck Notion. *Cognitive Psychology*, 44(3), 193–251. doi: 10.1006/cogp.2001.0767
- Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, 33(7), 1004–1023. doi: 10.1016/j.neubiorev.2009.04.001
- Nieuwland, M. S. (2019). Do ‘early’ brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews*, 96, 367–400. doi: 10.1016/j.neubiorev.2018.11.019
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaeert, K., Darley, E., Kazanina, N., ... Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. doi: 10.7554/eLife.33468
- Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, 53(2), 146–193. doi: 10.1016/j.cogpsych.2006.03.001
- Papesh, M. H., & Goldinger, S. D. (2012). Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics*, 74(4), 754–765. doi: 10.3758/s13414-011-0263-y
- Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, 66(3), 180–194. doi: 10.1007/s00426-002-0086-5
- Piai, V., Roelofs, A., Rommers, J., Dahlsätt, K., & Maris, E. (2015). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience*, 9. doi: 10.3389/fnhum.2015.00549
- Pickering, M. J., & Ferreira, V. S. (2008). Structural Priming: A Critical Review. *Psychological Bulletin*, 134(3), 427–459. doi: 10.1037/0033-2909.134.3.427
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and brain sciences*, 27(2), 169–190. doi:

- <https://doi.org/10.1017/S0140525X04000056>
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110. doi: 10.1016/j.tics.2006.12.002
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347. doi: 10.1017/S0140525X12001495
- Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 17(2), 291–321. doi: 10.1080/01688639508405124
- Pomeranz, A., Atkinson, J., & Heritage, J. (1984). Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In *Structures of Social Action* (pp. 53–101). Cambridge: Cambridge University Press.
- Pomeranz, A., & Heritage, J. (2012). Preference. In T. Stivers & J. Sidnell (Eds.), *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell.
- Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77(2), 97–132. doi: 10.1016/S0010-0277(00)00090-1
- Pylkkänen, L., Gonnerman, L., Stringfellow, A., & Marantz, A. (submitted). Disambiguating the source of phonological inhibition effects in lexical decision: an MEG study. , 27.
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rajah, M. N., Languay, R., & Grady, C. L. (2011). Age-Related Changes in Right Middle Frontal Gyrus Volume Correlate with Altered Episodic Retrieval Activity. *Journal of Neuroscience*, 31(49), 17941–17954. doi: 10.1523/JNEUROSCI.1690-11.2011
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., ... Acker, J. D. (2005). Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cerebral Cortex*, 15(11), 1676–1689. doi: 10.1093/cercor/bhi044
- Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The Neural Bases

- of Strategy and Skill in Sentence–Picture Verification. *Cognitive Psychology*, 40(4), 261–295. doi: 10.1006/cogp.2000.0733
- Reissland, J., & Manzey, D. (2016). Serial or overlapping processing in multitasking as individual preference: Effects of stimulus preview on task switching and concurrent dual-task performance. *Acta Psychologica*, 168, 27–40. doi: 10.1016/j.actpsy.2016.04.010
- Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: mechanisms and information sources. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00089
- Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, 133(6), EL471–EL477. doi: 10.1121/1.4802900
- Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of “trouble” in conversation. *Speech Communication*, 48(9), 1079–1093. doi: 10.1016/j.specom.2006.02.001
- Roberts, F., Margutti, P., & Takano, S. (2011). Judgments Concerning the Valence of Inter-Turn Silence Across Speakers of American English, Italian, and Japanese. *Discourse Processes*, 48(5), 331–354. doi: 10.1080/0163853X.2011.558002
- Roberts, S. G., & Levinson, S. C. (2017). Conversation, cognition and cultural evolution: A model of the cultural evolution of word order through pressures imposed from turn taking in conversation. *Interaction Studies*, 18(3), 402–442. doi: 10.1075/is.18.3.06rob
- Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00509
- Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: a review. *Frontiers in Psychology*, 2. doi: 10.3389/fpsyg.2011.00307
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382–386. doi: 10.1016/S1364-6613(02)01958-7
- Sassenhagen, J., & Alday, P. M. (2016). A common misapplication

- of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, 162, 42–45. doi: 10.1016/j.bandl.2016.08.001
- Sauppe, S. (2017). Symmetrical and asymmetrical voice systems and processing load: Pupillometric evidence from sentence production in Tagalog and German. *Language*, 93(2), 288–313. doi: 10.1353/lan.2017.0015
- Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics*, 11(3), 243–257.
- Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, 97(5), 1295–1345. doi: 10.1086/229903
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1), 1–63.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, 53(2), 361–382. doi: 10.2307/413107
- Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.00137
- Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the Phonological Level during Sentence Production. *Journal of Psycholinguistic Research*, 35(2), 189–213. doi: 10.1007/s10936-005-9011-6
- Schriefers, H., Meyer, A. S., & Levelt, W. (1990). Exploring the Time Course of Lexical Access in Language Production: Picture-Word Interference Studies. *Journal of Memory and Language*, 29, 86–102. doi: 10.1016/0749-596X(90)90011-N
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012a). Shared Syntax in Language Production and Language Comprehension—An fMRI Study. *Cerebral Cortex*, 22(7), 1662–1670. doi: 10.1093/cercor/bhr249
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012b). Shared syntax in language production and language comprehension — an fmri study. *Cerebral Cortex*, 22(7), 1662–1670. doi: 10.1093/cercor/bhr249



- Selting, M. (2007). Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics*, 39(3), 483–526. doi: 10.1016/j.pragma.2006.07.008
- Shipp, T., Izdebski, K., & Morrissey, P. (1984). Physiologic Stages of Vocal Reaction Time. *Journal of Speech Language and Hearing Research*, 27(2), 173. doi: 10.1044/jshr.2702.173
- Shitova, N., Roelofs, A., Coughler, C., & Schriefers, H. (2017). P3 event-related brain potential reflects allocation and use of central processing capacity in language production. *Neuropsychologia*, 106, 138–145. doi: 10.1016/j.neuropsychologia.2017.09.024
- Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, 111(43), E4687–E4696. doi: 10.1073/pnas.1323812111
- Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(6), 679–692. doi: 10.1002/wcs.1323
- Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, 136, 304–324. doi: 10.1016/j.cognition.2014.10.008
- Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, 32, 25–38.
- Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425–436. doi: 10.1037/0278-7393.32.2.425
- Stephens, G. J., & Beattie, G. (1986). On Judging the Ends of Speaker Turns in Conversation. *Journal of Language and Social Psychology*, 5(2), 119–134. doi: 10.1177/0261927X8652003
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430. doi: 10.1073/pnas.1008662107
- Stevens, K. N. (1960). Toward a Model for Speech Recognition. *The Journal of the Acoustical Society of America*, 32(1), 47–55. doi: 10.1121/1.1907874
- Stivers, T. (2012). Sequence Organization. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 191–209). Chichester:

- Wiley-Blackwell.
- 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), 10587–10592. doi: 10.1073/pnas.0903616106
- Strijkers, K., & Costa, A. (2011). Riding the Lexical Speedway: A Critical Review on the Time Course of Lexical Selection in Speech Production. *Frontiers in Psychology*, 2(356), 1–16. doi: 10.3389/fpsyg.2011.00356
- Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of Conversational Pressures on Speech Planning. *Discourse Processes*, 50(1), 23–51. doi: 10.1080/0163853X.2012.727719
- Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: evidence from eye-movements. *Language and Cognition*, 6(01), 12–44. doi: 10.1017/langcog.2013.5
- Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297–314. doi: 10.1016/j.jml.2012.12.002
- Tanaka, H. (2015). Action-projection in Japanese conversation: topic particles wa, mo, and tte for triggering categorization activities. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01113
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. doi: 10.1126/science.7777863
- Tanenhaus, M. K., Carlson, G., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, 4(3-4), SI211–SI234. doi: 10.1080/01690968908406368
- Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye Movements and Lexical Access in Spoken-Language Comprehension: Evaluating a Linking Hypothesis between Fixations and Linguistic Processing. *Journal of Psycholinguistic Research*, 29(6), 557–580.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, Mass.: Harvard Univ. Press. (OCLC: 247736399)

- Tombu, M., & Jolicoeur, P. (2005). Testing the Predictions of the Central Capacity Sharing Model. *Journal of Experimental Psychology: Human Perception and Performance*, 31(4), 790–802. doi: 10.1037/0096-1523.31.4.790
- Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00284
- Tromp, J., Hagoort, P., & Meyer, A. S. (2016). Pupillometry reveals increased pupil size during indirect request comprehension. *Quarterly Journal of Experimental Psychology*, 69(6), 1093–1108. doi: 10.1080/17470218.2015.1065282
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467. doi: 10.1037/0278-7393.31.3.443
- Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the Psychological Refractory Period effect? *Journal of Experimental Psychology: Human Perception and Performance*, 25(5), 1268–1283. doi: 10.1037/0096-1523.25.5.1268
- von Essen, O. (1956). *Grundzüge der Hochdeutschen Satzintonation* (1st ed.). Düsseldorf: A. Henn.
- Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 423–440. doi: 10.1037/a0018619
- Walker, M. B., & Trimboli, C. (1984). The Role of Nonverbal Signals in Co-Ordinating Speaking Turns. *Journal of Language and Social Psychology*, 3(4), 257–272. doi: 10.1177/0261927X8400300402
- Ward, N. (2019). *Prosodic patterns in English conversation*. Cambridge ; New York: Cambridge University Press.
- Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language*, 85(2), 231–244. doi: 10.1016/S0093-934X(02)00587-4
- Wells, B., & Macfarlane, S. (1998). Prosody as an Interactional Resource: Turn-projection and Overlap. *Language and Speech*, 41(3-4), 265–294. doi: 10.1177/002383099804100403

- Wesseling, W., & Son, R. J. J. H. v. (2005). Timing of Experimentally Elicited Minimal Responses as Quantitative Evidence for the Use of Intonation in Projecting TRPs. In *Proceedings of Interspeech 2005* (Vol. 6, pp. 3389 – 3392). Lisbon.
- Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346(3), 165–168. doi: 10.1016/S0304-3940(03)00599-8
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288. doi: 10.1162/0898929041920487
- Wilshire, C., Singh, S., & Tattersall, C. (2016). Serial order in word form retrieval: New insights from the auditory picture–word interference task. *Psychonomic Bulletin & Review*, 23(1), 299–305. doi: 10.3758/s13423-015-0882-8
- Yngve, V. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting Chicago Linguistic Society* (pp. 567–577). Chicago.
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility. *Ear and Hearing*, 31(4), 480–490. doi: 10.1097/AUD.0b013e3181d4f251
- Zipf, G. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley Press.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin*, 123(2), 162–185.



---

## CONCISE SUMMARY

---

When humans have a conversation with one-another, they generally take turns speaking one after the other without overlapping each others talk or leaving silence between turns for long stretches of time. Previous research has shown that conversation is a structured practice following rules that help interlocutors to manage the flow of conversation interactively. While at the beginning of a conversation it remains open who will speak when about what and for how long, interlocutors regulate the flow of conversation as it unfolds. One basic set of rules that interlocutors operate with governs the allocation of speaking turns, with the central rule stating that whoever starts speaking first at a point in time when speaker change becomes relevant has the rights and obligations to produce the next turn. The organization of turn allocation, therefore, is one reason for conversational turn taking to be so remarkably fast, with the beginnings of turns most often being quite accurately aligned with the ends of the previous turns. Observations of this outstanding speed of turn taking gave rise to a number of questions concerning language processing in conversational situations. The studies presented in this thesis investigate some of these questions from the perspective of the current listener preparing to be the next speaker who will respond to the current turn.

The study presented in Chapter 2 investigates when next speakers begin to plan their own turn with respect to two points in time, (i) the moment when the incoming turn's message becomes clear enough to make response planning possible and (ii) the moment when the incoming turn terminates. Results of previous studies were inconclusive about the timing of language planning in conversation, with evidence in favour of both late and early response planning. Furthermore, previous studies presented both evidence as well as counter evidence indicating that response planning depends or does not depend on an accurate prediction of the timing of the incoming turn's end. The study presented here makes use of a novel experimental paradigm which includes a dialogic task that participants need to fulfil in response

to critical utterances by a confederate. These critical utterances were structured, on the one hand, so that their message became clear either only at the end of the turn or before the end of the turn, and, on the other hand, so that it was either predictable or not predictable when exactly the turn would end. Participant's eye-movements as well as their response latencies indicated that they always planned their next turn as early as possible, irrespective of the predictability of the incoming turn's end. The presented results provide evidence in favour of models of turn taking that predict speech planning to happen in overlap with the incoming turn.

Having established that next speakers begin to plan their turn in overlap, the study presented in Chapter 3 goes more into detail investigating to which depth language planning progresses while the incoming turn is still unfolding. To this end, a number of psycholinguistic paradigms were combined. In the study's main experiment, participants had to fulfil a switch-task in which they switched from picture naming in response to an auditorily presented question to making a lexical decision. By manipulating the relatedness of the word for lexical decision with the picture that was prepared to be named before the task-switch it was possible to draw inferences on which processing stages were entered during the speech production process in overlap with the incoming turn. Participants' behavioural responses in the lexical decision task revealed that they entered the stage of phonological encoding while the incoming turn was still unfolding, showing that planning in overlap is not limited to conceptual preparation but includes all sub-processes of formulation.

Given that speech production regularly enters the stages of formulation in overlap with the incoming turn, as shown in Chapters 2 and 3, the question arises whether planning the next turn in overlap is cognitively more demanding than during the gap between turns. This question is approached in the study presented in Chapter 4 by measuring pupillometric responses of participants in a dialogic task. An increase in pupil diameter during a cognitive task is indicative of increased processing load, and pupillometric responses to planning in overlap with the incoming turn were found to be greater than responses to planning in the gap between turns. These results show that planning in overlap is more demanding than planning during the gap, even though it is highly practiced by speakers.

After Chapters 2 to 4 investigated the timing and mechanisms of speech planning in conversation, Chapter 5 turns towards the timing of articulation of a planned turn, asking the question what sources of information next speakers use to time the articulation of a planned utterance to start closely after the incoming turn comes to an end. In this Chapter's study, participants taking turns with a confederate responded to utterances containing or not containing different cues to the location of the incoming turn's end. Participants made use of lexical and turn-final intonational cues, but not of turn-initial intonational cues, responding faster when the relevant cues were present than when they were not present. These results show that the timing of turn initiation in next speakers depends on the recognition of the incoming turn's point of completion and not merely on the progress in planning the next turn.

All evidence presented in Chapters 2 to 5 is summed up and bundled together in a cognitive model of turn taking, which is being presented in Chapter 6. This model assumes, centrally, that the planning of a turn and the timing of its articulation are separate cognitive processes that run in parallel in any next speaker during conversation. Planning generally starts as early as possible, often in overlap with the incoming turn, while the timing of articulation depends on the next speaker's level of certainty that speaker change has become relevant at a particular moment, with a number of cues to the end of the incoming turn leading to an increase of certainty. Next turns are assumed to often be planned down to fully formulated utterance plans including their phonological form as early as possible on the basis of anticipations of the incoming turn's message, which are created with the help of the general and situational knowledge about the world, the current speaker and her intentions, as well as the input that has been received so far. The level of certainty that speaker change becomes relevant rises or decreases as lexico-syntactic, prosodic, and pragmatic projections about the development of the current turn are fulfilled or not fulfilled. As the incoming turn progresses towards its end as was projected by the current listener, he becomes certain that speaker change becomes relevant and will initiate articulation of the prepared next turn. Viewing these two processes, planning a next turn and timing of its articulation, as separate makes it possible to explain the observable fast timing of turn taking while still modelling the allocation of turns as interactionally managed by interlocutors — a considerable advantage of the presented



model compared to more traditional perspectives on turn taking and conversation.





---

## SAMENVATTING

---

Wanneer mensen een gesprek met elkaar hebben, spreken ze meestal om beurten, de een na de ander, zonder elkaar te overlappen of lang stil te staan tussen de beurten. Eerder onderzoek heeft aangetoond dat conversatie gestructureerd is volgens regels die gesprekspartners helpen interactief de gespreksstroom te beheren. Terwijl het aan het begin van een gesprek onduidelijk is, wie wanneer en hoelang zal spreken, bepalen de gesprekspartners de gespreksstroom gedurende het gesprek. Een basisset van regels waarmee gesprekspartners werken bepaalt de toewijzing van spreekbeurten, met de centrale regel dat degene die begint te spreken op het moment dat sprekerverandering relevant wordt, de rechten en plichten heeft om de volgende beurt te produceren. De organisatie van beurttoewijzing is daarom een van de redenen waarom spreken om de beurt zo opmerkelijk snel gaat, waarbij het begin van de beurt meestal vrij nauwkeurig is uitgelijnd met het eind van de vorige beurt. Observaties van deze buitengewone snelheid van beurten gaven aanleiding tot een aantal vragen met betrekking tot taalverwerking in conversationele situaties. De gepresenteerde studies in dit proefschrift onderzoeken enkele van deze vragen vanuit het perspectief van de huidige luisteraar die zich erop voorbereidt de volgende spreker te worden die zal reageren op de huidige beurt.

De studie gepresenteerd in hoofdstuk 2 onderzoekt wanneer volgende sprekers beginnen met het plannen van hun beurt met betrekking tot twee tijdstippen, (i) het moment waarop de boodschap van de inkomende beurt duidelijk genoeg wordt om responsplanning mogelijk te maken en (ii) het moment waarop de inkomende beurt wordt beëindigd. Resultaten van eerdere studies waren niet doorslaggevend over de timing van taalplanning in conversatie, met bewijs voor zowel late als vroege responsplanning. Bovendien presenteerden eerdere onderzoeken zowel bewijs als tegenbewijs dat aangeeft dat responsplanning al dan niet afhankelijk is van een nauwkeurige voorspelling van de timing van het einde van de inkomende beurt. De hier gepresenteerde studie maakt gebruik van een nieuw experimenteel paradigma

dat een dialogische taak omvat die deelnemers moeten vervullen als reactie op kritische uitingen van een gesprekspartner. Deze kritische uitingen waren enerzijds zo gestructureerd dat hun boodschap pas aan het einde van de beurt of vóór het einde van de beurt duidelijk werd, en anderzijds dat het voorspelbaar of niet voorspelbaar was wanneer precies de beurt zou eindigen. De oogbewegingen van de deelnemers en hun reactietijden gaven aan dat ze hun volgende beurt altijd zo vroeg mogelijk hadden gepland, ongeacht de voorspelbaarheid van het einde van de inkomende beurt. De gepresenteerde resultaten leveren bewijs voor modellen die voorspellen dat spraakplanning tegelijk met de inkomende beurt plaatsvindt.

Nadat is vastgesteld dat volgende sprekers hun beurt tegelijkertijd beginnen te plannen, gaat het onderzoek in hoofdstuk 3 verder in op het onderzoeken van de diepte van taalplanning terwijl zich de inkomende beurt nog steeds ontwikkelt. Daarvoor werden een aantal psycholinguïstische paradigma's gecombineerd. In het hoofdexperiment van de studie moesten de deelnemers een wissel-taak uitvoeren waarin ze van de naamgeving van de foto op een auditief gepresenteerde vraag overgingen naar een lexicale beslissing. Door de verwantschap van het woord voor lexicale beslissing te manipuleren met de afbeelding die was voorbereid om te worden benoemd vóór de taakwisseling, was het mogelijk om conclusies te trekken over de ingevoerde verwerkingsfasen tijdens het spraakproductieproces in overlap met de inkomende beurt. De gedragsreacties van deelnemers in de lexicale beslissingstaak openbaarden dat ze het stadium van fonologische codering binnengingen terwijl de inkomende beurt zich nog steeds ontvouwde, wat aantoonde dat planningsoverlap niet beperkt is tot conceptuele voorbereiding maar alle subprocessen van formulering omvat.

Gegeven dat spraakproductie regelmatig de fasen van formulering binnensluipt die overlappend zijn met de inkomende beurt, zoals weergegeven in hoofdstukken 2 en 3, rijst de vraag of het plannen van de volgende beurt in overlap cognitief veeleisender is dan tijdens de kloof tussen beurten. Deze vraag wordt benaderd in de studie gepresenteerd in hoofdstuk 4 door het meten van pupillometrische reacties van deelnemers in een dialogische taak. Een toename van de pupildiameter tijdens een cognitieve taak is indicatief voor een verhoogde verwerkingsbelasting, en pupillometrische reacties op planning in overlapping met de inkomende beurten bleken groter te zijn

dan antwoorden op planning in de spleet tussen beurten. Deze resultaten tonen aan dat planningsoverlapping veeleisender is dan plannen tijdens de kloof, ook al wordt deze door sprekers zeer goed beoefend. Nadat de hoofdstukken 2 tot en met 4 de timing en mechanismen van spraakplanning in conversatie hebben onderzocht, keert hoofdstuk 5 zich uit naar de timing van de articulatie van een geplande beurt, met de vraag welke informatiebronnen volgende sprekers gebruiken om de articulatie van een geplande uiting vlak na het einde van de inkomende beurt te laten beginnen. In de studie van dit hoofdstuk reageerden deelnemers om de beurt met een bondgenoot op uitingen die al dan niet verschillende aanwijzingen bevatten naar de locatie van het einde van de inkomende beurt. Deelnemers maakten gebruik van lexicale en beurt-finale intonele signalen, maar niet van beurt-initiële intonele signalen, en ze reageerden sneller wanneer de relevante signalen aanwezig waren dan wanneer ze niet aanwezig waren. Deze resultaten laten zien dat de timing van de beurtinitiatie bij volgende sprekers afhankelijk is van de herkenning van het startpunt van de inkomende beurt en niet alleen van de voortgang bij het plannen van de volgende beurt.

Alle bewijsmateriaal gepresenteerd in hoofdstukken 2 tot 5 wordt samengevat en gebundeld in een cognitief model van beurtwisseling, dat wordt gepresenteerd in hoofdstuk 6. Dit model veronderstelt centraal dat de planning van een beurt en de timing van de articulatie afzonderlijke cognitieve processen zijn die parallel lopen bij volgende sprekers tijdens een gesprek. Planning begint over het algemeen zo vroeg mogelijk, vaak in overlap met de inkomende beurt, terwijl de timing van de articulatie afhangt van de mate van zekerheid van de volgende spreker dat sprekerverandering op een bepaald moment relevant is geworden, met een aantal aanwijzingen tot het einde van de inkomende beurt die leiden tot een toename van zekerheid. Verwacht wordt dat volgende beurten zo snel mogelijk worden gepland in volledig geformuleerde uitingsplannen, inclusief hun fonologische vorm, op basis van anticipaties van de boodschap van de inkomende beurt, die zijn gemaakt met behulp van de algemene en situationele kennis over de wereld, de huidige spreker en haar bedoelingen, evenals de input die tot nu toe is ontvangen. De mate van zekerheid dat sprekerverandering relevant wordt, neemt toe of af naarmate lexico-syntactische, prosodische en pragmatische projecties over de ontwikkeling van de huidige beurt worden vervuld of niet worden vervuld. Naarmate de

inkomende beurt vordert naar het einde zoals geprojecteerd door de huidige luisteraar, wordt hij er zeker van dat de verandering van de spreker relevant wordt en de articulatie van de voorbereide volgende beurt begint. Het bekijken van deze twee processen, het plannen van een volgende wending en de timing van de articulatie als afzonderlijk, maakt het mogelijk om de waarneembare snelle timing van het wisselen van de beurt te verklaren, terwijl de toewijzing van beurten nog steeds wordt gemodelleerd als interactief beheerd door gesprekspartners — een aanzienlijk voordeel van het gepresenteerde model vergeleken met meer traditionele perspectieven op beurtwisseling en conversatie.







---

## ZUSAMMENFASSUNG

---

Wenn Menschen sich miteinander unterhalten, folgen ihre Redebeiträge (Turns) im Regelfall sehr schnell aufeinander, ohne größere zeitliche Überlappung oder lange Pausen zwischen den Turns. Die bisherige Forschung zeigt, dass Konversation eine menschliche Praxis ist, die Strukturen und Regeln befolgt, welche es den Teilnehmern einer Unterhaltung ermöglichen, den Fluss der Konversation interaktiv zu organisieren. Während es zu Beginn einer Konversation noch offen steht, wer wann im Laufe der Unterhaltung wie lange sprechen und was gesagt werden wird, regeln die Teilnehmer der Konversation miteinander im Laufe der Konversation wie diese sich entwickelt. Ein dabei grundlegendes Regelsystem das die Teilnehmer benutzen regelt die Zuweisung von Redebeiträgen. Die hierbei zentrale Regel lautet, dass der Teilnehmer, der zu einem Zeitpunkt, an dem ein Sprecherwechsel relevant wird, zuerst beginnt zu sprechen, das Recht und die Pflicht erhält, den nächsten Redebeitrag beizusteuern. Die Aufteilung von Redebeiträgen ist daher ein wichtiger Grund dafür, dass Konversationen so bemerkenswert schnell verlaufen und der Anfang von Redebeiträgen zeitlich meist nahtlos an das Ende des vorhergehenden Redebeitrags anknüpft. Die Beobachtung dieser enormen Geschwindigkeit des Turn Taking wirft eine Reihe von Fragen bezüglich der Sprachverarbeitung in Konversationen auf. Die in dieser Doktorarbeit vorgestellten Studien untersuchen einige dieser Fragen aus der Perspektive eines Zuhörers, der sich darauf vorbereitet, die Rolle des nächsten Sprechers zu übernehmen und auf den momentanen Turn zu antworten.

Die Studie in Kapitel 2 untersucht wann ein nächster Sprecher beginnt den eigenen Turn in Relation zu zwei Zeitpunkten zu planen. Zeitpunkt 1 ist der Moment, in dem der Inhalt des momentan gehörten Turns klar genug wird um mit der Planung einer Antwort zu beginnen; Zeitpunkt 2 ist der Moment, in dem der momentane Turn endet. Die Ergebnisse vorangegangener Studien kamen zu keinem eindeutigen Ergebnis bezüglich der zeitlichen Strukturierung von Sprachplanung in Konversation. So gab es sowohl Evidenz dafür, dass der nächste

Turn eher sehr spät - das heißt, zum Ende des momentan gehörten Turns - oder so früh wie möglich, bereits während des momentanen Turns, geplant wird. Auch waren in den Ergebnissen vorheriger Studien sowohl Evidenz als auch Gegenevidenz enthalten bezüglich der Annahme, dass die Planung des nächsten Turns anhängig ist von einer akkuraten Vorhersage, wann der momentane Turn zum Ende kommen wird. Die in Kapitel 2 vorgestellte Studie benutzt ein neues experimentelles Paradigma, in welchem Versuchsteilnehmer eine Dialogaufgabe erfüllen müssen, in der sie auf Äußerungen eines Gesprächspartners reagieren. Diese Äußerungen waren einerseits so strukturiert, dass ihr Inhalt entweder bereits Mitten im Turn oder erst am Ende des Turns klar wurde, und andererseits so, dass es entweder vorhersagbar oder nicht vorhersagbar war, wann genau sie zum Ende kommen würden. Sowohl die Blickbewegungen als auch die verbalen Reaktionslatenzen der Versuchsteilnehmer weisen darauf hin, dass sie so früh wie möglich beginnen ihren nächsten Turn zu planen, unabhängig davon, ob vorhersagbar ist, wann der Turn enden wird oder nicht. Die präsentierten Ergebnisse stützen daher Turn Taking Modelle, die vorhersagen, dass Sprachplanung in Überlappung mit dem momentan gehörten Turn stattfindet.

Nachdem im 2. Kapitel etabliert wurde, dass nächste Sprecher ihren Turn während des momentanen Turns planen, wird in Kapitel 3 näher untersucht in welcher Tiefe die Sprachplanung voranschreitet, während der momentane Turn noch nicht zuende ist. Zu diesem Zweck wurden mehrere psycholinguistische Paradigmen in einer Studie miteinander verwoben. Im Hauptexperiment der Studie müssen Versuchspersonen einen Aufgabenwechsel bewältigen, in welchem sie von einer Bildbenennungsaufgabe als Antwort auf eine auditiv präsentierte Frage zu einer lexikalen Entscheidungsaufgabe wechseln. Indem die Relation des Wortes in der lexikalen Entscheidungsaufgabe zu dem Bild der Bildbenennungsaufgabe manipuliert wurde, war es möglich Rückschlüsse darauf zu ziehen, welche Verarbeitungsschritte des Sprachproduktionsprozesses in Überlappung mit dem momentan gehörten Turn stattfinden. Die Entscheidungslatenzen in der lexikalen Entscheidungsaufgabe zeigen, dass die Versuchspersonen ihre Antwort auf die Frage, also die Benennung des Bildes, bereits phonologisch kodieren, während der momentane Turn noch nicht zum Ende kommt. Die Ergebnisse zeigen, dass Sprachplanung in Über-

lappung mit dem momentanen Turn nicht auf konzeptuelle Planung reduziert ist, sondern alle Subprozesse der Formulierung des nächsten Turns einschließt.

Wenn Sprachproduktion in Überlappung mit dem momentanen Turn die Teilprozesse der Formulierung inkludiert, wie in den vorangegangenen Kapiteln 2 und 3 gezeigt, stellt sich die Frage, ob das Planen des nächsten Turns in Überlappung mit dem momentanen Turn kognitiv anspruchsvoller ist als das Planen in Stille nach dem Ende des momentanen Turns. Diese Frage wird in Kapitel 4 untersucht, indem Pupillenreaktionen von Versuchspersonen während einer Dialogaufgabe gemessen wurden. Eine Erweiterung der Pupillen während einer kognitiven Aufgabe zeigt erhöhten Verarbeitungsaufwand an, und die Befunde der Studie in diesem Kapitel zeigen, dass Pupillenreaktionen bei Sprachplanung in Überlappung mit dem momentanen Turn größer sind, als Pupillenreaktionen bei Sprachplanung in Stille nach dem momentanen Turn. Diese Resultate zeigen, dass Sprachplanung in Überlappung kognitiv anspruchsvoller ist, als Sprachplanung in Stille zwischen den Turns, obwohl Sprachplanung in Überlappung eine viel geübte, alltägliche Aufgabe für Sprecher darstellt.

Nachdem die Kapitel 2 bis 4 das Timing und die Mechanismen der Sprachplanung in Konversation untersucht haben, widmet sich die Studie in Kapitel 5 der Problematik des Timings der Artikulation eines geplanten Redebeitrags. Hierbei wird die Frage untersucht, welche Informationsquellen nächste Sprecher nutzen, um die Artikulation des nächsten Redebeitrags möglichst nahe dem Ende des momentanen Turns zu beginnen. Zur Untersuchung dieser Frage antworten Versuchspersonen auf Äußerungen eines Dialogpartners, welche entweder bestimmte Hinweise auf das herannahende Turnende enthalten oder nicht. Die Versuchspersonen antworteten schneller, wenn lexikale Hinweise und turnfinale Intonationshinweise im momentanen Turn enthalten sind als wenn diese Hinweise nicht enthalten sind. Diese Ergebnisse zeigen, dass das Timing der Artikulation des nächsten Turns von der Identifizierung des Endpunktes des momentanen Turns und nicht ausschließlich vom Planungsfortschritt des eigenen, nächsten Turns abhängt.

Die Ergebnisse und Erkenntnisse der Studien der Kapitel 2 bis 5 werden zusammengefasst und gebündelt in einem kognitiven Turn Taking Modell, welches in Kapitel 6 vorgestellt wird. Eine zentrale Annahme

dieses Modells ist es, dass das Planen eines Turns und das Timing der Artikulierung separate kognitive Prozesse darstellen, welche bei einem Sprecher während einer Konversation parallel ausgeführt werden. Dabei beginnt die Planung des nächsten Turns generell so früh wie möglich während des momentanen Turns und oft in Überlappung mit diesem, während das Timing der Artikulation abhängig ist vom Grad der Gewissheit, dass ein Sprecherwechsel zu einem bestimmten Zeitpunkt relevant wird. Hierbei dienen eine Reihe von Hinweisen auf ein nahendes Turnende der Erhöhung der Gewissheit, sodass die Artikulation des nächsten Turns in dem Moment initiiert werden kann, wenn ein Sprecherwechsel relevant ist. Das Modell nimmt an, dass der nächste Turn so früh wie möglich bis hin zu einem vollständig formulierten Äußerungsplan mit einer spezifizierten phonologischen Form erstellt wird. Die Relevanz des nächsten Turns basiert auf Antizipationen der im momentan gehörten Turn enthaltenen Nachricht, welche mit Hilfe von generellem und situativem Wissen über die Welt, den Gesprächspartner und seine Intentionen, sowie aus dem bisher gehörten Input geformt werden. Die Gewissheit, dass ein Sprecherwechsel relevant wird erhöht sich oder fällt ab, wenn lexico-syntaktische, prosodische und pragmatische Projektionen des Verlaufs des momentanen Turns erfüllt oder nicht erfüllt werden. Wenn sich der momentane Turn zu seinem vom momentanen Hörer erwarteten Ende hin entwickelt, wird die Gewissheit erreicht, dass ein Sprecherwechsel relevant wird, woraufhin der nächste Sprecher die Artikulation seines vorbereiteten Turns initiiert. Diese zwei Prozesse, Sprachplanung und Artikulationstiming, als separat zu betrachten ermöglicht es einerseits, das beobachtbar schnelle Timing beim Turn Taking zu erklären und andererseits dennoch die Verteilung von Redebeiträgen als interaktional durch die Gesprächspartner organisiert zu beschreiben – ein nennenswerter Fortschritt des vorgestellten Modells gegenüber vorherigen Sichtweisen auf Turn Taking und Konversation.





---

## CURRICULUM VITAE

---

Mathias Barthel did his Abitur in Bad Döben, Germany in 2005, majoring in languages and finishing with a focus work on Interlinguistics and Planned Languages. He went on to study Language and Cognitive Psychology at the Universities of Leipzig, Germany, Wollongong, Australia, and Nijmegen, the Netherlands. He received his Magister Artium in English Studies and General Linguistics from Leipzig University in 2012, specializing in Psychology of Language as well as Semantics and Pragmatics. He then moved to the Max Planck Institute for Psycholinguistics in Nijmegen for his graduate studies on the Psycholinguistics of Dialogue and received his PhD from Radboud University Nijmegen in early 2020. Mathias currently works as a post-doc in Experimental Pragmatics at Humboldt University Berlin, Germany.





---

MPI SERIES IN PSYCHOLINGUISTICS

---

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turenhout*
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. *Margolein 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 Seyfed-dinipur*
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. *Ji-dong 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 Dingemans*
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 Savvotia Asaridou*
97. Incrementality 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*
105. Modelling Multimodal Language Processing. *Alastair Smith*
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*
107. Situational variation in non-native communication. *Huib Kouwenhoven*
108. Sustained attention in language production. *Suzanne Jongman*
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntactic, and orthographic effects. *Malte Viebahn*
110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*
111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*
112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*
113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*
114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*
115. Music and language comprehension in the brain. *Richard Kunert*
116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*
118. Language processing in a conversation context. *Lotte Schoot*
119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*
120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*
121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*
122. Sensorimotor experience in speech perception. *Will Schuerman*
123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*
124. Influences on the magnitude of syntactic priming. *Evelien Heyse-laar*
125. Lapse organization in interaction. *Elliott Hoey*
126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*
127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*
128. The role of voice and word order in incremental sentence processing. Studies on sentence production and comprehension in Tagalog and German. *Sebastian Sauppe*
129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*
130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*
131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*
132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*
133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*
134. Indirect request comprehension in different contexts. *Johanne Tromp*
135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*



136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*
137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*
138. Age and hearing loss effects on speech processing. *Xaver Koch*
139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*
140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*
141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*
142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašnáková*
143. From Kawapangan to Shawi: Topics in language variation and change. *Luis Miguel Rojas-Berscia*
144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. *Linda Drijvers*
145. Understanding temporal overlap between production and comprehension. *Amie Fairs*
146. The role of exemplars in speech comprehension. *Annika Nijveld*
147. A network of interacting proteins disrupted in language-related disorders. *Elliot Sollis*
148. Fast speech can sound slow: Effects of contextual speech rate on word recognition. *Merel Maslowski*
149. Reason-giving in everyday activities. *Julija Baranova*
150. Speech planning in dialogue — Psycholinguistic studies of the timing of turn taking. *Mathias Barthel*





---

## PUBLICATIONS

---

Barthel, M., & Levinson, S. C. (in press). Phonological planning is done in overlap with the incoming turn: Evidence from gaze-contingent switch task performance. *Language, Cognition and Neuroscience*.

Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science*, *43*(7), e12768. DOI: 10.1111/cogs.12768

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next Speakers Plan Their Turn Early and Speak after Turn-Final “Go-Signals.” *Frontiers in Psychology*, *8*, 393. DOI: 10.3389/fpsyg.2017.00393

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm. *Frontiers in Psychology*, *7*, 1858. DOI: 10.3389/fpsyg.2016.01858