

How gesture and speech interact during production and comprehension

Isabella Fritz



A thesis submitted to
The University of Birmingham
for the degree of
Doctor of Philosophy

Department of English Language and Applied Linguistics
School of English, Drama and American and Canadian Studies
College of Arts and Law
The University of Birmingham
September 2017

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

This thesis investigates the mechanisms that underlie the interaction of gesture and speech during the production and comprehension of language on a temporal and semantic level. The results from the two gesture-speech production experiments provide unambiguous evidence that gestural content is shaped online by the ways in which speakers package information into planning units in speech rather than being influenced by how events are lexicalised. In terms of gesture-speech synchronisation, a meta-analysis of these experiments showed that lexical items which are semantically related to the gesture's content (i.e., semantic affiliates) compete for synchronisation when these affiliates are separated within a sentence. This competition leads to large proportions of gestures not synchronising with any semantic affiliate. These findings demonstrate that gesture onset can be attracted by lexical items that do not co-occur with the gesture. The thesis then tested how listeners process gestures when synchrony is lost and whether preceding discourse related to a gesture's meaning impacts gesture interpretation and processing. Behavioural and ERP results show that gesture interpretation and processing is discourse dependent. Moreover, the ERP experiment demonstrates that when synchronisation between gesture and semantic affiliate is not present the underlying integration processes are different from synchronous gesture-speech combinations.

Keywords: gesture, motion events, planning units, synchronisation, ERP, discourse

ACKNOWLEDGMENT

First and foremost, I would like to thank my supervisors Andrea Krott, Sotaro Kita and Jeannette Littlemore. I am very grateful for your guidance, valuable advice and interest in my project. I couldn't have wished for better supervisors. I'd like to give special thanks to Andrea Krott – not only for being a wonderful advisor on my PhD project but also for your patience and constant encouragement through the ups and downs of the last four years.

My PhD studies would have not been the same without my office mates and friends Zheni and Beinan. I could always count on your support and for that I would like to thank you. It has been great working in Hills 1.08.

I would also like to thank my fellow PhD students and friends for being part of my PhD journey and special thanks to my friends from back home for their support from the distance. Thank you, Thitima, Rawan, Martine, Asma, Mahmoud, Alper, Magdalena, Betty and Eva.

To the members of the psycholinguistics lab, thanks for your valuable feedback and comments through the different stages of my PhD project.

The College of Arts and Law, I would like to thank for funding which allowed me to undertake this PhD. The gesture lab of RWTH Aachen University and the Institut für Sprachen und Literaturen at the University of Innsbruck, I would like to thank for letting me collect data in their departments. Further, I would like to thank Effie Pearson for her help with the intercoder-reliability.

Finally, I couldn't have done this PhD without my parents Josefina and Walter, my sister Julia, my aunt Erika and my grandma Elisabeth. I greatly appreciate your constant support and patience throughout my PhD studies. Dankeschön!

TABLE OF CONTENTS

Introduction	1
Research Topic	1
Overview of the Thesis	2
Chapter 1 The Study of Co-speech Gestures	4
1.1. Aren't Gestures Non-verbal?	4
1.2. Classification of Gestures	5
1.3. Gesture Components	5
Chapter 2 Speech-Gesture Production	8
2.1. Motion Events	9
2.2. Semantic Coordination	10
2.2.1. Motivation & Research Questions (Semantic Coordination)	12
2.3. Temporal Coordination	13
2.3.1. Motivation & Research Questions (Temporal Coordination)	16
2.4. Research Paradigms in Gesture-speech Production	16
2.4.1. Methods	17
2.4.2. Statistics	20
Chapter 3 Information Packaging in Speech Shapes Information Packing in Gesture: The Role of Speech Planning Units in the Coordination of Speech-Gesture Production.	26
3.1. Abstract	26
3.2. Introduction	27
3.3. Methods – Experiment 1	33
3.3.1. Participants	33
3.3.2. Material	34
3.3.3. Design	34
3.3.4. Procedure	36

3.3.5. Data Coding and Analysis	37
3.4. Results	41
3.5. Discussion.....	43
3.6. Experiment 2	44
3.7. Methods – Experiment 2	45
3.7.1. Participants	45
3.7.2. Material.....	46
3.7.3. Design.....	46
3.7.4. Procedure.....	47
3.7.5. Data Coding and Analysis	49
3.8. Results	50
3.9. Discussion.....	54
3.10. General Discussion.....	55
Chapter 4 Gesture-speech Synchronisation: Surface Locations of Semantic Affiliates within a Sentence Predict Gesture Onset and Gesture Duration	60
4.1. Abstract.....	60
4.2. Introduction	62
4.3. Present Study.....	67
4.4. Methods	70
4.4.1. Participants	70
4.4.2. Procedure.....	71
4.4.3. Data Coding and Analyses	71
4.5. Results	74
4.5.1. Gesture Onset Analyses.....	77
4.5.2. Gesture Duration Analyses.....	85
4.6. General Discussion.....	90

Chapter 5 Speech-gesture Comprehension.....	100
5.1. Do gestures influence language comprehension?.....	100
5.2. Motivation and Research Questions	101
5.3. ERP Technique (Event-related potential).....	103
5.4. What can ERP studies tell us about language processing?.....	106
5.4.1. N400	107
5.4.2. P600	109
5.4.3. N400 & P600 Revisited.....	111
5.4.4. Nref.....	113
5.5. ERP Research and Co-speech Gesture	114
5.5.1. Research Paradigms.....	114
5.5.2. ERP Components.....	117
Chapter 6 Multimodal Language Processing: How Preceding Discourse Constrains Gesture Interpretation and Influences Gesture Processing	119
6.1. Abstract.....	119
6.2. Introduction	120
6.3. The Present Study.....	127
6.4. Experiment 1 – Behavioural Experiment	128
6.5. Methods	129
6.5.1. Participants	129
6.5.2. Material.....	129
6.5.3. Stimuli Recordings & Editing	132
6.5.4. Procedure.....	134
6.6. Results	136
6.7. Discussion Experiment 1 - Behavioural Experiment	139
6.8. Experiment 2 – ERP Experiment	141

6.8.1. Relevant ERP components: Nref, N400, P600.....	141
6.8.2. Relevant ERP components and speech-gesture processing.....	143
6.8.3. ERPs time-locked to the gesture's onset	145
6.8.4. ERPs time-locked to the gesture's semantic affiliate	146
6.9. Methods	147
6.9.1. Participants	147
6.9.2. Material.....	148
6.9.3. Procedure	149
6.9.4. EEG Recording and Analysis	150
6.10. Results	152
6.11. Discussion Experiment 2 - ERP Experiment.....	159
6.11.1. Interpretation of ERPs time-locked to the gesture's onset	160
6.11.2. Interpretation of ERPs time-locked to the semantic affiliate's onset	161
6.11.3. Interpretation of posterior and anterior P600 effects.....	162
6.11.4. Ecological validity of the stimuli.....	163
6.12. General Discussion.....	164
6.13. Conclusions	168
Chapter 7 General Discussion	169
7.1. Summary of Findings & Conclusions	169
7.2. Implications	172
7.2.1. Planning Units in Speech and Gesture Production.....	172
7.2.2. The Attraction Point Hypothesis of Gesture-speech Synchronisation	173
7.2.3. Gesture-speech Production	175
7.2.4. Discourse and Gesture Processing.....	175
7.2.5. Gesture-speech Synchrony	176
7.3. Open Questions & Future Directions	177

7.3.1. Attraction Point Hypothesis	177
7.3.2. Gesture and Discourse	180
List of References	181
APPENDIX 1	199
Particle Verbs from Experiment 1	199
Particle Verbs from Experiment 2	200
APPENDIX 2	201
Training Phase – Experiment 2	201
APPENDIX 3	203
Example Stimuli Gesture-speech Comprehension Study	203

LIST OF TABLES

Table 2.1. Summary of studies on speech-gesture production using motion events as test domain.....	22
Table 3.1. Summary of the fixed effects of the full mixed logit model for gestural depiction in Experiment 1	42
Table 3.2. Summary of the fixed effects in the mixed logit models in English for gestural depiction in Experiment 1.	43
Table 3.3. Summary of the fixed effects in the mixed logit models in German for gestural depiction in Experiment 1.	43
Table 3.4. Summary of the fixed effects in the mixed logit model for the occurrence of pauses in the Inserted Clause Condition versus the Inserted Phrase Condition in Experiment 2.	51
Table 3.5. Summary of the fixed effects (Construction Type) in the mixed logit model for gestural depiction in Experiment 2.....	53
Table 3.6. Summary of the fixed effects (Pause) in the mixed logit model for gestural depiction in Experiment 2 (Inserted Clause and Inserted Phrase Condition only).	53
Table 3.7. Summary of the fixed effects (Pause and Clause Type) in the mixed logit model for gestural depiction in Experiment 2 (Inserted Clause and Inserted Phrase Condition only).	54
Table 4.1. Gesture synchronisation categories	72
Table 4.2. Means (ms) of the duration of Gesture Onset for Path Gestures synchronising with the Verb, the Particle and gestures that were placed between the verb and particle in the Inserted Clause Condition (including Main Clause Condition from Experiment 1) and Inserted Phrase Condition.....	77
Table 4.3. Summary of fixed effects in the mixed linear model for Gesture Onset (Path Gestures that synchronised with the Verb)	78
Table 4.4. Summary of fixed effects in the mixed linear model for Gesture Onset (Path Gestures falling between Verb & Particle)	79
Table 4.5. Summary of fixed effects in the mixed linear model for Gesture Onset (Path Gestures that synchronised with the Particle)	79

Table 4.6. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated Gestures that synchronised with the Verb and Verb & Particle)	82
Table 4.7. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated Gestures that fell in between verb and particle)	83
Table 4.8. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated Gestures that synchronised with the Particle)	83
Table 4.9. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated and Path Gesture that either fell between Verb and Particle or synchronised with the Particle)	83
Table 4.10 Summary of fixed effects in the mixed linear model for Gesture-Particle Asynchrony in the Inserted Clause Condition from Experiment 2 plus the Main Clause Condition from Experiment 1 and the Inserted Phrase Condition from Experiment 2.	84
Table 4.11. Means (ms) of the duration of Gesture Duration for Path Gestures synchronising with the Verb, the Particle and gestures that were placed between the verb and particle in the Inserted Clause Condition (including Main Clause Condition from Experiment 1) and Inserted Phrase Condition.	86
Table 4.12. Summary of fixed effects in the mixed linear model for Gesture Duration (Path Gestures that synchronised with the Verb)	87
Table 4.13. Summary of fixed effects in the mixed linear model for Gesture Duration (Path Gestures that fell between the Verb and the Particle)	87
Table 4.14. Summary of fixed effects in the mixed linear model for Gesture Duration (Path Gestures that synchronised with the Particle)	87
Table 4.15. Summary of fixed effects in the mixed linear model for Gesture Duration (Conflated Gestures that synchronised with the Verb and Verb & Particle)	88
Table 4.16. Summary of fixed effects in the mixed linear model for Gesture Duration (Conflated Gestures that fell between the Verb and the Particle)	88
Table 4.17. Summary of fixed effects in the mixed linear model for Gesture Duration (Conflated Gestures that synchronised with the Particle)	88
Table 4.18. Correlations between Gesture Duration and Gesture-Particle Asynchrony for Path Gestures and Conflated Gestures for the synchronisation categories Particle and Between Verb and Particle	90

Table 5.1. Example Response from the speech-gesture production study presented in Chapter 4 where gesture-speech synchrony is lost.....	102
Table 6.1. Example stimulus including all four stimuli sets. For the behavioural experiment the stimuli were cut before the second word in bold in the target sentence column. The onset of the first word in bold co-occurred with the onset of the gesture. For the ERP study, ERPs were time-locked to the onset of the words in bold (target sentence). The four stimuli sets were counter-balanced across participants. Target-verbs and speech elements between the gesture and the target-verb were counter-balanced as follows: Stimuli Set 1 and Stimuli Set 2 differ in terms of the elements between the gesture and the target-verb, the same applies to Stimuli Set 3 and Stimuli Set 4. The target-verbs are the same in Stimuli Set 1 and Stimuli Set 2 but differ from Stimuli Set 3 and Stimuli Set 4.....	133

LIST OF FIGURES

Figure 1.1. Example of a Gesture Unit.....	7
Figure 3.1. Example stimulus for the English Subordinate Clause condition. A fourth slide illustrated the sentence that the participants were expected to produce, i.e. I can see in the video that the elephant is climbing up the rainbow. Participants started to describe the video once the slide turned blank.....	37
Figure 3.2. Mean Proportions of Separate Response (i.e., responses in which gesture separated Manner and/or Path) and Conflated Only Responses (i.e., responses that include only gestures that conflated Manner and Path) in Main Clauses versus Subordinate Clause in English and German in Experiment 1. Error bars represent standard errors. The bottom of the figure shows an example sentence of each of the four conditions. For the German conditions, a word-by-word translation is provided.....	42
Figure 3.3. Example stimulus for the Inserted Clause Condition in Experiment 2.....	48
Figure 3.4. Mean proportions of responses with a pause produced either before or after the inserted element (phrase/clause) versus responses without any pauses in Experiment 2.	51
Figure 3.5. Mean Proportions of responses with Conflated Gestures and Separated Gestures across conditions of Experiment 2. Error bars represent standard errors. The bottom of the figure shows an example sentence of each of the three conditions including a word-by-word translation of the examples.	52
Figure 3.6. Mean Proportions of responses with Conflated Gestures and Separated Gestures with and without a pause in the Inserted Clause and Inserted Phrase Condition of Experiment 2. Error bars represent standard errors.....	54
Figure 4.1. Top: Mean proportion of Path Gestures that synchronised with the Verb, Particle, or Between Verb & Particle (see Table 4.1 for definitions), split into responses with and without an Additional Path Preposition. Note that no Path Gestures synchronised with Verb & Particle. Bottom: Means (ms) of the duration of Sentence Onset to Verb Onset (VO), Verb Onset to Particle Onset (VOtoPO), Sentence Onset to Gesture Onset (GO) and Gesture Onset to Particle Onset (GOtoPO) for Path Gestures whose stroke onset occurred after the verb.	75

Figure 4.2. Top: Mean proportion of Conflated Gesture which synchronised with the Verb, Particle, Between Verb & Particle, or Verb & Particle, split into responses with and without an Additional Path Preposition.	76
Figure 4.3. Relationship between Gesture Onset and Verb Onset for Path Gestures synchronising with the Verb, the Particle, and gestures that were placed between the Verb & Particle.....	80
Figure 4.4. Relationship between Gesture Onset and Particle Onset for Path Gestures synchronising with the Verb, the Particle, and gestures that were placed between the verb and the particle.	80
Figure 4.5. Relationship between Gesture Onset and Verb Onset for Conflated Gestures synchronising with the Verb (including gestures that synchronised with Verb & Particle), the Particle and gestures that were placed between the verb and the particle.	82
Figure 4.6. Relationship between Gesture Onset and Particle Onset for Conflated Gestures Synchronising with the Verb (including gestures that synchronised with Verb & Particle), the Particle and gestures that were fell between the verb and the particle. ...	82
Figure 4.7. Relationship between Stroke Duration and Gesture Onset – Particle onset asynchrony for Path Gestures synchronising with the Verb, the Particle and for gestures that fell between Verb and Particle.	87
Figure 4.8. Relationship between Stroke Duration and Gesture Onset – Particle Onset asynchrony for Conflated Gestures synchronising with the Verb (including Verb & Particle), the Particle and for gestures that fell between Verb and Particle.	89
Figure 6.1. Example of a trial used in the behavioural study.	136
Figure 6.2 Distribution of response to the open question: “What does the gesture represent?”, split into Related and Unrelated Discourse conditions. The Y-axis shows the number of different responses per item across all participants. The light and dark lines show the Discourse Related and Unrelated conditions, respectively. Items in the X-axis are ordered from the lowest number of different responses to the highest number of different responses in the Discourse Related Conditions.....	136
Figure 6.3. Mean ratings for the question “How easy was it to interpret the gesture?” for the Related and Unrelated Discourse Condition. Error bars represent the standard error of the means.....	137

Figure 6.4. Effects of Gesture Match and Discourse Relatedness on mean ratings for gesture match/mismatch for Question 3: “How well does your interpretation of the gesture match the following word?” Error bars represent the standard error of the means. ...	138
Figure 6.5. Mean ratings for Question 4: “Could the gesture (also) represent [verb]?” for matching/mismatching gestures and Related and Unrelated Discourse conditions. Error bars represent the standard error of the means.	139
Figure 6.6. Stimulus presentation in Experiment 2 (ERP experiment)	150
Figure 6.7. ERP responses at a subset of electrodes in the Related Discourse Condition compared to the Unrelated Discourse Condition time-locked to gesture onset.	154
Figure 6.8. Topographical map of the ERP difference between Unrelated Discourse Condition and the Related Discourse Condition in the time-window 800 -950 ms post gesture onset. Black dots show electrode sites and circles the ROIs included in the statistical analysis.	155
Figure 6.9. Gesture Match and Gesture Mismatch waveforms for selected electrodes in the Related Discourse Condition time-locked to the target verb onset.	157
Figure 6.10. Gesture Match and Gesture Mismatch wave forms for selected electrodes in the Related Discourse Condition time-locked to the target verb onset.	158
Figure 6.11. Topographical maps of the ERP difference between mismatching and matching gestures in the Related Discourse Condition and the Unrelated Discourse Condition for the 800-1200 ms time-window. Black dots refer to the electrode sites and circles to the ROIs included in the statistical analysis.	159

Introduction

Research Topic

When speaking, we inevitably use gestures. Gestures accompanying speech are considered to be universal because to date there is no report of a culture whose members do not gesture (Kita, 2009). Traditionally gestures have not been seen as part of our language system, but in recent years the view that language and gesture comprise two completely separate channels has started to be challenged (McNeill, 2015). Research from the growing field of gesture studies has produced increasing evidence that demonstrates a tight link between gestures and speech during production and comprehension. As Goldin-Meadow (2003, p. 3) points out, “[t]o ignore gesture is to ignore part of the conversation”.

Gestures play an important role in how we communicate and in how we think. The influence of gestures on thinking processes benefit speakers in different ways and situations. When we translate our thoughts into speech, gestures help break down complex spatial information in order to package them into units that are suitable for speech production (Kita, 2000; Kita & Özyürek, 2003). Recently it has been found that pupils gain a deeper understanding of mathematical concepts when solving a maths problem with the help of gestures (Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014). Studies of gesture in the language classroom show that the use of iconic gestures (gestures depicting an object or motion) aid language learners when remembering new vocabulary (Macedonia, 2014; Macedonia & Knösche, 2011).

In terms of the communicative functions of gestures, an increasing number of studies suggest that co-speech gestures influence language comprehension. Behavioural studies have demonstrated that information solely conveyed via gesture is available for retelling in speech (Cassell, McNeill, & McCullough, 1999; Goldin-Meadow, Wein, & Chang, 1992).

Electrophysiological research (using ERPs) has shown that difficulties of gesture integration (i.e., gesture-speech mismatch) elicit a similar effect (N400 effect) as when speakers have difficulties in semantically integrating a word into a sentence or discourse context (Kelly, Kravitz, & Hopkins, 2004). Taken together, the bulk of research on gesture-speech comprehension supports the claim that listeners use information conveyed by gestures to form a unified representation of a message (Hostetter, 2011; McNeill, 1992).

The field of gesture studies has grown rapidly in recent years (<http://www.gesturestudies.com>). Gesture researchers are working in various disciplines including linguistics, psychology, neuroscience and anthropology. This thesis is situated in the field of psycholinguistics and investigates cognitive processes involved in gesture-speech production and comprehension. Despite the growing research on the relationship between gesture and speech, the exact nature of this relationship is still controversial. Although it is generally accepted that gestures are coordinated with speech both on a semantic and on a temporal level (McNeill, 1992, 2005), the exact mechanisms underlying these two main features of co-speech gestures are unclear. The aim of this thesis is to contribute to the understanding of how speech and gesture interact on a semantic and on a temporal level during production and comprehension.

Overview of the Thesis

This section provides an outline of the thesis. The thesis has two main parts: A gesture-speech production part that includes two studies on how gesture and speech are coordinated on a semantic level and on a temporal level. The second part of the thesis is on the gesture-speech relationship in comprehension. This part includes a study on whether discourse information can guide gesture interpretation and gesture integration into a listener's discourse model. The structure of this thesis is as follows: Chapter 1 consists of a general introduction to major

concepts in gesture studies with a focus on the definition of relevant terms. Chapter 2 reviews research on how gesture and speech are coordinated on both a temporal (i.e., gesture-speech synchrony) and on a semantic level in production. This review leads up to the research questions and the motivation of the first two studies presented in this thesis. Since both studies look at the coordination of gesture and speech in the retellings of motion events, the focus of the literature review is on motion event gestures. Chapter 2 also includes a review of methods used in previous speech-gesture production experiments that investigated motion event gestures as well as a justification of the study design used in this thesis. Chapter 3 reports the first study that investigated on which linguistic level (planning unit level or lexicalisation level) gesture and speech are coordinated. Chapter 4 deals with the question of gesture-speech synchronisation. In this chapter a new hypothesis that aims to explain the mechanisms underlying gesture-speech synchronisation within a sentence context is put forward (*Attraction Point Hypothesis*). Chapter 5 introduces the second main topic of this thesis: the relationship between speech and gesture during comprehension. This chapter bridges the production and comprehension studies of this thesis. It also includes an introduction to the ERP technique and a literature review that focusses on research using the ERP technique to investigate gesture-speech processing. Chapter 6 introduces the final study of this thesis. The study reported in Chapter 6 was designed to investigate the role of discourse information on gesture-speech comprehension. In this study a behavioural experiment and an ERP experiment were conducted. Finally, Chapter 7 brings together all experiments and discusses the findings in relation to the existing literature. Chapter 7 also includes theoretical implications of this thesis as well as suggestions for future research directions.

Chapter 1

The Study of Co-speech Gestures

Chapter 1 provides an introduction to the field of gesture studies. In this chapter, important terms and concepts will be defined in order to provide a theoretical framework for the studies presented in this thesis.

1.1. Aren't Gestures Non-verbal?

For approximately 2000 years the study of gestures focussed on how gestures express emotions and inner states and already in Antique rhetoric, gestures were analysed in terms of how they are perceived by an audience. (see Müller, 2002 for a history of gesture studies). Moving forward to the twentieth century, linguists and psychologists did not show much interest in the study of gestures. Exceptions were Wilhelm Wundt, the “father of experimental psychology”, who discussed the topic of gesture in detail at the beginning of the century and David Efron who was probably the first researcher who investigated the link between speech and gesture (see Kendon, 2004 for an overview of Efron's & Wundt's work on gestures). Only in the 1970s researchers became increasingly interested in gestures as a research subject (McNeill, 2005). Important for the field of psycholinguistics is McNeill's (1985, 1992) pioneering work on the relationship between gesture and speech. Only his claim that gestures and speech constitute an integrated process during production introduced gestures into the field of psycholinguistics (Kita, 2014). In numerous publications, including his article “So you think gestures are nonverbal?” from 1985, McNeill promoted the importance of gesture research within the field of linguistics. Within the last three decades, researchers within the newly established field of gesture studies found striking evidence that co-speech gestures are so much more than just meaningless hand-waving (cf. Goldin-Meadow, 2003).

1.2. Classification of Gestures

The term “gesture” in this thesis refers to hand movements that co-occur with speech. These gestures are assumed to be part of our communication. Thus, they are in contrast to so-called adaptors which do not have any communicative functions (Goldin-Meadow, 2003). Adaptors include movements such as holding your chin or soothing your hair. In everyday communication, we use different types of hand gestures while speaking. In an influential gesture classification put forward by McNeill (1992), he distinguishes between four types of speech accompanied gestures, i.e. iconic gestures, metaphoric gestures, deictic gestures and beat gestures. Iconic gestures depict something concrete including an action, event or object (e.g., speech: the ball is rolling down the hill; gesture: the index finger is pointing away from the body and makes circles as it moves downwards) which may or may not be encoded in speech (Kita & Özyürek, 2003). Metaphoric gestures are similar to iconic gestures but they represent an abstract idea (e.g., a music teacher talks about high and low notes while using corresponding path gestures indicating high is up and low is down). Deictic gestures are pointing gestures that can either refer to something concrete or abstract. Finally, beat gestures are small rhythmic movements that align with prosodic peaks in speech. Beat gestures have been found to be important for structuring discourse (McNeill, 1992). The outline of the relationship between speech and gesture in this literature review chapter is based on research investigating iconic gestures with a focus on motion event gestures. Motion event gestures are the subject of the first two studies in this thesis.

1.3. Gesture Components

Gestures usually consist of multiple components that serve different functions. The gesture anatomy introduced in this section, is based on McNeill’s (1992) work. According to McNeill every gesture is embedded in a so-called Gesture Unit. A Gesture Unit begins when

the limbs move away from their resting position and ends with the return of the limbs to a resting position. This unit includes one or more gestures. Within a Gesture Unit we find Gesture Phrases that again can contain up to five different Gesture Phases (Kendon, 1980; McNeill, 1992). Not all of these five components of a gesture phrase are obligatory. As the name indicates, during the preparation phase (optional) the speaker's limb is getting ready to perform the stroke (e.g., arm moves up next to the head to perform a downwards path gesture). After the limb is in position for the stroke, a pre-stroke hold (optional) can occur (Kita, 1990). This often happens if the gesture is in the right position to perform the stroke before the corresponding element in speech has caught up. The stroke (obligatory) is the meaningful part of the gesture (e.g., downwards movement to indicate path). When the stroke is completed, a post-stroke hold (optional) can be important in terms of gesture-speech synchrony. Because if the element in speech has not completed its encoding yet, the post-stroke hold helps to achieve a gesture-speech alignment (Kita, 1990). Finally, at the end of a Gesture Unit the limbs return to their resting position. However, the endpoint of this so-called retraction phase does not necessarily have to be the same as the starting point of the Gesture Unit (McNeill, 2005). Although all components of a gesture have their specific functions, the most important one is the stroke which carries the meaning of the gesture (McNeill, 2005). Figure 1.1 shows an example of a Gesture Unit that depicts an iconic gesture.¹ The Gesture Unit includes a resting position, preparation phase, stroke phase and retraction phase. Semantically, the gesture depicts a downward path movement (bold face indicates when the participant performed the stroke in relation to speech). In the remainder of this thesis, when gesture-speech synchronisation is mentioned, this synchronisation refers exclusively to the gesture's stroke if not otherwise stated.

¹ For all foreign language examples presented in this thesis a word-by-word translation and a free translation are provided.



Resting Position

Preparation

Preparation

Stroke

Retraction

German: Der Kreis hüpfte, wie im Video gesehen, den **Berg hinunter**.
 Literal translation: The circle jumps how in the video seen the mountain down.
 Free translation: The circle is jumping down the mountain, as seen in the video.

Figure 1.1. Example of a Gesture Unit.

Chapter 2

Speech-Gesture Production

In this chapter two vital features of co-speech gestures will be introduced; i.e. the gestures' temporal and semantic relationship with speech. Researchers investigating the semantic and temporal link between these two modalities aim for a better understanding of how gesture and speech exchange information during production. This information exchange refers to the feedback channelling between gesture and speech during the different stages of production, i.e. from the pre-linguistic stage to the execution phase of the gesture. First, research in this area will be discussed and gaps in the literature highlighted. Second, the literature review leads up to the research questions of the two gesture-speech production studies presented in this thesis. The first study looks at the interaction of gesture and speech on a semantic level, i.e. how is the meaning expressed in gesture coordinated with the meaning expressed in speech (Chapter 3). In the second study it was tested how gesture and speech are coordinated on a temporal level. In particular, it was investigated whether the surface occurrence of the speech elements that are closest to the gesture's meaning have an influence on where within a sentence a gesture occurs (i.e., gesture onset) and for how long a gesture is prolonged (i.e., gesture duration) (Chapter 4). Furthermore, motion events will be introduced since they are an important test domain in gesture studies. Moreover, the current chapter provides a review of research on gesture-speech production studies where motion event gestures were used as test domain. The focus of this review is on methods used in previous studies. Besides a review of previous methods including their drawbacks for statistical analyses, also the study design used in this thesis will be introduced and a justification why previous research methods were not suitable to answer the current research questions will be provided.

2.1. Motion Events

How are gesture and speech coordinated on a semantic and on a temporal level? In the last two decades this question has received much research attention (see P. Wagner, Malisz, & Kopp, 2014 for a review). Researchers who have investigated this issue have focused on a cross-linguistic perspective. More specifically, cross-linguistic studies on motion event gestures tested whether the content and the synchronisation patterns of iconic gestures are shaped by language (e.g., Kellerman & van Hoof, 2003; Kita & Özyürek, 2003; Stam, 2006; Wessel-Tolvig & Paggio, 2016). One important observation of these studies is that imagery for speaking and gesturing has been found to vary systematically across languages. This is evident in differing synchronisation patterns but also in the gestural content (McNeill, 2009; McNeill & Duncan, 2000). Here, motion events have proven to be appropriate subject matter for cross-linguistic research because they are vital in all languages and their linguistic encoding varies greatly across languages (cf. Slobin, 2003). According to Talmy's (2000) influential typology, languages generally fall into one of two categories depending on how they lexicalise motion events. In so-called satellite-framed languages the manner component of a motion event is linguistically encoded within the verb whilst path is encoded in a particle ("satellite") outside of the verb (1). Satellite-framed languages include Germanic languages like English and German. Although the lexicalisation pattern is the same in German and English, the word order of the utterance differs.

Satellite-framed Language – English

(1) The balls	is rolling	down	the hill.
	MANNER+MOTION	PATH	

Satellite-framed Language – German

(2) Der	Ball	rollt	den	Hügel	hinunter.
The	ball	rolls	the	hill	down.
	MANNER+MOTION			PATH	

In verb-framed languages on the other hand, the path component is encoded within the verb whilst the encoding of manner is not obligatory and can be omitted. If manner is linguistically encoded it is subordinated to the main verb in either a gerund, new verb or a clause. In the category of verb-framed languages fall, for example, Romance languages, Japanese and Turkish.

*Verb-framed Languages – French*²

(3) Le	chien	est	entré	dans	la	maison	(en courant).
The	dog	has	entered	in	the	house	(by running).
	FIGURE		PATH + MOTION			GOAL	MANNER
	The dog ran into the house.						

Based on this typological distinction, numerous studies have been conducted to test how the linguistic encoding of motion events influences how speakers gesture when describing such events (see Table 2.1 for a list of studies).

2.2. Semantic Coordination

How motion event gestures differ across typologically different languages has been shown in studies where narratives from speakers across typologically different languages were analysed. These analyses have shown that in English (a satellite-framed language) speakers tend to break down the path component of a motion event into multiple path gestures. In contrast, speakers of Spanish (a verb-framed language) tend to represent path more holistically by indicating path with a single gesture. This reflects the linguistic structure used in the retellings (McNeill, 2009, p. 522ff.) that is illustrated in the example below. In this example an English and a Spanish speaker explain a scene from the Sylvester and Tweety cartoon (Freleng, 1950) where Tweety drops a bowling ball into a drainpipe in which Sylvester is trying to climb up. The bowling ball and Sylvester meet half way and Sylvester falls down and together with

² The French example was taken from Slobin (2003, p. 162).

the bowling ball rolls out of the drainpipe and onto the street. The English speaker breaks the scene down into six path gestures whilst the Spanish speaker describes the same event with one gesture indicating a curved trajectory.

English³

- (1) and it **goes** down
- (2) but it **rolls him out**
- (3) down **the** (4) **rainspo**
- (5) ut **out** into
- (6) **the sidewalk** into a bowling **alley**

Spanish

entonces **SSS**⁴
 then he-falls ONOM
 then SSS he falls

McNeill explains these differences in gestural encoding with his Growth Point Theory (1992; 2000). Growth Points (GP) are seen as theoretical idea units which underlie speech and gesture. More specifically, a GP is a “minimal unit of an imagery-language dialectic” (McNeill, 2005, p. 105). From this GP speech and gesture emerge in the course of a dynamic process. A GP can be identified through the semantic content of a gesture as well as the gesture’s synchrony with speech. In terms of characteristics of the GP, McNeill & Duncan (2000, p. 154) argue that “[t]hinking emerges via the GP with language categories built in.” Thus, different encoding possibilities of a languages (e.g., verb-framed vs. satellite-framed) encourage different forms of thinking. The idea of a Growth Point fits nicely into Slobin’s (1987, 2000, 2003) influential thinking-for-speaking hypothesis. The thinking-for-speaking hypothesis proposes that in the online process of speech production our thoughts have to be put into the given linguistic encoding possibilities of a language. Slobin’s thinking-for-speaking hypothesis has often been considered as a weak version of the Sapir-Whorf Hypothesis (cf. Whorf, 1956) in which language has an influence on our “habitual thought”. Slobin (2003) does not claim

³ Annotation has been simplified. Only the stroke phase is indicated (bold face)

⁴ Onomatopoeia which is a frequent substitute of verbs in Spanish (McNeill, 2009)

that language has an influence on our thinking per se, rather grammatical encoding possibilities of a language guide our thinking while we are speaking. The claim that while speaking visuospatial cognition varies across languages is also evident in varying gestural patterns across typologically different languages (McNeill, 2009).

From a more psycholinguistic perspective, Kita & Özyürek (2003) found that gesture and speech are coordinated on a clausal level. In their study they also made use of linguistic differences between verb-framed and satellite-framed languages. Besides different lexicalisation patterns (i.e., verb plus particle versus two verbs), typologically different languages also differ in terms of clausal packaging of motion events. Motion events in satellite-framed languages are encoded within one single clause (e.g. “rolling down”), whilst in verb-framed languages manner and path is distributed across two clauses (e.g. “going down while rolling”). Kita & Özyürek (2003) found that when speakers use a one-clause construction, they also tend to package gestural information within one gesture; i.e. conflating manner and path gesturally. Since Kita & Özyürek’s (2003) paper, numerous studies have investigated the relationship between lexicalisation patterns, clausal encoding and gestural content. These studies include a wide range of languages as well as different participant groups (e.g., L1-L2, children and aphasic patients). Methods used in these studies are discussed later in this chapter.

2.2.1. Motivation & Research Questions (Semantic Coordination)

Despite the presence of a large body of research that investigates the semantic coordination of speech and gesture, researchers thus far have not been able to clarify on which linguistic level gestures and speech are coordinated. The main reason for this is that in studies so far, whenever the lexicalisation patterns differed (verb-framed vs satellite-framed constructions) also the clausal packaging differed (one-clause vs. two-clause structure). Thus, the differences in gestural content could either stem from different lexicalisations of motion

events (i.e., one-verb versus two-verb constructions) or from the encoding of the motion event in either one clause (satellite-framed) or two clauses (verb-framed). Since a clause is assumed to be a good proxy for planning units (Bock, 1982; Levelt, 1989), the coordination on a clausal level would indicate that planning units in speech shape gestural planning units. However, this “Planning Unit Account” has not been tested directly. In the first study described in this thesis (Chapter 3), two experiments were designed to test directly on which linguistic level speech and gestures are coordinated during production. The studies were conducted in German (Experiment 1 and Experiment 2) and English (Experiment 1). Both languages can be classified as satellite-framed languages. Thus, the visuospatial cognition that according to McNeill (2009) influences gestural content should not play a role in the conducted studies.

2.3. Temporal Coordination

The close temporal coordination of gesture and speech is another important concept within the field of gesture studies. The importance of the temporal coordination lies in the claim that speech and gesture emerge from the same idea unit which is evident in their synchrony on the linguistic surface structure (McNeill, 1985, 1992). Generally, gestures synchronise with speech elements that are closest to the gesture’s meaning which is also known as McNeill’s (1992, 2005) *Semantic Synchrony Rule*. These lexical items that reflect the gesture’s meaning are often referred to as a gesture’s “lexical affiliate” (introduced by Schegloff, 1984). However, the term “lexical affiliate” implies that there is a one-word-one-gesture mapping which is often not the case. More often, more than one word constitutes a gesture’s lexical affiliate (de Ruiter, 1998). In this thesis the term “semantic affiliate” will be used to refer to lexical items that are semantically related to the meaning of an iconic gesture. The term “semantic affiliate” is similar to the term “conceptual affiliate” which has been used by de Ruiter (2000). However, the term “semantic affiliate” focusses on the meaning relationship between gesture and specific speech

elements whilst the term “conceptual affiliate” suggests a more loose relationship between the two modalities.

McNeill (2005, p. 22) suggests that when we are speaking “our mind is doing the same thing in two ways”. For example, if a speaker talks about a cake, she might use a gesture indicating the shape of the cake whilst this information is not mentioned in speech. However, close examination of where exactly a gesture is placed in relation to its semantic affiliate has shown that the concept of synchrony is not straight-forward. As de Ruiter (1998, p. 17) has aptly pointed out “the issue of temporal synchronization is a nebulous one”, however defining the temporal relationship between gesture and speech has already proven to be problematic for researchers. One reason for this unresolved issue lies in the way meaning is conveyed across the gesture and speech channels. The meaning of a gesture is conveyed globally whilst speech conveys the meaning of a message in a more linear fashion (McNeill, 1992; Özyürek, 2014).

In terms of synchronisation, this leads to a technical problem; i.e. uttering the semantic affiliate might take longer/shorter than representing the underlying idea unit gesturally. Due to these restrictions, gesture and speech cannot be in complete synchrony (cf. de Ruiter, 1998). Thus, de Ruiter (1998, 2000) suggests that a gesture’s onset is only roughly time-locked to its semantic affiliate. Indeed, research suggests that the onset of a gesture tends to precede its semantic affiliate (Beattie & Aboudan, 1994; Morrel-Samuels & Krauss, 1992; Schegloff, 1984). However, the precise nature of these asynchronies and their underlying cause are not clear yet.

Another problem with defining synchrony arises when investigating gestures within a sentence or discourse context. As mentioned above, not every gesture can be allocated one semantic affiliate. Rather, gestures can have multiple lexical items which function as an affiliate. If the claim holds that a gesture is temporally aligned with its semantic affiliate, where

would a speaker place a downward path gesture produced while uttering the following sentence: *The leaf is floating down into pond?* In this case the manner verb, the path particle, the path preposition could function as the gesture's semantic affiliate. Given multiple lexical items that can function as a gesture's affiliate, what factors determine which lexical item(s) a gesture co-occurs with? According to research on motion event gestures, gesture placement is influenced by different factors. This includes the focus of the speaker (McNeill & Duncan, 2000) and discourse development (what is newsworthy) (Chui, 2005). As mentioned above, it has also been found that the imagery for speaking and gesturing varies systematically across languages. This is evident in different synchronisation pattern across languages (McNeill, 2009; McNeill & Duncan, 2000).

To sum up, researchers who focus on gesture-speech synchrony are primarily interested in testing exactly where and when a gesture is placed in relation to speech. The question of “*where*” concerns the semantic and pragmatic synchronisation which indicates a common idea unit (McNeill, 1992, 2005). In other words, researchers are interested in the semantic relationship between the lexical items the gesture synchronises with and the gesture's content within a discourse context. The “*when*” is a more psycholinguistic question, dealing with the production mechanisms of these two modalities. In particular, it concerns the precise coordination between speech and gesture during different stages of production. However, this research question has been widely neglected when it comes to iconic gestures, especially within a sentence context. The few studies undertaken in this area focussed on the synchronisation of deictic gestures and speech when one of the two modalities is interrupted (Chu & Hagoort, 2014; Levelt, Richardson, & La Heij, 1985).

2.3.1. Motivation & Research Questions (Temporal Coordination)

As the review of the literature on gesture-speech synchronisation research has shown, the question of how synchronisation between gesture and speech is achieved remains unresolved. Moreover, the exact nature of when an iconic gesture has its onset in relation to its semantic affiliate is far from clear. For the synchronisation study presented in Chapter 4 a meta-analysis of the experiments of Chapter 3 was conducted. In this meta-analysis of motion event gestures it was tested whether gesture onset and gesture duration are influenced by where within a sentence the gestures' semantic affiliates are encoded (i.e., the manner verb and the path particle and possible additional path prepositions). This detailed analysis of the relation between semantic affiliates, gesture onset and gesture duration allowed us to infer how gesture and speech are coordinated during the different stages of production as well as the gesture's execution phase.

2.4. Research Paradigms in Gesture-speech Production

In order to explain the methodology employed in this thesis, in this section methods used in previous studies on gesture-speech production that used motion events as test domain are reviewed. More specifically, material, procedure and analyses used in previous studies will be discussed and their advantages and drawbacks will be highlighted. This assessment of the value of previous study designs will then lead up to the justification of the study design and the statistical analysis employed in this thesis.

The review includes 21 studies (see Table 2.1) on motion event gestures using speech production tasks. The scope of this review is limited to journal papers and book chapters. (except one conference paper by Mol and Kita (2012) because of its methodological relevance). Furthermore, the review is limited to studies conducted with adult participants. Studies conducted with participants from a special population were excluded (e.g., studies with people

with aphasia: Dipper, Pritchard, Morgan, & Cocks, 2015). The reviewed papers include cross-linguistic studies (spoken languages only) as well as studies conducted with monolinguals and/or bilinguals. In terms of the research topic, the review includes studies that investigate the effect of linguistic encoding (clausal structure and/or lexicalisation) on the gesture's content and synchronisation patterns. Most of the studies tested the effect of clausal structure or lexicalisation on gestural content (e.g., type of motion event gesture: manner path conflated versus manner and path separated). Only 8 out of 21 studies investigated the influence of linguistic encoding on gesture-speech synchrony.

2.4.1. Methods

In terms of the studies' procedure, a large number of studies (10 studies) on motion event gestures follow McNeill's⁵ (1992, 2005) approach of collecting narratives through cartoon retellings. In his studies, McNeill used an approximately six minute long Sylvester and Tweety Bird cartoon as stimulus (Freleng, 1950) which the participants had to retell to a confederate. This type of study design elicits a comprehensive narration whilst it is up to the participant which information s/he wants to include.

In more recent studies, also shorter animated clips/movies were used as stimuli for an elicitation task. This includes the Tomato Man Movies (Özyürek, Kita, & Allen, 2001) which were specifically designed to elicit different syntactical packaging of manner and path in speech. These short clips that depict single motion events do not elicit comprehensive narratives. This makes them especially useful for eliciting specific sentence structures but also to keep the influence of discourse and context to a minimum.

⁵ McNeill's work is not included in the review table since the methodology he used in his studies is described in detail in this section.

Other researchers also created animated stimuli that depict motion events (e.g., Akhavan, Nozari, & Göksun, 2017; Wessel-Tolvig & Paggio, 2016). Occasionally, *The Frog Story* (Mayer, 1969) was used as stimulus material (Kellerman & van Hoof, 2003; Negueruela, Lantolf, Jordan, & Gelabert, 2004) despite McNeill's (2005) observation that this static-picture story lowers gesture frequency.

As stimuli for the experiments in this thesis, short clips taken from a German educational programme for children "Die Sendung mit der Maus" (The Programme with the Mouse) (WDR 1971-2014)⁶ were used. In every clip, a mouse, an elephant or a duck is performing a motion event that the participants (English and German speaking) could encode with a particle verb (e.g., rolling down). Furthermore, clips from the *Tomato Man Movies* (Özyürek et al., 2001) were part of the stimuli set. Compared to previous studies on motion events, the participants were trained to use specific sentence structures and also the particle verbs describing the motion event were given. By doing so, it was possible to better control speech outcome. Furthermore, this enabled us to compare gesture use across different sentence structures.

In McNeill's tradition, the experimenter does not mention that the purpose of the study has anything to do with gestures (McNeill, 2005: for Methods of gesture recording, transcription and coding). The reason for deceiving the participants in terms of the study's purpose is to minimise the participants' awareness of their gesture production (Gullberg, 2010). Thus, the aim is to collect spontaneously produced co-speech gestures. In terms of statistical analysis these rather unguided elicitation tasks can be problematic. First, a long stimulus movie such as the *Sylvester and Tweety Bird* cartoon makes it difficult to control for the content of the retelling. Although such a study design enhances gesture production (McNeill, 2005), not all participants use gestures describing specific target events and some participants do not

⁶ We got permission from the WDR (Westdeutscher Rundfunk) to use clips from "Die Sendung mit der Maus" for this research project.

gesture at all. In studies where the focus is on specific constructions in speech or specific gestures such as motion event gestures, it can be particularly difficult to reach a sufficient sample size. Another issue arising from individual differences in gesture frequency is the unequal distribution of gestures across items (Gullberg, 2010). Possibly, that is one reason why a number of the studies listed in Table 2.1 (6 in total) did not include any statistical analysis but analysed their data qualitatively.

One option to control better for gesture production is to encourage participants to use gestures. This method has been used in studies investigating the role of gestures in problem solving tasks (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Chu & Kita, 2011; Novack et al., 2014) or in the production of metaphorical speech (Argyriou, Mohr, & Kita, 2017). In speech-gesture production studies where motion events were the test domain, this method has rarely been used. For co-speech gesture studies, only Özçalışkan, Lucero, and Goldin-Meadow (2016, p. 12) instructed their participants to describe the stimuli scene “while using their hands as naturally as possible”. In another study on motion event gestures, Mol and Kita (2012) told the participants which gestures they have to use during their retellings of the Tomato Man Movies. With this study design they were able to test whether a given gestural content (verb-framed vs. satellite-framed gesture pattern) has an influence on how the participants package the motion event syntactically (one clause structure vs. two clause structure).

To sum up, most of the studies on motion event gestures are using a quasi-experimental method and do not mention gestures in the participants’ instructions. On the other hand there are some studies that use more experimental methods (e.g., Mol & Kita, 2012; Özçalışkan et al., 2016) where participants were encouraged to gesture. In the study design developed for the experiments presented in this thesis, the participants were also encouraged to use gestures. This decision was based on a pilot study for Experiment 1 which showed that without telling the

participants to use their hands, barely any gestures were produced. Consequently, this would have made a sound statistical analysis problematic.

2.4.2. *Statistics*

As already mentioned, in a number of studies on speech-gesture production listed in Table 2.1 the data was analysed qualitatively. If statistical analyses were used, studies tended to conduct ANOVAs and t-tests on proportions (e.g., Kita et al., 2007; Özyürek, Kita, Allen, Furman, & Brown, 2005) or they used their non-parametric counterparts (e.g., Brown & Chen, 2013). Only one study by Özçalışkan et al. (2016) used mixed effects statistical model.

This statistical method is also used in this thesis since it offers numerous advantages for analysing gesture data. In a nutshell, in mixed effects models subject and item can be included as random effects in a single analysis. Having multiple random factors in the same analysis contrasts with traditional repeated-measures ANOVAs where subject and item have to be tested individually (Quené & van den Bergh, 2008). In this regard mixed models are more powerful since they enable researchers to generalise across both items and subjects. Apart from combining by-subject and by-item within one analysis, mixed effects models have further advantages for analysing gesture data.

One advantage of mixed effects models is how missing data is treated. If the missing cells occur randomly, mixed models are robust against these missing cells (Quené & van den Bergh, 2008). As mentioned above, in gesture studies it is very difficult to control for gesture use (e.g. participants gesture rate differs or participants do not produce gestures which are relevant for research question) leading to an increased number of missing values. Thus, a statistical method that tolerates these missing data points is appropriate for gesture research.

Another advantage of mixed models is that categorical data can be included in the model as explanatory variables (i.e., predictors). Hence, we do not have to run ANOVAs on

proportions which despite applying the arcsine-square-root transformation are prone to Type I and Type II errors (Jaeger, 2008). A detailed explanation of how models were fitted in this thesis will be given in the relevant methods chapters.

Table 2.1. Summary of studies on speech-gesture production using motion events as test domain.

Study	Languages	Stimuli	Methods	Analysis	Instructions (gestures mentioned?)	Temporal Coordination	Semantic Coordination
Akhavan et al. (2017)	Farsi	short movie clips, depicting different motion events	elicitation task	quantitative	No	-	clause-level analysis, syntactic packaging
Brown and Chen (2013)	Mandarin-Chinese English Japanese	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task (McNeill (1992))	quantitative	No	-	v-framed vs. s-framed languages, clause-level analysis
Brown and Gullberg (2008)	Japanese-English English Japanese	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task (McNeill (1992))	quantitative	No	-	v-framed vs. s-framed languages, clause-level analysis, manner expression in gesture
Brown and Gullberg (2010)	Japanese English Japanese-English	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task (McNeill (1992))	quantitative	No	-	v-framed vs. s-framed languages, expressions of path in L1-L2
Brown (2008)	Japanese English Japanese-English	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task (McNeill (1992))	quantitative	No	-	gesture view-point in the description of motion
Brown (2015)	Mandarin Mandarin-English Japanese-English	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task (McNeill (1992))	quantitative	No	-	v-framed vs. s-framed languages, construal of manner in speech/gesture in L1-L2
Choi and Lantolf (2008)	Korean-English English-Korean Korean	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task	qualitative	No	v-framed vs. s-framed languages in terms of synchronisation in L1-L2	v-framed vs. s-framed languages, semantic coordination of gesture-speech in L1-L2
Chui (2009)	Chinese	episode of the Mickey Mouse and Friends series	narrative retelling task	qualitative	not mentioned in the paper	synchronisation patterns in Chinese discourse	
Chui (2012)	Chinese	episode of the Mickey Mouse and Friends series & free conversation	narrative retelling task, daily face-to-face conversations	qualitative	No	-	motion event gestures in Chinese discourse

Duncan (2002)	English Mandarin	three elicitation tasks: 1. Sylvester and Tweety Bird cartoon (Canary Row), 2. action sequences involving small plastic characters or inanimate objects (1-1.5 sec.) (Supalla et al., n.d.) 3. early feature film by Alfred Hitchcock	narrative retelling task, elicitation task	qualitative	No	gesture duration in relation to the speech element the gesture synchronises with	influence of verb aspect in discourse context on gestural content and gesture duration
Kellerman and van Hoof (2003)	English Dutch-English Spanish-English	The Frog Story (static picture story)	retelling the picture story	qualitative	No	v-framed vs. s-framed languages in terms of synchronisation in L1-L2	v-framed vs. s-framed languages, semantic coordination of gesture-speech in L1-L2
Kita and Özyürek (2003)	English Turkish Japanese	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task	quantitative	No	-	v-framed vs. s-framed languages, clause-level analysis, semantic coordination of gesture-speech
Kita et al. (2007)	English	Tomato Man Movies	elicitation task	quantitative	not mentioned in the paper	-	v-framed vs. s-framed constructions within a language, clause-level analysis, semantic coordination of gesture-speech
Mol and Kita (2012)	Dutch	Tomato Man Movies	elicitation task	quantitative	participants were told to use specific gestures during their retellings: Condition 1: manner-path conflated gestures Condition 2: one manner	-	testing influence of gesture on speech production

					and one path gesture		
Negueruela et al. (2004)	English Spanish English-Spanish Spanish-English	The Frog Story (static picture story)	retelling the picture story	qualitative	No	v-framed vs. s-framed languages in terms of synchronisation in L1- L2	v-framed vs. s-framed languages, semantic coordination of gesture-speech in L1-L2
Özçalışkan (2016)	English Turkish Turkish-English	animated motion clips	elicitation task	quantitative	condition 1: No instructions on gestures Condition 2: Just gestures no speech	-	v-framed vs. s-framed languages, (co-speech gesture vs. silent gesture)
Özçalışkan et al. (2016)	English Turkish	three dimensional scenes that depict motion along three different types of paths	elicitation task	generalized linear mixed-effect modelling	condition 1: use your hands as naturally as possible while speaking (co-speech gestures) Condition 2: use only your hands without any speech (silent gestures)	-	v-framed vs. s-framed languages, (co-speech gesture vs. silent gesture)
Özyürek et al. (2005)	Turkish English	Tomato Man Movies	elicitation task (if participants did not mention the target event, they were encouraged to do so by a question about manner/path component)	quantitative	not mentioned in the paper	synchronisation pattern across languages.	syntactical packaging and gestural content
Stam (2006)	Spanish English Spanish-English	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task	quantitative	not mentioned in the paper	gesture speech co-occurrence in L1-L2	gestural content in L1-L2; expression of path in gesture and speech

Stam (2015)	Spanish-English	Sylvester and Tweety Bird cartoon (Canary Row)	narrative retelling task	qualitative	not mentioned in the paper	-	gesture speech coordination L1-L2; longitudinal case study
Wessel-Tolvig and Paggio (2016)	Danish Italian	Tomato Man Movies and additional clips similar to the Tomato Man Movies	elicitation task	quantitative	not mentioned in the paper	co-expressivity of speech-gesture within a clause	syntactic packaging and gesture-speech coordination (verb/satellite-framed constructions)

Chapter 3

Information Packaging in Speech Shapes Information Packing in Gesture: The Role of Speech Planning Units in the Coordination of Speech-Gesture Production.

Preliminary results of Experiment 1 presented in this chapter have been published in the following Conference Proceedings:

Fritz I., Kita S., Littlemore J., & Krott A. (2015) The Influence of Clause Structure on Gestural Depiction of Motion Events. In G. Ferré, & M. Tutton (Eds.), Proceedings of the 4th GESPIN - Gesture & Speech in Interaction Conference. Nantes: Université of Nantes, 113-117.

3.1. Abstract

Linguistic encoding influences gestural manner and path depiction of motion events. Gestures depict manner and path of motion events differently across languages, either conflating or separating manner and path, depending on whether manner and path are linguistically encoded within one clause (e.g., “rolling down”) or multiple clauses (e.g., “descends as it rolls”) respectively. However, it is unclear whether such gestural differences are affected by how speech packages information into planning units or by the way information is lexicalised (as verb plus particle or as two verbs). In two experiments, we manipulated the linguistic encoding of motion events in either one or two planning units while lexicalisation patterns were kept constant (i.e., verb plus particle). It was found that separating manner (verb) and path (particle) into different planning units also increased gestural manner and path separation. Thus, lexicalisation patterns do not drive gestural depiction of motion events. Rather gestures are shaped online by how speakers package information into planning units in speech production.

Keywords: gesture, motion events, syntax, planning units, event conceptualization

3.2. Introduction

When we speak, we often spontaneously produce gestures. Gestures are tightly linked to how we encode information linguistically at the temporal and semantic levels (McNeill, 1992, 2005). In terms of synchronisation, gestures co-occur with the element in speech that is closest to the gesture's content, often initiated before their semantic affiliates (McNeill, 1992; Morrel-Samuels & Krauss, 1992; Schegloff, 1984). From a semantic perspective speech-gesture coordination is reflected by linguistic choices. This coordination is evident on both a lexical and structural level. For instance, gestures have been found to adapt to fine-grained differences in verb semantics (Gullberg, 2011; Gullberg & Narasimhan, 2010; Kita & Özyürek, 2003), and it has been suggested that the content of a gesture is linked to how clauses package information in speech (Kita, 2000; Kita & Özyürek, 2003).

Evidence for the influence of linguistic packaging on gestural content stems from cross-linguistic studies on motion event gestures (Kita & Özyürek, 2003; Özçalışkan, 2016; Özyürek et al., 2008; Özyürek et al., 2005; Wessel-Tolvig & Paggio, 2016). Motion events are linguistically encoded differently across languages. Based on Talmy's typology (2000), languages generally fall into two different categories. In so-called satellite-framed languages (e.g., German and English) the manner component of a motion event is usually encoded within the verb, while the path component is encoded in a "satellite" (a particle or an affix). Both components together often form a so-called particle verb (e.g., "to roll down" or "to climb up"). In verb-framed languages (e.g., Spanish and Japanese), path is encoded in the verb, while manner, if encoded in speech at all, is encoded in a phrase, gerund or in a separate clause (e.g., "he descends the hill as he rolls").

Gestures that accompany descriptions of motion depict manner and path in different ways in satellite-framed and verb-framed languages. In satellite-framed languages, where the motion event is linguistically encoded within one clause, manner and path tend to be conflated in one

gesture. In contrast, in verb-framed languages, where manner and path are encoded in two separate clauses, speakers also tend to separate manner and path gesturally (Kita & Özyürek, 2003; Kita et al., 2007; Özçalışkan, 2016; Özçalışkan et al., 2016; Özyürek et al., 2008; Özyürek et al., 2005; Wessel-Tolvig & Paggio, 2016). Thus, linguistic encoding in speech influences gestural content; but what is the relevant linguistic level?

These cross-linguistically varying gestural patterns may stem from differences in information packaging during speech production planning (Kita & Özyürek, 2003). We call this view the Planning Unit Account. More specifically, when manner and path are linguistically encoded in a single planning unit for speech production, a single gesture expresses both manner and path. When manner and path are linguistically encoded in two planning units, two gestures separately express manner and path. Because a clause is a good proxy for planning units in speech production (Bock, 1982; Levelt, 1989) clausal packaging of manner and path (one clause in satellite-framed languages, two clauses in verb-framed languages) is related to gestural packaging of manner and path. However, Kita and Özyürek (2003, p. 17) do not claim that gesture-speech coordination is bound to a clausal scope. Rather they argue that gesture-speech coordination is linked to a more general processing unit⁷ defined as a unit that “corresponds to what can be processed within one processing cycle for the formulation of speech”.

Alternatively, the cross-linguistically varying gestural patterns might stem from differences in motion event conceptualisation based on lexicalisation of motion concepts and its implication for clausal structure. We call this view the Lexicalisation Account. This hypothesis is based on the argument that clauses are not just important units for speech production but also “conceptual units” (Pawley, 1987, 2010). For example, Pawley (1987) points out that a “conceptual event” is constituted of an “event classifier”, which usually forms

⁷ In this thesis the term “planning unit” will be used synonymously to Kita and Özyürek’s (2003) “processing unit”.

one grammatical clause. Hence, this view can also account for Kita and Özyürek's (2003) cross-linguistic finding. According to this account, satellite-framed languages, which encode manner and path in a single clause, represent manner and path within a single conceptual unit. In contrast, verb-framed languages represent manner and path in two separate conceptual units. What is expressed in a single conceptual unit is expressed as a single gesture. The influence of lexicalisation patterns on motion event conceptualisation is also in line with Slobin's (2003, 2006) thinking-for-speaking hypothesis, which states that during speech production we have to filter our thoughts through linguistic encoding possibilities. Hence, lexicalisation patterns guide the speaker's attention to different aspects of the motion event (Slobin, 2000). Slobin (2000, 2003) argues that due to the obligatory encoding of manner in satellite-framed languages (within the main verb) in combination with the path component which is governed by the verb (i.e., particle verbs), speakers of these languages tend to perceive motion events as a "single conceptual event". Since it is not obligatory to encode manner linguistically in verb-framed languages, speakers of these languages do not perceive manner as inherent to the motion event. Manner is rather perceived as an activity that accompanies the path element of the motion event which is encoded in the main verb (e.g., exit, enter). These differences in conceptualisation would also explain the tendency of manner and path conflated gestures in satellite-framed languages and manner and path separated gestures in verb-framed languages. In sum, the crucial difference between the Planning Unit Account and the Lexicalisation Account is the type of unit (i.e., conceptual unit vs. planning unit) that shapes event conceptualisation which further drives gestural depiction of events. The example below (taken from Kellerman & van Hoof, 2003, p. 267) illustrates this difference, i.e. in the example the motion event consists of only one conceptual unit because it is encoded with one verb only (Pawley, 2010). However, the verb and the particle presumably fall into two separate planning units because the complexity

of the clause (i.e. verb and particle are separated by a number of words) makes the processing of the whole motion event within a single planning unit unlikely (see Ferreira, 1991; V. Wagner, Jescheniak, & Schriefers, 2010 for sentence complexity and advanced planning).

- (1) Op een gegeven moment gaat 't hondje **tegen** die groene boom die
 On a given moment goes the little dog **against** that green tree which
 daar groeit staan **springen**.
 there growing stand **jumping**.
 At a certain point the dog starts **jumping at** the green tree growing there.

Importantly, Kita et al. (2007) found that linguistic encoding only has an online effect on motion event conceptualisation and that gestural content is not bound to a habitual way of thinking based on how a language predominantly encodes motion events (satellite-framed versus verb-framed construction). In their study, Kita et al. (2007) compared gestures accompanying the two types of constructions within English: a satellite-framed construction (one verb framing, e.g. “he rolled down the hill”) or a verb-framed construction (two verb framing, e.g., “he went down as he rolled”). If habitual (dominant) language-specific event conceptualisation shaped motion event gestures, conflated manner and path gestures were expected regardless of the construction type. However, in Kita et al.’s study (2007) the participants’ gestures differed between satellite-framed and verb-framed constructions. When participants used satellite-framed constructions they accompanied speech with the expected conflated manner and path gesture. But when participants used verb-framed constructions manner and path were not only linguistically but also gesturally separated. Furthermore, essentially the same effect of construction types on gesture was also found in Dutch, another satellite-framed language (Mol & Kita, 2012). Hence, these studies suggest that gestures are shaped during speech production based on on-line linguistic choices and not on habitual language-specific event conceptualisations. Thus, conceptual events, which the Lexicalization

Account associates with gestural information packaging, must be generated online at the moment of speaking.

One shortcoming of studies on motion events so far is that they could not disentangle whether differences in gestural patterns stem from differences in how speech is packaged into planning units or differences in lexicalisation patterns (what information is encoded in a clause). Hence, different accounts could explain the gestural differences between verb-framed and satellite-framed constructions. Thus, the main aim of the present study is to provide unambiguous evidence for the Planning Unit Account. To this end, we manipulated the linguistic distance between manner and path components while keeping the lexicalisation pattern constant.

In Experiment 1 we tested whether increasing the linguistic distance between manner and path elements within the same clause of a satellite-framed construction can break up the planning unit into a manner and a path component, and consequently separate manner and path into two different gestures. Crucially, we asked German speakers to insert a sub-clause (“as seen in the video”) between manner and path elements, which should make speakers plan manner and path in two different planning units.

It is possible in German to insert extra linguistic elements between manner and path elements. German particle verbs (e.g. “hinunterrollen” – “to down-roll”) can be linguistically combined into one word or split up into two (potentially) distant words, depending, among other factors, on the clause type (main clause versus subordinate clause). German main clauses have an S-V-O structure where the verb always has to be placed in the second position of the clause and the particle comes in the final position. As seen in (2), the verb (*klettert*, “climbs”) and the particle (*hinauf*, “up”) can be separated by inserting elements such as prepositional

phrases, direct objects (*einen Regenbogen*, “a rainbow”) or even whole clauses (*wie im Video gesehen*, “as seen in the video”).

- (2) Der Elefant **klettert** wie im Video gesehen einen Regenbogen **hinauf**.
 The elephant **climbs** how in video seen a rainbow **up**.
 The elephant is climbing up the rainbow, as seen in the video.

In German subordinate clauses, the verb and the particle are in reverse order compared to main clauses. Importantly, these two elements are contracted (e.g., *hinaufklettert*, “up-climb”) in the final position of the clause (3).

- (3) Ich sehe im Video, dass der Elefant einen Regenbogen **hinaufklettert**.
 I see in the video that the elephant a rainbow **up-climbs**.
 I can see in the video that the elephant is climbing up a rainbow.

As a control, we also tested English native speakers in Experiment 1. Just like in German, English motion events are linguistically encoded with a particle verb. However, manipulating the clause type does neither change the distance nor the word order of the particle and the verb (4 and 5). Thus, testing English native speakers allowed us to keep planning units constant across clause types compared to our planning unit manipulation in German (i.e., two planning units in main clauses vs. one planning unit in subordinate clauses).

- (4) The elephant is **climbing up** the rainbow, as seen in the video. (Main Clause)
 (5) I can see in the video that the elephant is **climbing up** the rainbow. (Subordinate Clause)

The Lexicalisation Account and Planning Unit Account predict different patterns of results across the constructions in (2) - (5). First, the Lexicalisation Account (Pawley, 1987, 2010; Slobin, 2000, 2003) predicts that gestures should be shaped by what information is encoded within a single clause (one clause construction versus two clause construction). Since German and English are both satellite-framed languages where manner and path are encoded within the same clause (one clause construction), very similar motion event gestures should occur in both languages and across clause types. Second, the Planning Unit Account predicts

that co-speech gestures should be shaped by how information is packaged into planning units in speech. In German main clauses, manner and path elements are separated from each other, thus it is more likely that manner and path are encoded in different planning units. Consequently, this account predicts that gesture separates manner and path more often in German main clauses compared to German subordinate clauses, where the particle verb is linguistically contracted. In English, by contrast, gestures should be similar for both main and subordinate clauses.

In Experiment 2, we sought further evidence for the Planning Unit account by inserting different linguistic elements between manner and path expressions to see if they influence gestural packaging of information differently. More specifically, we inserted either a clause (“as seen in the video”) or a phrase (“in this short video”), while controlling for the length of the inserted element in terms of syllables. As mentioned above, clauses are assumed to constitute planning units (Bock, 1982; Levelt, 1989). Thus, manner and path are more likely to be in different planning units when a clause is inserted than when a phrase is inserted. According to the Planning Unit Account, we should find a higher rate of gestural separation of manner and path for inserted clauses than inserted phrases.

3.3. Methods – Experiment 1

3.3.1. Participants

25 native English speakers and 23 native German speakers took part in the study. Participants either received course credits or a £3 voucher for participation. They all gave written consent to have their data included in the study. Two participants (one English and one German) were excluded from the coding and analysis because they did not use any iconic gestures which depicted the target motion events. Another two participants were excluded because their sentence structure differed from the target sentence structure in most of the

responses. More specifically, one English participant used additional embedded clauses apart from the given ones (“as seen in the video”) which might have influenced gesture production. One German participant was excluded because she used the past perfect tense in the main clause condition where the particle verb was not split but contracted in the final position of the clause, similar to the subordinate condition. The language and gestures produced by the remaining 23 English and 21 German speakers were coded. Because only participants that used a reasonable amount of both separated gestures and/or conflated gestures can inform us about the effect of syntactic structure on gesture separation, we included only participants who produced at least two gestures per experimental condition which depicted both manner and path gesturally (either separated or conflated). Based on these exclusion criteria, 17 English participants (average age 21.2, SD = 3.9) and 15 German participants (average age 23.3, SD = 2.6) were included in the analyses.

3.3.2. *Material*

13 short cartoons taken from the German children’s series “Die Sendung mit der Maus” (“The programme with the mouse”) were used as stimuli (WDR 1974-2015). The cartoon sequences ranged from 3-8 seconds and all trials included a character (mouse, duck or elephant) which performed a motion event. To control speech output, participants were given a particle verb to describe the target motion event, for instance “roll into” (hineinrollen) and “spin up” (hinaufdrehen). All 13 particle verbs are listed in Appendix 1.

3.3.3. *Design*

The experiment had a 2x2 design with Language (English vs. German; between participant) and Clause Type (main vs. subordinate; within participant) as independent variables. The dependent variable was a binary variable as to whether gesture depicted manner

and path in two separate gestures or in a single conflated gesture for a given response to a stimulus video (see the Coding and Analysis subsection for more details).

The clause type was manipulated so that the distance between the manner element and the path element in a sentence is different in German (but not in English). In the subordinate clause condition participants were instructed to begin their retellings with the element “I can see in the video that” (German: “Ich sehe im Video, dass”) (see Examples in 6 and 7). Initiating a sentence with this clause forced the participants to continue with a subordinate clause. In both English and German, the manner element and the path element are adjacent with each other.

(6) German Subordinate Clause Condition

Ich sehe im Video, dass der Elefant in eine Sandgrube **hineinrollt**.
I see in the video that the elephant in a sandpit **in-rolls**.

(7) English Subordinate Clause Condition

I can see in the video that the elephant is **rolling into** a sandpit.

In the main clause condition, German participants were instructed to insert the clause “wie im Video gesehen” (“as seen in the video”) between verb and particle as in (8). To ensure that the participants produced this grammatical structure, they were instructed to start the sentence with the subject (the mouse, the elephant or the duck), followed by the verb in second position. Unlike in the subordinate condition, the manner element and the path element were separated by other words. The total distance between verb and satellite could vary, depending on how many other elements the participants chose to include (e.g., “with an umbrella”). Although inserting a clause between verb and particle is possible in English with emphatic stress on the particle (e.g., down), it is a more marked construction in English than it is in German. Moreover, for the purpose of the study, i.e., manipulating planning units in one group (German) whilst keeping them constant in a control group (English), English participants were instructed to place the clause “as seen in the video” at the end of their sentence as in (9). Thus,

we could keep the speech output and the overall complexity of the sentences in both languages as similar as possible.

(8) German Main Clause Condition

Die Maus **schwebt**, wie im Video gesehen mit einem Regenschirm in den Pool **hinunter**.

The mouse **floats** how in the video seen with an umbrella in the pool **down**.

(9) English Main Clause Condition

The mouse is **floating down** into the pool, as seen in the video.

3.3.4. *Procedure*

Participants were tested in a lab at the University of Birmingham. They were told that the purpose of the study was to investigate how different sentence structures of a language influence our speech production in narrations. They were instructed to retell the cartoon clips within a single sentence using either a main or a subordinate clause construction. The participants retold the cartoons to a third person (confederate) in the room who was not able to see the cartoons. The participants were told that the confederate's task was to write down one keyword for each clip, which she/he thinks is the most vital information of the story. Participants' responses were video and audio recorded for later analyses.

The experiment was set up in a PowerPoint presentation and presented to the participants on a laptop. Main clause and subordinate clause conditions were blocked and the order of conditions was counter-balanced across participants. The 13 stimuli were repeated in the two conditions. The experimenter explained the task to the participants with an example stimulus (see Figure 3.1.) followed by a practice stimulus. Each condition was explained and introduced with the same example and practice stimulus. The second condition was not explained until the first condition was completed. Each trial started with a slide showing the particle verb and either the initial main clause in the subordinate condition or the embedded clause (German)/final clause (English) in the main clause condition. This slide was displayed for three seconds before

the actual clip started. After the clip had been shown, the screen turned blank. In order to keep advanced sentence planning to a minimum, participants were told that they should start their retelling as soon as they saw the blank screen. For the example clip, an example answer was presented after the blank slide. This example answer illustrated the correct sentence structure to make it easier for the participants to re-produce it in their own retellings. Concerning the form of the verb (progressive form, tense), no limitations were given and the participants were told that it was up to them which form of the verb they used. Also, participants were instructed to use their hands while describing what the characters were doing, but it was not specified what type of hand movements they should produce. If participants asked how to use their hands or whether a certain type of gesture was correct, they were told that it was up to them what to do with their hands. If participants did not use their hands twice in a row, they were reminded to do so.

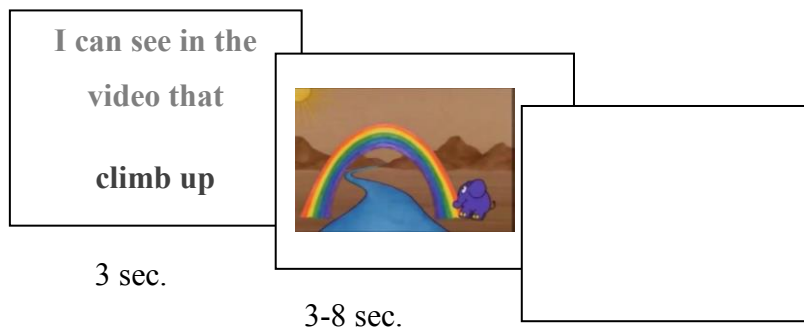


Figure 3.1. Example stimulus for the English Subordinate Clause condition. A fourth slide illustrated the sentence that the participants were expected to produce, i.e. I can see in the video that the elephant is climbing up the rainbow. Participants started to describe the video once the slide turned blank.

3.3.5. Data Coding and Analysis

The recordings were coded using the video annotator ELAN (Lausberg & Sloetjes, 2009). All speech was transcribed, but only responses where the participants produced the trained sentence structure were considered in the analyses. Hence, manner and path had to be

linguistically encoded within one clause for the utterance to be included. The responses in which participants failed to follow the instructions (e.g., they forgot to include the given clause “as seen in the video” or they produced a gesture after speech) were excluded from the analyses. Participants sometimes used other verbs than the ones presented on the slides. We only included responses with particle verbs into the analyses and only if they were semantically similar to the given verb. These responses contained different particles (e.g. “emporkriechen” – “crawl upwards”; instead of “hervorkriechen” “crawl out”) or different manner verbs (“climb out” instead of “crawl out”). For three of the 13 cartoon clips (slide down, jump over, jump into), it turned out that participants found it very difficult to gesturally express manner without using whole body gestures. Participants seemed to avoid such whole body gestures, meaning that they usually produced path gestures only. We therefore excluded these items in any analyses. They were in the stimulus set because they are useful for the research questions investigated in Chapter 4.

The gestural coding of motion events was based on the “Cross-linguistic Motion Event Project” coding manual (used in: Kita et al., 2007; Özyürek et al., 2008) and was adapted and elaborated for the stimuli used in the current experiment. Only strokes depicting the target motion event, i.e. the gestural depiction of the given particle verb of each trial, were coded (Kita, van Gijn, & van der Hulst, 1997; McNeill, 1992). In a first step all target event gestures were classified either as manner, path, conflated, or hybrid. Path gestures depict only the direction of the event (e.g., for the motion “to float down” this might involve a downward movement with (an) open palm(s) but without any movements to the left or right which would indicate manner). Manner gestures depict only the manner aspect of the motion event (e.g., for the motion “to climb up” this might involve the participant opening and closing their palm(s) without moving their arms upwards). Conflated gestures depicted motion and manner of the

motion event in a single gesture (e.g., for the motion “to roll into”, rotating one’s wrist(s) with a simultaneous change of location away from the body). Hybrid gestures are combinations of path or manner gestures with a conflated gesture, produced within a single stroke.

Next, we classified each response to each stimulus video into three types, based on the types of gestures produced: Separated Responses, Conflated Responses, and Singleton Responses. In Separated Responses, both manner and path were expressed gesturally and in a separate fashion. This included a) responses with either one manner and one path gesture and b) responses with a conflated gesture plus a manner or a path gesture, either produced separately or combined in hybrid gestures. Conflated Responses contained conflated gestures only. Finally, Singleton Responses included either manner or path gestures (but not both). These typically contained either a single manner gesture or a single path gesture, but also cases where participants produced two manner only or two path only gestures in one response. To ensure coding reliability, a second coder blind to the research question, coded 19 % of the trials (including responses to all 13 stimuli movies). Only responses including a gesture were considered. The second coder was trained to identify target event gestures and to classify them as either Manner, Path, Conflated or Hybrid Gesture. Based on these annotations, interrater reliability was calculated on whether the gesture(s) identified for each trial fell into the three categories (Singleton, Conflation, Separation) used for the analyses. The two coders agreed in 88 % of the responses (Cohen’s $\kappa = .769$, $p < .001$).

Since we were interested in whether participants conflate or separate manner and path gesturally within one response, Singleton Responses were excluded from the analyses. Singleton Responses comprised the following mean percentages: 52 % in German Main Clauses, 58 % in German Subordinate Clauses, 67 % in English Main Clauses, 63 % in English Subordinate Clauses. Any responses which could not clearly be categorised in any of the three

categories (Conflated Responses, Separated Responses or Singleton Responses) were excluded from the analysis. This made up 1.69 % of all data collected from the participants included in the analyses (15 German participants, 17 English participants). Furthermore, responses of the participants included in the analyses that did not contain any target event gesture were excluded; the percentage of such trials were as follows: 7 % in German and 6 % in English.

The resulting data were analysed by fitting linear generalised mixed effect models in RStudio (R Core Team, 2014) using the `glmer` function (Bates, Mächler, Bolker, & Walker, 2015) We started by fitting a full model with Gestural Depiction (Separated Responses vs. Conflated Responses) as dependent variable, Language (German, English) and Clause Type (Main Clause, Subordinate Clause) plus their interaction as fixed factors and items and subjects as random factors. Parameter fitting of models with random slopes for Clause Type did not converge for any of our analyses. But instead of dropping random slopes completely, we followed Bates (2009) and chose a model structure that, like one with random slopes, included random intercepts for Clause Type and allowed for an additive shift for each combination of subject/item and Clause Type. However, in contrast to a model with random slopes, this model assumes that the co-variances among the combinations are constant (Bates, 2009). Such a model has fewer variance/covariance parameters to be estimated and can therefore be fitted to smaller datasets (for a discussion on model fitting convergence see Chapter 4, procedure section: 4.4.3).

In one case this simplified model did not converge (see below). Here, we dropped this more complex random intercept structure for item but kept it for subject. The significance of a factor or an interaction between factors was determined by comparing models with and without these factors/interactions, using a maximum likelihood method (i.e., ANOVAs). We will report the chi-square statistics, degree of freedom and p-value for each model comparison.

3.4. Results

Results are summarised in Figure 3.2 and Table 3.1. First we examined the interaction between Language and Clause Type. Model comparison yielded a significant interaction between Language and Clause Type ($\chi^2 = 4,72$, $df = 1$, $p = .03$). To test whether there was an effect of Clause Type for both participant groups, the English and German data were analysed separately. For the English data, model convergence was reached by dropping the more complex random intercept described above for items (but keeping them for subjects) (see Table 3.2). In German, we found a significant effect of Clause Type ($\chi^2 = 8.31$, $df = 1$, $p = .004$) (see Table 3.3). The mean proportions of Separated Responses (i.e., responses with separated gestures) were higher in Main Clauses, i.e. when manner and path were separated linguistically, than in the Subordinate Clauses, i.e. when manner and path were linguistically expressed together. In English where linguistic encoding of the manner and path element did not change across clause types, no effect of Clause Type was found ($\chi^2 = 0.14$, $df = 1$, $p = .710$).

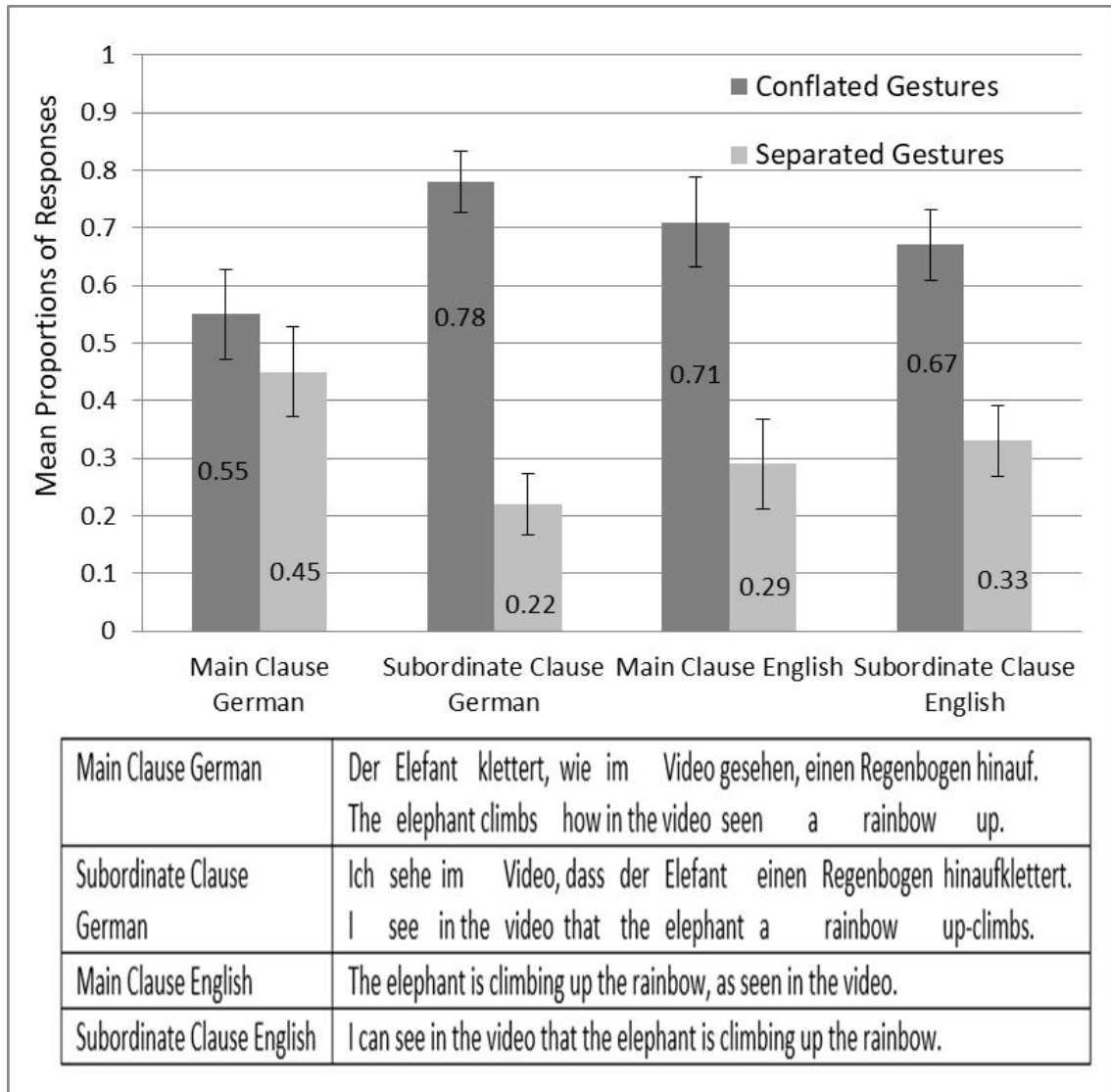


Figure 3.2. Mean Proportions of Separate Response (i.e., responses in which gesture separated Manner and/or Path) and Conflated Only Responses (i.e., responses that include only gestures that conflated Manner and Path) in Main Clauses versus Subordinate Clause in English and German in Experiment 1. Error bars represent standard errors. The bottom of the figure shows an example sentence of each of the four conditions. For the German conditions, a word-by-word translation is provided.

Table 3.1. Summary of the fixed effects of the full mixed logit model for gestural depiction in Experiment 1

Predictor	Coefficient	SE	Wald Z	p
Intercept	-0.9301	0.4314	-2.156	.03109 *
Clause Type	0.1294	0.4491	0.288	.77324
Language	1.1460	0.4158	2.756	.00586 **
Clause Type * Language	-1.2814	0.5910	-2.168	.03014 *

Note: N = 264, log-likelihood = - 150.7. **<.01; *<.05.

Table 3.2. Summary of the fixed effects in the mixed logit models in English for gestural depiction in Experiment 1.

Predictor	Coefficient	SE	Wald Z	p
Intercept (English)	-1.0873	0.4412	-2.465	.0137 *
Clause Type (English)	0.1782	0.4926	0.362	.7175

Note: N = 132, log-likelihood = -76.7. * < .05.

Table 3.3. Summary of the fixed effects in the mixed logit models in German for gestural depiction in Experiment 1.

Predictor	Coefficient	SE	Wald Z	p
Intercept (German)	0.3465	0.5666	0.612	.54085
Clause Type (German)	-1.2822	0.4530	-2.830	.00465 **

Note: N = 132, log-likelihood = -73.8. ** < .01.

3.5. Discussion

The results show that clause type affected gestural depiction of manner and path of motion events in German but not in English. Gestures separated manner and path more often in German main clauses where manner and path are also linguistically separated by an inserted clause and other elements, compared to German subordinate clauses where manner and path are adjacent with each other (the path particle is a prefix to the manner verb). In English, where manner and path are always adjacent with each other independent of clause type, gestures separated manner and path equally often in the two clause types, and at about the same frequency as in German subordinate clauses. Because a clause is a good proxy for planning units in speech (Bock, 1982; Levelt, 1989), inserting a clause into German main clauses is likely to have separated the manner verb and the path particle into separate planning units (see V. Wagner et al., 2010 for evidence that the scope of planning changes when processing load increases). Thus, the findings in Experiment 1 support the Planning Unit Account; that is, gestures package information that is encoded within each planning unit for speech production.

These results do not support the Lexicalisation Account, which predicted no difference in the usage of gestures across clause types and languages. According to the Lexicalisation Account the lexical items which are used to describe an event influence how we conceptualise this event (Pawley, 1987, 2010; Slobin, 2000, 2003), and in turn how gesture depicts manner

and path. Because German and English express manner and path in a single clause (manner verb + path particle), Slobin (2000, p. 132) argues that speakers of these languages tend to conceptualise manner and path as a “single conceptual event”. The current study did not provide evidence that such conceptual events shape gestural expressions.

Though the current results support the Planning Unit Account, the manipulation of main vs. subordinate clauses in German may involve some confounding. First, it is possible that mere surface distance between the manner element and the path element may be causing separation of path and manner in gestures. Second, the main vs. subordinate clauses may have different degree of processing demand. Therefore, in Experiment 2, we focused on German and manipulated planning units all within the main clause that contains manner and path, while controlling for the distance between the manner element and the path element.

3.6. Experiment 2

We compared how often gesture separate manner and path in three constructions, which are all main clauses in German: a) inserting an embedded clause (“wie im Video gesehen” – “as seen in the video”) between the manner verb (after the subject NP) and the path particle (at the end of the sentence) in present tense, b) inserting a prepositional phrase (“in diesem kurzen Video” – “in this short video”) in present tense, c) without any inserted clause or phrase but in the present perfect tense, which would force the path particle to be prefixed to the manner verb at the end of the sentence (e.g., ”hinaufgeklettert” – “up-climbed”).

If a clause is a good proxy for a planning unit, the manner verb and the path particle are separated into two planning units in the Inserted Clause Condition but not in the Inserted Phrase condition and the Verb Final condition. Thus, the Planning Unit Account predicts that gestures should separate manner and path *more often* in the Inserted Clause Condition than the other two conditions. If mere distance between the manner verb and the path particle determines how

often gestures separate manner and path, then gestures should separate manner and path *less often* in the Verb Final Condition than the other two conditions

We also tested whether our manipulation of planning units was effective for the two key conditions, the Inserted Clause Condition and the Inserted Phrase Condition by examining frequencies of pauses. Previous studies have analysed the occurrence of pauses as indicators for syntactical planning units (Ford & Holmes, 1978; Goldman-Eisler, 1958, 1972). Although pauses tend to cluster around clause boundaries (Goldman-Eisler, 1972; Pawley & Syder, 1983), Ferreira (1991) found that when a clause is more complex, speakers pause more often within clauses. She concluded that if a clause gets too complex it has to be broken down into two separate “performance units”, which gives the speaker time to plan the next phrase. Hence, in order to detect processing units in speech, we coded all pauses in the responses in Experiment 2 (filled and unfilled, cf. Zellner, 1994).

3.7. Methods – Experiment 2

3.7.1. Participants

26 German native speakers took part in Experiment 2. Participants were tested at the Natural Media Lab at the RWTH Aachen University and at the University of Innsbruck. For participating in the study they received compensation in form of a €5 voucher. All participants gave written consent to have their data included in the study. Three participants were excluded because they had learned German Sign Language (Deutsche Gebärdensprache). Previous studies found that learning a sign language has an influence on co-speech gesture production in terms of gesture frequency (Casey, Emmorey, & Larrabee, 2012) and in terms of the production of signs while speaking (Casey & Emmorey, 2008; Casey et al., 2012). Another two participants were excluded because they did not use any iconic gestures depicting the target motion events, and one participant had to be excluded due to technical problems. The language and gestures

produced by the remaining 20 participants were coded for analyses. Like in Experiment 1, the analyses only included participants who produced at least two gestures per condition which depicted both manner and path (conflated or separated). Consequently, all analyses below include 15 participants (mean age = 27.2, years, SD = 3.5).

3.7.2. *Material*

We increased the number of experimental stimuli to 15 to minimise the number of participants excluded for lack of sufficient gesturing. In all experimental stimuli, manner and path could be easily gesturally separated. Ten clips were taken from the German children's series "Die Sendung mit der Maus" ("The programme with the mouse") (WDR 1971-2015). Nine of those had also been used in Experiment 1. In addition, five new experimental stimuli were taken from the "Tomato Man movies" (Özyürek, Kita, & Allen, 2001). An additional stimulus from The Programme with the Mouse movies ("slide down") was used as an example to explain the task, and two additional stimuli were used as practice clips ("climb up" and "ride around"). All particle verbs used in Experiment 2 are listed in Appendix 1.

3.7.3. *Design*

The experiment manipulated construction type (inserted clause, inserted phrase, verb final; within participant). See Examples 10-12. The dependent variable was a binary variable as to whether gesture depicted manner and path in two separate gestures or in a single conflated gesture for a given response to a stimulus video (as in Experiment 1).

The construction type manipulated the likely number of planning units: two planning units for the Inserted Clause Condition, one planning unit in the Inserted Phrase Condition and the Verb Final Condition. The construction type also manipulated the distance between the manner verb and the path particle: far in the Inserted Clause Condition and the Inserted Phrase Condition, and adjacent in the Verb Final Condition. The inserted clause and the inserted phrase

separated the manner verb and the path particle by the same number of syllables. The Verb Final Condition also included the same inserted phrase as in the Inserted Phrase Condition (i.e., “in this short video”) to keep the planning complexity equivalent. Note that the Inserted Clause Condition is the same as the Main Clause condition in Experiment 1.

3.7.4. Procedure

Participants came to the lab and they were told that the study was about sentence production. For Experiment 2, we decided to create a more natural speech production task. We therefore did not present the verb that participants had to use right before each video clip as in Experiment 1, but participants were familiarised with all of the stimuli clips together with the particle verbs to be used before the experimental task. This way, participants had to retrieve the particle verbs from their mental lexicon during the retelling of the cartoons as during natural speech production. In the familiarisation phase, stimuli were presented in a PowerPoint presentation. All participants saw the stimuli in the same order. Then, participants were tested to see if they had memorised the particle verbs to be used. For that all screenshots of the cartoon clips were shown to the participant on an A4 paper (see Appendix 2). The order of the screenshots differed from the order of the stimuli in the PowerPoint presentation. Participants were asked to produce for each screenshot the particle verb they had seen before. If a participant’s response was different from the expected one, the experimenter reminded them of the particle verb to be used before the participant continued with the next screenshot.

After this familiarisation phase, the participants were instructed as to what their retellings should look like. They were told that there would be three different ways of retelling the cartoons and that the first one will be introduced shortly. Firstly, the participants were shown how the stimuli were presented on the screen. A booklet was created to illustrate how the stimuli would be presented. They were told that every trial would start with a fixation cross for 1000

ms followed by the stimulus. Then certain linguistic elements would be shown for two seconds to ensure that the participant would use the correct phrase or clause. Note that the given elements occurred after instead of before the cartoon in Experiment 2 to reduce the error in participants' construction choice. To further control the linguistic outcome, the participants were told during the instructions how to refer to the characters (the Mouse, the Elephant, the Tomato and the Triangle). Participants were also instructed as to how to use the correct sentence structure and how to use their hands while retelling what the characters are doing. These instructions stayed the same as in Experiment 1.

Two practice trials followed the instructions to assure that the participants would be able produce the correct sentence structure and, if necessary, to enable the experimenter to repeat some of the instructions. Figure 3.3 shows the example stimulus in Experiment 2.

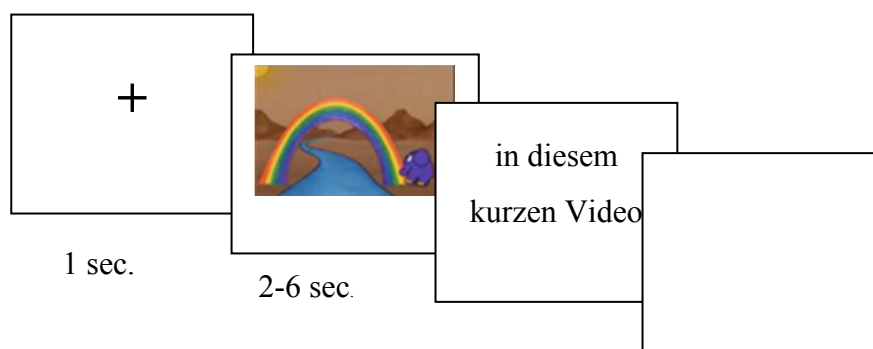


Figure 3.2. Example stimulus for the Inserted Clause Condition in Experiment 2

Stimuli were presented using E-Prime. Construction type conditions were blocked and instructions for every new condition were given after the previous condition was completed. The order of the three conditions was counterbalanced across participants and the trials within each block were presented in a randomised order. After the participants finished retelling one cartoon clip, the experimenter initiated the next trial with a button press. In this experiment, the experimenter was sitting opposite the participant and functioned as a listener.

Further cues were given to make it easier for participants to produce desired constructions. In the Verb Final condition, the present perfect tense was used. In German the present perfect tense is formed with the auxiliary “sein” (“to be”) or “haben” (“to have”) in its conjugated form (e.g. ist – 3rd Person Singular of “sein” – to be). Whether “sein” or “haben” is used depends on the particular verb. To help the participants, the correct auxiliary was displayed together with the given phrase that appeared after the cartoon clip (e.g. “ist in diesem kurzen Video” – “is in this short video”). The reflexive pronoun “sich” (oneself) was also given if it was obligatory to form a grammatically correct sentence (e.g. “hat sich in diesem kurzen Video” – “has itself in this short video”)

3.7.5. Data Coding and Analysis

As in Experiment 1, responses which could not be clearly categorised as either Conflated Response, Separated Response or Singleton Response, were not included in the analysis. This made up 2.5 % of the gestures produced by the 15 participants who were included in the analyses. Gesture and speech coding was done in the same way as in Experiment 1. Singleton Responses were again excluded from the analysis. Singleton Responses comprised the following mean percentages: 43 % in the Inserted Clause Condition, 45 % in the Inserted Phrase Condition, 46 % in the Verb Final Condition. Responses that did not include any gesture or no target-event gesture (i.e., gesture including manner or path) comprised 9 % of the whole dataset.

Inter-coder reliability was assessed as in Experiment 1. 19 % of the all trials which included the trained sentence structure and also included a gesture were considered. The second coder was trained to identify target event gestures and to classify them as either Manner, Path, Conflated or Hybrid Gesture. Based on these annotations, interrater reliability was calculated on whether the gesture(s) identified for each trial fell into the three categories (Conflated

Response, Separated Response, Singleton Response) used for the analyses. The two coders agreed in 87% of the responses (Cohen's $\kappa = .775, p < .001$).

For the pause analysis, only pauses that were longer than 200 ms (Smith & Wheeldon, 1999) and that were produced either before the inserted clause/phrase or after were considered. Only these two phrasal boundaries were considered because Ferreira (1991) found that pauses which presumably indicate processing boundaries usually do not violate major syntactic/semantic boundaries. Furthermore, pauses occurring within clauses (e.g., between article and object; e.g. “the [pause] rainbow”) are assumed to be caused by lexical retrieval difficulties (Kircher, Brammer, Levelt, Bartels, & McGuire, 2004; Levelt, 1983). For each response, we determined whether or not a pause, as defined above, was produced (a binary variable).

3.8. Results

We first tested whether the nature of the inserted element (clause vs. phrase) influenced the number of planning units for speech production. We fitted a generalised linear mixed effect model with Pauses (Pause, No Pause) as the dependent variable, Inserted Element as fixed factor (Inserted Clause, Inserted Phrase). Due to convergence problems we used the same random effect structure as in Experiment 1. To test for differences in the occurrence of pauses between the Inserted Clause Condition and the Inserted Phrase Condition, a model comparison yielded an overall effect of Inserted Element ($\chi^2 = 9.88$ $df = 1, p = .002$). Participants produced significantly more pauses in the inserted clause condition compared to the inserted phrase condition (Figure 3.4 and Table 3.4).

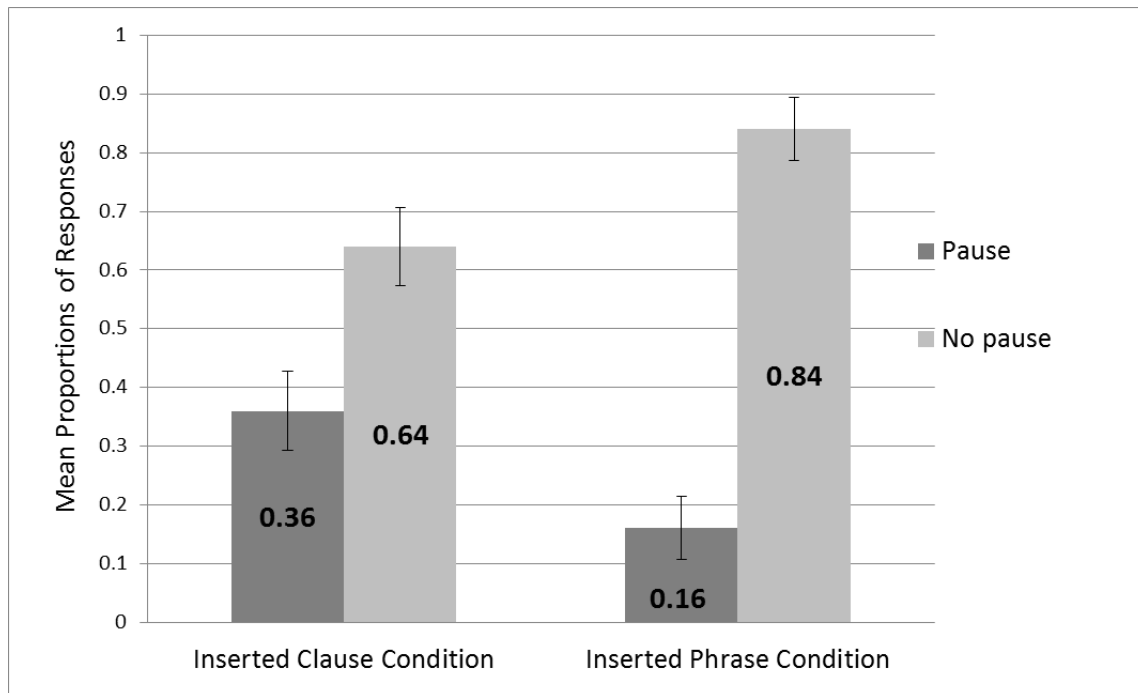


Figure 3.4. Mean proportions of responses with a pause produced either before or after the inserted element (phrase/clause) versus responses without any pauses in Experiment 2.

Table 3.4. Summary of the fixed effects in the mixed logit model for the occurrence of pauses in the Inserted Clause Condition versus the Inserted Phrase Condition in Experiment 2.

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Wald Z</i>	<i>p</i>
Intercept	-0.8093	0.3656	-2.213	.026876 *
Inserted Element	-1.4377	0.3949	-3.640	.000272 ***

Note: N = 212, log-likelihood = -103.2. * < .05; *** < .001

To test for effects of our manipulation Construction Type on gestural separation of manner and path, the same linear generalised mixed effect modelling as in Experiment 1 was applied, except that there was only one fixed factor (Construction Type) with three levels (Inserted Clause, Inserted Phrase, Verb Final). The results on gestural conflation and separation of manner and path are summarised in Figure 3.5 and Table 3.5. Model comparison yielded an overall effect of Construction Type ($\chi^2 = 7.83$, $df = 2$, $p = .020$). To compare three levels in Construction Type against each other, we initially set the Verb Final Condition as our baseline condition. To assess statistical difference between Inserted Clause and Inserted Phrase Conditions we refitted the same model with the Inserted Clause Condition as baseline. We

found that responses of the Inserted Clause Condition differed significantly from the Inserted Phrase Condition ($\beta = -0.85, p = .012$), and the Verb Final Condition ($\beta = 0.83, p = .015$), with more conflated gestures in the Inserted Clause Condition than in the two other conditions. No statistical difference was found between the Inserted Phrase Condition and the Verb Final Condition ($\beta = -0.016, p = .964$).

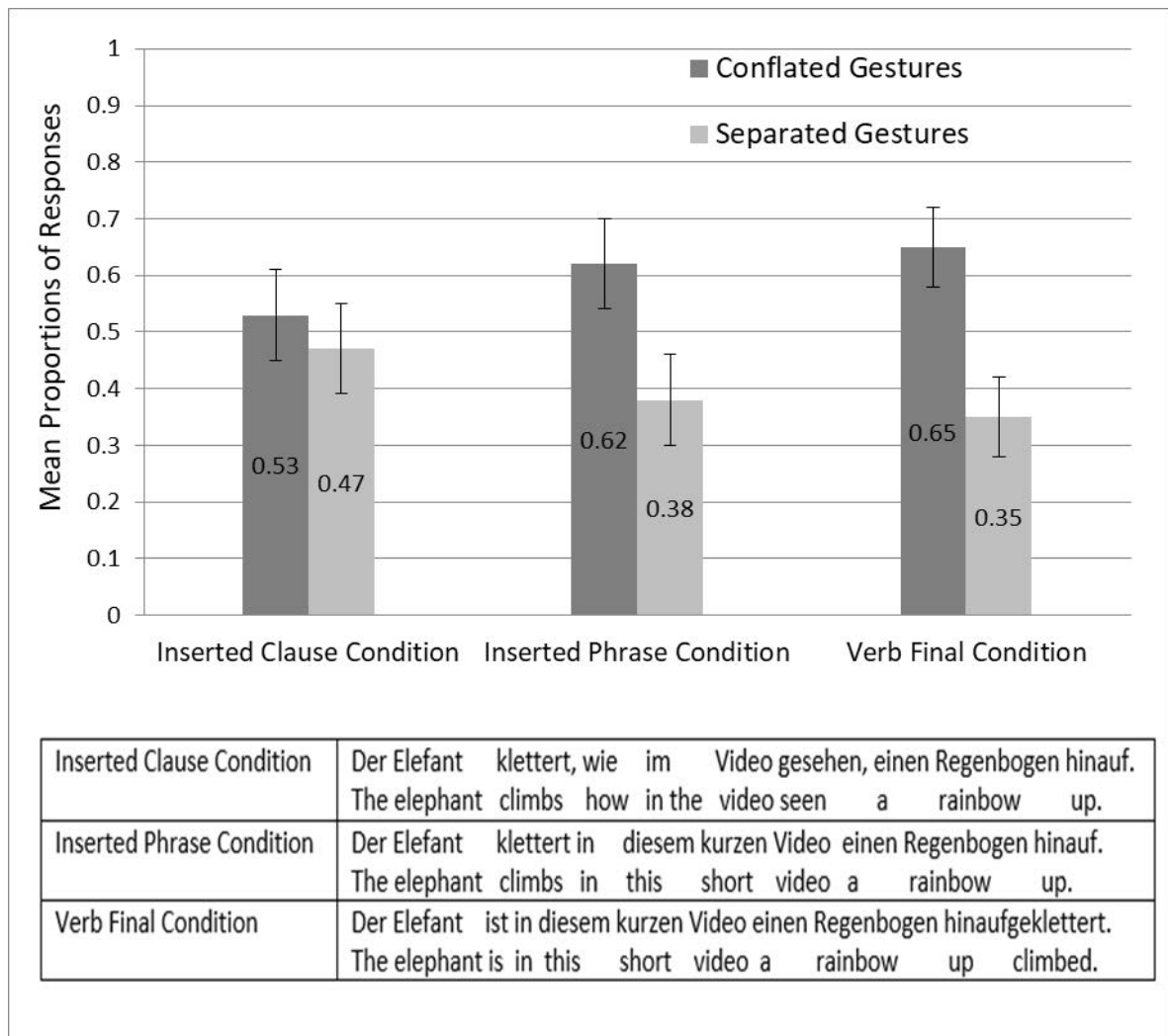


Figure 3.5. Mean Proportions of responses with Conflated Gestures and Separated Gestures across conditions of Experiment 2. Error bars represent standard errors. The bottom of the figure shows an example sentence of each of the three conditions including a word-by-word translation of the examples.

Table 3.5. Summary of the fixed effects (Construction Type) in the mixed logit model for gestural depiction in Experiment 2.

<i>Predictor (baseline Verb Final Condition)</i>	<i>Coefficient</i>	<i>SE</i>	<i>Wald Z</i>	<i>p</i>
Intercept	-0.81723	0.42983	-1.901	.0573
Inserted Phrase versus Verb Final	-0.01616	0.36110	-0.045	.9643
Inserted Clause versus Verb Final	0.83326	0.34219	2.435	.0149 *
<i>Predictor (baseline Inserted Clause Condition)</i>	<i>Coefficient</i>	<i>SE</i>	<i>Wald Z</i>	<i>p</i>
Intercept	0.01604	0.41390	0.039	.9691
Inserted Clause versus Inserted Phrase	-0.84939	0.33919	-2.504	.0123 *

Note: N = 311, log-likelihood = -175.1. *<.05.

Consistent with our assumption that a pause in a response indicates two separate planning units, when we replace Construction Type in the above analysis with Pause (present or absent), we obtained essentially the same result, i.e. separated gestures are more likely when there is a pause present. More specifically, we analysed only the data from the Inserted Clause Condition and Inserted Phrase Condition where pauses were coded. We used linear generalised mixed effect modelling to test whether gestural separation of manner and path is predicted by a fixed factor (Pause) with two levels (absent, present) and with the same random effect structure as in previous analysis but due to convergence issues random intercept-only structure for item and the more complex random-intercept structure for subject. Dropping the factor Pause from the model yielded a significant effect of Pause ($\chi^2 = 4.77$, $df = 1$, $p = .028$). Results are summarised in Table 3.6 and Figure 3.6.

When we put Pause and Construction Type and their interaction in a single model, it does not yield any significant results, which is not surprising as Pause and Construction Type are correlated with each other. Summary of a full model including the interaction between Pause and Construction Type are summarised in Table 3.7.

Table 3.6. Summary of the fixed effects (Pause) in the mixed logit model for gestural depiction in Experiment 2 (Inserted Clause and Inserted Phrase Condition only).

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Wald Z</i>	<i>p</i>
Intercept	-0.703	0.403	-1.745	.08105
Pause	1.165	0.416	2.800	.00511 **

Note: N = 212, log-likelihood = -118.3. **<.01.

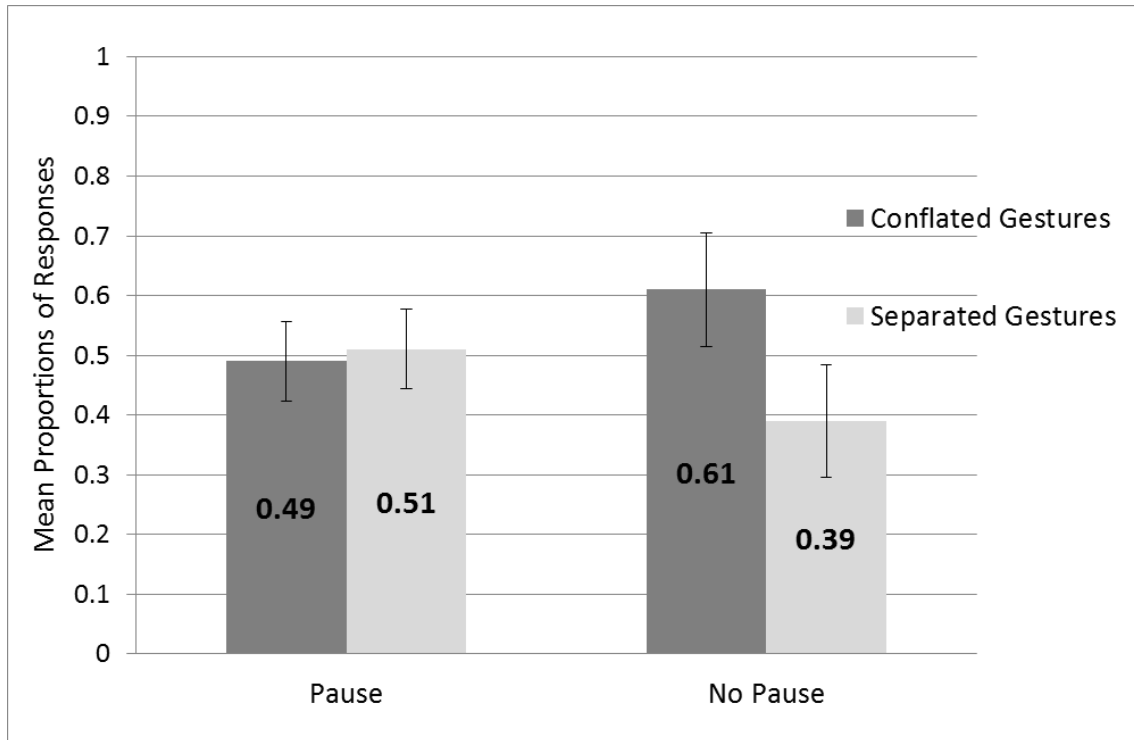


Figure 3.6. Mean Proportions of responses with Conflated Gestures and Separated Gestures with and without a pause in the Inserted Clause and Inserted Phrase Condition of Experiment 2. Error bars represent standard errors.

Table 3.7. Summary of the fixed effects (Pause and Clause Type) in the mixed logit model for gestural depiction in Experiment 2 (Inserted Clause and Inserted Phrase Condition only).

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Wald Z</i>	<i>p</i>
Intercept	-0.342	0.459	-0.744	.4567
Pause	0.922	0.504	1.831	.0672
Clause Type	-0.683	0.418	-1.634	.1023
Pause * Clause Type	0.0618	0.911	0.068	.9459

Note: N = 212, log-likelihood = -116.7.

3.9. Discussion

Participants produced more pauses between the manner verb and the path particle when they inserted a clause than when they inserted a phrase. Based on the assumption that pauses at phrase boundaries reflect processing unit boundaries (Ferreira, 1991), these results confirm our assumption that embedding a clause makes it more likely that the manner verb and path particle fall into different planning units.

The key finding was that participants produced significantly more separated gestures when they had to embed a clause (“as seen in the video”) (Inserted Clause Condition) compared to when they had to embed a phrase (“in this short video”) (Inserted Phrase Condition) or when they produced the manner verb and path particle together at the end of a sentence (Verb Final Condition). The significant difference between the Inserted Clause Condition and the Verb Final Condition dovetails with the results of Experiment 1; but unlike in Experiment 1, in this experiment, the manner verb and the path particle were always in the same main clause. The significant difference between the Inserted Clause Condition and the Inserted Phrase Condition showed that the linguistic nature of what separates the manner verb and the path particle matters (while controlling for the lengths of inserted elements). The findings overall indicate that when a manner verb and a path particle are produced in different planning units, gestures are likely to depict manner and path in separate gestures. Furthermore, the likelihood of manner-path separation in gestures was comparable between the Inserted Phrase Condition and the Verb Final Condition. This suggests that mere surface distance between verb and particle does not influence gestural content. Thus, the results from Experiment 2 provide further evidence for the Planning Unit Account.

3.10. General Discussion

The main aim of this study was to provide evidence that supports the Planning Unit Account of how speech and gesture coordinate their contents, which cannot be explained by the Lexicalisation Account. In Experiment 1, manner and path were more likely to be depicted in separate gestures (as opposed to in a single gesture) when German speakers produced the manner verb and the path particle in a main clause that are separated by an inserted clause than when they produced the manner verb and the path particle together in a contracted form in a subordinate clause. In English, no difference in gestural depiction was found between the Main

Clause Condition and the Subordinate Clause Condition because the verb and the particle were always produced one after the other. In Experiment 2, the manner verb and the path particle were always in the same main clause. Manner and path were more likely to be depicted in separate gestures when German speakers produced the manner verb and the path particle that are separated by an inserted clause than when they produced the manner verb and the path particle together in a contracted form (i.e., Verb Final condition) or when they produced the manner verb and the path particle that are separated by an inserted phrase. Because a clause is a good proxy for a planning unit (Bock, 1982; Levelt, 1989), when an inserted clause separates the manner verb and the path particle, the verb and the particle are likely to be processed in two separate planning units. And, this was exactly the situation in which manner and path were more likely to be depicted separately in two gestures. This supports the Planning Unit Account, which states that a gesture tends to depict spatial information contained within in a single planning unit for speech production. In both Experiments, the lexicalisation of manner and path (manner expressed as a verb and path expressed as a particle within the same clause) was kept constant; thus, the Lexicalisation Account cannot explain the current findings.

Comparing the two conditions which were kept the same in Experiment 1 and Experiment 2 (i.e., Main Clause Condition and Inserted Clause Condition), shows that the proportions of Conflated and Separated Gestures are very similar across these two experiments. However, in the conditions where the particle verbs were linguistically attached at the sentence final position in Experiment 1 (i.e., Subordinate Clause Condition) and Experiment 2 (i.e., Verb Final Condition), we found more Separated Gestures in Experiment 2 (mean percentages: 22 % vs 35 %). Although these two Conditions are very similar, they do differ in terms of clause type and in terms of clause length, i.e., shorter subordinate clause in Experiment 1 compared to longer main clause in Experiment 2. Possibly, the larger proportion of separated gestures stems

from the longer clause. Although the particle verbs are linguistically attached, it is possible that the manner component and the path component of the motion event occasionally fell into different planning units in Experiment 2 which led to an increased gestural separation of manner and path. However, this explanation is only speculative.

What are the implications of these results for the existing literature on motion event gestures? The current study, for the first time, provided evidence that unambiguously support the Planning Unit Account for coordination of speech and gesture production. Studies so far (Kita & Özyürek, 2003; Kita et al., 2007; Özyürek et al., 2008; Özyürek et al., 2005; Wessel-Tolvig & Paggio, 2016) always compared satellite-framed and verb-framed constructions. Hence, it was not possible to untangle whether lexicalisation pattern (one verb framing vs. two verb framing) caused gestural differences across verb-framed and satellite-framed constructions or whether these differences were caused by information packaging into planning units for speech production. That is, the current study provided unambiguous evidence for the key component of the Interface Model for speech-gesture production (Kita & Özyürek, 2003). Furthermore, the pause analysis of Experiment 2, for the first time, provided evidence that clauses are a good proxy for planning units for speech production for the type of sentences used in this line of studies. This is in line with the literature on speech production (Bock, 1982; Levelt, 1989).

The current study did not directly refute the Lexicalisation Account. However, all results in the literature regarding manner and path depiction in co-speech gesture, along with similar results regarding the impact of the richness of verb meaning on gesture (placement verbs: Gullberg, 2011; Gullberg & Narasimhan, 2010; the motion verb "swing": Kita, 1993; Kita & Özyürek, 2003), can be explained by the Planning Unit Account, and there is no finding that

can only be explained by the Lexicalisation Account. Thus, the Planning Unit Account is the most parsimonious account for the coordination of speech and gesture production.

The idea that the information packaging in speech influences information packaging in gesture is further supported by comparison of co-speech and silent gestures (Özçalışkan, 2016; Özçalışkan et al., 2016). Özçalışkan et al. (2016) investigated how speakers of English (satellite-framed) and Turkish (verb-framed) depict manner and path in co-speech gestures and silent gestures. The gestural depiction differed cross-linguistically as in earlier studies (e.g., Kita & Özyürek, 2003) for co-speech gestures, but not for silent gestures. Speakers of both languages conflated manner and path in one gesture for a large majority of the time, when producing silent gestures. These results indicate that conflated gestures may reflect the “default setting” (Gullberg, 2011, p. 185) for event conceptualisation and more importantly that the speech production process triggers reconceptualisation of events. What are the theoretical implications of such a reconceptualisation process for the gesture-speech production literature? This interpretation contradicts a strong modular view of speech production processes which assumes that there is no online interaction between the preverbal message produced in the Conceptualiser and the Formulator which is responsible for grammatical encoding and for building the surface structure (Levelt, 1989). This modular view is also incorporated in a gesture-speech production model put forward by de Ruiter (2000). According to his Sketch Model which is based on Levelt’s (1989) speech production model, speech and gesture are produced in parallel but independently after the preverbal message has been created in the Conceptualiser. Hence, this modular view on speech-gesture production cannot explain a reconceptualisation of motion events based on packaging into speech-gesture planning units. If motion events are reconceptualised due to information packaging in speech, this would suggest that during speaking event representations on a conceptual level and event representations on a

syntactical level are generated interactively. Hence, this would allow online interaction between the planning of the pre-verbal message and the planning of the surface structure (i.e., syntactical planning) as suggested by Kita (1993), Kita and Özyürek (2003) and Vigliocco and Kita (2006).

One question that the Planning Unit Account does not address is why large proportions of responses in the conditions where manner and path of the motion event were produced within the same planning unit included either a Path only gesture or a Manner only gesture (i.e., Subordinate Clause Condition and both English Conditions in Experiment 1 and Verb Final Condition and Inserted Phrase Condition in Experiment 2). Thus, these Singleton responses, which overall comprise approximately half of the dataset, show that the gesture plan did not always include both motion event components although they are generated together in the same speech planning unit. Previous research has suggested that only information that is important in discourse is expressed gesturally (Kita & Özyürek, 2003; McNeill, 1992). For instance, for the Tomato Man Movies (which were also used in the present study), Kita et al. (2007) suggest that the path component (i.e., change of location) is more important for the development of the story since it is the event's closing event. Thus, more path gestures were produced than manner gestures. Similarly, participants might have chosen to gesturally express only the component that was important for their own description of the video clips.

In sum, our study provides evidence that planning units for speech production plays an important role in how speech and gesture production are coordinated; more specifically, what information is encoded in each gesture tends to correspond to what information is linguistically encoded within each speech planning unit. This highlights the interactive nature of gesture-speech production – probably starting from a pre-verbal event conceptualisation all the way to the linguistic and gestural planning processes.

Chapter 4

Gesture-speech Synchronisation: Surface Locations of Semantic Affiliates within a Sentence Predict Gesture Onset and Gesture Duration

4.1. Abstract

Speech and gesture production are two highly interactive systems. However, it has not been tested whether linguistic surface structure influences synchronisation of speech and iconic gestures. We analysed onset and duration of iconic gestures elicited by motion events in relation to their semantic affiliates within a sentence context. The data stem from two experiments (see Chapter 3) conducted in German where manner verb (e.g., “rolling”) and path particle (e.g., “down”) can be separated linguistically by other speech elements. A high proportion of gestures were placed between the verb and particle resulting in varying degrees of gesture-semantic affiliate asynchrony. For gestures placed after the verb, the surface distance (ms) between the verb and the particle predicted gesture onset; i.e. the larger the distance was between *rolling* and *down*, the later the gesture was initiated. This was the case independently of the type of gesture, i.e. conflated (encoding both manner and path simultaneously) or path gesture. We argue that semantic affiliates function as attraction points for gesture onsets. If they are separated on the surface structure, they compete for gesture synchronisation and thus the gesture onset falls somewhere between the two parts of the semantic affiliate. It was also found that gesture duration is partly determined by the location of semantic affiliates. When the onset of a gesture falls between two semantic affiliates, then the duration of gesture is longer when the second affiliate is further away from the gesture onset. That is, gestures are lengthened to make overlap with the upcoming semantic affiliate more likely. This indicates that the speech and gesture production system exchange information even after the gesture’s onset. The results

suggest that both the onset and offset of iconic gestures are pulled by semantic affiliates on the surface structure.

Keywords: gesture, gesture synchronization, gesture production, motion events, iconic gestures

4.2. Introduction

Speech accompanying gestures are a universal aspect of communication (Kita, 2009). They are usually not redundant with information conveyed in speech, but express additional semantic information. Speech and gestures are linguistically tightly coordinated at a phonological, pragmatic and semantic level (see for overview: P. Wagner et al., 2014). They are also coordinated at a temporal level (McNeill, 1992, 2005). This study investigates the mechanism with which iconic gesture and speech are synchronised with each other. In the current study two theoretical debates surrounding the temporal coordination of gesture and speech are addressed. The first debate concerns the influence of linguistic surface structure on the timing of gesture initiation. The second debate concerns the stages during which the gesture and speech production systems exchange information. These questions are crucial in specifying how speech production and gesture production are interlinked with each other.

According to McNeill (1992, 2005, 2015), the close link between these two modalities on the surface structure is based on the so-called Growth Point that represents a theoretical idea unit underlying gesture and speech. In a dynamic process the Growth Point is then unpacked in order to fit into a grammatical and lexical frame of a given language. Since gesture and speech share the same origin in production (i.e., the Growth Point), gestures are initiated in such a way that they temporally overlap with their semantic affiliates ("conceptual affiliate" in de Ruiter, 2000; "lexical affiliate" in Schegloff, 1984), i.e. the portion(s) of speech that semantically correspond most closely to the gestural content. Thus, when a gesture is initiated depends on where on the surface utterance its semantic affiliates are produced. This dominant view on how the linguistic surface structure influences when gesture is initiated in relation to speech is the "Semantic Synchrony Rule" (McNeill, 1992). This idea is supported by analyses of how motion event gestures in narratives overlap with semantically co-expressive words (Chui, 2005, 2009; Duncan, 2006; Kellerman & van Hoof, 2003; McNeill, 1992, 2005; McNeill & Duncan, 2000;

Stam, 2006). Usually more than one lexical item is semantically related to a motion event gesture. Consistent with the lack of a one-word-one-gesture mapping, previous studies on motion event gestures found that gestures synchronise with different portion(s) that are part of the gesture's semantic affiliate (e.g., manner verb: *rolling* and/or path particle: *down*). To account for the variance in these gesture synchronisation patterns, different factors have been proposed that influence gesture synchronisation. For example, McNeill & Duncan (2000) argue that a gesture synchronises with the part of the motion event that the speaker focusses on. Furthermore, studies highlighted the importance of language typology (how a language lexically encodes motion events) and the developing thought of the speaker to be a driving force for synchronisation patterns (Stam, 2006).

An alternative view on how surface structure of speech influences the timing of gesture initiation can be found in de Ruiter's (1998, 2000) model of speech-gesture production, the Sketch Model. De Ruiter argues that linguistic surface structure does not have an influence on the temporal coordination of gesture and speech. The Sketch Model predicts that the gesture's onset is roughly locked to the onset of its affiliate in speech because the content of a gesture ("sketch") and the pre-verbal message for speech are constructed simultaneously in the conceptualiser (in Levelt's sense (1989)). It predicts that the onset of a gesture roughly coincides with the onset of a semantically co-expressive phrase. However, gesture and speech are processed independently after the conceptual stage; thus when a gesture is initiated is not sensitive to exactly where in the utterance its semantic affiliates are produced.

Both the Semantic Synchrony Rule and the Sketch Model claim synchrony of iconic gestures and semantic affiliate in speech. But in both cases, this is only an approximation as the onset of a gesture does not perfectly align with the onset of the semantic affiliate. More specifically, researchers generally acknowledge that the onset of a gesture usually precedes its

affiliate (Morrel-Samuels & Krauss, 1992; Nobe, 2000; Schegloff, 1984). For example, Morrel-Samuels and Krauss (1992) analysed gestures and semantic affiliates in their dataset (60 gestures collected from narratives) and found that gesture onsets either preceded onsets of semantic affiliates or were initiated at the same time as the affiliates. Furthermore, they found evidence that asynchrony is related to speech production process. The amount of gesture-speech asynchrony was negatively correlated with the familiarity of the semantic affiliate: The more familiar a word was (faster lexical access), the smaller was the gap (in ms) between gesture onset and the onset of the semantic affiliate.

Why does the gesture onset tend to precede the semantic affiliate in speech? Three different explanations have been given. First, McNeill (1985) has suggested that, though a gesture and its semantic affiliate get activated simultaneously, a gesture can sometimes be initiated before its semantic affiliate in the actual utterance because grammatical encoding may put the affiliate later in the utterance. Second, according Morrel-Samuels and Krauss (1992), gesture precedes speech because the motoric representation necessary for gesture production is accessed faster than semantic information necessary for lexical retrieval. Third, de Ruiter (2000) has suggested that the conceptual-level message representation for speech is not passed onto linguistic formulation processes until the motor plan for the co-expressive gesture is completed. Because linguistic formulation processes plus articulation processes for speech take time, the gesture onset precedes the semantic affiliates. All these explanations predict that the gesture onset may precede, but never follow, the onset of the semantic affiliates.

Not only gesture onset is a crucial measurement in order to infer how gesture and speech interact during production but also the duration of a gesture's stroke (i.e., the meaningful part of a gesture (McNeill, 1992)) has to be taken into account. The Semantic Synchrony Rule (McNeill, 1992, 2005) and de Ruiter's Sketch Model make different assumptions on how

gesture duration is influenced by the linguistic surface structure. The Semantic Synchrony Rule as a part of McNeill's (2005, p. 24) Growth Point Theory assumes that the idea unit from which gesture and speech emerges results in an almost "unbreakable bond" between these two modalities. This bond persists until the execution phase. Thus in order to achieve synchronisation, not only gesture onset but also gesture duration is influenced by the linguistic surface structure. In contrast, the Sketch Model assumes that gesture and speech run independently after the conceptualisation stage. Thus, this model predicts that linguistic surface structure does not impact gesture duration (de Ruiter, 1998, 2000).

Studies on pointing gestures have directly tested whether the information exchange between the two production systems continues after the onset of the gesture (interactive view) or whether the two systems become independent as soon as the gesture is initiated (ballistic view) (Chu & Hagoort, 2014; Feyereisen, 1997; Levelt et al., 1985). In a pioneering study, Levelt et al. (1985) asked participants to point at lights which were randomly lit and say "this/that light". Occasionally, the participants' gestures were interrupted by a 1600-gram weight which was tied to the participant's wrist. They found that gesture and speech are planned interactively before the speech and gesture onsets, but once the gesture is launched, just prior to speech onset (i.e., between 300-370 ms before speech onset), the two systems become modular. Thus, from this point onwards the two systems run independently and no adaptations to the gesture's duration can be made. Levelt et al. (1985) refer to this view of gesture-speech independence, which begins roughly at the onset of the gesture, as the "ballistic view". However, de Ruiter (1998) found evidence for the interactive view in a similar study. He found that pointing gestures had a longer duration when speech becomes disfluent. That is, the gesture execution process knows the trouble in speech formulation/articulation processes. In a more recent study on pointing gestures, Chu and Hagoort (2014) found further evidence for the

interactive view. Using virtual reality and motion tracking technology, they found that when the speech production system is interrupted through a self-repair, the gesture is prolonged even if the self-repair occurred after the onset of both speech and gesture. This indicates that speech and gesture exchange feedback not only through all stages of gesture production but also gesture execution.

Morrel-Samuels & Krauss' study (1992) described above suggests that this interactive view may also hold for iconic gestures. They found that the greater the asynchrony of speech and gesture (i.e., gesture precedes speech), the longer the gesture's duration. These results indicate that a gesture produced within a sentence context aims to synchronise with its semantic affiliate and, when the gesture onset is "too early", the gesture is prolonged. Because they analysed gestures produced during narrations, we assume that most of the gestures were iconic gestures; however, they do not report the types of gestures. Thus, it remains unclear whether the interactive view holds for iconic gestures.

Another important issue to consider when drawing conclusions from previous studies on the interactive nature of gesture and speech is the linguistic context in which gestures were embedded in. In all previous studies that conducted a detailed analysis (in ms) on gesture onset and gesture duration in relation to its semantic affiliate, participants produced very simple utterances (single noun: Feyereisen (1997); noun phrase: Chu & Hagoort (2014); Levelt et al. (1985); colour term: Chu and Hagoort (2014)) or the analysis did not take into account where in the sentence semantic affiliates occurred (e.g. in relation to the sentence onset, or to other semantic affiliates) (Church, Kelly, & Holcombe, 2014; Morrel-Samuels & Krauss, 1992). Hence, it is unclear whether or to what extent linguistic surface structure has an impact on gesture onset and gesture duration.

4.3. Present Study

In the present study the temporal relationship between gestures and speech within a sentence context was investigated. First, we tested whether iconic gestures always precede the semantic affiliate, as predicted by all theories in the literature. Second, the gesture's onset and its duration were analysed in relation to its semantic affiliates in speech. For these analyses we took advantage of the syntactical properties of German particle verbs whose two components (i.e., manner verb and path particle) can be separated on the surface structure. This separation of the gesture's semantic affiliate on the linguistic surface structure allowed us to directly test the two most influential views on how gesture and speech are synchronised, i.e. the Semantic Synchrony Rule (McNeill, 1992, 2005) and the Sketch Model (de Ruiter, 1998, 2000).

The data in the current study stems from two previous experiments that investigated the influence of planning units on the content of motion event gestures (see Chapter 3). In these experiments, German native speakers were asked to retell a short video clip about a motion event that includes manner and path of motion (e.g., "the tomato is rolling down the hill"). The participants were instructed to produce gestures during the description, but it was up to the participants what kind of gesture they produced and when within the sentence they produced the gesture(s). We analysed iconic gestures that depicted path (e.g., a downward movement) and those that conflated manner and path in a single gesture stroke (e.g., henceforth "conflated gestures", simultaneous gestural depiction of rolling and down) because these two types of gestures were frequent enough for statistical analysis. For manner-path conflated gestures, both the manner verb and the path particle are semantic affiliates. For path gestures, path particle is clearly a semantic affiliate. We would argue that the verb is also a semantic affiliate because, according to Talmy's semantic analysis (2000), path particles encode only the direction and verbs encode change of location (motion itself). However, some may argue that a path gesture's

semantic affiliate is only the path particle. Thus, we analysed manner-path conflated gestures and path gestures separately.

As for speech, the participants were asked to use one sentence to describe a motion event, following a specific sentence structure with a specific particle verb (e.g., hinaufklettern, “to climb up”). Furthermore, we asked participants to insert “as seen in the video” (in Experiment 1 & 2) or “in this short video” (Experiment 2) between the manner verb and the path particle, which is grammatical in German, to increase the distance between the manner verb and the path particle. In most of the trials, participants also spontaneously added a so-called ground element (e.g., the rainbow) between the two motion event components. This created a linguistic surface distance between manner and path expressions, as in (1).

- (1) Der Elefant **klettert**, wie im Video gesehen, einen Regenbogen **hinauf**.
 The elephant **climbs**, how in the video seen a rainbow **up**.
 The elephant is climbing up a rainbow, as seen in the video.

Furthermore, depending on the particle verb, German speakers can express the path of a motion event not only with a particle, but also with a preposition. In these cases, particles and prepositions can have the identical form as in (2). Such constructions are referred to as “double-framing” (Croft, Barðdal, Hollmann, Sotirova, & Taoka, 2010). The particle and the preposition can also be different, with slightly different semantics as in (3). Approximately 40% of the responses in our dataset included more than one path element. These additional path prepositions constitute an additional semantic affiliate for motion event gestures and hence have been considered in our analyses.

- (2) Die Maus **bohrt** sich, wie im Video gesehen, **durch** den Globus **durch**.
 The mouse **bores** itself how in the video seen **through** the globe **through**.
 The mouse is boring through the globe, as seen in the video.
- (3) Die Maus **schwebt**, wie im Video gesehen, **in** einen Pool **herunter**.
 The mouse **floats** how in the video seen **in** a pool **down**.
 The mouse is floating down into a pool, as seen in the video.

Given this sentence structure, the Semantic Synchrony Rule and the Sketch Model make different predictions on gesture onset and gesture duration of iconic gestures. According to de Ruiter's Sketch Model (1998, 2000), the onset of both Path Gestures and Conflated Gestures should be roughly time-locked to (and slightly precede) the onset of the verb phrase (i.e., the manner verb). This is based on the model's assumption that gestures are planned with the pre-verbal message, whose minimal chunks comprise of noun phrases and verb phrases. Thus, the iconic gesture onset roughly coincides with the onset of the first word that is semantically affiliated with the gesture. However, when the manner verb and the path particle are separated as in (1-3), the iconic gesture may not extend to the sentence final path particle because the gesture production process is blind to linguistic formulation (i.e., blind to the fact that particle is placed at the sentence final position). McNeill's (1992, 2005) *Semantic Synchrony Rule* predicts that the onset of iconic gestures should be roughly time-locked (and slightly precede) to the gesture's semantic affiliate that was part of the speaker's idea unit (i.e., verb, path preposition, particle or a combination of the affiliates). Unlike the Sketch Model, the Semantic Synchrony Rule predicts that iconic gestures are extended to overlap with their affiliate if they are initiated prior to the affiliate's onset. Thus, if a gesture onset falls between the verb and the particle, the *Semantic Synchrony Rule* predicts a positive correlation between gesture-semantic affiliate asynchrony and gesture duration; i.e. the larger the asynchrony, the longer is the gesture. That is, because McNeill's theory assumes feedback between the gesture and speech channels during gesture execution. When a gesture is initiated before its semantic affiliate, it should be prolonged to achieve synchronisation. However, whether the gesture can actually overlap with the semantic affiliate may differ between path gestures and conflated gestures. Due to their repetitive nature, it might be easier to lengthen conflated gestures compared to path gestures. Thus, even when a conflated gesture is initiated much before the semantic affiliate is

encoded in speech, synchronisation might still be achieved. This might not be the case for path gestures which are more difficult to prolong. This prediction is in line with Kita's (1990) observation that repetitive gestures are less likely to have a hold phase after the stroke than non-repetitive gestures. Although the Semantic Synchrony Rule does not predict any gestures falling between the verb and the particle, this type of gesture synchronisation would not necessarily be in favour of the ballistic view. Whether our data is in line with the ballistic or the interactive view depends on the presence of a correlation between the gesture-semantic affiliate asynchrony and the gesture's duration. If there is no correlation in our dataset between gesture onset and gesture duration, this would then be in favour of the ballistic view as in the Sketch Model.

Contrary to the predictions by the Semantic Synchrony Rule and the Sketch Model, iconic gestures may "fall between" the two affiliates and thus do not overlap with either affiliate. This might be the case because German allows the manner verb and the path particle to be separated by many words as in (1-3) creating a large distance between the semantic affiliates. In such cases, semantic affiliates may still serve as "attractors" for the gesture. That is, the verb, the path preposition and the path particle may all pull the onset of path and conflated gestures towards them. If the semantic affiliates are pulling the gesture towards them, the occurrence of all semantic affiliates (i.e., manner verb, additional path preposition and path particle) within the sentence should predict the gesture's onset.

4.4. Methods

4.4.1. Participants

The same German participants from the study in Chapter 3 were included in the current synchronisation analysis. 21 participants from Experiment 1 (mean age = 23.6 years, SD = 3.9) and 20 participants (mean age = 27.3 years, SD = 3.9) from Experiment 2.

4.4.2. Procedure

The full procedure of both experiments is described in Chapter 3.

4.4.3. Data Coding and Analyses

For the analyses, data from Experiment 1 and Experiment 2 were collapsed. The data was coded with the linguistic annotator ELAN (Lausberg & Sloetjes, 2009). In a first step, speech was transcribed. But only responses where the participants used the expected sentence structure were considered for the analyses. The gestural coding of motion events was based on the “Cross-linguistic Motion Event Project” coding manual (Kita et al., 2007; Özyürek et al., 2008) and was adapted and elaborated for the stimuli used in the current experiment in the following way: Only strokes depicting the target motion event, i.e. the gestural depiction of the given particle verb of a response, were coded (Kita et al., 1997; McNeill, 1992). In a first step, all target event gestures were classified either as path, manner, conflated. To be classified as a path gesture the gesture had to only depict the direction of the event (e.g., for the motion “to float down” this might involve a downward movement with (an) open palm(s) but without any movements to the left or right (which would indicate manner). Manner gestures were defined as depicting solely the manner aspect of the motion event (e.g., for the motion “to climb up” this might involve the participant opening and closing their palm(s) without moving their arms upwards). Conflated gestures were defined as depicting motion and manner of the motion event in a single gesture (e.g., for the motion “to roll into”, this might involve rotating one’s wrist(s) with a simultaneous change of location away from the body). Any other gestures were excluded from the analysis. These included combinations of path or manner gestures with a conflated gesture produced within a single stroke (i.e. hybrid gestures as defined by Özyürek et al. (2008)) and responses with multiple target event gestures (e.g., a path gesture combined with a manner gesture or two path gestures). In the latter cases different synchronisation strategies might

apply. The combination of two or more gestures with the same content (e.g., two path gestures) within one response was very rarely used by the participants (15 instances). These were excluded from the analysis as well. Manner Gestures were used infrequently (mean percentage 9 %). Hence, the dataset was too small to perform a sound analysis. We therefore examined Path Gestures (mean percentage 60 %) and Conflated Gestures (mean percentage 31 %) and analysed those separately.

Gesture strokes (McNeill, 1992) and stroke durations (in ms) were coded and the synchronising speech elements were annotated. Since the synchronisation analysis focusses on the components of the semantic affiliate (verb and particle), we categorised gestures on the basis of their overlap with the particle, verb, verb plus particle, or whether they fell in between verb and particle (see Table 4.1). This way of data presentation provides an overview of gesture-speech alignment in relation to the gestures' semantic affiliates. Gestures were defined as synchronising with the particle if the stroke overlapped with one syllable of the particle. The same was applied for verbs (cf. Stam, 2006).

Table 4.1. Gesture synchronisation categories

Categories	Description
Particle	Gestures that synchronised (i.e., overlapped) with the particle only or particle plus any other speech element except the verb
Verb	Gestures that synchronised with the verb only or verb plus any other speech element except the particle
Between Verb & Particle	Gestures that synchronised with speech elements between verb and particle (e.g., ground elements, inserted phrases/clauses), but not with verb or particle
Verb & Particle	Gestures that synchronised with both the verb and particle plus everything between these two elements, forming an “overarching gesture”

In order to test whether gesture onset is influenced by the surface location (in ms) of different motion event components (verb, particle and additional path preposition), we determined the following measures: Gesture onset (GO) and Verb onset (VO) were measured

from sentence onset. Particle onset was measured from verb onset (VOtoPO). We did not measure it from sentence onset due to high collinearity between a measure from sentence onset and VO. Furthermore, we measured the distance between gesture onset and particle onset (GOtoPO). Figure 4.1 and Figure 4.2 illustrate these measurements. We also included Additional Path Preposition in the analyses as a binary factor (yes/no). But we did not add their onset because it turned out that the onset of the additional path preposition (measured from sentence onset) correlated highly with the variable Particle Onset. Thus, it would have been difficult to assign independent variance to these two factors.

All responses that deviated from an individual's means by more than three SDs were considered as outliers and were excluded from all the analyses. This criterion led to the exclusion of 9 Path Gestures and 7 Conflated Gestures. The same criterion was applied for calculating outliers for the measurement of stroke durations (in ms). This led to the exclusion of 6 Path Gestures. For the gesture onset and gesture duration analyses we collapsed the categories Verb and Verb & Particle. Both of these categories are included in the "Verb" category.

The data were analysed by fitting linear mixed effect models in RStudio (R Core Team, 2014) using the `lmer` function (Bates, Maechler, Bolker, & Walker, 2014). To obtain approximate p-values and degrees of freedom when fitting our models with the `lmer` function, we used the Satterthwaite approximation, which is implemented in the `lmerTest` package (Kuznetsova, Brockhoff, & Christensen, 2015). The significance of a particular factor was determined by step-wise removing them and comparing models with and without these factors, using a maximum likelihood method (i.e., ANOVAs). We report the chi-square statistics, degree of freedom and p-value for each model comparison. For all analyses, we treated subject and item as random effects including random intercepts for both. Since the fitting of models

including random slopes did not converge in most analyses, the random effect structure of all but two models included only random intercepts for subject and item. How the models were fitted is explained in detail in the relevant results section. Bates, Kliegl, Vasishth, and Baayen (2015) have shown that dropping random slopes has a relatively minor effect on the significance of a factor. Thus, it can lead to wrong conclusions only when p-values are close to the critical level of 0.05. This was not the case for our results here, meaning that the results' reliability is not compromised.

4.5. Results

Figure 4.1 and Figure 4.2 show where in the sentence Path Gestures and Conflated Gestures were produced based on the synchronisation categories in Table 4.1 (Verb, Particle, Between Verb & Particle, Verb & Particle) and on whether or not the sentence included an additional path preposition. A substantial number of gestures did not overlap with the semantic affiliate. More specifically, more than half of the Path Gestures and over 30% of Conflated Gestures fell in between the verb and the particle (i.e., onset and offset of the gesture stroke both fell in between). Generally, very few gestures synchronised with the verb but not with the particle, and there were very few overarching gestures synchronising with both the verb to the particle (14 responses in total).

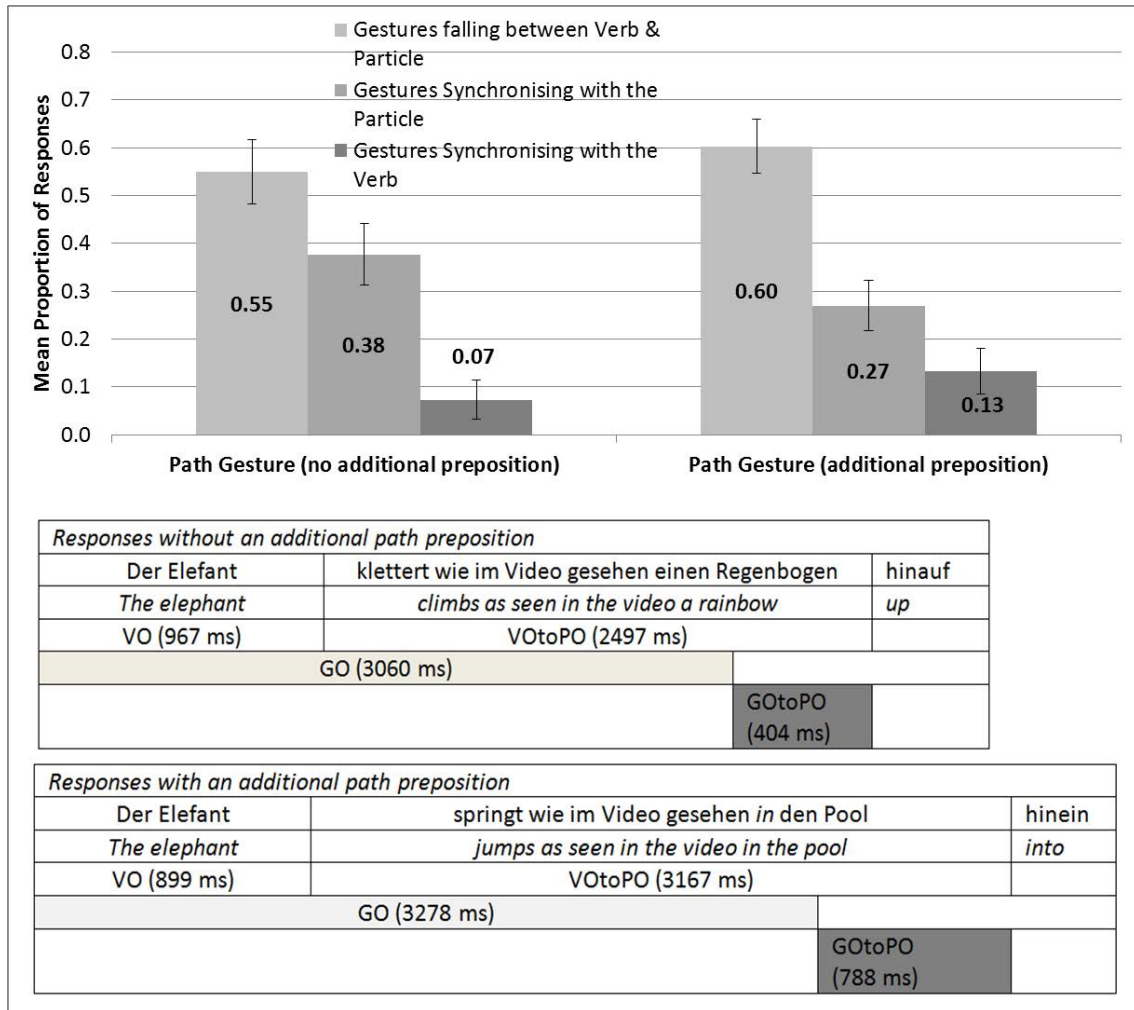


Figure 4.1. Top: Mean proportion of Path Gestures that synchronised with the Verb, Particle, or Between Verb & Particle (see Table 4.1 for definitions), split into responses with and without an Additional Path Preposition. Note that no Path Gestures synchronised with Verb & Particle. Bottom: Means (ms) of the duration of Sentence Onset to Verb Onset (VO), Verb Onset to Particle Onset (VOtoPO), Sentence Onset to Gesture Onset (GO) and Gesture Onset to Particle Onset (GotoPO) for Path Gestures whose stroke onset occurred after the verb.

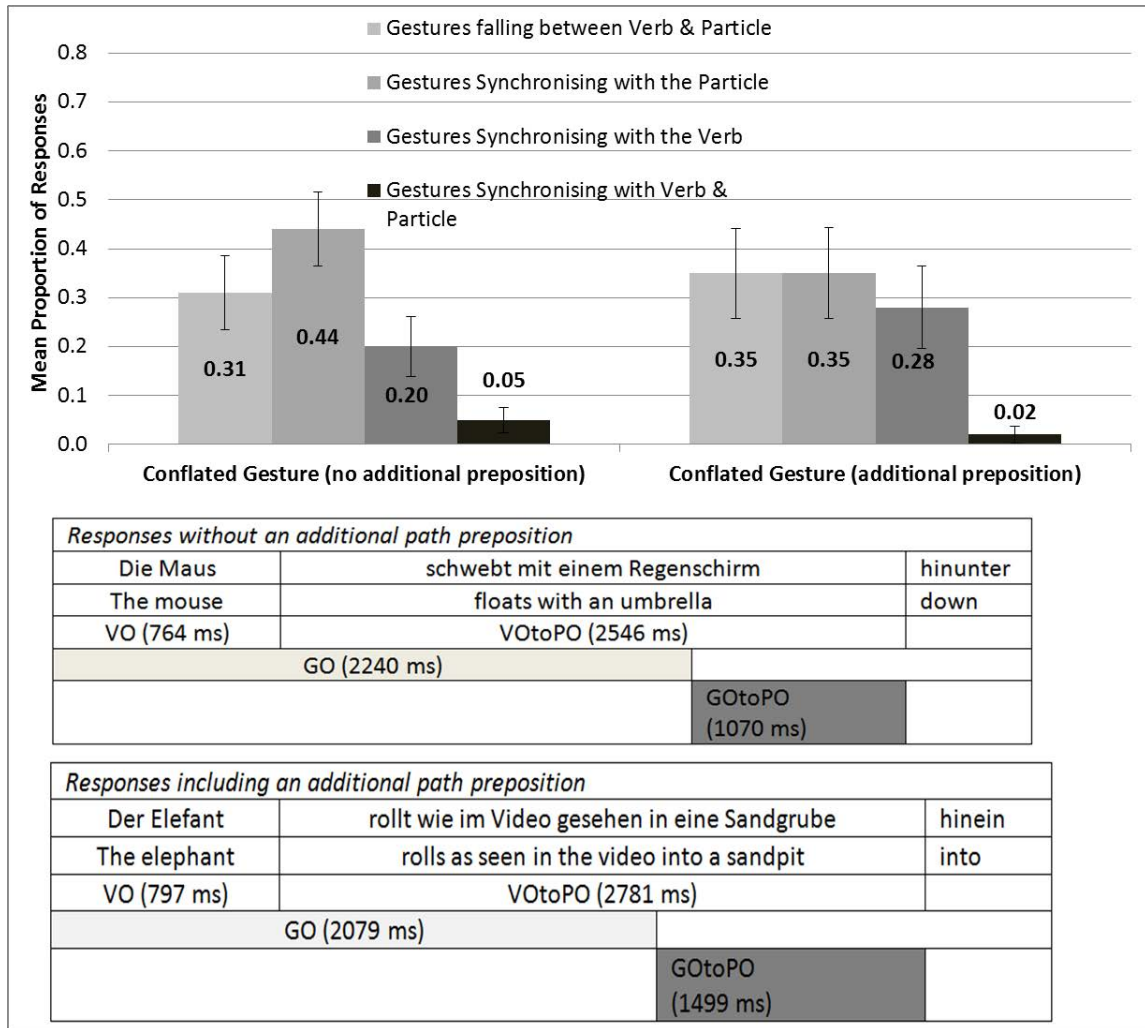


Figure 4.2. Top: Mean proportion of Conflated Gesture which synchronised with the Verb, Particle, Between Verb & Particle, or Verb & Particle, split into responses with and without an Additional Path Preposition.

Bottom: Means (ms) of the duration of Sentence Onset to Verb Onset (VO), Verb Onset to Particle Onset (VOtoPO), Sentence Onset to Gesture Onset (GO) and Gesture Onset to Particle Onset (GOTOPO) for Conflated Gestures whose stroke onset occurred after the verb.

4.5.1. Gesture Onset Analyses

We started with the gesture onset analysis. The dependent variable is Gesture-Onset (GO). In particular, we tested whether gesture onset is influenced by the surface location of the verb (VO) and the particle (VOtoPO). Because these effects might not be the same if the gesture synchronised with the verb, the particle or fell in between the verb and particle, we ran separate analyses for these three categories. Table 4.2 shows the means of the dependent variable (i.e., Gesture Onset) for Path Gestures and Conflated Gestures for the three different categories. Furthermore, the means of the dependent variable are shown across Conditions which we collapsed for the analyses. In particular, the “Inserted Clause Condition” includes the Main Clause Condition from Experiment 1 and the Inserted Clause Condition from Experiment 2 and the “Inserted Phrase Condition” only includes the Inserted Phrase Condition from Experiment 2.

Table 4.2. Means (ms) of the duration of Gesture Onset for Path Gestures synchronising with the Verb, the Particle and gestures that were placed between the verb and particle in the Inserted Clause Condition (including Main Clause Condition from Experiment 1) and Inserted Phrase Condition.

<i>Analysis</i>	<i>Means of the Duration of Gesture Onset in the Inserted Clause Condition</i>	<i>Means of the Duration of Gesture Onset in the Inserted Phrase Condition</i>
Path Gestures (Verb)	975 ms	1103 ms
Path Gestures (Between Verb & Particle)	2974 ms	2901 ms
Path Gestures (Particle)	3349 ms	3526 ms
Conflated Gestures (Verb, Verb & Particle)	701 ms	641 ms
Conflated Gestures (Between Verb & Particle)	2148 ms	1755 ms
Conflated Gestures (Particle)	2446 ms	2191 ms

Figure 4.3 and Figure 4.4 show the relationship of gesture onset with Particle Onset and Verb Onset respectively for Path Gestures. We first analysed Path Gestures that synchronised with the verb. We started our analysis by fitting a full model with Gesture Onset (GO) as dependent variable, Verb Onset (VO), Particle Onset (VOtoPO) and Additional Path

Preposition as fixed factors. Results of the final model are summarised in Table 4.3. Only Verb Onset survived as a predictor in the final model (Additional Path Preposition: $\chi^2 = 0.0147$, $df = 1$, $p = .903$; Particle Onset: $\chi^2 = 0.64$, $df = 1$, $p = .423$; Verb Onset: $\chi^2 = 12.35$, $df = 1$, $p < .001$). Although these results have to be interpreted with caution due to the small number of responses in this category, it suggests that when gestures synchronise with the verb, the particle and the additional path preposition do not have any effect, i.e. the gesture is strongly synchronised with the verb and the surface location of the particle or any other path preposition becomes irrelevant.

Table 4.3. Summary of fixed effects in the mixed linear model for Gesture Onset (Path Gestures that synchronised with the Verb)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	194.321	145.688	13	1.334	.250
VO (Sentence Onset to Verb Onset)	0.8600	0.128	16	6.739	<.001 ***

Note: $N = 17$, log-likelihood = -117.5; ***<.001

Next we fitted the same model with Path Gestures that fell between the verb and the particle to test whether in this case the verb and the particle's surface location as well as the presence of the additional path preposition influence gesture onset. First we dropped the factor Additional Path Preposition which did not significantly increase the model's fit ($\chi^2 = 1.6632$, $df = 1$, $p = .198$). Dropping the factors Verb Onset ($\chi^2 = 70.53$, $df = 1$, $p < .001$) or Particle Onset ($\chi^2 = 141.53$, $df = 1$, $p < .001$) from the model yielded a decreased model fit. We thus kept them in the model. Results of the final model are summarised in Table 4.4. These results support the idea that the surface location of both verb and particle influence onset of Path Gestures that fell between the verb and the particle. In other words, the onset of Path Gestures was pulled by both, the verb and the particle. However, the Additional Path Preposition does not show an effect on gesture onset.

Table 4.4. Summary of fixed effects in the mixed linear model for Gesture Onset (Path Gestures falling between Verb & Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	-579.572	195.444	102	-2.965	.004**
VO (Sentence Onset to Verb Onset)	1.116	0.117	113	9.569	<.001 ***
VOtoPO (Verb Onset to Particle Onset)	0.840	0.053	127	15.748	<.001 ***

Note: N = 144, log-likelihood = -1109.6; **<.01, ***<.001

In the last Gesture Onset analysis of Path Gestures, we investigated whether this pulling effect also occurred for gestures that synchronised with the particle. We fitted the same model as before but with gestures that synchronised with the Particle as dependent variable. Dropping the factor Additional Path Preposition ($\chi^2 = 15.60$, $df = 1$, $p < .001$), Verb Onset ($\chi^2 = 185.26$, $df = 1$, $p < .001$) and Particle Onset ($\chi^2 = 245.54$, $df = 1$, $p < .001$) led to a worse model fit. The final model is summarised in Table 4.5. The significant effect of Verb Onset and Particle Onset shows that regardless of whether the gesture actually synchronised with its semantic affiliate (i.e., the particle), the verb still had a pulling effect. Moreover, the significant effect of the additional path preposition shows that this additional attractor is pulling the gesture closer to the additional path preposition (i.e., to the sentence onset).

Table 4.5. Summary of fixed effects in the mixed linear model for Gesture Onset (Path Gestures that synchronised with the Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	-38.427	98.645	77	-0.390	0.698
Additional Path Preposition	0.965	0.043	102	22.293	<.001 ***
VO (Sentence Onset to Verb Onset)	0.993	0.033	93	30.471	<.001 ***
VOtoPO (Verb Onset to Particle Onset)	-212.370	51.949	86	-4.088	<.001 ***

Note: N = 117, log-likelihood = -797.0; ***<.001

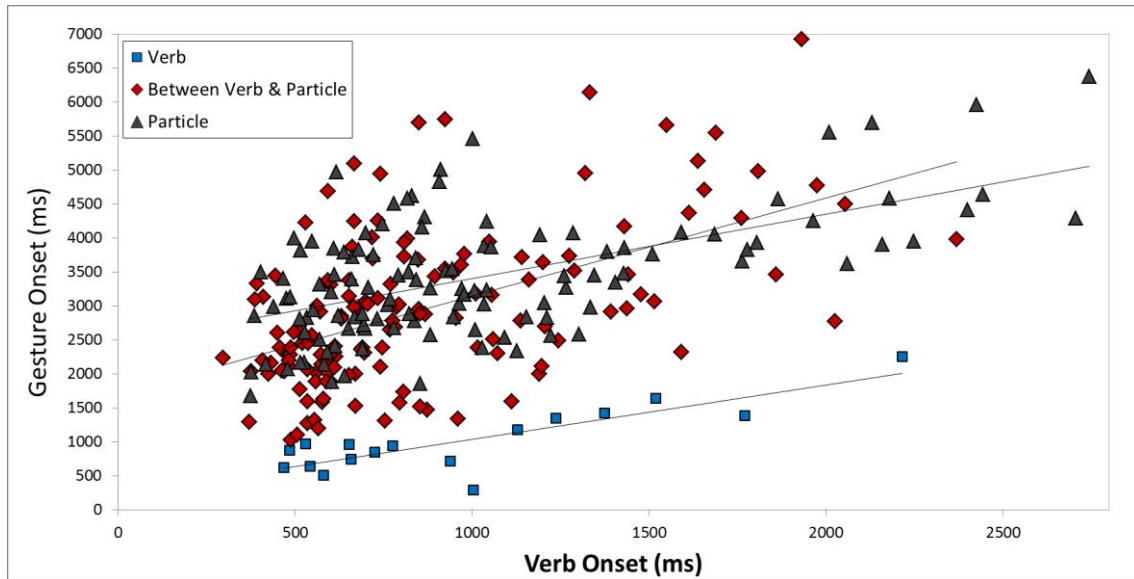


Figure 4.3. Relationship between Gesture Onset and Verb Onset for Path Gestures synchronising with the Verb, the Particle, and gestures that were placed between the Verb & Particle.

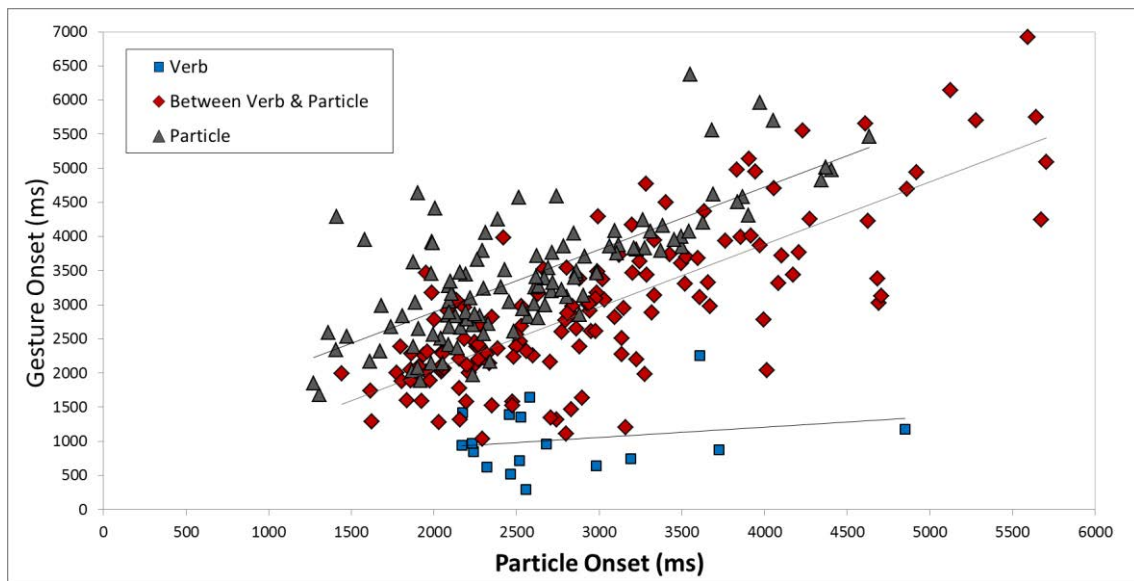


Figure 4.4. Relationship between Gesture Onset and Particle Onset for Path Gestures synchronising with the Verb, the Particle, and gestures that were placed between the verb and the particle.

Figure 4.5 and Figure 4.6 show the relationship of Conflated Gestures with particle onset and verb onset respectively in the different synchronisation categories (i.e., Verb, Between Verb and Particle, Particle). The same models as for Path Gestures were fitted for Conflated Gestures. First, we analysed Conflated Gestures that synchronised with the verb together with gestures

that synchronised with the verb and particle. Results of the final model are summarised in Table 4.6. Model comparisons did not yield a worse fit when dropping the factors Additional Path Preposition ($\chi^2 = 0.202$, $df = 1$, $p = .653$) and Particle Onset ($\chi^2 = 2.43$, $df = 1$, $p = .120$). Only dropping the factor Verb Onset led to a worse fit ($\chi^2 = 49.07$, $df = 1$, $p < .001$). These results are in line with those of the Path Gesture that synchronised with the verb: When gestures synchronise with the verb, the particle and the additional path preposition do not seem to pull the gesture. The gesture is linked exclusively to the verb and the surface location of the particle or the inclusion of an additional path preposition seems irrelevant.

Next, we fitted a model to Conflated Gestures that fell between Verb and Particle. Results are summarised in Table 4.7. Model comparisons did not show a significant effect of Additional Path Preposition ($\chi^2 = 0.737$, $df = 1$, $p = 0.391$) but a significant effect of Verb Onset ($\chi^2 = 10.671$, $df = 1$, $p = .001$) and Particle Onset ($\chi^2 = 12.919$, $df = 1$, $p < .001$). Finally, we fitted the same model to Conflated Gestures that synchronised with the Particle. Again, the factor Additional Path Preposition did not lead to a significantly better fit of the model ($\chi^2 = 2.755$, $df = 1$, $p = .097$). Dropping both other factors resulted in a worse fit of the model (Verb Onset: ($\chi^2 = 14.361$, $df = 1$, $p < .001$), Particle Onset: ($\chi^2 = 12.946$, $df = 1$, $p < .001$)). Results of the full model are summarised in Table 4.8. The results from Conflated Gestures that fell either between verb and particle or synchronised with the particle are similar to those for the Path Gestures: the surface locations of the verb and the particle had an effect on the gesture's onset, i.e. the later the onset of the verb and the particle, the later the gesture onset was. One difference between the two types of gestures is that, for Conflated Gestures synchronised with the particle, unlike the equivalent Path Gestures, the presence of additional path preposition did not predict the gesture onset.

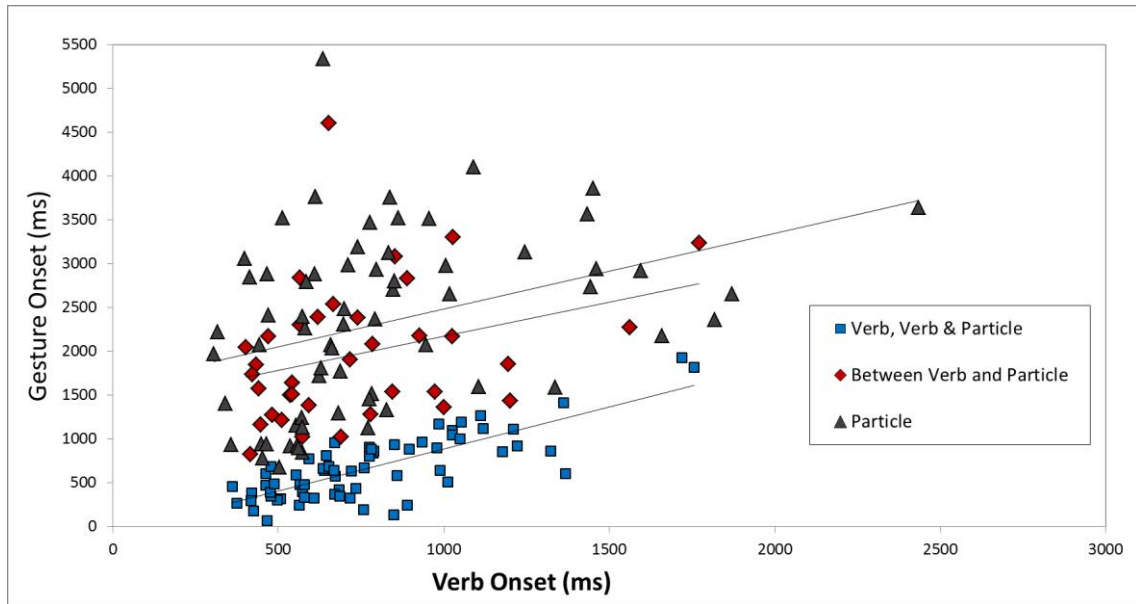


Figure 4.5. Relationship between Gesture Onset and Verb Onset for Conflated Gestures synchronising with the Verb (including gestures that synchronised with Verb & Particle), the Particle and gestures that were placed between the verb and the particle.

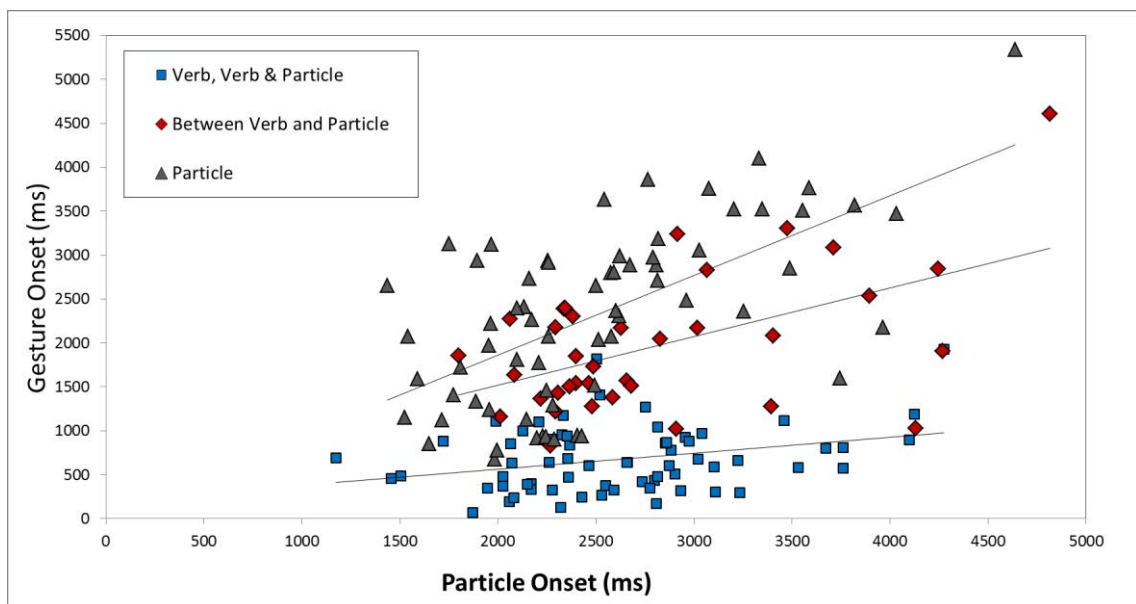


Figure 4.6. Relationship between Gesture Onset and Particle Onset for Conflated Gestures Synchronising with the Verb (including gestures that synchronised with Verb & Particle), the Particle and gestures that were fell between the verb and the particle.

Table 4.6. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated Gestures that synchronised with the Verb and Verb & Particle)

Predictor	Coefficient	SE	Df	t-value	Pr(> t)
Intercept	-47.043	81.930	35	-0.574	.570
VO (Sentence Onset to Verb Onset)	0.937	0.094	61	9.954	<.001 ***

Note: N = 66, log-likelihood = -452.4; ***<.001

Table 4.7. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated Gestures that fell in between verb and particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	-468.434	505.512	34	-0.927	.361
VO (Sentence Onset to Verb Onset)	1.210	0.295	30	4.100	<.001 ***
VOtoPO (Verb Onset to Particle Onset)	0.554	0.138	35	4.003	<.001 ***

Note: N = 36, log-likelihood = -279.5, ***<.001

Table 4.8. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated Gestures that synchronised with the Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	424.187	375.161	65	1.131	.262
VO (Sentence Onset to Verb Onset)	0.8761	0.205	61	4.277	<.001 ***
VOtoPO (Verb Onset to Particle Onset)	0.5365	0.132	64	4.051	<.001 ***

Note: N = 66, log-likelihood = -523.8, ***<.001

We then analysed whether the Gesture Onset for gestures that either fell between Verb and Particle or synchronised with the Particle differs between Path Gestures and Conflated Gestures, when the effects of the onset of the verb, the presence of additional path preposition, and the distance from the verb to the particle are controlled for. We fitted the same linear mixed effects model with Gesture Onset as dependent variable, Verb Onset, Particle Onset, Additional Path Preposition and Gesture Type (Path, Conflated) as independent variable and including Gesture Type as random slope for subject and item. Removing Gesture Type from the model leads to a significantly worse fit of the model ($\chi^2 = 10.2$, $df = 1$, $p = .001$). The results show that gesture onset was significantly earlier for Conflated Gestures than for Path Gestures (see Figure 4.1 and Figure 4.2). The full model is summarised in Table 4.9.

Table 4.9. Summary of fixed effects in the mixed linear model for Gesture Onset (Conflated and Path Gesture that either fell between Verb and Particle or synchronised with the Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	-496.170	175.424	69	-2.828	.006 **
Gesture Type	499.586	125.718	19	3.974	<.001 ***
Additional Path Preposition	1.057	0.073	222	14.493	<.001 ***
VO (Sentence Onset to Verb Onset)	0.790	0.043	262	18.532	<.001 ***
VOtoPO (Verb Onset to Particle Onset)	-271.413	76.456	64	-3.550	<.001 ***

Note: N = 363, log-likelihood = -2811.8; **<.01, ***<.001

Finally, in an exploratory analysis we tested whether the manipulation of planning units (i.e., by either inserting a clause or phrase between verb and particle) had an effect on Gesture-Particle Asynchrony for gestures with their onset after the Verb. For this analysis we compared Gesture-Particle Asynchrony between the Inserted Clause Conditions (i.e. Inserted Clause Condition in Experiment 2 and Main Clause Condition in Experiment 1) with the Inserted Phrase Condition in Experiment 2. If planning units have an effect on gesture-speech synchrony, we predict that when the verb and the particle are encoded within the same planning unit (i.e., in the Inserted Phrase Condition) the asynchrony between gesture onset and particle onset should be smaller compared to verb and particle being encoded across two planning units (i.e., in the Inserted Clause/Main Clause Condition). This difference would arise from a stronger pulling effect of the particle when the verb is not encoded in the same planning unit as the particle. We fitted the full model with Gesture-Particle Asynchrony as dependent variable and Condition (i.e., Inserted Clause/Main Clause Condition and Inserted Phrase Condition) as fixed factor and controlled for the effects of Additional Path Element, Verb Onset and Particle Onset. Furthermore, we included Condition as random slope for subject and item. Dropping the factor condition did not yield a significant effect ($\chi^2 = 0.119$, $df = 1$, $p = .730$). Thus, we did not find evidence that planning units have an effect on gesture-speech synchrony. The full model is summarised in Table 4.10.

Table 4.10 Summary of fixed effects in the mixed linear model for Gesture-Particle Asynchrony in the Inserted Clause Condition from Experiment 2 plus the Main Clause Condition from Experiment 1 and the Inserted Phrase Condition from Experiment 2.

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	132.514	157.682	138	0.840	.402
Condition (Clause vs. Phrase)	-30.111	71.397	58	-0.422	.675
Additional Path Preposition	261.801	81.162	71	3.226	.002 **
VO (Sentence Onset to Verb Onset)	0.167	0.044	183	3.791	<.001 ***
VOtoPO (Verb Onset to Particle Onset)	-0.090	0.078	185	-1.154	.250

Note: N = 261, log-likelihood = -1999.8; **<.01, ***<.001

In sum, we found a positive correlation between Gesture Onset and Verb Onset for Path Gestures and Conflated Gestures that synchronised with the verb. For these gestures the surface location of the particle and the additional path preposition did not predict Gesture Onset. Furthermore, we found a positive correlation between Gesture Onset and Verb Onset as well as Gesture Onset and Particle Onset for gestures that either fell between Verb & Particle or synchronised with the Particle. The effect of the additional path preposition on gesture onset was only marginally significant for path gestures that synchronised with the particle (i.e., the gesture being pulled towards the sentence onset). Furthermore, the onset of gestures that either fell between the Verb and the Particle or synchronised with the Particle was earlier for Conflated gestures than for Path gestures, indicating that the verb has a stronger pulling effect on Conflated Gestures than on Path Gestures. Finally, the manipulation of planning units (i.e., by either inserting a clause or phrase between verb and particle) did not have an effect on the asynchrony between gesture onset and particle onset.

4.5.2. Gesture Duration Analyses

In terms of gesture duration, it was tested whether the participants produced a longer gesture when the particle was produced in the sentence final position and the gesture's onset occurred earlier in the sentence. If this was the case, we would expect gesture strokes to be longer, the further their onset is away from the particle. We analysed gesture duration again by fitting linear mixed effects models. The full model included Gesture Duration as dependent variable and Gesture-Particle Asynchrony (distance in ms between gesture onset and particle onset) and Additional Path Preposition as fixed factors. Table 4.11 shows the means of the dependent variable (i.e., Gesture Duration) for Path Gestures and Conflated Gestures for the three different synchronisation categories. As for the Gesture Onset means, Gesture Duration means are shown across Conditions (i.e., Inserted Clause and Inserted Phrase Condition).

Table 4.11. Means (ms) of the duration of Gesture Duration for Path Gestures synchronising with the Verb, the Particle and gestures that were placed between the verb and particle in the Inserted Clause Condition (including Main Clause Condition from Experiment 1) and Inserted Phrase Condition.

<i>Analysis</i>	<i>Means of the Duration of Gesture Duration in the Inserted Clause Condition</i>	<i>Means of the Duration of Gesture Duration in the Inserted Phrase Condition</i>
Path Gestures (Verb)	912 ms	1815 ms
Path Gestures (Between Verb & Particle)	623 ms	540 ms
Path Gestures (Particle)	537 ms	519 ms
Conflated Gestures (Verb, Verb & Particle)	2070 ms	2506 ms
Conflated Gestures (Between Verb & Particle)	1334 ms	1385 ms
Conflated Gestures (Particle)	1398 ms	1463 ms

Figure 4.7 illustrates the relationship between Gesture Duration and Gesture-Particle Asynchrony for Path Gestures and Figure 4.8 for Conflated Gestures. We started our duration analysis with Path Gestures that synchronised with the Verb. Removing Additional Path Preposition yielded a marginally significant effect ($\chi^2 = 3.69$, $df = 1$, $p = .055$). Removing gesture-particle asynchrony from the model had no effect on the gestures' duration ($\chi^2 = .83$, $df = 1$, $p = 0.362$). The full model is summarised in Table 4.12. Similar to the gesture onset analysis on gestures that synchronised with the verb, the particle has no effect on the duration of the gesture. We then fitted the same model for Path Gestures that fell between verb and particle. Dropping the factor Additional Path Preposition did not lead to a worse fit of the model ($\chi^2 = 0.154$, $df = 1$, $p = .695$). Only Gesture-Particle Asynchrony survived as a significant factor ($\chi^2 = 57.22$, $df = 1$, $p < .001$). The larger the asynchrony was, the longer was the gesture. Results of the final model are summarised in Table 4.13. Path Gestures that synchronised with the Particle showed the same effect of Gesture-Particle Asynchrony on gesture duration ($\chi^2 = 49.60$, $df = 1$, $p < .001$) but no effect of the Additional Path Preposition ($\chi^2 = 0.394$, $df = 1$, $p = .530$). Results are summarised in Table 4.14.

Table 4.12. Summary of fixed effects in the mixed linear model for Gesture Duration (Path Gestures that synchronised with the Verb)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	823.275	716.188	17	1.150	.266
GO-PO (Gesture Onset – Particle Onset)	-0.1061	0.1996	17	-0.531	.602
Additional Path Preposition	904.903	359.174	17	2.519	0.022 *

Note: N = 17, log-likelihood = 277.3; *<.05

Table 4.13. Summary of fixed effects in the mixed linear model for Gesture Duration (Path Gestures that fell between the Verb and the Particle)

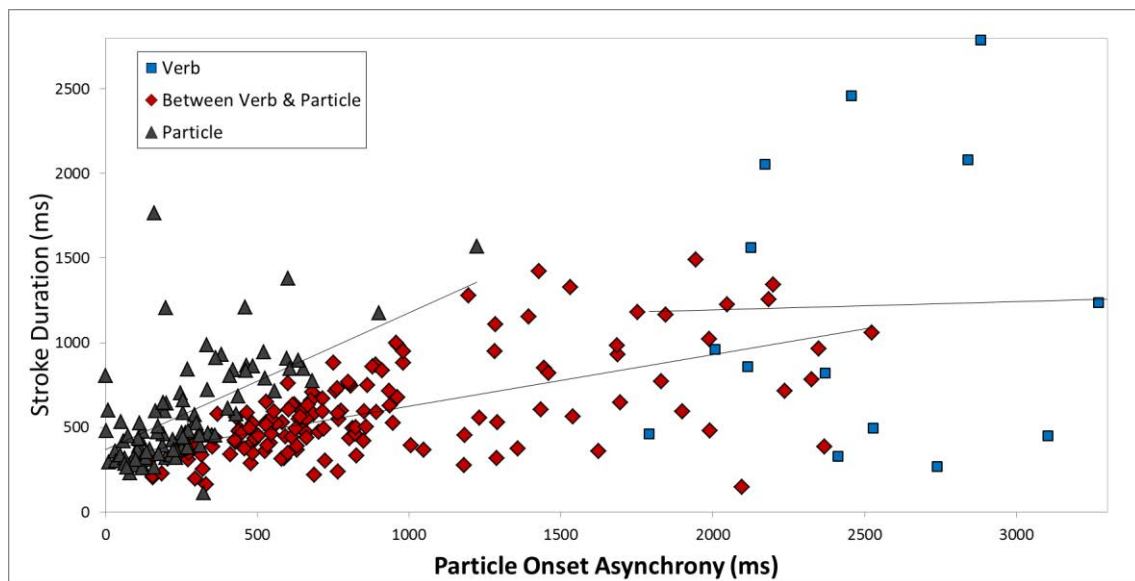
<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	341.389	43.952	64	7.767	<.001 ***
GO-PO (Gesture Onset – Particle Onset)	0.272	0.032	142	8.482	<.001 ***

Note: N = 144, log-likelihood = -976.9; ***<.001

Table 4.14. Summary of fixed effects in the mixed linear model for Gesture Duration (Path Gestures that synchronised with the Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	466.097	46.493	33	10.025	<.001 ***
GO-PO (Gesture Onset – Particle Onset)	0.564	0.069	96	8.177	<.001 ***

Note: N = 117, log-likelihood = -778.5; ***<.001

**Figure 4.7.** Relationship between Stroke Duration and Gesture Onset – Particle onset asynchrony for Path Gestures synchronising with the Verb, the Particle and for gestures that fell between Verb and Particle.

Finally, we ran the same three duration analyses with Conflated Gestures. We started again with Conflated Gestures that synchronised with the Verb together with the gesture that synchronised with the Verb and the Particle. Removing Additional Path Preposition ($\chi^2 = 0.507$,

df = 1, p = .477) had no effect on the gesture's duration. Only removing Gesture-Particle Asynchrony from the model led to a significant worse fit of the model ($\chi^2 = 5.155$, df = 1, p = .023), meaning that the asynchrony did affect gesture durations of Conflated Gestures synchronising with the verb and those gestures which synchronised with the verb and the particle.

For gestures that fell between the Verb and the Particle, dropping Gesture-Particle Asynchrony from the model led to a worse fit ($\chi^2 = 25.117$, df = 1, p < .001) but not dropping the Additional Path Preposition ($\chi^2 = 0.249$, df = 1, p = .618). The same results were found for gestures that synchronised with the Particle (Additional Path Preposition: $\chi^2 = 0.156$, df = 1, p = 0.692; Gesture-Particle Asynchrony: $\chi^2 = 134.82$, df = 1, p < .001). Thus, asynchronies did affect durations of gestures of all three synchronisation categories. Results are summarised in Table 4.15, Table 4.16 and Table 4.17.

Table 4.15. Summary of fixed effects in the mixed linear model for Gesture Duration (Conflated Gestures that synchronised with the Verb and Verb & Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	1046.563	383.340	56	2.730	.008 **
GO-PO (Gesture Onset – Particle Onset)	0.293	0.119	65	2.457	.017 *

Note: N = 66, log-likelihood = -514.7, * < .05, ** < .01

Table 4.16. Summary of fixed effects in the mixed linear model for Gesture Duration (Conflated Gestures that fell between the Verb and the Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	434.075	184.612	36	2.351	.024 *
GO-PO (Gesture Onset – Particle Onset)	0.611	0.100	32	6.082	<.001 ***

Note: N = 36, log-likelihood = -267.6; * < .05, *** < .001

Table 4.17. Summary of fixed effects in the mixed linear model for Gesture Duration (Conflated Gestures that synchronised with the Particle)

<i>Predictor</i>	<i>Coefficient</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>Pr(> t)</i>
Intercept	588.592	62.776	28	9.376	<.001 ***
GO-PO (Gesture Onset – Particle Onset)	0.831	0.039	58	21.358	<.001 ***

Note: N = 66, log-likelihood = -455.6; *** < .001

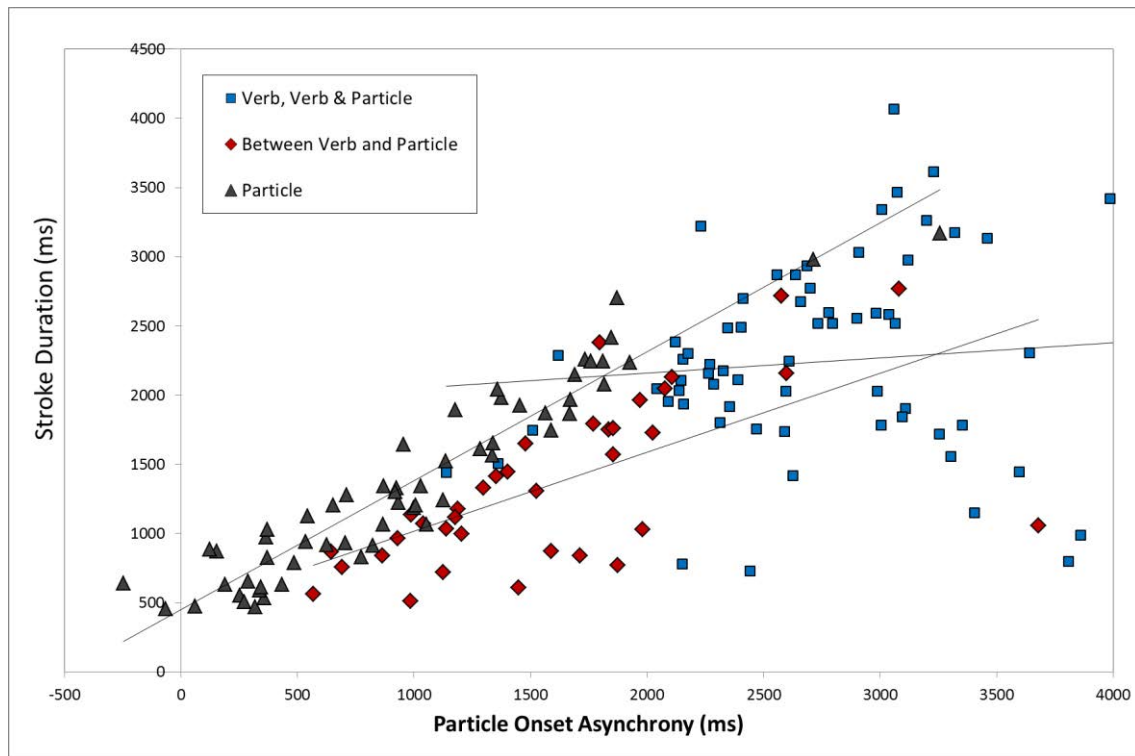


Figure 4.8. Relationship between Stroke Duration and Gesture Onset – Particle Onset asynchrony for Conflated Gestures synchronising with the Verb (including Verb & Particle), the Particle and for gestures that fell between Verb and Particle.

In sum, for Path Gestures, the stroke duration increased with gesture-speech asynchrony when the gesture either fell in between the verb and particle or synchronised with the particle. For Conflated Gestures, we found a significant increase of stroke duration with gesture-speech asynchrony for all three synchronisation categories.

In order to directly compare whether the strength of the relationships of Gesture Duration and Gesture-Particle Asynchrony is significantly different for Conflated Gestures and Path Gestures that either fell between Verb and Particle or synchronised with the Particle, it was necessary to use a different statistical test rather than mixed-effects modelling. In particular, for the comparison of correlation strength, we had to calculate correlation coefficients (see Table 4.18), i.e. using Fisher's r -to- z transformation. Not surprisingly, all correlations were strong. For gestures that fell between the verb and the particle, correlation strength did not differ for Path Gestures and Conflated Gestures ($z = -0.64$, $p = 0.261$). For gestures that synchronised

with the Particle, the relationship between gesture duration and gesture-particle asynchrony was weaker for Path than Conflated Gestures ($z = -5.01$, $p < .001$).

Table 4.18. Correlations between Gesture Duration and Gesture-Particle Asynchrony for Path Gestures and Conflated Gestures for the synchronisation categories Particle and Between Verb and Particle

<i>Gesture</i>	<i>N</i>	<i>Spearman's Rho</i>	<i>p-value</i>
Path Gestures (Between Verb & Particle)	144	0.613	<.001 ***
Conflated Gestures (Between Verb & Particle)	36	0.684	<.001 ***
Path Gestures (Particle)	117	0.658	<.001 ***
Conflated Gestures (Particle)	66	0.918	<.001 ***

4.6. General Discussion

The current study investigated the temporal coordination of gestures and speech within a sentence context by analysing the gestures' onset as well as the gestures' duration in relation with their semantic affiliates. The findings of this study can be summarised around five main points. First, a substantial proportion (30-60%) of conflated and path gestures did not synchronise with (i.e., temporally overlap with) semantic affiliates in German sentences in which the manner verb and the path particle were separated by other linguistic elements. Second, Gesture Onset positively correlated with Verb Onset for gestures (path and conflated) that synchronised with the first semantic affiliate within the sentence (i.e., the verb). However, for these gestures the surface location of the second semantic affiliate (i.e., the particle) or the occurrence of a potential additional path preposition was irrelevant. Third, gesture onset positively correlated with both Verb Onset and Particle Onset when the gesture either fell between Verb & Particle or when the gesture synchronised with the Particle. In particular, when the verb was initiated later in relation to sentence onset, the gesture was also initiated later. Furthermore, the further away the second semantic affiliate (i.e., particle) was encoded from the first semantic affiliate (i.e., verb), the later the gesture was initiated. The additional path preposition, however, only had a significant effect on gesture onset for path gestures that

synchronised with the particle, i.e., the additional path preposition pulled the gesture's onset towards the sentence onset. Fourth, the onset of gesture was earlier for Conflated gestures than for Path gestures (the mean onset fell between the manner verb and the path verb for both gesture types). Fifth, the degree of asynchrony between gesture onset and particle onset highly correlated with gesture duration for both, path gestures and conflated gestures that either fell between the verb and the particle or that synchronised with the particle. In other words, the larger the asynchrony was, the longer was the gesture. For gestures that were synchronised with the particle, this correlation was stronger for Conflated Gestures than for Path Gestures.

In an exploratory analysis, we tested whether the encoding of motion events into either one or two planning units had an influence on the asynchrony between the onset of Path Gestures and the onset of the Particle. We did not find a difference between the Conditions where the participants inserted a clause between the verb and the particle (i.e., two planning units) compared to an inserted phrase (i.e., one planning unit). Thus, we did not find any evidence that the encoding of the verb and the particle into either one or two planning units has an influence on gesture-speech synchrony. However, further studies are needed to test whether planning units have an effect on the synchrony between gesture and speech.

Our results on gesture onsets do not support de Ruiter's (1998, 2000) Sketch Model. The Sketch Model predicts that the onset of iconic gestures slightly precede or coincide with the onset of the first word in semantic affiliate phrases, but never follow the onset of the first word. According to Talmy's (2000) linguistic analysis the manner verb encodes both Manner and Motion. Consequently, the manner verb is a semantic affiliate for both types of gesture because both types encode Motion. Taken together, the Sketch Model predicts that the onset of both Path Gestures and Conflated Gestures should slightly precede or coincide with the onset of the verb, but never follow the onset of the verb. However, this synchronisation pattern was not

present in our data. The majority of Path Gestures synchronised with either the Particle or fell between Verb and Particle, and Conflated Gestures were initiated after the manner verb approximately two thirds of the time. Moreover, the Path Gestures and Conflated Gestures that were initiated *after* the manner verb were still time-locked to the manner verb: the onsets of the gestures were positively correlated with the onsets of the manner verbs, further supporting the idea that manner verb is a semantic affiliate. In summary, a large proportion of gestures were produced *after* the first semantic affiliate was encoded. Furthermore, the first semantic affiliate still had an impact on the gesture onset when the gesture's onset occurred after the verb has been encoded.

Our results on gesture duration did not support the Sketch Model either. According to the Sketch Model, the gesture system is blind to which words occur when in speech. The Sketch Model is in favour of the ballistic view (i.e., gesture and speech do not exchange any information after gesture onset). However, gesture duration was longer when the distance between the gesture onset and the path particle onset was longer, indicating that the gesture was prolonged when the particle was further away. This supports the interactive view of gesture-speech production with the feedback channel between gesture and speech staying open through all phases of gesture production and execution. Although in the Sketch Model the gesture system is blind to where within an utterance a word occurs, the adaption of gesture duration during the speech execution phase has been addressed by de Ruiter (2000). He suggests that repetitive-gestures such as Conflated Gestures are specified as a "loop" where the stroke is repeated until the preverbal message of which the gesture was a part of has been produced. Although the pre-verbal message has not specified the surface location of the gesture's semantic affiliate, gesture duration can be adapted online to some extent and for a specific type of gesture (i.e., repetitive gestures). However, this explanation does not account for our results from

Conflated Gestures because it would predict that strokes are always prolonged until the preverbal message has been completed. In the present study, this would be the production of the path particle. Moreover, this explanation cannot account for path gestures because de Ruiter (1998, 2000) does not specify where the duration of non-repetitive strokes, including Path Gestures, is determined.

Our data partly support McNeill's (1992, 2005) Semantic Synchrony Rule. Our findings suggest that gestures *aim* to synchronise with their semantic affiliates which is evident in the positive correlation between gesture-particle asynchrony and gesture duration (i.e., evidence for the interactive view). However, contrary to the prediction of the Semantic Synchrony Rule, gestures often did not manage to synchronise with any of the semantic affiliates: a substantial proportion of gestures fell between the manner verb and the path particle. One may argue that the speaker aimed the gesture to synchronise with a portion of speech that was encoded between verb and particle (e.g., with the path preposition or the ground element). However, this seems unlikely for two reasons. First, Figure 4.1 and Figure 4.2 show that if we consider only responses without any additional path preposition, we still find that half of the Path Gestures and almost one third of Conflated Gestures fell between Verb & Particle. Second, if the speakers' aim was for the gesture to synchronise with a ground element (e.g., the rainbow) for pragmatic reasons (e.g., put the focus on this element in speech), the surface locations of the verb and the particle should not have predicted gesture onset. Another current finding that is not predicted by the Semantic Synchrony rule is that the gesture onset can be predicted by non-overlapping semantic affiliates. McNeill (1992, 2005) proposed that the words that the gesture stroke synchronises with are the words whose meaning was a part of the underlying idea unit, the Growth Point, that generated both speech and gesture; thus, other words in the sentence are not directly linked to the gesture in the underlying representation. However, we found that semantically co-

expressive words that do not overlap with the gesture can still determine speech-gesture synchronisation. Taken together, the Semantic Synchrony Rule cannot explain why a large proportion of Path Gestures and Conflated Gestures fell between their semantic affiliates, but their onsets were predicted by the onsets of the separated semantic affiliates.

Our duration results are similar to Morrell-Samuels and Krauss' (1992) finding that an early gesture onset in relation to the semantic affiliate lengthens the gesture, but Morrell-Samuels and Krauss did not report what types of gestures were analysed. Thus, our findings go beyond their finding in that two subtypes of iconic gestures differ as to how long a gesture is prolonged and consequently whether a gesture synchronises with its semantic affiliate. Although the correlation between gesture-particle asynchrony and gesture duration was very strong for both Path Gestures and Conflated Gestures that synchronised with the Particle and for gestures that fell between Verb and Particle, this correlation was significantly stronger for Conflated Gestures than for Path Gestures that synchronised with the particle. The reason for this difference might be due to the stronger tendency of Conflated Gestures to have a repetitive stroke to encode the manner of movement. Generally, gestures can be classified as being repetitive when the duration of the stroke can be prolonged by repeating the same movement (e.g., manner of a motion event). As Kita (1990) suggested, repetitive strokes may have been repeated until they synchronise with their semantic affiliates. Thus, it was easier for participants to prolong a conflated gesture compared to a path gesture, though both types of gestures seem to have been lengthened towards the path particle as shown by a positive correlation between gesture-speech asynchrony and gesture duration.

The current findings are not compatible with either Semantic Synchrony Rule (McNeill, 1992) or the Sketch Model (2000). We thus propose a different theoretical model for synchronisation between speech and representational gestures (i.e., deictic and iconic gestures),

the Attraction Point Hypothesis of Speech-Gesture Synchronisation. The general idea is that the semantic affiliates of a representational gesture serve as “attraction points” and “pull” the gesture towards them. In other words, the gesture *aims* to overlap with its semantic affiliates, but it does not necessarily achieve this temporal overlap.

The three basic rules are as follows:

1) Gesture Onset Rule

The onset of a gesture is pulled towards semantic affiliates (words that are co-expressive with the gesture) in the sentence.

2) Gesture Offset Rule

The offset of a gesture is pulled towards semantic affiliates that follow the gesture onset.

(This determines gesture duration).

3) Semantic Overlap Rule

The onset and offset of a gesture is pulled more strongly by the semantic affiliate that shares more semantic features with the gesture.

The first two rules are motivated by the current finding that the onset and duration of gestures were predicted by the onset of their semantic affiliates. The third rule is motivated by the fact that the onset of Conflated Gestures was closer to the onset of the verb than the onset of the Path Gestures. Following Talmy (2000), manner verbs share two features with Conflated Gestures (Manner, Motion) but they only share one feature with Path Gestures (Motion). Thus, Conflated Gestures are attracted (and pulled) more strongly towards the manner verb than the Path Gestures.

Two further sub-rules can be postulated:

1b) *When there are multiple semantic affiliates in a sentence, the semantic affiliates produced later in the sentence may not exert any pulling effects.*

2b) *The duration can be lengthened more for gestures with a repetitive stroke than for gestures with a non-repetitive stroke.*

The sub-rule 1b) is motivated by the finding that when the gesture is synchronised with the manner verb, the onset of the gesture is not predicted by the onset of the path particle (not significant for Path Gestures; only a weak effect for Conflated Gestures). We speculate that these are the cases where forward planning of speech production specified only the surface location of the manner verb, but not the up-coming path particle. Thus, only the manner verb served as an attraction point. Consequently, the gesture synchronised with the manner verb.

The sub-rule 2b), initially proposed by Kita (1990), is motivated by the finding that Conflated Gestures which were more repetitive than Path Gestures were prolonged longer in order reach their semantic affiliate.

An important consequence of this Hypothesis is that a representational gesture may not overlap with semantic affiliates. When semantic affiliates are located far away from each other as in German utterances in the current study, the gesture may fall between semantic affiliates and thus does not overlap with any of them.

It is important to note that the Attraction Point Hypothesis is still a post-hoc explanation for the current findings. Thus, future research should directly test this hypothesis. One important area that requires future research is the influence of pragmatics and prosody. De Ruiter (1998) found that pointing gestures were pulled towards the word that carried contrastive stress. In this study, participants pointed at one of four pictures which is indicated by a small lamp, and named the colour and object depicted by the picture. When the colour was contrastive (the target was a green crocodile and distractors were all crocodiles in different colours), then the word “green” in “the green crocodile” carried the contrastive stress and the gesture onset was closer to the

onset of “green”, in comparison to when the object was contrastive (distractors were all green objects, but not crocodiles). Furthermore, Cooperrider (2014) found that self-points produced during spontaneous speech often co-occur with pronouns that carry the lexical stress. Thus, future studies should further test the influence of discourse significance of a word in the context and prosodic peaks on speech-gesture synchronisation (see also Nobe, 1996 regarding prosodic peaks). For example, an experiment could be designed where change of location (Path) is contrasted. This contrast could be created by the participants watching and retelling two different video clips where in the first one the ball is rolling *down* the hill whilst in the second the ball is rolling *up* the hill. The analysis of Path gestures produced during such retellings would be very informative regarding the relationship between stress and synchronisation.

Another aspect that future studies should take into account is the role of post-stroke holds in gesture-speech synchronisation. Post-stroke holds have been considered as evidence that the gesture is waiting for the semantic affiliate to catch up (Kita, 1990). Thus, when the stroke is completed a post-stroke hold occurs until the gesture production system receives a retraction signal from the speech production system. For de Ruiter (2000), this retraction signal is sent after the completion of the preverbal fragment has been produced. Based on the interactive view of gesture-speech production, we would expect that strokes which fell between the Verb and the Particle are followed by a post-stroke hold. These post-stroke holds are then prolonged until the particle has been produced. However, this prediction needs to be tested in future studies.

Regarding the *Attraction Point Hypothesis*, we propose that this hypothesis does not only apply to the specific grammatical construction (i.e., German particle verbs) tested in the current study. Rather, we suggest that the same pulling mechanism applies to motion event constructions in other languages (e.g., Dutch such as in Kellerman & van Hoof, 2003) but also to other domains where a semantic affiliate is linguistically expressed with more than one

lexical item. Such domains include “giving directions” (e.g., “At the corner turn left”) or placement verbs (e.g., “put the glass on the table” (cf. Gullberg, 2011 for placement gestures). It is important to test the hypothesis in different constructions in different languages in the future.

What are the implications of the present study on the synchronisation literature? First, the current study provides evidence that the interactive view of gesture and speech production as proposed by Chu and Hagoort (2014) for pointing gestures also holds for iconic gestures which are produced within a sentence context. Moreover, whether a gesture actually synchronises with its affiliate is partly determined by the nature of the gesture: i.e., repetitive gestures can be prolonged longer than non-repetitive gestures.

Furthermore, the results from this study have implications for the role of the semantic affiliate in gesture-speech production. As previously pointed out, an iconic gesture can often be linked to multiple semantic affiliates (or “lexical affiliates” in Schegloff (1984)), on the surface structure. This lack of a one-word-one-gesture mapping can be seen as a result of the unpacking of the “idea unit” (in McNeill’s (1992) sense) into the given grammatical and lexical structure of a language. Unlike McNeill (1992, 2005, 2015) who argues that the idea unit can be inferred by the element(s) of speech the gesture synchronises with, a different approach was employed in the current study: The semantic affiliates were defined based on the gesture’s content (i.e., conflated gesture and path gesture are both semantically related to the manner verb and the path particle/path preposition). This approach provided evidence that a gesture is not exclusively linked to the portion(s) in speech that the gesture synchronises with but also with other semantic affiliates.

To sum up, our analyses of motion event gestures showed that linguistic surface structure has an influence on gestures’ onset and gesture’s duration. The findings do not support de

Ruiter's Sketch Model (1998, 2000) and partly support McNeill's (1992, 2005) Semantic Synchrony Rule. They support the Semantic Synchrony Rule in a way that the gesture aims to synchronise with a semantic affiliate. However, it was found that the presence of multiple semantic affiliates within an utterance leads to a pulling effect of these affiliates that causes gesture-speech asynchronies. Based on our findings we proposed the *Attraction Point Hypothesis* of Gesture-speech Synchronisation which assumes that a gesture's semantic affiliates function as attractors for the gesture that results in a competition for synchronisation. This was evident in a detailed gesture onset analysis in relation to the gesture's semantic affiliates (measured in ms). This analysis revealed that the varying degrees of gesture-speech asynchronies are correlated with the surface distance between the two semantic affiliates. Moreover, we showed that gesture duration is positively correlated with gesture-speech asynchronies. This does not only support the interactive view of gesture-speech production but also indicates that speakers prolong their gestures in order to synchronise with the semantic affiliate.

Chapter 5

Speech-gesture Comprehension

The overall aim of this chapter is to provide the theoretical foundation for the study outlined in Chapter 6. First, the current state of the art in research on the role of gestures in language comprehension will be introduced. This is followed by an outline of how the findings of the gesture-speech synchronisation study presented in Chapter 4 motivated the research questions investigated in the gesture-speech comprehension study in chapter 6. The remaining sections will concentrate on four topics that provide the theoretical and methodological foundation of the study presented in Chapter 6. First, the importance of gesture-speech synchrony in language comprehension will be highlighted. Second, the ERP technique as a method to answer questions about language processing in general and speech-gesture processing in particular will be introduced. Third, the literature on how to interpret different components of the ERP waves in relation to language processing will be reviewed and how they can be linked to gesture processing will be pointed out. In the final part of this chapter, different approaches to ERP study designs in previous research on gesture-speech comprehension will be discussed.

5.1. Do gestures influence language comprehension?

Gestures are assumed to be multifunctional. From a gesture-speech production perspective, research has shown that gestures serve self-oriented functions. For example, they help us to package information into chunks readily available for speech production (Kita, 2000; Kita & Özyürek, 2003, chapter 3). Furthermore, it has been shown that we benefit from gestures when we are thinking without producing any overt speech (Chu & Kita, 2011; Kita, Alibali, & Chu, 2017). In terms of the listener's perspective, researchers have been interested to see whether gestures have any communicative functions. Apart from some exceptions (e.g., Krauss,

Chen, & Chawla, 1996; Krauss, Dushay, Chen, & Rauscher, 1995), there is consensus among researchers that gestures have an influence on how we process language (see Özyürek, 2014 for a review). For example, one line of research tested whether a message including a gesture is better understood compared to the same message without a gesture. In a meta-analysis of 63 datasets, Hostetter (2011) showed that the benefits of gestures differ depending on various factors such as the topic, the listener's age and how closely the gesture's meaning is to the meaning conveyed in speech. Another line of research has looked at the online processing of co-speech gestures using electroencephalography (EEG) to test whether listeners extract meaning from gestures and more generally how listeners process gesture-speech combinations. An increasing number of studies provided evidence that iconic gestures are processed similarly to spoken language and are also integrated into a wider sentence context and thus become part of a speaker's discourse model (e.g., Holle & Gunter, 2007; Kelly et al., 2004; Özyürek, Willems, Kita, & Hagoort, 2007). Moreover, this effect has not only been found for iconic gestures but also for metaphorical gestures (Cornejo et al., 2009; Ibáñez et al., 2010; Ibáñez et al., 2011) and abstract pointing gestures (Gunter & Weinbrenner, 2017; Gunter, Weinbrenner, & Holle, 2015).

5.2. Motivation and Research Questions

So far in this thesis, we have looked at the relationship between gestures and speech from a language production point of view. In the previous chapters it was investigated how speech and gesture systems interact during production. In the final study of this thesis, the role of gestures in language comprehension will be tested. This comprehension study has developed from the results of the synchronisation part of the thesis presented in Chapter 4. In particular, the finding that a high proportion of gestures did not synchronise with their semantic affiliates raised the question of how listeners would perceive gestures in cases where synchrony between

a gesture and its affiliate is lost. Previous studies suggest that some overlap of the gesture with its semantic affiliate is crucial for the listener to automatically integrate the gesture into its speech context (word or sentence) and thus to form a unified semantic representation (Habets, Kita, Shao, Özyürek, & Hagoort, 2011; Obermeier, Holle, & Gunter, 2010). Similar to gesture-speech production, McNeill (1992) suggested that the same idea unit must occur simultaneously in both gesture and speech in order to allow an automatic integration. Another reason why gesture-speech synchrony is important is the lack of transparency of iconic gestures. In other words, semantic affiliates often guide the interpretation of an otherwise ambiguous gesture (Feyereisen, Vandewiele, & Dubois, 1988; Hadar & Pinchas-Zamir, 2004; Krauss, Morrel-Samuels, & Colasante, 1991).

Table 5.1. Example Response from the speech-gesture production study presented in Chapter 4 where gesture-speech synchrony is lost.



preparation

preparation

stroke

retraction to resting position

Die Maus steigt in diesem kurzen Video **die Treppe** hinauf.
 The mouse climbs in this short video **the stairs** up.
 In this short video, the mouse is climbing up the stairs.

Following these arguments, one would not expect asynchronous gestures produced in the study presented in Chapter 4 to be integrated into the sentence context during comprehension and thus the gesture would not become part of the listener's discourse model. However, previous studies in the literature have not looked at whether the preceding context (sentence or

discourse) in which the gesture was placed in has an influence on how gestures are processed. If we take a closer look at an example (see Table 5.1) of gesture-speech asynchrony from the dataset of the synchronisation study presented in this thesis, it can be seen that the sentence context (i.e., the verb *climb*) might help the listener to construct the gesture's meaning without its semantic affiliate (i.e., *up*). But is having a context that constrains a gesture's interpretation enough for an automatic integration when the gesture does not synchronise with its semantic affiliate? To investigate this research question a study was designed where discourse information preceding the gesture was manipulated (see Chapter 6). In one condition the gesture's preceding discourse information would arguably help the listener to interpret the gesture, while in a second condition the preceding discourse was unrelated to the gesture's meaning. In a behavioural experiment, it was first tested whether discourse information can help interpret the gesture's meaning and whether the difficulty of interpreting a gesture is related to the gesture's preceding discourse information. Furthermore, the behavioural experiment was used to validate the stimuli in terms of the gesture match/mismatch manipulation which was used in the second experiment (an ERP study). In the ERP study it was tested whether preceding discourse information related to the gesture's meaning enables an automatic integration of iconic gestures into a listener's discourse model when the gesture does not synchronise with its semantic affiliate. Before describing the nature of this particular ERP study, a more general introduction to the ERP technique is provided as well as an overview of its use in language research.

5.3. ERP Technique (Event-related potential)

Electroencephalography (EEG) records electrical brain activity that the brain constantly generates. These electrical signals are picked up by electrodes which are placed on the participant's scalp. Usually participants wear a cap that looks similar to a swimming cap but

with placeholders for the electrodes to be clicked in. Via the electrodes the signals are transmitted to an amplifier which then converts information about changes in voltage (electrical potentials) into a digital signal. The amplifier patterns the on-going EEG into waves. Each wave can be described by its frequency and amplitude. The ongoing EEG consists of waves occurring at different frequencies which can tell us, for example, something about the mental state of the participant (e.g., whether she/he is sleepy or not) (Key, Dove, & Maguire, 2005). However, in linguistic studies, researchers are usually interested in how the brain processes specific stimuli. To pinpoint brain responses that reflect language processing is difficult from the ongoing EEG because the recorded brain activity comprises different processes related to a number of cognitive and sensory functions as well as self-regulation processes including the maintenance of body temperature or breathing (Key et al., 2005). For that reason, in linguistic research, researchers do not usually analyse the whole EEG but only small sequences (epochs) time-locked to the stimulus (usually stimulus onset) (Coulson, 2007; Kutas & Federmeier, 2007). The most common method includes the averaging of many either identical trials or trials which are assumed to elicit a similar response. The averaged brain response to a stimulus type is called an event-related potential (ERP) (Kuperberg, 2008). A large number of trials is needed to average out processes which are unrelated to the processing of the stimulus (noise). The idea behind the averaging of trials is that the response of the stimulus is systematic, but the noise is random. Thus, the more trials that are included in an experiment, the lower the signal-to-noise ratio (Kuperberg, 2008). The required number of trials depends, among other factors, on the noise level in the data and the size of the effect. The noise level varies, for example, across subject groups. Data obtained from children is usually noisier than data from healthy adults. Furthermore, the size of the effect is crucial in deciding how many trials are needed. Larger effects like those linked to language processing (e.g., N400 and P600) need fewer trials than

smaller effects (Luck, 2005b). Luck (2005b) suggests for larger effects like the ones investigated in this thesis, that 30-60 trials are necessary. However, more trials than those needed for a sound analysis need to be included in the experiment. This is because ERPs also include artifacts such eye-blinks and movements that can distort the results. Usually, trials that include artifacts are rejected before the ERPs are averaged. When the proportion of excluded trials is high the signal-to-noise ratio increases. In order to keep a reasonable number of trials for the analysis, it is possible to correct, for example, ocular artifacts. One method is to decompose the EEG signal into independent components that can be distinguished spatially and temporally (ICA – Independent Component Analysis). Thus, the component(s) which represent the blink artifact (or other artifacts like muscle artifacts or lateral eye-movements) can be identified and removed. In a next step the EEG is reconstructed without the removed component (Nunez et al, 2016, Hoffman & Falkenstein, 2008). Due to the nature of the stimulus of the ERP study in this thesis (i.e., videos lasting for a few seconds) a large number of blinks occurred in the data. Thus, an ICA was used to correct blink artifacts.

What do effects in the ERP signal look like? A waveform time-locked to a stimulus usually consists of negative and positive peaks. These peaks are also known as ERP components. Each component can be described by its polarity (negative or positive), latency (time point when the component reaches its peak) and topographical distribution of the component across the scalp (Coulson, 2007). The names of the components are usually based on their polarity and latency. For example, the N400 has a negative peak at around 400 ms after the onset of the stimulus. In ERP experiments, the waveforms of two or more conditions are compared to test whether the experimental manipulation has an influence on the participants' brain response (Kaan, 2007). If the observed difference of mean voltage between two fixed latencies (e.g., 300-500 ms post stimulus onset) that are assumed to comprise a specific

component (e.g., N400) is significantly different across conditions, this is referred to as an effect (e.g. N400 effect) (Kutas & Federmeier, 2011).

Single ERP components are known to reflect specific cognitive processes. In brief, early components up to approx. 200 ms after stimulus onset are thought to reflect processing of the physical nature of the stimulus (e.g., size, modality). They are often referred to as exogenous components. Later components are generally attributed to information processing since they depend on task properties (e.g., difficulty of the task, semantic processing) (Kuperberg, 2008; Ward, 2015). Usually these so-called endogenous components are the focus of linguistic research since they indicate cognitive processing rather than perceptual processing (Coulson, 2007). Importantly, exogenous and endogenous components do not comprise clear-cut categories rather they should be seen as dimensions with some degree of overlap (Key et al., 2005).

5.4. What can ERP studies tell us about language processing?

With regard to research on language processing, ERP studies enable us to answer questions about the time-course of language comprehension. Since EEG has an excellent time resolution, this method is used to test how language is being processed in real-time, i.e. in terms of milliseconds. In other words, the use of ERPs makes it possible to determine at which stage of processing an experimental manipulation has an impact. Another advantage of using ERPs is that they can show online processing of a stimulus without any response (Luck, 2005a). This makes the method extremely valuable for research on language processing but also gesture-speech processing. Although EEG has an excellent temporal resolution, from the EEG signal recorded from the scalp electrodes we cannot infer where exactly in the brain an effect occurs. Thus, if spatial resolution is important to the research question, other methods are preferred

which have a spatial resolution in the millimetre range (e.g., fMRI or PET) (Kaan, 2007; Luck, 2005a).

Several components have been found to be sensitive to language processing (see Kutas, Federmeier, Staab, & Kluender, 2007 for an overview). Two important components in linguistic research are the N400, which is known as the semantic processing component, and the P600, which has been attributed to syntactical processing. In this section, however, more recent research will be reviewed that suggests that the P600 is also a component that indexes integration processes of a stimuli into a discourse model which includes a reanalysis of an already existing mental representation of a message (see Brouwer, Fitz, & Hoeks, 2012 for a review). Both of these ERP components have also been used in gesture-speech comprehension studies. Moreover, they are relevant for the ERP study presented in this thesis. Besides the N400 and the P600, the Nref, another ERP component relevant to the current study will be discussed. The Nref has previously been linked to referent processing within a discourse context (see Van Berkum, Koornneef, Otten, & Nieuwland, 2007 for a review). Thus, it is important to outline what functions these three ERP components are attributed to and to define how these components are used in the ERP study presented in Chapter 6.

5.4.1. N400

The N400 was first discovered by Kutas and Hillyard (1980). They conducted a reading experiment where the last word of the sentence was manipulated. The sentences had three different endings that were compared to one another: congruent (e.g., “I shaved off my moustache and beard.”), odd (e.g., “He mailed the letter without a thought.”) and anomalous (e.g., “He planted string beans in his car.”). They found that the ERPs time-locked to the onset of the sentence-final word elicited a large negative deflection peaking at around 400 ms for the anomalous condition. This negativity was also observed, though it was not as large, in the

condition with the odd sentence ending when compared to the congruent condition. Based on this observation, the difficulty of semantic processing was thought to be shown by the N400 component (Kutas & Federmeier, 2011). Importantly, the N400 does not have to be negative in absolute terms. What is studied is the difference between ERPs elicited by two or more experimental conditions where semantic features (e.g., congruent vs incongruent) are manipulated. Only the subtraction of ERPs in Condition A from ERPs in Condition B can tell us something about differences in semantic processing costs. These differences between conditions is called the N400 effect which can be described as a negative, monophasic effect which is largest over centro-parietal sides (for written words in sentence contexts) and occurs between 200-600 ms after stimulus onset. Since Kutas & Hillyard's (1980) first study, more than 1 000 papers have been published that have used the N400 as a measurement of semantic processing costs (Kutas & Federmeier, 2011). In these studies the N400 was not only observed for semantic anomalies. Rather it was found that the amplitude of the N400 component is sensitive to various factors including word frequency and the predictability of a word in a sentential context (see Kutas et al., 2007 for a review). The N400 was also elicited by the manipulation of different contexts. Apart from the sentential context, this includes the manipulation of the discourse context (van Berkum, Zwitterlood, Hagoort, & Brown, 2003) and the manipulation of world knowledge (Hagoort & Van Berkum, 2007).

Although the N400 has been acknowledged to reflect semantic processing, its exact function is still controversial (see Lau, Phillips, & Poeppel, 2008 for an overview). Some researchers argue that the N400 can be associated with semantic integration of a word into a given context (e.g., sentence, discourse). According to this view, the N400 is a post-lexical effect; i.e. it occurs after the stage of lexical access. Two main arguments that support this interpretation are as follows. First, the N400 occurs too late after stimulus onset to be an

indicator for lexical access (Hauk, Davis, Ford, Pulvermuller, & Marslen-Wilson, 2006; Hauk & Pulvermuller, 2004). For example, Hauk and Pulvermuller (2004) found in their study that lexical access can occur as early as 150 ms post-stimulus. Second, it has been found that the more predictable a word within its context, the smaller is the observed N400 amplitude. Thus, this indicates that at the time of the N400, integration processes of a word into its context are reflected (Hagoort, 2008; van den Brink, Brown, & Hagoort, 2006). However, the very same argument is also used by researchers who support the lexical access view. In this view, the N400 is linked to lexical retrieval from long-term memory. Predictability is seen as a facilitating factor of lexical access that reduces the N400 amplitude. The more predictable a word is, the easier it is to retrieve it from the mental lexicon because the context already pre-activates it (Federmeier, 2007; Kutas & Federmeier, 2000). Also, generally speaking, low-frequency words elicit a larger N400 amplitude than high-frequency word (e.g., van Petten & Kutas, 1990). This correlation of word frequency and the N400 amplitude is another argument supporting the lexical retrieval view.

Importantly, the N400 is not limited to linguistic stimuli, but has been found in other modalities that involve some kind of semantic processing, including drawings, pictures, actions and gestures (Kutas & Federmeier, 2011). The current view of the research community is that the N400 should not be seen as a language processing component. Rather the N400 reflects meaning processing in general (Kutas & Federmeier, 2011). Although the N400 has been observed in countless contexts related to linguistic and non-linguistic semantic processing, it has not been observed in syntactic violation paradigms (Kutas & Federmeier, 2011).

5.4.2. P600

Syntactic and morphosyntactic processing has traditionally been attributed to the P600 (Key et al., 2005). The P600 has a positive deflection starting at around 500 ms post-stimulus.

Its typical time-window ranges from 500-800 in the form of a broad peakless shift (Kutas & Federmeier, 2007). This effect is strongest over posterior sites, but also anterior effects have been observed (Coulson, 2007). The P600 as a component elicited by syntactic anomalies was first reported by Hagoort, Brown, and Groothusen (1993) and Osterhout and Holcomb (1992), although the latter research group termed the effect *syntactic positive shift*. Generally, the amplitude of the P600 is known to reflect syntactic processing costs (Kutas & Federmeier, 2011). This is evident in so-called garden path sentences where at some point within a sentence the grammatical structure diverges from the expected one. In the following example taken from Kaan and Swaab (2003), the ERPs were time-locked to the onset of the word *were*. In this example, sentence number one is considered as the preferred sentential continuation whilst sentence number two is considered as grammatically unexpected. ERPs showed that the unexpected grammatical structure elicits a larger P600 compared to the expected structure. The increased processing cost in the unexpected sentence continuation is attributed to a reanalysis and repair of the already constructed mental representation of the sentence (Kaan & Swaab, 2003).

(1) The man borrowed the hammer but the pliers were in his toolbox.

(2) The man borrowed the hammer and the pliers were in his toolbox.

The underlying functions of syntactic processing that the P600 reflects is still debated (see Kutas et al., 2007 for a review). Moreover, in the last couple of years numerous studies have found a P600 effect in semantic and pragmatic violation paradigms (Kutas et al., 2007). Also, in many of these studies, semantic violations only elicited a P600 but no N400 (see Brouwer et al., 2012 for a review). For example, Hoeks, Stowe, and Doedens (2004) compared active and passive sentences where the final word of the active sentence was very predictable whilst the passive version included a semantic anomaly on the message-level. The manipulation

is illustrated in the example below, where the active sentence is semantically plausible but not the passive sentence because the javelin cannot throw the athlete. However, in both versions of the sentence the “lexico-semantic fit” was kept constant. In other words, the context provided did not change across conditions (active and passive) because before the critical word (i.e., thrown) the participants were presented with both words which were crucial for building a mental representation of the message (i.e., athlete and javelin).

(1) Active sentence

Dutch: De speer werd door de atleten geworpen.

lit: The javelin was by the athletes thrown.

(2) Passive sentence

Dutch: De speer heeft de atleten geworpen.

lit: The javelin has the athletes thrown.

Although the violation in this experiment was semantic in nature, Hoeks et al. (2004) did not find a difference in the N400 amplitude across conditions but rather a P600 effect. Findings like the ones presented from Hoeks et al.’s (2004) study pose a problem for traditional accounts that assume that the N400 reflects semantic processing while the P600 is seen as an index for syntactic processing (see also: Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nieuwland & Van Berkum, 2005). In the next section two different theories will be presented that attempt to explain a P600 elicited by semantic anomalies. Furthermore it will be pointed out which of these theories have been used to explain findings in the field of gesture-speech processing.

5.4.3. N400 & P600 Revisited

One question that arose from findings like the ones illustrated from Hoeks et al.’s (2004) study is why semantic violations not necessarily elicit an N400 effect but a P600 instead. One

account assumes that the participants were under the “temporary semantic illusion” that the given sentence is semantically plausible (Hoeks et al., 2004, p. 72). However this illusion only holds briefly which is evident by the P600 effect. In this case, the P600 presumably reflects additional syntactic processing because the participant tries to repair the incorrectly assigned meaning of the message. The “Semantic Illusion Effect” led to the proposal of numerous complex processing models that try to explain why the semantic processor can reach a meaningful representation that violates the given syntactical structure of the message (see Brouwer et al., 2012 for an overview). These so-called multi-stream models have the assumption in common that there are different processing streams in comprehension that run independently. In particular, the semantic processor analyses incoming information independently from the syntactical processor. Moreover, these processors work in parallel to reach a unified interpretation of the given message (e.g., Bornkessel-Schlesewsky & Schlewsky, 2008; Kolk et al., 2003; Kuperberg et al., 2003). In multi-stream models, the N400 reflects semantic integration of a word into the wider context. Thus, they explain the absence of an N400 effect in semantic violation experiments with the “Semantic Illusion Effect” introduced above (Brouwer, Crocker, Venhuizen, & Hoeks, 2017).

In contrast to the multi-stream models, Brouwer et al. (2012) proposed that semantic processing is biphasic. In the so-called Retrieval Integration Account (Brouwer et al., 2017; Brouwer et al., 2012; Brouwer & Hoeks, 2013) the N400 is seen as indexing the retrieval of conceptual knowledge. In particular, this means lexical retrieval for verbal stimuli and semantic retrieval for non-verbal stimuli from long-term memory. The P600 reflects the integration of new information into the already existing mental representation/discourse model. Thus, the P600 can be linked to problems in interpreting a given message. These problems do not only arise from semantic implausibility but also from newly introduced information into a discourse.

Moreover, the Retrieval Integration Account interprets the P600 elicited by syntactic violations also as problems of updating the current mental representation of a message.

The biphasic Retrieval Integration Account is relevant to the current study because it has been used to explain findings from two studies where abstract pointing gestures violated the already established discourse (Gunter & Weinbrenner, 2017; Gunter et al., 2015). Following these two gesture studies, in the current study the N400 is seen as an indicator for the retrieval of conceptual knowledge (i.e., lexical retrieval for verbal stimuli; semantic retrieval for non-verbal stimuli) and the P600 as an index for integrating a gesture/word into an existing mental representation/discourse model. However, as pointed out later in this chapter, this interpretation of the N400 is not consistent across ERP studies on gesture-speech comprehension.

5.4.4. *Nref*

The *Nref* component reflects processing problems caused by referential ambiguity within a sentence or discourse context (e.g., Boudewyn et al., 2015; Nieuwland, Otten, & Van Berkum, 2007; Nieuwland & Van Berkum, 2008). In particular, if the word that the ERP is time-locked to is ambiguous in relation to the already established message, then the search for the “conceptually suitable referents” elicits a frontal negativity (Van Berkum, 2009, p. 287). This increased frontal ERP negativity elicited by referent ambiguity compared to ERPs time-locked to an unambiguous word, has been found to be a sustained effect that starts at around 300-400 ms post-stimulus onset (for written and spoken words) (Van Berkum et al., 2007). For example, Boudewyn et al. (2015) found that when two entities introduced in a discourse context can be linked to an anaphor in the target-sentence (i.e., S-4, time-locked to OAK), an *Nref* effect is elicited when compared with an unambiguous discourse context.

S-1: A lumberjack hiked into a forest carrying a chainsaw.

S-2: He was going to cut down a tree.

S-3: Unambiguous/ambiguous: In a clearing he found an **oak** that had a mushroom on it, and an **elm/oak** that had birds in its branches.

S-4: The lumberjack cut down the **OAK**...(Boudewyn et al., 2015, p. 2311)

This example also illustrates that the Nref is not elicited by an anomaly but rather by an increased processing cost due to an ambiguous word within its context.

Regarding gesture-speech processing, we hypothesise that if a gesture does not synchronise with its semantic affiliate, the gesture interpretation relies even more on referential cues. Thus, finding the correct referent to relate the gesture to is presumably more effortful when the discourse information preceding the gesture is not related to the gesture's meaning. Since in the ERP study in Chapter 6 preceding discourse information was manipulated, the Nref will be used as an index of whether discourse information is taken into account during gesture processing.

5.5. ERP Research and Co-speech Gesture

Only recently have gesture researchers started using the ERP technique to answer questions about how co-speech gestures are processed. In this section, research paradigms and ERP components used in these studies will be reviewed. Although some researchers (e.g., Kelly et al., 2004) also looked at early components (0 – 200 ms post-stimulus), the focus of this section will be on the N400 and P600. As mentioned above, to our knowledge, the Nref has not been investigated in any previous studies on gesture-speech processing. But previous studies have not manipulated discourse information and thus the Nref was not an expected component.

5.5.1. Research Paradigms

In studies so far, the most commonly used study design was a match/mismatch paradigm where the semantic fit of a gesture with speech was manipulated. In one of the first gesture-

speech ERP studies, Kelly et al. (2004) tested whether iconic gestures have an influence on how we process language. In addition to the gesture match condition (gesture: indicating the *tallness* of a glass, speech: “tall”) and a gesture mismatch condition (gesture: indicating the *shortness* of a dish, speech: “tall”), they had two additional conditions: A complementary gesture condition (gesture: indicating the *thinness* of a thin and tall glass; speech: “tall”) and a no gesture condition as baseline. Also, in more recent studies, match/mismatch paradigms were used to investigate gesture-speech processing on different linguistic levels (gesture-speech synchrony on a word level: Habets et al., 2011; sentence level: Özyürek et al., 2007). In contrast to Kelly et al.’s study (2004), none of these studies included a “no gesture condition” or a “complementary gesture condition”. Furthermore, match/mismatch paradigms have not only been used in studies on iconic gestures but also for metaphorical gestures (Cornejo et al., 2009; Ibáñez et al., 2010) and abstract pointing gestures (Gunter & Weinbrenner, 2017; Gunter et al., 2015).

In another study on the integration of gestures into its discourse, a homonym disambiguation paradigm was used (Holle & Gunter, 2007). In this study, it was tested whether a gesture which synchronised with a homonym (e.g., “ball”) disambiguated the homonym’s meaning (e.g., *dancing gesture* vs. *playing volleyball gesture*). Whether the gesture became part of the listener’s discourse model was tested downstream the sentence where the target word which disambiguated the homonym in speech (e.g., “game”/”dance”) either matched or mismatched the gesture’s meaning (Holle & Gunter, 2007).



playing volleyball gesture



dancing gesture

- (1) Sie kontrollierte den Ball, was sich im *Spiel* beim Aufschlag deutlich zeigte.

She controlled the ball, which during the *game* at the serve clearly showed.

- (2) Sie kontrollierte den Ball, was sich im *Tanz* mit dem Bräutigam deutlich zeigte.

She controlled the ball, which during the *dance* with the bridegroom clearly showed.

Obermeier and Gunter (2014, p. 298)

The same stimuli set was used to investigate different aspects of gesture-speech processing including gesture-speech synchronisation (Obermeier & Gunter, 2014; Obermeier et al., 2010), gesture comprehension in hearing impaired participants (Obermeier, Dolk, & Gunter, 2012) and the comparison of meaningful versus meaningless (grooming) gestures (Obermeier, Kelly, & Gunter, 2015).

The study presented in this thesis (Chapter 6) also used a match/mismatch paradigm to test whether preceding discourse information related to the gesture's meaning functions as cue for gesture interpretation and thus enables gesture integration into a listener's discourse model when gesture-speech synchrony is not present. More specifically, an iconic gesture that depicted an action verb was placed at the beginning of the sentence while the gesture's semantic affiliate (i.e., the verb) was encoded later within the sentence and either matched or mismatched the gesture. Thus, the stimuli sentences in the current study are somewhat similar to the stimuli used in the homonym studies. However, as outlined in the next section, the ERP study presented in this thesis does not only look at the N400, as they did in the homonym studies, but also at the P600 and the Nref.

5.5.2. ERP Components

As mentioned above, the current study focussed on the N400, P600 and the Nref components to investigate gesture-speech processing. In particular, the N400 will be seen as an indicator for lexical retrieval (verbal stimuli) and semantic retrieval (non-verbal stimuli), whilst the P600 is assumed to reflect the integration of a word or gesture into a listener's discourse model. However, it is important to point out that this interpretation of the N400 and P600 functions is not consistent across gesture studies.

Surprisingly, only a handful gesture studies have looked at the P600 as reflecting the participants integration of new information (i.e., gesture or word) into an already existing mental representation of a message/discourse model. Two of these studies investigated abstract pointing (Gunter & Weinbrenner, 2017; Gunter et al., 2015). Three other studies investigated the processing of metaphorical gestures (Cornejo et al., 2009; Ibáñez et al., 2010; Ibáñez et al., 2011).

Interestingly, in the homonym disambiguation study introduced above, Holle and Gunter (2007) did not look at a possible P600 effect. They only stated that potential effects after 500 ms post-stimulus were beyond the scope of the current paper. Also, two further papers that used the same stimuli set as that by Holle and Gunter (2007) did not look at late positivity effects (Obermeier & Gunter, 2014; Obermeier et al., 2010). In all three of these homonym studies, the authors adapted the view that the N400 reflects integration of a word into its speech context. The same interpretation of the N400 was used in a study conducted by Özyürek et al. (2007), which tested if a gesture's meaning is integrated into the previous sentence context.

A study that looked at the P600, but did not find an effect is that by Kelly et al. (2004). They presented gestures as primes to a target word. Gestures either matched or mismatched the target word. It is possible that the absence of a P600 effect is due to a lack of a wider discourse that the gesture could be integrated in. Thus, the gesture might only have had an influence on

lexical retrieval processes. Similarly, Habets et al. (2011) only used the N400 as dependent measurement in a match/mismatch paradigm where synchronisation of a gesture in relation to a single word (i.e., verb) was manipulated. This supports the conclusion that the P600 is only relevant when a sentence or discourse context is presented (cf. Gunter & Weinbrenner, 2017). Since numerous previous ERP studies did not look at the P600 time-window, we do not know whether there was a P600 effect present in these studies (Holle & Gunter, 2007; Obermeier & Gunter, 2014; Obermeier et al., 2010; Özyürek et al., 2007). Finally, due to the different functions attributed to the N400 in previous ERP studies, one has to be cautious when drawing general conclusions about how co-speech gestures are processed in relation to speech. This issue will be further discussed in the next chapter where the ERP study is presented.

Chapter 6

Multimodal Language Processing: How Preceding Discourse Constrains Gesture Interpretation and Influences Gesture Processing

6.1. Abstract

Previous studies have found that during language processing the meaning of a gesture is immediately integrated into a discourse model. However, this integration process is assumed to be automatic only if the gesture synchronises with its semantic affiliate in speech. We investigated whether preceding discourse related to the gesture's meaning constrains gesture interpretation and thus impacts gesture processing when the gesture does not synchronise with its semantic affiliate. First, a behavioural experiment showed that listeners take discourse information into account when interpreting such gestures. Second, results from an ERP experiment showed that synchronisation between gesture and semantic affiliate is not essential in order for the gesture to become part of a discourse model. However, the underlying integration processes are different from when gestures synchronise with their semantic affiliate. Based on these ERP findings, we distinguished three different gesture integration processes within a discourse context: search for a referent in preceding discourse (Nref), context driven meaning construction/semantic lexical retrieval (N400), post-semantic integration into a discourse model (P600).

Keywords: gesture, ERP, gesture-speech synchrony, discourse

6.2. Introduction

Communication is multimodal. In everyday conversation information is not only conveyed via speech but also through gestures that accompany speech (Goldin-Meadow, 2003; McNeill, 1992, 2005). Generally, researchers agree that gestures play a role in language comprehension (see Goldin-Meadow & Alibali, 2013 for an overview). For example, behavioural studies found that information presented only in gesture was picked up by listeners (Cassell et al., 1999; Goldin-Meadow et al., 1992). Moreover, meaning conveyed via gesture and meaning conveyed via speech is combined automatically and thus indicates that gesture and speech comprise an integrated system during comprehension (Kelly, Özyürek, & Maris, 2010).

However, this gesture-speech integration process is not obligatory. It is modulated by various factors, including gesture-speech synchrony (see Özyürek, 2014 for a review). The synchrony between a gesture and its semantic affiliate (i.e., portion(s) of speech closest to the gesture's meaning) is crucial in order to form a unified representation of a message because gestures are highly ambiguous without a semantic affiliate in speech (see Kelly, Manning, & Rodak, 2008 for a review). However, it is not clear if and how a gesture that does not synchronise (overlap) with its semantic affiliate can be integrated with gesture. We investigated if the discourse that precedes the gesture may play an important role in such instances. This research question is also relevant for the speech-gesture production literature because previous studies demonstrated that such instances of asynchronies (i.e., gestures preceding their semantic affiliate) frequently occur in speech-gesture production (e.g., chapter 4; Morrel-Samuels & Krauss, 1992; Schegloff, 1984).

The information in discourse that precedes a gesture might be important in gesture processing because speech comprehension studies (spoken and written) have shown that

comprehension is discourse-dependent (Van Berkum, 2008). In order to construct meaning of an utterance, we relate incoming information to already established referents (e.g., Van Berkum et al., 2007), discourse information is used to predict upcoming words (e.g., Otten & Van Berkum, 2008) and we are integrating incoming information into the already existing discourse model (e.g., Brouwer et al., 2012). Thus, how a word is processed depends on its preceding discourse. The present study explores whether the same applies to gesture processing when a gesture does not synchronise with its semantic affiliate.

In order to test whether gestures and speech form a unified representation of a message in comprehension, previous studies often used electrophysiology, more specifically event-related potentials (ERPs). ERP studies so far focussed on a particular ERP component, namely the N400 (e.g., Habets et al., 2011; Holle & Gunter, 2007; Kelly et al., 2004; Özyürek et al., 2007). The N400 is a component with a negative deflection peaking at around 400 ms after the onset of a target stimulus and is sensitive to semantic processing cost (see Kutas et al., 2007 for a review). The N400 was first reported by Kutas and Hillyard (1980). They manipulated the final word of a sentence in terms of how well it fitted into the sentence context. A semantically anomalous word (e.g., “He took a sip from the *transmitter*”) elicited a larger negativity peaking at around 400 ms after the onset of the final word, compared to a more predictable sentence ending (e.g., “He took a sip from the *tap*”). This so-called N400 effect is not only elicited by such strong mismatch cases, but also subtle manipulations of how well a word semantically fits into a given context leads to an N400 effect. For instance, while *waterfall* in the sentence “He took a sip from the *waterfall*” is a semantically congruent sentence ending, it is unexpected and therefore leads to a larger N400 component compared to “He took a sip from the *tap*”

(Kutas & Hillyard, 1980). Importantly for the current study, the N400 has not only been observed in linguistic contexts but also in other domains that require some form of semantic processing. This includes drawing (e.g., Ganis, Kutas, & Sereno, 1996), actions (e.g., Proverbio & Riva, 2009) and gestures.

Kelly et al. (2004) demonstrated that semantic processing of iconic gestures is very similar to semantic processing of speech. In this study, two objects were placed on a table (i.e., a *short wide dish* and a *tall thin glass*). A person verbally described one of the two objects after providing a gesture that either fit the description or not. When the gesture was incongruous with speech (e.g., speech: “tall”; gesture indicating *shortness*) the observed N400 was larger compared to when gesture and speech were congruent (e.g., both expressing *shortness*).

Similar cross-modal semantic integration has been observed on the sentence level (e.g., Cornejo et al., 2009; Holle & Gunter, 2007; Özyürek et al., 2007). Özyürek et al. (2007) found evidence that gestures are integrated into the preceding sentence. They used a gesture-speech mismatch paradigm where the gesture was synchronous with a verb. In a baseline condition, the verb and the gesture both fit the preceding sentence context (e.g., “He slips on the roof and rolls down”, with “rolls” accompanied by a *rolling down* gesture). In three further conditions, the sentence had either the verb (“walking” instead of “rolling”) or the gesture (*walking across* gesture instead of *rolling down* gesture), or both (verb and gesture depicting “walking across”) mismatching the preceding sentence context. It was found that all mismatch conditions elicited a larger negativity at around 400 ms time-locked to the onset of the verb (which was also the gesture’s onset) compared to the baseline condition. The authors concluded that speech and gestures are

simultaneously integrated into the preceding sentence context and thus become part of a unified representation of message (i.e., a listener's discourse model).

A different approach to test gesture integration into a discourse model was taken by Holle and Gunter (2007). They tested whether an iconic gesture (e.g., “ball” = dance or game) encoded early in the sentence can disambiguate a homonym (e.g., a *dancing* gesture or a *playing volleyball* gesture) before it was disambiguated in speech later in the sentence.

Example Stimulus from Obermeier and Gunter (2014, p. 298):



(1) Sie kontrollierte den Ball, was sich im *Spiel* beim Aufschlag deutlich zeigte.

She controlled the ball, which during the *game* at the serve clearly showed.

(2) Sie kontrollierte den Ball, was sich im *Tanz* mit dem Bräutigam deutlich zeigte.

She controlled the ball, which during the *dance* with the bridegroom clearly showed.

In their study, the gesture synchronised with the homonym. ERPs were time-locked to a target word (e.g., *game/dance*) downstream in the sentence that disambiguated the earlier homonym. If information from gestures was used to disambiguate the homonym, this should be evident in the ERPs time-locked to the target word. Indeed, Holle and Gunter (2007) found a larger N400 on the target word when the gesture mismatched with the target word compared to when it matched. The authors concluded that the gesture

together with the homonym formed an idea unit in the discourse model that modulated sentence processing downstream.

Can such an idea unit also be formed without gesture-speech synchrony? Previous studies suggest that gesture-speech synchrony is a crucial factor for whether a gesture becomes part of a listener's discourse model or not (see Özyürek, 2014 for a review). Iconic gestures are often ambiguous. Although they depict (features of) objects, actions or motions, their interpretation is guided by speech and often even depends on speech (Beattie & Shovelton, 1999; Feyereisen et al., 1988; Hadar & Pinchas-Zamir, 2004; Krauss et al., 1991). Furthermore, ERP studies concluded that a certain amount of temporal overlap of gesture and its corresponding speech element is necessary for an automatic integration of gestures. For example, Habets et al. (2011) manipulated the degree of synchrony between iconic gestures and corresponding verbs. They found that when speech and gesture did not overlap, the gesture was not integrated with the semantic affiliate. This was evident in the absence of an N400 effect on the verb, when comparing matching and mismatching gestures. Habets et al. (2011) hypothesised that when synchrony is lost, listeners interpreted the gestures before processing the verb. And because gestures are generally very ambiguous, the interpretation of the gestures was likely to be different from the verbs presented subsequently. This conclusion was underlined by the results of a pretest that investigated whether the gestures presented in their study were semantically transparent without the corresponding speech element. Only in 11 % of the cases did the participants' gesture interpretation match the verb used in the experiment.

An overlap between gesture and speech is also important for gestures within a sentence context (Obermeier et al., 2010). Using the homonym disambiguation paradigm

described above (Holle & Gunter, 2007; see example stimulus above), they found that when gesture and their semantic affiliates (homonyms) did not overlap, the gesture that mismatched with the target word (the disambiguating word) did not elicit an N400 effect. Their results confirmed that when speech-gesture synchrony is lost, a gesture does not automatically become part of a listener's discourse model. However, they also found that integration is not completely impossible. They ran the same experiment but changed the instructions: In particular, participants were explicitly *instructed* to judge whether speech and gestures matched. In this case, listeners did integrate the gestures into the discourse model which was evident in an N400 gesture mismatch effect on the disambiguating word later in the sentence. This suggests that gesture-speech synchrony is needed for an automatic integration process. But when synchrony is lost, "effortful gesture-related memory processes" are required in order to combine the gesture and speech channels (Obermeier et al., 2010, p. 1660).

In a similar study (i.e., using the same homonym stimuli), Obermeier and Gunter (2014) found preliminary evidence that synchronisation with a semantic affiliate might not be necessary for a gesture to be integrated with the affiliate and become part of a listeners' discourse model. They explored whether there is a specific time-window that enabled automatic integration of a gesture. In all but one condition, did gestures overlap with their semantic affiliates. In the non-overlapping condition the gesture preceded the homonym (the affiliate) and synchronised with the verb (e.g., the verb "controlled"). In contrast to Obermeier et al. (2010), all conditions showed a mismatch effect (i.e., an N400 effect) on the disambiguating word later in the sentence. This was seen as an indicator of a successful integration of the gesture into a discourse model.

Based on these results, Obermeier and Gunter (2014) suggested that a gesture does not necessarily have to synchronise with its semantic affiliate, but that there might be multiple positions within an utterance that allow gesture integration. Furthermore, the authors hypothesised that content words might constitute such positions. Because in an earlier study where asynchrony eliminated a mismatch effect, the gesture synchronised with a pronoun (e.g., she) (Obermeier et al., 2010). Thus, Obermeier and Gunter (2014) hypothesised that content words might constitute anchors for gesture integration but function words do not. This hypothesis has not been tested directly.

Moreover, other factors might have contributed to the integration of asynchronous gestures. One such factor, we hypothesise, is the discourse information preceding the gesture's onset. More specifically, preceding discourse might function as a cue for constraining a gesture's meaning. The potential effect of the preceding discourse for the integration of ambiguous gestures becomes apparent when comparing Kelly et al.'s (2004) and Habets et al.'s (2011) studies introduced above. In both studies, quite ambiguous gestures were presented before speech affiliates. However, participants in Kelly et al.'s (2004) study were able to infer the gestures' meaning from the discourse, namely the objects presented together with the gestures. In contrast, in Habets et al.'s (2011) study no discourse was presented. Kelly et al. (2004) did find a mismatch effect, whilst Habets et al. (2011) did not. Furthermore, in the homonym disambiguation studies, Obermeier et al. (2010) had placed gestures on pronouns, while Obermeier and Gunter (2014) had placed them on verbs. It is possible that the verbs provided enough information to construct the gesture's meaning, leading to integration effects later in the sentences, while the pronouns did not. Thus, verbal (and non-verbal) discourse might be able to affect gesture interpretation and integration processes.

6.3. The Present Study

The current study tested if preceding verbal discourse a) constrains a listener's interpretation of a gesture and b) enables gesture to be integrated into a listener's discourse model, when the gesture does not synchronise with its semantic affiliate. To answer our research questions, a behavioural experiment and an ERP experiment were conducted.

In order to test both hypotheses, we manipulated discourse information that preceded the gesture. More specifically, we used a mismatch paradigm where semantic relation between the gesture and its preceding discourse was manipulated by two types of introductory sentences (semantically related versus unrelated to the gesture). These were presented before the sentence that included the gesture. In the Unrelated Discourse Condition, the introductory sentence was chosen so that it did not provide any information to constrain the interpretation of the gesture (e.g., "At the beginning of the week the weather was dreadful" for a following *picking strawberry* gesture). In contrast, in the Related Discourse Condition, the introductory sentence provided the listener with information that was related to the gesture's meaning ("Some of the strawberries in the garden were already ripe" for the following *picking strawberry* gesture) and thus could constrain the possible interpretations of the gesture. The gesture was not placed on its semantic affiliate (the verb *picking*), but on a content word earlier in the sentence that, by itself (i.e. without a related preceding discourse), could not disambiguate the gesture (e.g., the *picking* gesture synchronised with the word *uncle*). By placing the gesture on a noun that could not help interpret the gesture, we tested Gunter et al.'s (2014) hypothesis that gestures can be integrated with non-overlapping semantic affiliates when the gestures are placed on any content word. In our study, the gesture's semantic affiliate was always the

verb, which appeared further downstream the sentence. We used the same material for both the behavioural (Experiment 1) and the ERP experiment (Experiment 2), apart from the fact that in the behavioural experiment, sentences were presented up to, but not including, the gestures' semantic affiliates (i.e., the verbs). In order to test the integration of the gestures into the discourse model in the ERP experiment, we manipulated the match of the gesture with the semantic affiliate (e.g., verb "picking" (match) versus verb "watering" (mismatch)).

6.4. Experiment 1 – Behavioural Experiment

We first conducted the behavioural experiment. This experiment tested whether our discourse manipulation had an effect on how participants interpret the meaning of iconic gestures.

Previous studies have found that iconic gestures are difficult to interpret without speech (cf. Beattie & Shovelton, 1999; Habets et al., 2011; Holle & Gunter, 2007). Also, gesture interpretation is far from perfect even in forced-choice paradigms, where participants are asked to select a gesture's meaning from a list of options (Feyereisen et al., 1988; Hadar & Pinchas-Zamir, 2004), and even if participants are presented with the gesture in combination with speech (i.e., the phrase in speech with which the gesture synchronised) (Hadar & Pinchas-Zamir, 2004). The authors argued that participants performed poorly because they only heard one single phrase. Possibly, a larger discourse extract was needed in order to accurately identify the gesture's semantic affiliate (Hadar & Pinchas-Zamir, 2004).

In our behavioural experiment, participants heard and watched the stimuli videos up to the target verb that was the gesture's semantic affiliate. If preceding discourse that is related to the gesture's meaning has an influence on gesture interpretation, the number

of different gesture interpretations that different participants come up with should be lower in the Related Discourse Condition compared to the Unrelated Discourse Condition. Also, gesture interpretation should be easier in the Related Discourse Condition than in the Unrelated Discourse Condition.

Furthermore, the behavioural experiment provided a validation of the gesture match/mismatch manipulation of our stimuli for the ERP experiment. If our manipulations worked well, then participants' gesture interpretations (e.g., of the *picking* gesture) should match the semantic affiliate in the Match Condition (e.g., the verb *picking*) than the Mismatch Condition (e.g., the verb *watering*).

6.5. Methods

6.5.1. Participants

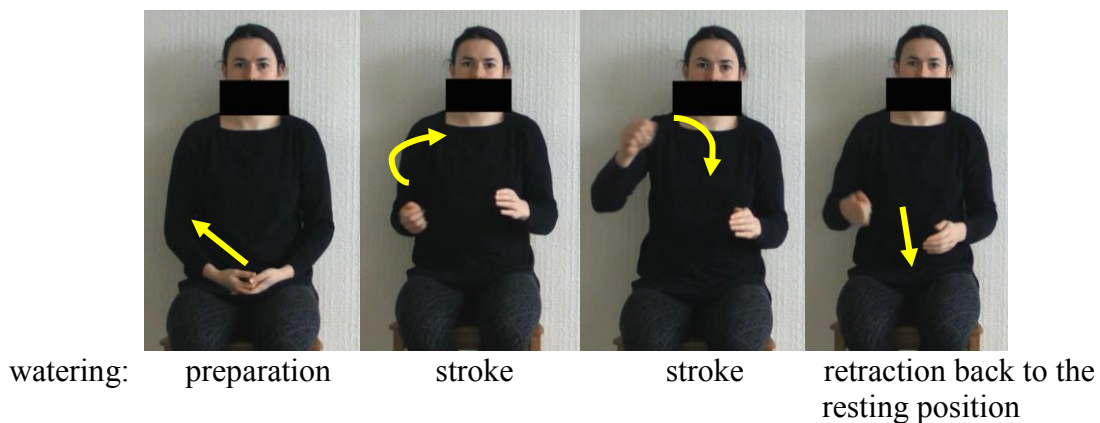
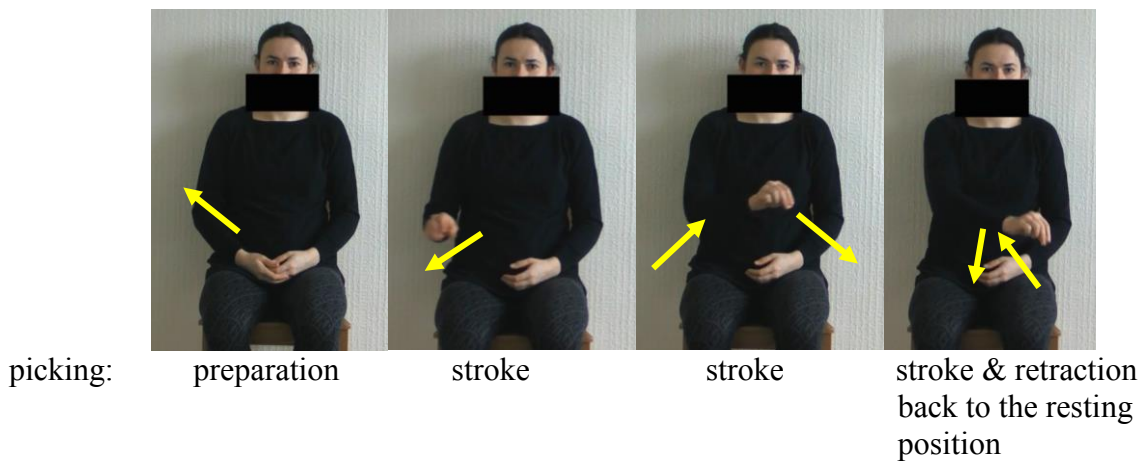
Seventeen English native speakers took part in the behavioural experiment. One participant was excluded from the analysis due to severe visual impairment. The remaining sixteen participants were all women (mean age = 19 years, SD = 0.8). All studied Psychology at the University of Birmingham and received course credits for their participation.

6.5.2. Material

We created 53 basic stimuli, each consisting of an introductory sentence and a target sentence. The gestures and their semantic affiliates (verbs) were part of the target sentences. Discourse information was manipulated by varying whether the introductory sentences provided helpful information for the interpretation of the gesture or not. In the Related Discourse Condition the introductory sentence provided information that was related to the gesture's meaning and could therefore help to interpret it (see Table 6.1 and

Appendix 3 for example stimuli). For the Unrelated Discourse Condition this was not the case.

In both conditions, the gesture was placed on a content word (2-3 syllables long) towards the beginning of the target sentence. This word could not disambiguate the gesture by itself (e.g. a *watering gesture* synchronising with the word “uncle”). The semantic affiliate was always the verb of the target sentence, which was linguistically encoded further downstream the sentence (e.g., watering, picking). The screenshots below show the gestures that depicted the verbs “picking” and “watering”.



A mismatch paradigm was used in this study. In the Gesture Match Condition, the gesture matched the target-verb (e.g., speech: “picking”; gesture: *picking*) and in the Gesture Mismatch Condition the gesture did not match the target-verb (e.g., speech:

“picking”; gesture: *watering*). Importantly, up to the verb the semantic fit of the Match Gesture and the Mismatch Gesture did not differ in either the Related Discourse Condition or the Unrelated Discourse Condition. To assure that there was no overlap between the gesture and the verb, we separated them by inserting a number of words with in total 5-7 syllables (number of words varied). The words placed between gesture and verb were kept neutral in relation to the verb’s and the gesture’s meaning. This was to prevent the participants from predicting the critical verb or providing more information that disambiguates the gesture.

The structure of the target sentences was very similar across stimuli. They all started with one of the following constructions: “I could see/hear”, “I saw”, “I noticed”, “I was told”. Despite the target verb occurring towards the end of the sentence, it never appeared in the sentence final position. This was to avoid sentence-final wrap-up effects, which increases processing cost due to global processing of the sentence (Hagoort, 2003; Osterhout, 1997).

Every participant was presented with the introductory sentence from the Related Discourse Condition and the Unrelated Discourse Condition twice, once with a matching and once with a mismatching gesture, whilst she/he was presented with the target sentence four times (i.e., matching and mismatching gesture in both discourse conditions). Hearing the target sentence four times throughout the experiment, might have resulted in the participants memorising parts of the sentences which could help them with predicting the verb. We therefore changed the words between the gesture and the verb across conditions, keeping the total number of syllables the same (see Table 6.1 for an example stimulus).

The words between gesture and verb were counter-balanced across participants. Furthermore, we counter-balanced across participants which verb of the match/mismatch

combination the participants heard (e.g., half of the participants heard “watering” the other half “picking”). In sum, this led to four different stimuli lists (see Table 6.1).



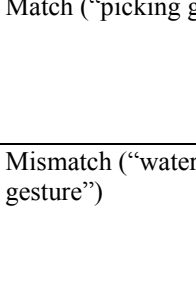
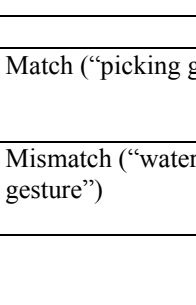

6.5.3. *Stimuli Recordings & Editing*

Speech was recorded in a sound proof booth by a female native speaker of English. Introductory sentences and target sentences were recorded separately and were later combined using Adobe Premiere Pro (www.adobe.com). The introductory sentence and the target sentence were recorded separately because due to counter-balancing of the target-verb and the speech elements between the gesture and the target-verb, we needed four versions of each target sentence (see counter-balancing in Table 6.1).

For gesture, videos of a woman performing the gestures were recorded (with 29 frames per second). Although a few verbs occurred twice in the stimuli, we used different gestures that fit the discourse context (e.g., to pick strawberries vs. to pick an apple). This avoided repetitions of the same gestures. To match the gesture recordings with the recorded spoken sentences, the actress was instructed to utter the beginning of the target sentence and to produce the gesture while uttering the word with which the gesture will later synchronise. After executing the gesture, her hands returned to the resting position (i.e., her lap) while a few more seconds were recorded of her sitting still.

For the temporal alignment of gesture and speech, stroke onset was synchronised with the onset of the target noun. On average, strokes lasted for 922 ms. ($SD = 364$ ms.). Finally, the mouth of the actress was covered with a black box because her mouthing did not match the audio track. In total 848 stimuli were created.

Table 6.1. Example stimulus including all four stimuli sets. For the behavioural experiment the stimuli were cut before the second word in bold in the target sentence column. The onset of the first word in bold co-occurred with the onset of the gesture. For the ERP study, ERPs were time-locked to the onset of the words in bold (target sentence). The four stimuli sets were counter-balanced across participants. Target-verbs and speech elements between the gesture and the target-verb were counter-balanced as follows: Stimuli Set 1 and Stimuli Set 2 differ in terms of the elements between the gesture and the target-verb, the same applies to Stimuli Set 3 and Stimuli Set 4. The target-verbs are the same in Stimuli Set 1 and Stimuli Set 2 but differ from Stimuli Set 3 and Stimuli Set 4.

<i>Condition</i>	<i>Introductory Sentence</i>	<i>Target Sentence</i>	<i>Gesture (match/mismatch)</i>
Stimuli Set 1			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my uncle and his lovely wife were picking strawberries.	Match (“picking gesture”) 
	At the beginning of the week the weather was dreadful.	I saw that my uncle and his lovely wife were picking strawberries.	Mismatch (“watering gesture”) 
Related Discourse Condition	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his two children were picking strawberries.	Match (“picking gesture”) 
	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his two children were picking strawberries.	Mismatch (“watering gesture”) 
Stimuli Set 2			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my uncle and his two children were picking strawberries.	Match (“picking gesture”) 
	At the beginning of the week the weather was dreadful.	I saw that my uncle and his two children were picking strawberries.	Mismatch (“watering gesture”)

Related Discourse Condition	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his lovely wife were picking strawberries.	Match (“picking gesture”)
	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his lovely wife were picking strawberries.	Mismatch (“watering gesture”)
Stimuli Set 3			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my uncle and his lovely wife were watering the strawberries.	Match (“watering gesture”)
	At the beginning of the week the weather was dreadful.	I saw that my uncle and his lovely wife were watering the strawberries.	Mismatch (“picking gesture”)
Related Discourse Condition	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his two children were watering the strawberries.	Match (“watering gesture”)
	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his two children were watering the strawberries.	Mismatch (“picking gesture”)
Stimuli Set 4			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my uncle and his two children were watering the strawberries.	Match (“watering gesture”)
	At the beginning of the week the weather was dreadful.	I saw that my uncle and his two children were watering the strawberries.	Mismatch (“picking gesture”)
Related Discourse Condition	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his lovely wife were watering the strawberries.	Match (“watering gesture”)
	Some of the strawberries in the garden were already ripe .	I saw that my uncle and his lovely wife were watering the strawberries.	Mismatch (“picking gesture”)

6.5.4. Procedure

Participants were randomly assigned to one of the four stimuli lists. The participants’ task was to listen and watch the videos carefully and afterwards answer four questions about the gesture they had seen in the video (see Figure 6.1 for study design). The first question was an open question, asking the participants to type in what they thought the gesture represents. Participants were instructed that their answer can range from one to four words. This open question was analysed as follows. We counted the

number of different responses for a particular gesture given by different participants. Responses with the same verb in different grammatical forms were collapsed (e.g., fly, flying, to fly were all counted as ‘to fly’ responses). So were combinations of the verb with other words (e.g. bird flying, flying around). Answers without a verb (e.g., down, wide, above) were counted as separate responses.

In Questions 2 to 4, participants gave ratings on a scale from 1 (not at all) to 5 (very). Question 2 asked participants how difficult it was to interpret the gesture. In Questions 3, the participants rated how similar their interpretation of the gesture (response to Question 1) was to the target-verb. The target-verb either matched or mismatched the gesture. Which of the two target-verbs (e.g., “picking” or “watering”) the participants saw depended on the stimuli list they were assigned to (see Table 6.1). Question 4 asked participants to rate how well the target-verb fit the gesture regardless of their interpretation. Thus, Question 4 asked the participant to reanalyse the gesture’s meaning. Due to the presentation of the wrong verb in Question 3 and Question 4, 2.5 % of the responses had to be excluded from the analysis for these two questions.

Participants were presented with the verb in the infinitive (e.g., to swim) to avoid confusion with nouns (e.g., to iron vs the iron). The study was self-paced and the participants could take breaks whenever needed. In total, each participant responded to 212 stimuli, presented in four blocks. The order of the blocks was counter-balanced and the trials within the blocks randomised. The experiment took approximately one hour to complete.

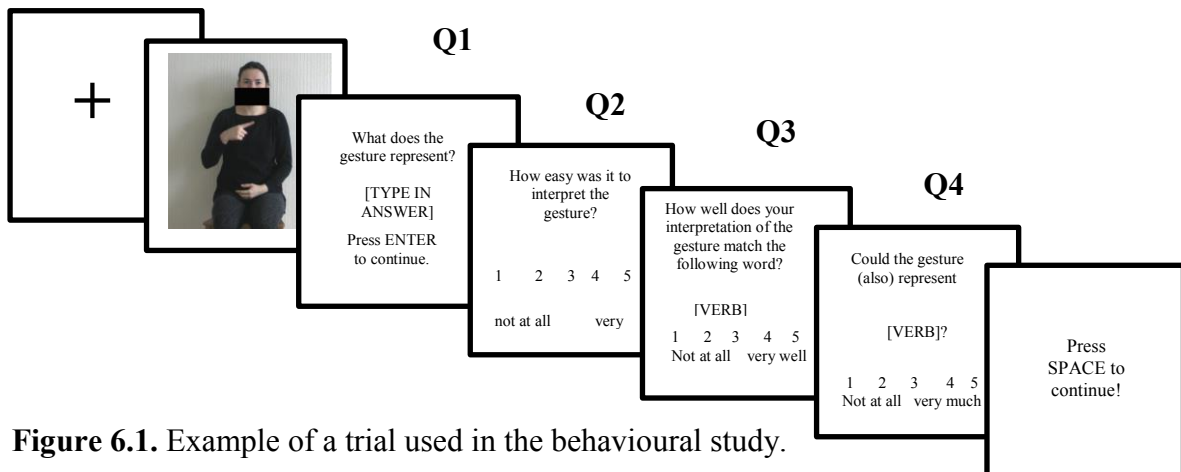


Figure 6.1. Example of a trial used in the behavioural study.

6.6. Results

First, we analysed responses to Question 1, which asked participants to write down their interpretation of the gesture (see Figure 6.2). For this question, a by-item analysis was conducted. A paired-samples t-test showed that there were significantly more different responses per item in the Unrelated Discourse Condition compared to the Related Discourse Condition ($t(52), -5.561, p < .001$).

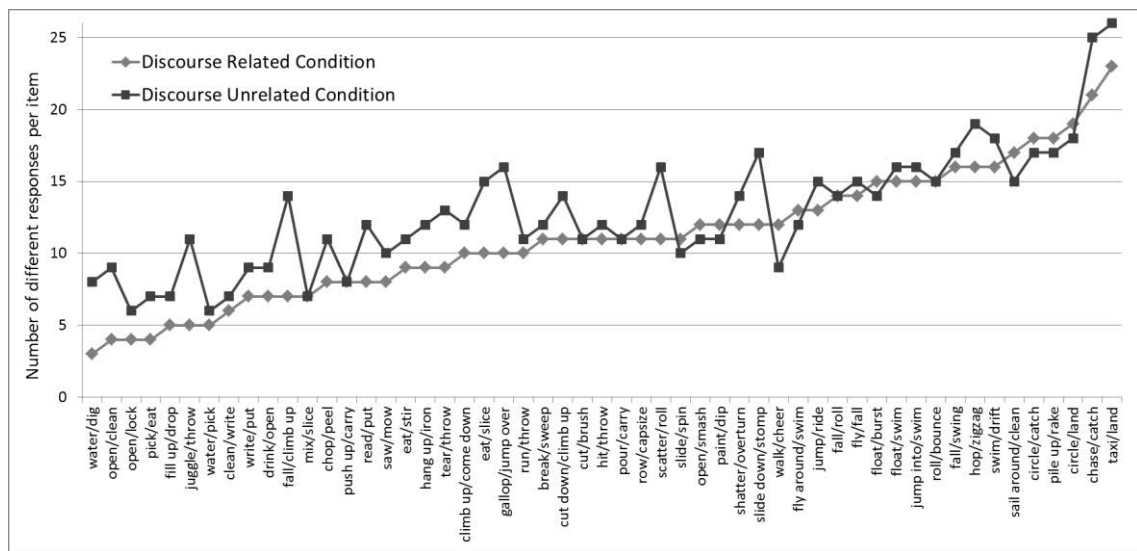


Figure 6.2 Distribution of response to the open question: “What does the gesture represent?”, split into Related and Unrelated Discourse conditions. The Y-axis shows the number of different responses per item across all participants. The light and dark lines show the Discourse Related and Unrelated conditions, respectively. Items in the X-axis are ordered from the lowest number of different responses to the highest number of different responses in the Discourse Related Conditions.

For the second question, we compared the responses from the Related Discourse Condition with the Responses from the Unrelated Discourse Condition. Since the stimuli were cut right before the target verb, Gesture Match responses and Gesture Mismatch responses were collapsed. In particular, Question 2 asked participants how difficult it was to interpret the gesture. We found a significant effect of Discourse type ($t(15) = 4.842$, $p < .001$), with participants rating gestures shown in the Related Discourse Condition easier to interpret than those shown in the Unrelated Discourse Condition (see Figure 6.3).

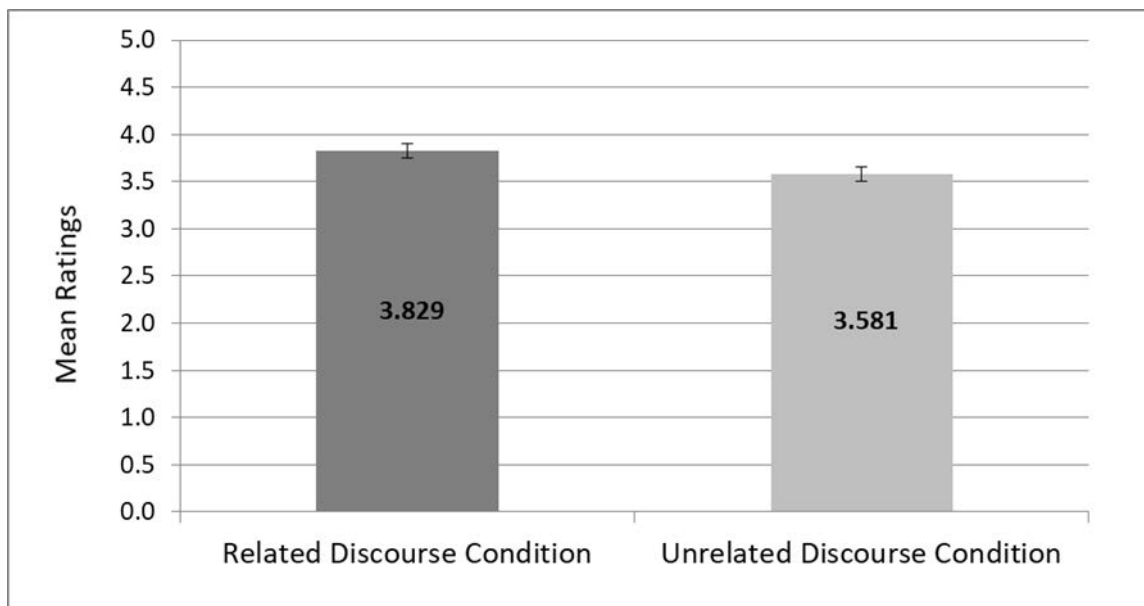


Figure 6.3. Mean ratings for the question “How easy was it to interpret the gesture?” for the Related and Unrelated Discourse Condition. Error bars represent the standard error of the means.

Results for Question 3 (“How well does your interpretation of the gesture match the following word?”) are summarised in Figure 6.4. The rating was analysed in a repeated measures ANOVA with Gesture Match (match/mismatch) and Discourse (Related Discourse/ Unrelated Discourse) as within-subject variables. A significant interaction between Discourse and Gesture Match was found ($F(1,15) = 5.631$, $p = .031$), as well as a significant effect of Gesture Match ($F(1,15) = 646.627$, $p < .001$) and Discourse

($F(1,15) = 9.316, p = .008$). Post-hoc paired t-tests revealed a significantly higher rating for matching gestures compared to mismatching gestures for both Related ($t(15) = 21.664, p < .001$) and Unrelated Discourse ($t(15) = 25.907, p < .001$), confirming the gesture-verb matching manipulation. Importantly, matching gestures in the Related Discourse condition were rated as a better match with the verb than the same gestures in the Unrelated Discourse conditions ($t(15) = 3.349, p = .004$). This was not the case for mismatching gestures ($t(15) = .741, p = .470$). Thus, regardless of the context, if a gesture mismatched the verb the preceding discourse was irrelevant. But a related discourse guided the interpretation of the gesture the way we intended.

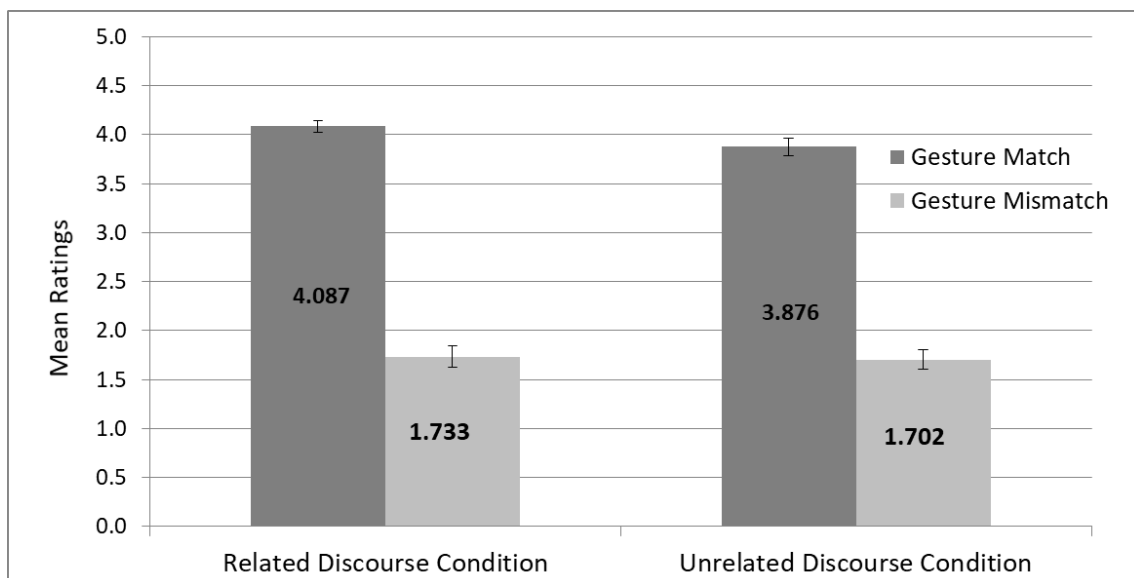


Figure 6.4. Effects of Gesture Match and Discourse Relatedness on mean ratings for gesture match/mismatch for Question 3: “How well does your interpretation of the gesture match the following word?” Error bars represent the standard error of the means.

Finally, for question four (“Could the gesture (also) represent [verb]?”) we did not find a significant effect of Discourse ($F(1,15) = 0.093, p = .765$), nor a significant interaction between Gesture Match and Discourse ($F(1,15) = 3.181, p = .095$). Only Gesture Match yielded a significant effect ($F(1,15) = 347.133, p < .001$), with matching

gestures being rated to be a better representation of the verbs than mismatching gestures.

Results are summarised in Figure 6.5.

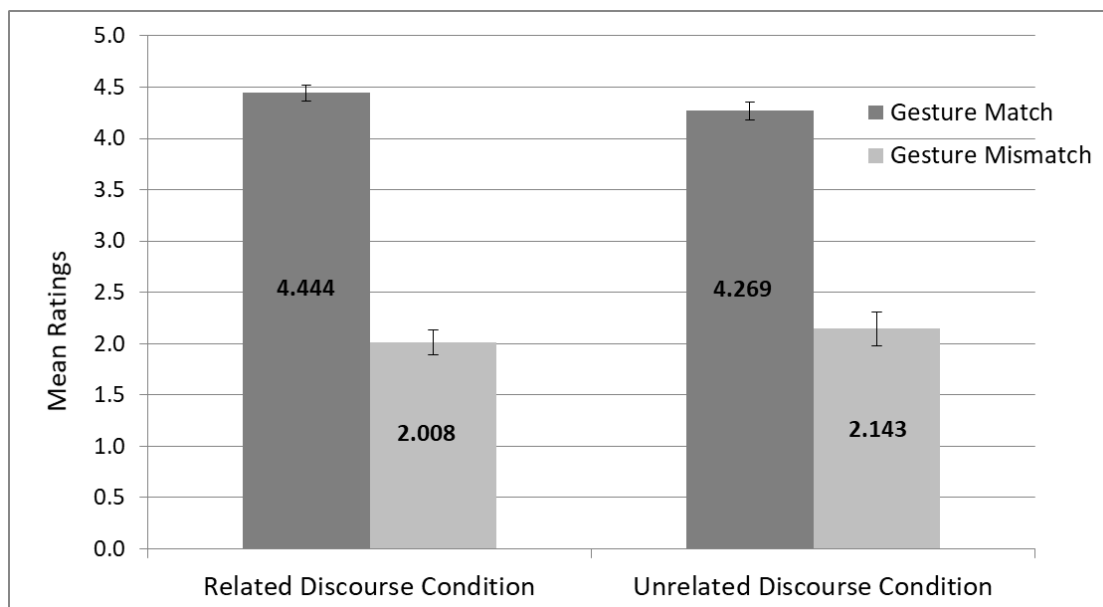


Figure 6.5. Mean ratings for Question 4: “Could the gesture (also) represent [verb]?” for matching/mismatching gestures and Related and Unrelated Discourse conditions. Error bars represent the standard error of the means.

6.7. Discussion Experiment 1 - Behavioural Experiment

The current behavioural experiment tested whether preceding discourse semantically related to a gesture’s meaning constrains gesture interpretation when the gesture does not synchronise with its semantic affiliate. Three results showed that the participants’ interpretation of the gesture was influenced by our discourse manipulation. First, a preceding discourse related to the gesture’s meaning led to reduced response variability in the open question (Question 1) that asked participants to interpret the gesture. That is, participants agreed with each other in gesture interpretation more in the Related Discourse Condition than in the Unrelated Discourse Condition. Second, participants rated gestures that were presented in the Related Discourse Condition less difficult to interpret than in the Unrelated Discourse Condition (Question 2). Third, the participants rated gesture interpretations to be more similar to the target verb in the

Related Discourse Condition than the Unrelated Discourse Condition when the gestures matched the verb. For the mismatch gesture, we did not find a difference between the two discourse conditions (Question 3).

However, when participants were asked if the gestures could represent the target verb independently of their own interpretation, this reanalysis of the gesture did not elicit a discourse effect (Question 4).

What do these results tell us about the role of preceding discourse information in gesture interpretation? Experiment 1 showed that similar to speech processing (cf. Van Berkum, 2008), meaning construction of gestures is also discourse-dependent. In particular, meaning construction is not exclusively linked to the gesture's concurrent speech element(s), but also preceding discourse information functions as cues for gesture interpretation. However, these cues provided in the Related Discourse Condition were not strong enough to disambiguate the gesture's meaning but rather constrained the number of possible interpretations.

Previous studies on the role of speech context in interpreting gestures showed participants gestures with speech being muted compared to gesture-speech combinations (i.e., the gesture's co-occurring phrase or clause) (Beattie & Shovelton, 1999, 2002; Hadar & Pinchas-Zamir, 2004) or just gestures without speech (Feyereisen et al., 1988; as part of a pretest: Habets et al., 2011; Holle & Gunter, 2007). The discourse manipulation in our study adds to the literature that gesture interpretation is discourse-dependent. Previous studies could not provide a full picture of how gestures are interpreted because they took the gestures out of their discourse context (i.e., narrative) and presented them only with their co-occurring speech elements (i.e., phrase or clause) (Beattie & Shovelton, 1999, 2002; Hadar & Pinchas-Zamir, 2004).

6.8. Experiment 2 – ERP Experiment

Experiment 2 investigated speech-gesture integration when the gesture does not synchronise with its semantic affiliate. In particular, it examined whether the information in preceding discourse that is loosely related to a gesture's meaning can influence processing of the gesture, and whether the information encoded in such a gesture can influence processing of its semantic affiliate later in the sentence. In order to investigate our research questions, we were interested in three ERP components reflecting different processes during discourse comprehension: the Nref, the N400 (mentioned in the introduction) and the P600.

6.8.1. Relevant ERP components: Nref, N400, P600

The Nref has been associated with discourse processing (spoken and written language). In particular, the Nref reflects referential processing within a discourse context (see Van Berkum et al., 2007 for a review). Nref is a sustained anterior negativity starting at around 300-400 ms for ERPs time-locked to a word that has more than one possible antecedents in a previously established discourse model when compared to unambiguous referents (e.g., “David shot at John as he jumped over the fence.”, “*he*” being ambiguous within the sentence context; (Van Berkum, Zwitserlood, Bastiaansen, Brown, & Hagoort, 2004)). Thus, it has been proposed that the Nref reflects the assignment of the correct referent to an ambiguous word rather than indexing any semantic anomaly (Van Berkum, 2009).

As we discussed in detail in the introduction, N400, a negative deflection at around 400 ms, has been taken to indicate processing load on semantic integration processes (Kutas & Hillyard, 1980). This component is often seen when the meaning of a word is

anomalous in relation to its preceding context. But already subtle manipulations of the degree of semantic congruency impacts the N400 amplitude (Kutas & Federmeier, 2011).

The P600 is a positive deflection peaking at around 600 ms after the presentation of a target stimulus. Although this component is traditionally associated with syntactic processing, more recent research suggests that the P600 more generally reflects the integration of the stimulus into the existing mental representation/discourse model (Brouwer et al., 2017; Brouwer et al., 2012; Brouwer & Hoeks, 2013). A conflict between new information and the existing discourse model elicits a larger P600 regardless of whether this conflict arises from syntactic anomalies or semantic anomalies (see Brouwer et al., 2012 for a review).

Based on these recent findings, Brouwer and colleagues (Brouwer et al., 2012) proposed the Retrieval Integration Account which suggests that the N400 reflects the retrieval of conceptual knowledge. For linguistic stimuli, this means retrieval of lexical semantic representations, and for non-linguistic stimuli, semantic retrieval from long-term memory (Hoeks, Brouwer, & Holtgraves, 2014). In the Retrieval-Integration Account, the P600 is associated with a particular type of integration processes; i.e. making the meaning of a stimulus and an already existing discourse model compatible with each other (e.g. through re-analysis/re-structuring of the discourse model). Thus, the P600 is a post-semantic component.

The Retrieval Integration Account is based on studies where semantic anomalies elicited a P600 effect but no N400 (Brouwer et al., 2012). For example, comparing ERPs time-locked to “eat” in the following sentences, did not elicit an N400 effect but a P600 effect: “For breakfast the eggs would eat toast and jam” and “For breakfast the boys would eat toast and jam” (Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007). According

to the Retrieval Integration Account the absence of an N400 effect can be explained by the semantic relationship between “breakfast” and “eat”. In this case, the preceding discourse (i.e., breakfast) primes the word “eat” regardless of its thematic violation. Thus, in both sentences lexical retrieval of “eat” is very similar and does not elicit an N400 effect (cf. Brouwer et al., 2012). However, a P600 effect was elicited because after lexical retrieval when the target word (i.e., “eat”) was integrated into the listener’s discourse model, the reanalysis of the discourse model elicited by the integration process is more effortful when the sentence included a “thematic role violation” (Kuperberg et al., 2007).

6.8.2. Relevant ERP components and speech-gesture processing

Taken together, the Retrieval Integration Account and the Nref are particularly useful for the present study because it allows us to differentiate between three gesture-speech integration processes; i.e., search for a referent (Nref), context driven meaning construction/semantic lexical retrieval (N400), post-semantic integration into a discourse model (including reanalysis of the discourse model) (P600).

Although previous ERP studies on gesture-speech processing used the N400 as indicator for a gesture’s integration into a discourse model (or other speech contexts such as word or phrase level), findings from previous studies could also be interpreted using the Retrieval Integration Account. According to the Retrieval Integration Account, N400 modulations (time-locked to the gesture) in previous studies could be interpreted as difficulties in constructing gesture’s meaning and/or retrieving conceptual knowledge necessary for the meaning construction (e.g., Cornejo et al., 2009; Özyürek et al., 2007). Such difficulties arise when the gesture does not semantically match its concurrent speech element(s) (Cornejo et al., 2009; Habets et al., 2011) or preceding speech (Özyürek et al., 2007) because mismatching speech elements make the construction of meaning and the

retrieval of relevant conceptual knowledge more effortful. N400 modulations time-locked to a target-word after the gesture has occurred (later in the sentence) could be interpreted as a priming effect of the gesture which facilitated lexical retrieval when the gesture matched the subsequent target-word (e.g., Holle & Gunter, 2007; Obermeier & Gunter, 2014).

Although most ERP studies on gesture-speech processing relevant to the current study looked exclusively at the N400 (Habets et al., 2011; Holle & Gunter, 2007; Obermeier et al., 2012; Obermeier & Gunter, 2014; Obermeier et al., 2010; Özyürek et al., 2007), the P600 has been observed in studies that investigated whether the gesture's meaning becomes part of a discourse model. A P600-like effect has been observed for metaphoric gestures (Cornejo et al., 2009; Ibáñez et al., 2010; Ibáñez et al., 2011) and abstract pointing gestures (Gunter & Weinbrenner, 2017; Gunter et al., 2015). For example, Cornejo et al. (2009) found a P600-like effect time-locked to the gesture's onset when gesture and its semantic affiliate (i.e., metaphorical expression), which was synchronised with the gesture, mismatched compared to when they matched. They interpreted this effect as a reanalysis of the existing discourse model elicited by the mismatching gesture. This interpretation is in line with Brouwer et al.'s (2012) Retrieval Integration Account.

The Nref, to our knowledge, has not been used in any previous gesture-speech processing study. However, for the current study design, the Nref is an informative component because it can tell us whether the participants searched preceding discourse for relevant information when constructing the meaning of a gesture that does not synchronise with its semantic affiliate and thus is ambiguous in meaning.

In order to investigate the effect of preceding discourse information on gesture-speech integration into a discourse model, we examined ERPs time-locked to the onset of gestures as well as time-locked to the onset of semantic affiliates (verbs) later in the sentence. The gestures were synchronised with a word that do not provide any information about the meaning of a gesture. We manipulated whether the discourse before the gesture provides information related to the gesture's meaning or not, and whether the gesture's meaning matched or mismatched with their semantic affiliates (verbs) later in the sentence.

6.8.3. ERPs time-locked to the gesture's onset

We first compared ERPs time-locked to the gesture's onset in the Related Discourse Condition with those in the Unrelated Discourse Condition. Experiment 1 showed that gesture interpretation was more constrained (but not fully disambiguated) by preceding discourse information in the Related Discourse Condition than in the Unrelated Discourse Condition. Moreover, participants rated the interpretation of the gesture to be more difficult in the Unrelated Discourse Condition. Hence, for the ERPs time-locked to the gesture, we predict an Nref with a more negative amplitude in the Unrelated Discourse Condition. This Nref might be elicited because in the Unrelated Discourse Condition, the listener should more intensively look for information in the already established discourse model to facilitate gesture interpretation, but might not be successful. Thus, in the Unrelated Discourse Condition the processing costs for the attempt to find information in the preceding discourse that could constrain the meaning of the gesture is higher than in the Related Discourse Condition.

We would not expect a strong N400 effect on gesture when comparing the two discourse conditions because the preceding discourse only roughly constrains but does

not disambiguate gestures, thus even the preceding discourse in the Related Discourse Condition should not prime semantic retrieval (i.e., construction of the gesture's meaning).

We would not expect a P600 effect on gesture when comparing the two discourse conditions either because the P600 is a post-semantic effect that restructures the discourse model. Because the preceding discourse only roughly constrains but does not disambiguate gestures, post-semantic processes are not fully triggered. That is, the information in the preceding discourse is not sufficient, even in the Related Discourse condition, to trigger the process of making the meaning of the gesture and the discourse model compatible with each other.

6.8.4. ERPs time-locked to the gesture's semantic affiliate

In order to investigate whether the information provided by a gesture can influence processing of the semantic affiliate later in the sentence, we examined ERPs time-locked to the onset of the gesture's semantic affiliate (i.e., the verb). There are three possible outcomes.

First, if synchrony of gesture and its semantic affiliate is crucial for gesture integration (e.g., Habets et al., 2011; McNeill, 1992), a mismatch between gesture and semantic affiliate (i.e., the verb) should not elicit any effects (N400, P600) on the semantic affiliate. This should be the case regardless of the discourse preceding the gestures.

Second, if Obermeier and Gunter's (2014) hypothesis holds that gestures that are not synchronous with their semantic affiliates *can* be integrated (at least if they are synchronous with a content word), then we would expect mismatch N400 effects time-locked to the target-verb, regardless of the discourse preceding the gestures. However,

from Obermeier and Gunter's (2014) study we cannot make any predictions about the P600 because they did not examine the time-window for the P600.

Third, we propose different predictions of the ERPs time-locked to the semantic affiliate. In particular, Experiment 1 showed that discourse information constrains gesture interpretation but the interpretation often does not converge on the semantic affiliate (the action verb in the second sentence): interpretations of a gesture included the correct target-verb used in the stimulus only 47% of the time in the Related Discourse Condition, and 41 % in the Unrelated Discourse Condition. Thus, in many cases gestures were still ambiguous, and thus cannot prime semantic retrieval of their lexical affiliates; therefore, we would not predict an N400 in either discourse condition.

However, we would expect a P600 effect on semantic affiliates of gestures. After the semantic representation of the semantic affiliate has been retrieved, the semantic affiliate is integrated into the discourse model. The integration of the semantic affiliate then triggers a reanalysis of the gesture's meaning which up to this point is still vague. P600 is elicited in the mismatch condition by this reanalysis of gesture which would both be more effortful in the mismatch condition than in the match condition. However, it is not clear whether the P600 effect differs according to the discourse preceding gesture (related vs. unrelated). However, how constraining the preceding discourse is may have an impact on the reanalysis process, and thus on P600 effects.

6.9. Methods

6.9.1. Participants

In total 38 participants took part in Experiment 2. One participant was excluded from the analysis because of being bilingual. Furthermore, five participants were excluded due to excessive artifacts (more than 25% of all trials). Artifacts are electrical

activities that are generated from sources other than the brain (e.g., muscle activity).⁸ Thus, 32 participants were included in the analyses (mean age = 20.1, SD = 3.4, 24 female, 29 right-handed). These were all monolingual English native speakers with normal or corrected-to-normal vision and did not report any hearing impairment. Note that we kept left-handed participants in the analyses to better represent the wider population, which consists of approximately 10% left-handers (cf. Willems, van der Haegen, Fisher, & Francks, 2014). All participants gave written informed consent and received either course credits or £ 7 per hour as compensation.

6.9.2. *Material*

Although the overall result of Experiment 1 showed a clear match/mismatch effect, we excluded some of the trials from the ERP experiment. First, responses to Question 3 (“How well does your interpretation of the gesture match the following word?”) showed that three items only showed a very small gesture match effect (i.e., the score for match minus that for mismatch) in the Related Discourse condition (<0.5 on the Likert Scale) (e.g., swimming gesture versus floating gestures for the verb ‘to swim’). Second, another item (verb pair scatter/roll) was excluded because its mismatching gesture (i.e., “rolling marbles across the floor”) was not perceived as a mismatching gesture (rated as 3.625 out of 5). Finally, another item was excluded because the rating for the match gesture was lower than for the mismatch gesture (i.e., for the verb pair catch/chase). In sum, forty-eight out of the 53 stimuli used in the behavioural experiment were included in the ERP experiment.

⁸ For an overview of the different types of artifacts see Nunez, Nunez, and Srinivasan (2016).

6.9.3. Procedure

Participants were told before the study that it investigated mechanisms underlying sentence comprehension. After the experiment, they were debriefed and the true purpose was revealed. The experiment was conducted in a sound-proof booth. The videos were presented on a computer screen placed 90 cm away from the participants. Audio was presented via speakers. Participants were instructed to watch and listen carefully. Furthermore, they were told when they were allowed to blink their eyes. Since each trial was very long (2 sentences, approximately 9 seconds), participants were told that they can blink normally when they see the fixation cross and during the first sentence of the trial, but should not blink during the second sentence.

In order to keep participants' attention, a comprehension task was included. After approximately five trials (randomised), participants were asked a yes/no comprehension question about the sentence they had just heard. For instance, for the example stimulus presented in Table 6.1, the question was "Did the speakers' uncle pick strawberries?". In this case, the answer was *yes*. For questions where the correct answer was *no*, one word was changed from the actual stimulus sentence (e.g., "Did the speaker's **aunt** pick strawberries?"). All questions were about the target sentence and not the introductory sentence.

The stimulus presentation is illustrated in Figure 6.6. First, participants saw a fixation cross for 1,000 ms, followed by the stimulus video. If a comprehension question followed the video, it was displayed in the centre of the screen. No time limit was given for responding to the question. Since the behavioural data was rather small (39 out of 192 items) we decided to include all trials into the ERP analysis regardless of whether the participants' response was correct or not (cf. Obermeier & Gunter, 2014).

Before the actual experiment, participants were presented with five practice trials. The purpose of the practice was twofold. First, the participants had the chance to get used to how the stimuli were presented. Second, the experimenter had a chance to check whether the participant understood the instructions on when to blink and when not to blink. The experiment was divided into four blocks, each containing 12 stimuli from each of the four conditions, randomly presented. These blocks were counter-balanced across participants. Taken together the four different lists of stimuli introduced above and the counter-balancing of the stimulus order resulted in 16 different versions of the experiment. Each block took approx. 9 minutes to complete. After each block the participant could take a break. In total, the experiment lasted about 50 minutes.

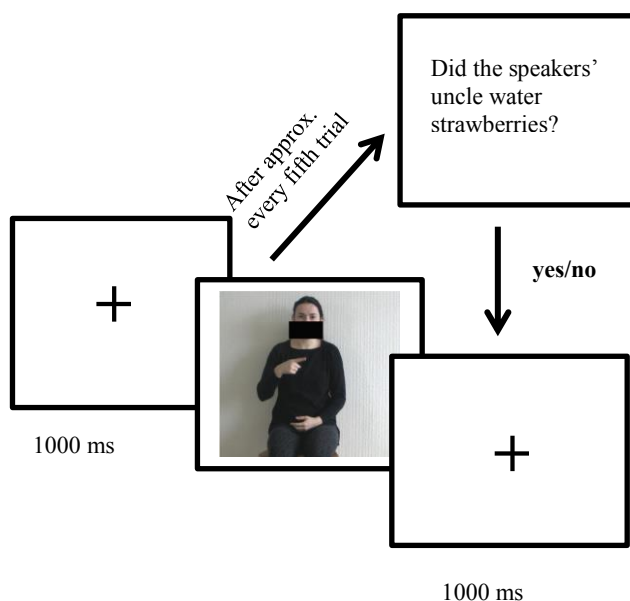


Figure 6.2. Stimulus presentation in Experiment 2 (ERP experiment)

6.9.4. EEG Recording and Analysis

The electroencephalograms (EEG) were recorded with a 64 BioSemi ActiView EEG system. Electrodes were placed and labelled according to the international 10-10 system. EEG data were sampled at 512 Hz. All electrodes were re-references offline to averaged mastoids. Electrooculograms (EOG) monitored vertical and horizontal eye

movements. Data were filtered offline using a band-pass filter of 0.1-30 Hz. Additionally, for presentation purposes only, a low pass filter was applied (10 Hz). For each trial, two epochs were created. The first epoch was time-locked to the onset of the gesture; the second epoch was time-locked to the onset of the verb of the second sentence. Epochs started 200 ms pre-stimulus onset and lasted 1200 ms. Since the stimuli videos were quite long and since we could not tell the participants where exactly not to blink, our datasets included large proportions of ocular artifacts. We therefore corrected eye-blinks using an Independent Component Analysis and following the procedure described in Nunez et al. (2016). After eye-blink correction, trials were rejected according to the following criteria: For the automatic artifact rejection a 200 ms sliding (50 ms steps) window was used. If any of the electrodes (except EOG) exceeded a ± 100 mV threshold within an epoch, the epoch was excluded from the analyses. Additionally, all EEG data were again inspected for artifacts. For the analysis of ERPs time-locked to gesture onset (out of a total 96 trials), match and mismatch trials were collapsed because at this point participants were not affected yet by the gesture–verb match. Based on the rejection criteria described above, on average, 93.8 % of the epochs time-locked to the gesture onset were included in the analysis. The average number of trials included in the Related Discourse condition was 89.7 (SD = 6.5) and 90.5 (SD = 5.1) in the Unrelated Discourse condition. For the analysis of ERPs time-locked to the verb, 92.7 % of all epochs were included in the analysis. This meant for each condition (out of a total 48 trials): 44.6 trials (SD = 3.9) in the Related Discourse condition and Matching Gestures, 44.5 trials (SD = 3.7) in the Related Discourse condition and Mismatching Gestures, 44.3 trials (SD = 4.0) in the Unrelated Discourse condition and Matching Gestures, and 44.6 trials (SD = 3.6) in the Unrelated

Discourse condition and Mismatching Gestures. Artifact-free trials were averaged for each participant and each condition.

For the analyses time-locked to the gesture, we defined seven ROIs with seven electrode sites each: Anterior Left (Fp1, AF7, AF3, F7, F5, F3, F1), Anterior Right (Fp2, AF8, AF4, F8, F6, F4, F2), Centre (FC1, FCz, FC2, C1, C2, Cz, CPz), Posterior Left (P1, P3, P5, P7, PO3, PO7, O1), Posterior Right (P2, P4, P6, P8, PO4, PO8, O2), Centre Left (FT7, FC5, FC3, T7, C5, C3, CP3) and Centre Right (FT8, FC6, FC4, T8, C6, C4, CP4). For the analysis time-locked to the verb, we excluded 2 ROIs (i.e., Centre Right and Centre Left).

Time-windows for the statistical analyses were chosen based on visual inspection of the wave forms. Mean amplitudes between two fixed latencies were submitted to repeated measures ANOVAs including the within-subject factors Gesture Match (Match, Mismatch), Discourse (Related Discourse, Unrelated Discourse) and ROI (Anterior Left, Anterior Right, Centre, Posterior Left, Posterior Right; for gesture analyses also Centre Left and Centre Right). Where appropriate, Greenhouse-Geisser (1959) correction was applied. Significant three-way interactions (Gesture Match x Discourse x ROI) were followed-up by two-way ANOVAs (Gesture Match x ROI) for each Discourse condition. For the verb-locked analysis, only effects and interactions that include the relevant factor Gesture Match are reported.

6.10. Results

Behavioural Results

The average response accuracy to the comprehension questions was 94.0 % (SD = 5.2). This high accuracy rate indicates that participants paid attention to the stimuli.

ERP Results

As can be seen in Figure 6.7, ERPs time-locked to gesture onset were not affected by the Discourse (Related versus Unrelated Discourse Condition) in the typical N400 time-window (350-550 ms). Therefore, no statistical analysis was conducted for this time-window. But Figure 6.7 suggests a more negative deflection for the Unrelated Discourse Condition compared to the Related Discourse condition, with the largest deviation of the two conditions from 800 ms up to 950 ms post-stimulus. Although later than in previous studies (Van Berkum et al., 2007), possibly due to the fact that it takes longer to interpret a gesture than a word, the anterior distribution of this deviation of the two Discourse Conditions suggests an Nref effect. We submitted the mean amplitudes between 800 and 950 ms to a repeated measures ANOVA with Discourse (2) and ROI (7) as within-subject factors. Figure 6.8 shows the topographic distribution of the difference between the Related and the Unrelated Discourse Conditions during this time window. No main effect of Discourse ($F(1, 31) = 2.222, p = .146$) was found, but a trend for a Discourse x ROI interaction ($F(4, 117) = 2.180, p = .088$). Post-hoc comparisons showed a significant difference between the Related Discourse Condition and the Unrelated Discourse Condition in the Centre Left region ($t(31) = 2.245, p = .032$) and trends for differences in the two anterior regions (Anterior Left: $t(31) = 1.834, p = .076$; Anterior Right: $t(31) = 1.839, p = .076$), but no significant differences in the other regions (Centre: $t(31) = 1.105, p = .278$; Centre Right: $t(31) = .942, p = .353$; Posterior Left: $t(31) = .638, p = .528$; Posterior Right: $t(31) = .206, p = .838$). While the effect is not very strong, it indicates that the Unrelated Discourse indeed led to more negative deflections than the Related Discourse at left channels and weakly at frontal sites.

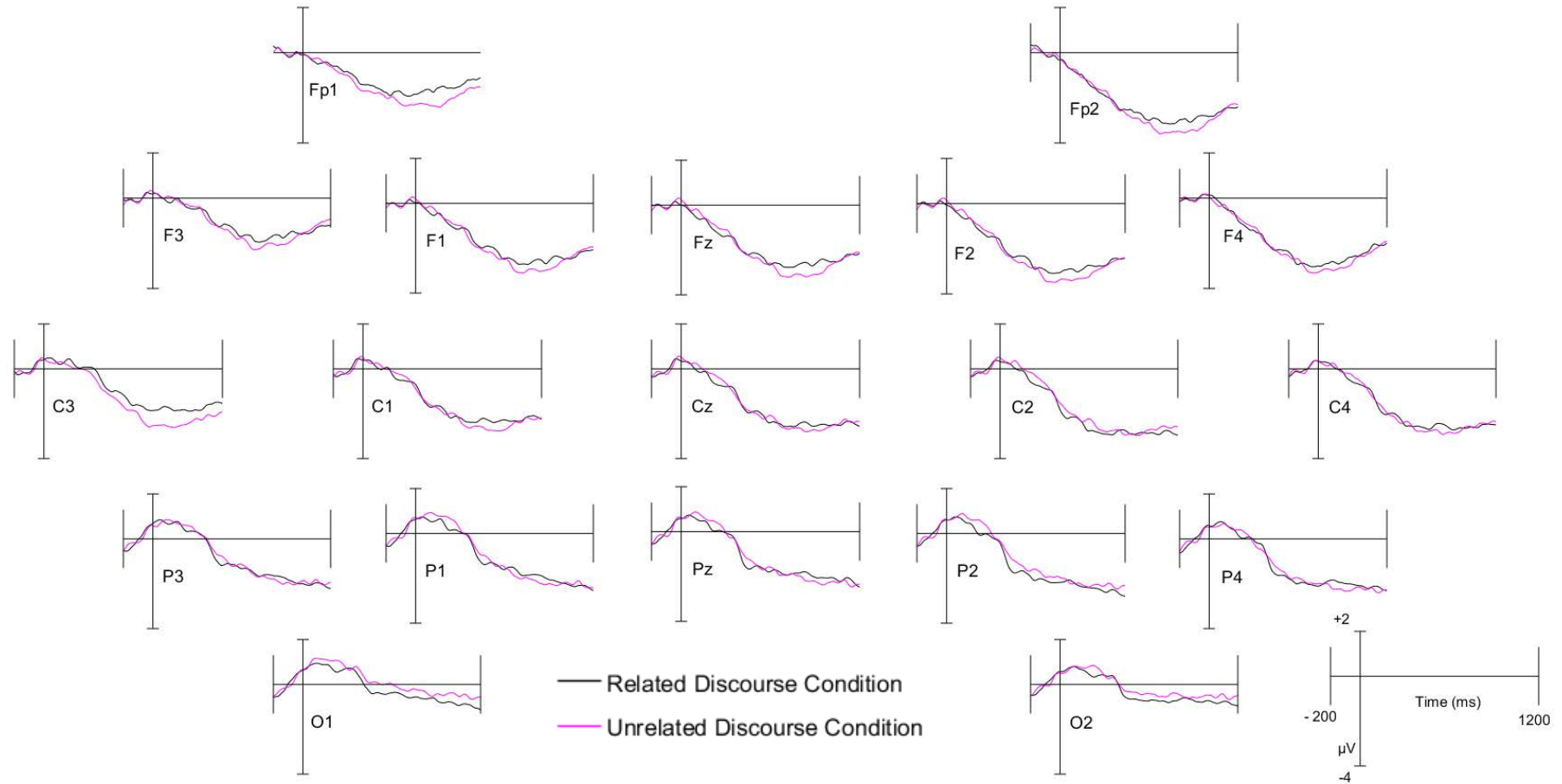


Figure 6.3. ERP responses at a subset of electrodes in the Related Discourse Condition compared to the Unrelated Discourse Condition time-locked to gesture onset.

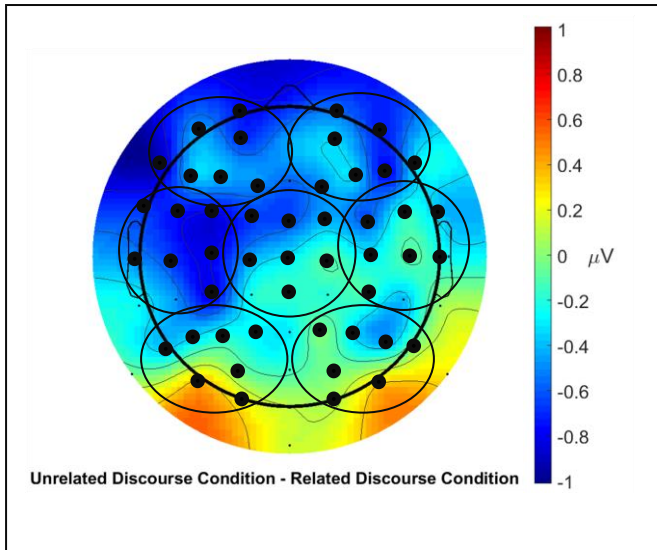


Figure 6.4. Topographical map of the ERP difference between Unrelated Discourse Condition and the Related Discourse Condition in the time-window 800 -950 ms post gesture onset. Black dots show electrode sites and circles the ROIs included in the statistical analysis.

Figure 6.9 and Figure 6.10 show the ERPs for a selection of electrodes time-locked to the target verb. Mismatching gestures led to a more positive deflection than matching gestures, particularly from 800-1200ms. This effect was found over parietal and occipital sites for the Related Discourse Condition and over anterior sites in the Unrelated Discourse Condition (see Figure 6.11).⁹ We submitted mean amplitudes between 800 and 1200 ms to a repeated measures ANOVA with Discourse (2), Gesture Match (2) and ROI (5) as within-subject factors. We found a significant main effect of Gesture Match ($F(1, 31) = 6.076, p = .019$) and a significant three-way interaction between Discourse, Gesture Match and ROI ($F(2, 68) = 4.932, p = .008$). We therefore conducted separate ANOVAs for the Related and Unrelated Discourse Conditions, testing the effects of the factors Gesture Match and ROI. For the Related Discourse Condition, no main effect of Gesture Match ($F(1, 31) = 1.611, p = .214$), but a significant interaction between Gesture Match

⁹ Visual inspection of the waveforms further indicated that there might be an early mismatch effect in the Related Discourse Condition (50-300 ms). Although we found a significant interaction between ROI and Gesture Match, follow-up t-tests did not yield a significant difference of Gesture Match and Gesture Mismatch in any of the five ROIs ($p > .05$).

and ROI ($F(3, 82) = 3.767, p = .018$) was found. T-tests for each ROI showed that mismatching gestures led to a significant positivity compared to matching gestures at Posterior Left ($t(31) = -2.709, p = .011$) and Posterior Right sites ($t(31) = -2.523, p = .017$). No significant difference was found for the other three regions (Centre: $t(31) = -.533, p = .598$; Anterior Left: $t(31) = .423, p = .675$; Anterior Right: $t(31) = -.158, p = .875$). For the Unrelated Discourse Condition, a marginally significant main effect of Gesture Match ($F(1, 31) = 3.061, p = .090$) and a marginally significant interaction of Gesture Match and ROI was found ($F(2, 72) = 2.649, p = .070$). Comparisons for each ROI yielded a significant ERP difference between mismatching and matching gestures in the Centre region ($t(31) = -2.531, p = .017$) and a trend in the Anterior Right region ($t(31) = -1.739, p = .092$). Gesture Match did not affect ERPs in any of the other three regions (Anterior Left: $t(31) = -1.576, p = .125$; Posterior Left: $t(31) = -.249, p = .805$; Posterior Right: $t(31) = -.306, p = .762$). These analyses show that the manipulation of Gesture Match led to a posterior P600 effect in the Related Discourse Condition and a central-frontal P600 effect in the Unrelated Discourse Condition.

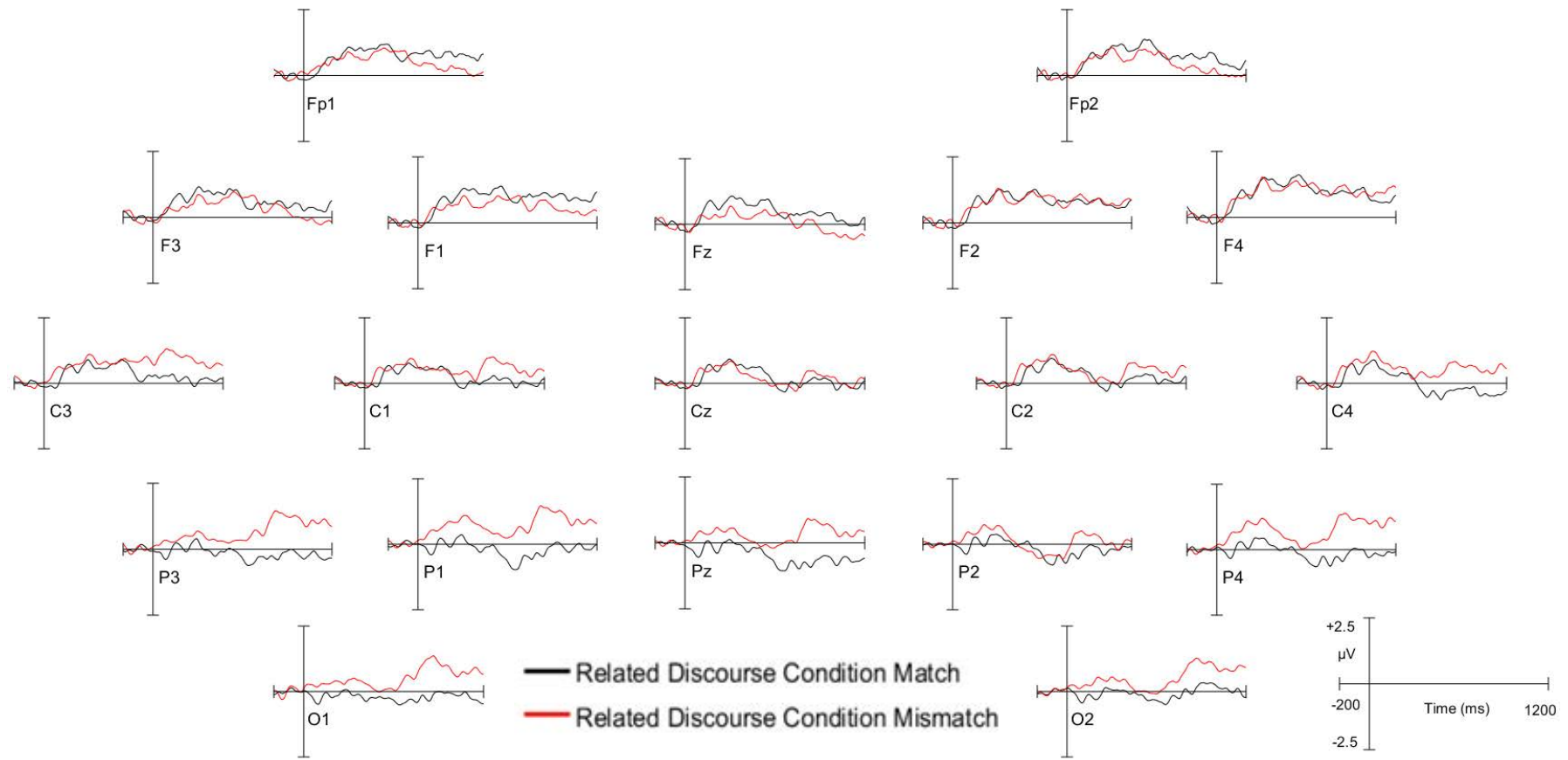


Figure 6.5. Gesture Match and Gesture Mismatch waveforms for selected electrodes in the Related Discourse Condition time-locked to the target verb onset.

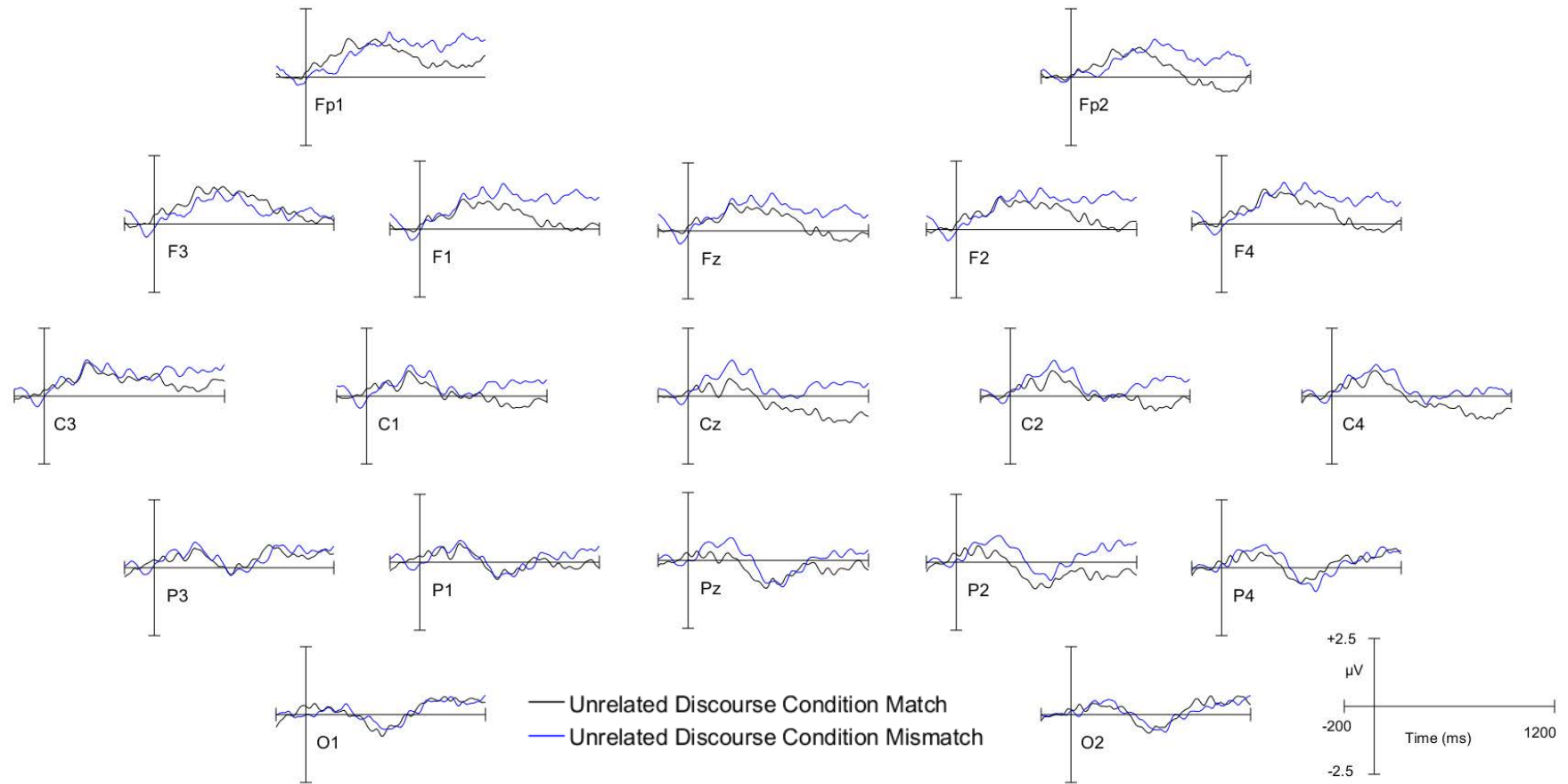


Figure 6.6. Gesture Match and Gesture Mismatch wave forms for selected electrodes in the Related Discourse Condition time-locked to the target verb onset.

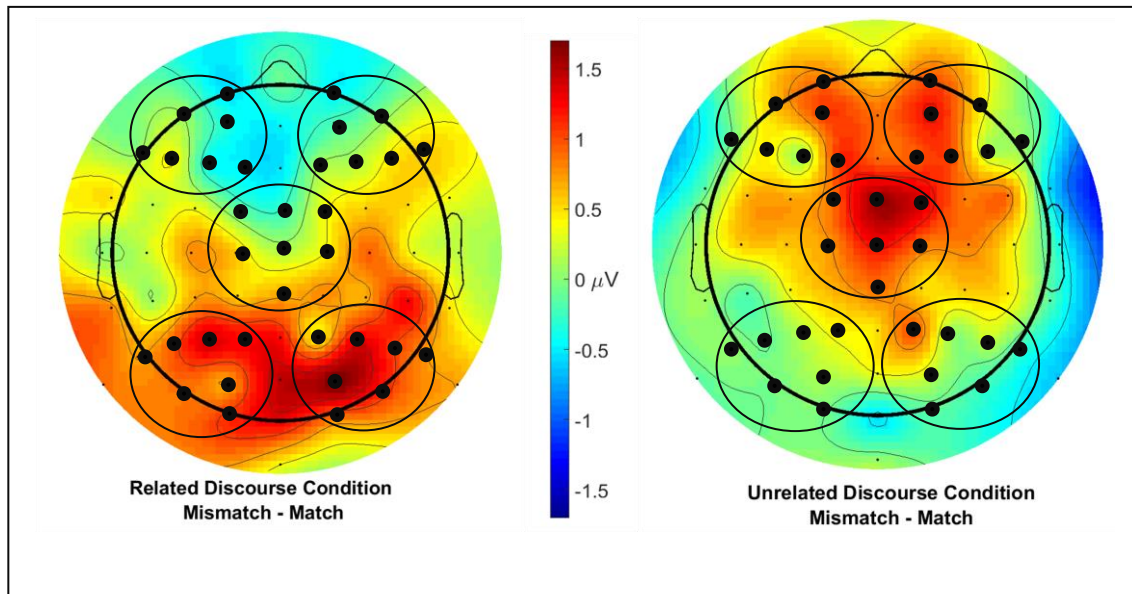


Figure 6.7. Topographical maps of the ERP difference between mismatching and matching gestures in the Related Discourse Condition and the Unrelated Discourse Condition for the 800-1200 ms time-window. Black dots refer to the electrode sites and circles to the ROIs included in the statistical analysis.

6.11. Discussion Experiment 2 - ERP Experiment

The ERP experiment tested whether discourse information enables an integration of gesture into a listener's discourse model when the gesture does not synchronise with its semantic affiliate. ERPs time-locked to gesture onset showed an Nref effect, namely more negative ERPs in the Unrelated Discourse Condition compared to the Related Discourse Condition at left sites in a quite late time-window (i.e., 800-950). ERPs time-locked to the gesture's semantic affiliate, the target verb later in the sentence did not show a gesture mismatch effect in the N400 time-window, but showed a P600 effect. More specifically, in the Related Discourse Condition we found a posterior P600 effect (mismatch more positive than match) distributed over both posterior hemispheres, starting at around 800 ms post-stimulus. No such posterior effect was found in the Unrelated Discourse Condition. Instead, in the same time-window (800-1200 ms), ERPs in the Unrelated Discourse Condition showed a significantly

more positive deflection for mismatching gestures than matching gestures distributed over central-frontal electrode sites.

6.11.1. Interpretation of ERPs time-locked to the gesture's onset

ERPs time-locked to the gesture's onset showed the expected negative deflection over anterior sites in the Unrelated Discourse Condition compared to the Related Discourse Condition. Although this effect occurred later than previously reported Nref effects, i.e. 300-400 ms post-stimulus (Van Berkum et al., 2007), we would still argue that this negative deflection in the Unrelated Discourse Condition compared to the Related Discourse Condition reflects the allocation of a referent to the gesture. First, the scalp distribution is very similar to previously reported Nref effects; i.e. across anterior left electrodes (Boudewyn et al., 2015; Nieuwland & Van Berkum, 2006). Second, in previous studies that found an Nref effect, the ambiguous word was either a pronoun or noun (Boudewyn et al., 2015; Nieuwland & Van Berkum, 2008) which presumably was semantically and referentially processed faster than a gesture. This argument is supported by findings from previous studies on gesture and action processing where unusually late latencies have been observed (Cornejo et al., 2009; Ibáñez et al., 2011; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008). The late onset of ERP components arise from the gesture's or action's meaning unfolding over time (i.e., a few hundred milliseconds) (e.g., Cornejo et al., 2009). If so, the characteristics of the video stimuli can explain the late onset of the Nref effect observed in our study.

As predicted, we did not find an N400 effect. According to the Retrieval Integration Account (Brouwer et al., 2012), a more negative deflection in one of the discourse conditions would have indicated difficulties in constructing the gesture's meaning which involves retrieval of conceptual knowledge from the listener's long term memory. The absence of an N400 effect time-locked to the gesture's onset can be explained by the unpredictability of the gesture in

both discourse conditions (i.e., related and unrelated). Thus, the preceding discourse information does not prime the construction of the gesture's meaning.

Also in the P600 time-window, we did not find an effect of discourse either. Following the Retrieval Integration Account by Brouwer and colleagues (Brouwer et al., 2012), the lack of a P600 effect suggests that the gesture's meaning was too vague to complete post-semantic processing of a gesture and the discourse. In particular, post-semantic processing was only completed once the semantic affiliate has been retrieved from the mental lexicon which then triggered a reanalysis of the gesture's meaning.

6.11.2. Interpretation of ERPs time-locked to the semantic affiliate's onset

In terms of the ERPs time-locked to the semantic affiliate (the target-verb), we did not find a gesture mismatch effect in the N400 time-window in either of the two discourse conditions, as predicted. Following the Retrieval Integration Account (Brouwer et al., 2012) this means that the gesture did not prime the verb's lexical retrieval. We argue that we did not observe an N400 effect because the meaning of gestures was too vague in order to prime the semantic affiliate (even in the Related Discourse Condition). This is in line with our behavioural experiment where we found that although preceding discourse information can constrain the gesture's interpretation, it does not necessarily disambiguate the gesture.

The lack of an N400 effect time-locked to the target-verb further supports our hypothesis that gesture integration into a discourse model has not been completed before the semantic affiliate was encoded later in the sentence. The post-semantic integration process of the semantic affiliate (after its lexical retrieval) triggered the reanalysis of the gesture's meaning, which resulted in a P600 effect in both discourse conditions (although with different topographical distributions). Thus, the P600 mismatch effect elicited time-locked to the semantic affiliate indicates that the gesture has been fully disambiguated.

6.11.3. Interpretation of posterior and anterior P600 effects

The different topographical distributions of the P600 effects in the Related Discourse Condition (i.e., posterior) and in the Unrelated Discourse Condition (i.e., anterior) suggest that preceding discourse information had an influence on how gestures and consequently the gesture's semantic affiliates were processed. In previous sentence processing studies (i.e., congruent/incongruent sentence completion experiments) that found a "semantic P600 effect", a posterior (i.e., parietal) P600 effect was more common. This P600 effect was attributed to a reanalysis of the incoming signal (see van Petten & Luka, 2012 for a review). However, P600s with an anterior (i.e., frontal) distribution have been observed in similar experiments as well (e.g., Thornhill & Van Petten, 2012). Moreover, van Petten and Luka (2012) pointed out that the semantic manipulation of the sentence final word had an influence on the topographical distribution of the P600 effects. In particular, in experiments where congruous sentence completions were compared with incongruous completions, mainly posterior P600 effects were observed (e.g., congruent (original stimuli in Dutch): "The painter colored the details with a small paint *brush*"; incongruent: "The painter colored the details with a small *labyrinth* (van den Brink et al., 2006)). In contrast, experiments that compared sentences with high-cloze probability endings (i.e., most preferred endings established in a pretest) versus low-cloze probability endings (i.e., semantically congruent but unexpected endings) primarily reported frontal P600 effects (e.g., high-cloze: "On his vacation, he got some much needed *rest*"; congruent but low-cloze: "On his vacation, he got some much needed *sun*" (Thornhill & Van Petten, 2012)). According to van Petten and Luka (2012), the posterior effect reflects reanalysis and the anterior effect reflects disconfirmed predictions. For the results of the present study, we propose a related but slightly different interpretation of the underlying mechanism of the anterior P600 effect.

As discussed above, in both discourse conditions the participants' meaning constructed for a given gesture was vague until the semantic affiliate was retrieved from the mental lexicon (i.e., matching gesture and mismatching gesture) which can explain the absence of an N400 effect. After the semantic affiliate's meaning had been retrieved, the integration of the semantic affiliate into the discourse model triggered the reanalysis of the gesture's meaning. This reanalysis process was less effortful when the gesture actually matched the verb compared to the mismatch gesture. Moreover, this re-analysis of the match/mismatch gesture led to a *mismatch effect* in the Related Discourse Condition (i.e., posterior P600). In the Unrelated Discourse Condition the reanalysis process of the gesture's meaning did not elicit a strong enough contrast to be perceived as mismatch but rather as not fitting well (or unexpected) which was evident in an anterior P600 effect. In other words, the preceding discourse information in the Related Discourse Condition constrained the gesture's possible interpretations which influenced the reanalysis process.

6.11.4. *Ecological validity of the stimuli*

In order to gain a better understanding of multimodal language processing, it is important to study gesture-speech processing in cases where gesture and its semantic affiliate are not in synchrony. This is important because such cases commonly occur in spontaneous gesture-speech production. Although McNeill's proposed "Semantic Synchrony Rule" states that gestures synchronise with their semantic affiliates, this rule is mainly based on qualitative analysis of English narratives (McNeill, 1992). Other qualitative analyses have suggested that iconic gestures can be produced (and terminated) prior to the encoding of the gestures' semantic affiliates. For instance, Kita (2000, pp. 173-175) discussed an example where an adult English speaker produced an iconic gesture depicting a complex motion sequence with a filled pause "aaaa" and the word "assume" in the utterance, "because it **aaaa**. Well actually what happens

is, he I, you, **assume**”, before the actual motion event description started in speech (“that he swallows this bowling ball and he comes rolling out...”). Similarly, McNeill (2005, pp. 138-141) discussed a case where an adult Turkish speaker produced a gesture depicting manner of motion one phrase before the co-expressive manner word was produced in speech. Another example discussed by McNeill (1985, pp. 361) further shows that gesture-speech synchronisation is not always present. In this example, a speaker produced a gesture indicating a location of a character while producing the sentence “they keep on **flashing back** to Alice just sitting there”. The gesture anticipated its semantic affiliate and even terminated several words before the semantic affiliate (i.e., “there”). Quantitative analyses further suggest that in many cases gestures do not synchronise with their semantic affiliates. In Chapter 4, we showed that in motion event narrations encoding both manner and path in speech, more than half of the path gestures were produced (and terminated) before the semantic affiliate (i.e., the path particle). Although the stimuli in the present study are not exact replica of the responses from the production experiments presented in this thesis, there is qualitative evidence of gesture asynchronies similar to the synchrony manipulation in our stimuli. Thus, our stimuli do reflect natural speech production. Future studies could further investigate varying degrees of asynchrony and their influence on gesture-speech processing.

6.12. General Discussion

In both experiments we investigated how gestures that did not synchronise with their semantic affiliates can be integrated with speech and whether preceding discourse related to the gesture’s meaning influences integration processes. Our findings showed that in terms of synchrony language processing is more flexible than often assumed (cf. Habets et al., 2011; McNeill, 1992). In particular, we found that gestures *can* be integrated into a discourse model even when the gesture does not synchronise (i.e., overlap) with its semantic affiliate. Moreover,

gesture interpretation and processing is discourse-dependent. In particular, behavioural and electrophysiological evidence that information encoded in the preceding discourse of an iconic gesture is taken into account when interpreting a gesture which further has an impact on how the gesture's semantic affiliate is processed later in the sentence.

The key novel contribution of this study is to distinguish three different types of speech-gesture integration processes, based on research on discourse integration (Van Berkum et al., 2007) and the Retrieval Integration Account (Brouwer et al., 2012): search for a referent in preceding discourse (Nref), context driven meaning construction/semantic lexical retrieval (N400), post-semantic integration into a discourse model (P600). In particular, we found an Nref and P600 effects in our study. Moreover, the present study has important implications for the role of gesture-speech synchrony on language processing. That is, synchronisation of a gesture and its semantic affiliate is not essential for a gesture to become part of a discourse model. However, the underlying cognitive processes involved in this integration differ from synchronous gesture-speech combinations which we could show by distinguishing between three different integration processes reflected by the Nref, N400 and P600.

First, to our knowledge, this is the first gesture study reporting an Nref effect in gesture-speech integration. More specifically, an Nref effect was found in the Unrelated Discourse condition in comparison to the Related Discourse condition. This indicates that the participants looked for information relevant for gesture interpretation in the preceding discourse because the Nref reflects the search for a referent in the previously established discourse model (Van Berkum et al., 2007).

Second, contrary to previous studies (e.g., Holle & Gunter, 2007; Özyürek et al., 2007), we did not find an N400 effect time-locked to the gesture's semantic affiliate later in the sentence. We argue that because the meaning of gesture was vague (in both discourse

conditions), it did not prime the lexical retrieval of the semantic affiliate. According to the Retrieval Integration Account (Brouwer et al., 2012) no N400 should be elicited in such cases. Moreover, the lack of an N400 effect is compatible with studies that manipulated the synchronisation of gesture and its semantic affiliate. When gesture precedes and does not overlap with the semantic affiliate, the gesture's meaning is left vague; consequently, no mismatch N400 effect is observed on the semantic affiliate (Habets et al., 2011; Obermeier et al., 2010).

Furthermore, our results contradict Obermeier and Gunter's (2014) proposal on the interpretation of gesture integration processes when gestures do not synchronise with (i.e., precede) its semantic affiliate. They proposed (based on findings in Obermeier et al., 2010 and Obermeier and Gunter, 2014) that such gestures are automatically integrated into a discourse model when they synchronise with a content word but not when they synchronise with a function word. Content words can at least partially be integrated with gestures' meaning, but function words cannot. This is because function words have very little semantic contents. However, the current study did not find a N400 gesture mismatch effect time-locked to the semantic affiliate but rather a P600 effect. In line with the Retrieval Integration Account (Brouwer et al., 2012), we argue, that in both studies by Obermeier and colleagues (Obermeier & Gunter, 2014; Obermeier et al., 2010), an N400 gesture mismatch effect was elicited when the gesture was disambiguated by discourse information related to the gesture's meaning which then primed the semantic affiliate. When gesture synchronised with a pronoun and was disambiguated, then, based on our findings, we would have expected a P600 effect due to reanalysis of the gesture at the target word. However, neither of the two studies looked at this time-window.

Third, we found gesture mismatch effects in the P600 time-window in both discourse conditions (although with different topographical distributions). So far only one gesture study reported similar ERP findings (i.e., “semantic P600 effect”). In particular, in Gunter and Weinbrenner (2017), participants watched an actor placing two referents (Shakespeare vs. Goethe) left vs. right in gesture space with an abstract pointing gesture. Later in the discourse, an abstract pointing gesture (i.e., pointing to the left or right) was used synchronous with Shakespeare or Goethe in speech, and indicated a location matched or mismatched the previously established locations for the referent (e.g., Goethe was on the right, Shakespeare on the left). A P600 effect was elicited when the gesture did not match the target-word. They argued that participants tried to reanalyse gestural information when the target-word is integrated into a discourse model which elicited a P600 gesture mismatch effect.

We hypothesise that the constructed meaning of the gesture was vague (regardless of the discourse condition) because the gesture did not synchronise with its semantic affiliate. Thus, semantic processing of the gesture was not completed until after lexical retrieval of the semantic affiliate. We argue that the gesture has been re-accessed from the working memory once the semantic affiliate has been retrieved from the mental lexicon (i.e., elicitation of a P600 effect).

In terms of the distinction between different gesture integration processes discussed above, we not only provide an explanation for the findings of the current study, but also provide an alternative explanation of the N400 effects found in previous studies that only looked at the N400 time-window (e.g., Habets et al., 2011; Holle & Gunter, 2007; Obermeier & Gunter, 2014; Obermeier et al., 2010; Özyürek et al., 2007), i.e., the N400 reflects context driven meaning construction/semantic lexical retrieval (N400) whilst the P600 indicates post-semantic integration into a discourse model. So far, the Retrieval Integration Account (Brouwer et al., 2012) has only been used by Gunter et al. (2015) and Gunter and Weinbrenner (2017) to explain

the processing of abstract pointing gestures. Our study showed that this account is also informative for interpreting integration processes of iconic gestures.

6.13. Conclusions

To sum up, previous research on gestures asynchronous with their semantic affiliate in speech showed that gesture processing is influenced by gesture transparency and gesture synchrony with the semantic affiliate (Habets et al., 2011; Kelly et al., 2004; Obermeier & Gunter, 2014; Obermeier et al., 2010). The current study adds to the literature that preceding discourse also influences how listeners interpret and process gestures that do not synchronise with their semantic affiliates. Furthermore, when the speech context, including preceding discourse, does not fully disambiguate gesture's meaning, semantic processing of the gesture is only completed once the semantic affiliate downstream in the sentence has been encoded. In particular, our results indicate that when the gesture precedes its semantic affiliate, a reanalysis of the gesture's meaning is triggered once the semantic affiliate is processed semantically. Thus, in the current study, gestures had no facilitating effect on language comprehension. Rather, the manipulation of synchrony resulted in different integration processes compared to synchronous gesture-speech combinations. Different processes are reflected in different ERP components.

Chapter 7

General Discussion

The temporal and the semantic relationship between speech and gesture has been the subject of an increasing body of research, and findings from this research have been used as basis for numerous speech-gesture production and comprehension models/theories (e.g., de Ruiter, 2000; Kita & Özyürek, 2003; McNeill, 1992). Despite growing research efforts, the ways in which speech and gesture interact during the different stages of production and comprehension are, in many ways, still poorly understood. This thesis aimed to contribute to a better understanding of the mechanisms that underlie gesture-speech production and comprehension. In this final section of the thesis, the findings of the individual chapters are brought together. Moreover, the contributions of this thesis to the literature are discussed. Finally, future directions and open questions related to the topics investigated will be presented.

7.1. Summary of Findings & Conclusions

Chapter 3 investigated on which linguistic level gesture and speech are coordinated in production. Two production experiments (conducted in German and English) were designed to answer this question. In particular, the study tested whether previously found differences in the gestural packaging of motion events can be linked to differences in how motion events are lexicalised (one-verb versus two-verb constructions) in speech or how these events are packaged into planning units (e.g., Kita & Özyürek, 2003; Kita et al., 2007; Özçalışkan et al., 2016; Wessel-Tolvig & Paggio, 2016). This manipulation was possible, because in German the manner verb and the path particle can be separated on the surface structure by other linguistic elements. The findings from both experiments showed that when manner (i.e., verb) and path (i.e., particle) are linguistically encoded within two planning units, these two motion event

components are more likely to be separated gesturally, compared to particle verbs being encoded within one planning unit. In this case, conflated manner and path gestures were more likely. The results from the two experiments are most parsimonious with the Planning Unit Account which assumes that gesture and speech are coordinated on a planning unit level. Thus, the study could demonstrate that the amount of information we process within one planning cycle in speech shapes gestural content.

In Chapter 4 a meta-analysis of the German data from the gesture-speech production tasks was conducted. The study aimed to answer how speech and gesture are coordinated during production as well as their coordination during the gesture's execution phase. These issues were tackled by analysing whether the surface location of a gesture's semantic affiliate that is separated on the surface structure (e.g., literal translation: *rolling in this short video down*) predicts gesture onset and gesture duration. The results show that the surface location of a semantic affiliate has an influence on both, gesture onset and gesture duration. In particular, if a gesture (i.e., manner and path conflated gesture or path only gesture) was placed after the first semantic affiliate, both the verb and the particle pulled the gesture towards them which resulted in varying degrees of asynchrony between the gesture's onset and the onset of the semantic affiliate (i.e., path particle). Consequently, a large proportion of gestures did not synchronise with either of their semantic affiliates. Moreover, the pulling effect of the manner verb was stronger for conflated gestures than for path gestures. This indicates that the strength of the pulling effect varies depending on how many semantic features a gesture shares with an affiliate (i.e., MANNER + MOTION for conflated gestures; MOTION for path gestures (Talmy, 2000)). The analyses also revealed that the speakers tried to balance these asynchronies by lengthening the gesture if its onset preceded the semantic affiliate. Thus, the study presented in the

synchronisation chapter (Chapter 4) showed that semantic affiliates of a gesture function as attraction points that pull the gesture towards them.

Based on the asynchronies between gestures and their semantic affiliates in speech found in Chapter 4, the study in Chapter 6 tested how gestures that do not synchronise with their semantic affiliates are processed by listeners. In particular, the study presented in Chapter 6 tested whether preceding discourse related to the gesture's meaning has an influence on gesture interpretation (behavioural experiment). Moreover, an ERP experiment tested whether discourse preceding the gesture influences gesture integration into a listener's discourse model when synchrony between gesture and its semantic affiliate is not present. The findings from the behavioural experiment showed that listeners consider discourse information when they interpret a gesture's meaning in a way that discourse information related to the gesture's meaning constrains gesture interpretation.

The ERP study showed an Nref effect time-locked to the gesture's onset with a more negative deflection in the condition where discourse information was unrelated to the gesture's meaning. This indicates difficulties in finding an appropriate referent for the gesture. ERP results are in line with the behavioural results. That is, gesture interpretation is easier when preceding discourse constrains interpretation. However, the ERP study also showed that the referents provided in the related discourse that constrain gesture interpretation, were not strong enough to elicit an N400 effect time-locked to the semantic affiliate when comparing match and mismatch gestures. This suggests that the meaning construction of the gesture has not been completed since the gesture's meaning was too vague. Following the Retrieval Integration Account (Brouwer et al., 2012), the absence of an N400 effect was interpreted as the gesture not having a facilitating effect on the lexical retrieval of the semantic affiliate.

In the P600 time-window, a P600 gesture mismatch effect was found time-locked to the semantic affiliate in both discourse conditions (i.e., related and unrelated). This led to the conclusion that only after the semantic affiliate has been retrieved, the gesture was integrated into the discourse model. In particular, the integration process of the semantic affiliate into the already existing discourse model triggered the reanalysis of the gesture's meaning which was more effortful when the gesture did not match the semantic affiliate. At this point, also preceding discourse information came into play which was evident in the topographically different P600 gesture mismatch effects. Topographically different P600 effects suggest that the gesture was assigned a more specific meaning during the reanalysis of the gesture when preceding discourse was related to the gesture's meaning.

7.2. Implications

7.2.1. Planning Units in Speech and Gesture Production

A large number of previous studies have compared gestures produced by speakers of either a satellite-framed or a verb-framed language (see Chapter 3, Table 2.1 for a list). The differences found across two typologically different languages (i.e., verb-framed: manner and path gesturally separated, satellite-framed: manner and path gesturally conflated) have been attributed to differences in clausal packaging (i.e., one clause construction versus two clause construction). However, these studies were unable to establish why the clause plays an important role in gesture-speech production. Thus, the finding that information packaging in speech (i.e., into planning units) has an influence on gestural information packaging has important implications for the gesture-speech production literature. First, it provides evidence that the way in which a motion event is lexicalised (i.e., one-verb constructions versus two-verb constructions) does not drive gestural depiction. What matters is whether the two motion event components (i.e., manner and path) are encoded within one or within two planning units. This

supports the Interface Model proposed by Kita and Özyürek (2003). On the other hand, these results refute the idea that clauses represent “conceptual units” (Pawley, 1987, 2010; Pawley & Syder, 1983) as well as Slobin’s (2000, 2003) claim that in satellite-framed languages motion events are perceived as “single conceptual events” and thus shape gestural content. Second, the study also provides evidence that the way in which a motion event is packaged into planning units can lead to a reconceptualisation of this event. Recent research on silent gestures (Özçalışkan, 2016; Özçalışkan et al., 2016) suggests that conflated gestures might be the “default setting” (Gullberg, 2011, p. 185) for how we conceptualise motion events if no language is involved. The study presented in this thesis provides evidence that information packaging into planning units can reconceptualise this pre-linguistic conceptualisation. This view contradicts a strict modular view of speech production that suggests that there is no feedback channelling between the conceptualiser and the formulator where the surface structure (i.e. syntactical structure) is generated (de Ruiter, 2000; Levelt, 1989). Rather, the findings are in line with the idea that the pre-verbal message and the surface structure are constructed interactively (Kita, 1993; Kita & Özyürek, 2003; Vigliocco & Kita, 2006).

7.2.2. The Attraction Point Hypothesis of Gesture-speech Synchronisation

The findings from the synchronisation study presented in Chapter 4 are not/only partly in line with current theories of gesture-speech synchronisation. First, the finding that semantic affiliates that are separated on the surface structure pull the gesture in an aim for synchronisation, does not support de Ruiter’s (1998, 2000) Sketch Model which predicts that a gesture is initiated roughly at the same time as its affiliate in speech. This assumption is based on the idea that the pre-verbal message for speech and the gesture are planned simultaneously in the conceptualiser. Hence, the Sketch Model does not predict that the surface locations of the semantic affiliates have an influence on the gesture’s onset or duration. Second, the results are

partly in line with McNeill's Semantic Synchrony Rules (1992, 2005) which proposes that gestures synchronise with their semantic affiliate. However, according to the Semantic Synchrony Rule, gestures were not expected to fall between the two semantic affiliates (i.e., the verb and the particle) which was the case for a large proportion of gestures (Path gestures and Conflated gestures). More specifically, the finding that semantically co-expressive words with which the gesture does not synchronise still impacts gesture onset, is not in line with McNeill's (1992, 2005) theory of "idea units". McNeill assumes that an idea unit can only be inferred via the speech elements the gesture is in synchrony. The results in Chapter 4 show that a gesture does aim to synchronise with one or multiple of its semantic affiliates but having more than one affiliate or "attractor" within an utterance can cause asynchronies due to the pulling effect of these attractors.

The findings that an early onset of a gesture lengthened the gesture's stroke provides further evidence for the interactive view of gesture-speech production (but not the ballistic view). The interactive view assumes that the feedback channel between gesture and speech stays open even after the gesture's onset. The interactive view of gesture-speech production has previously been found for deictic gestures (Chu & Hagoort, 2014) and for gestures produced in narratives (Morrel-Samuels & Krauss, 1992). The synchronisation study presented in Chapter 4 adds to the literature that the interactive view also holds for iconic gestures and that the type of iconic gesture (repetitive, non-repetitive) influences for how long a gesture is prolonged (cf. Kita, 1990).

Since no existing model or theory of gesture-speech synchronisation could explain the synchronisation patterns found in the study presented in this thesis, a new theory of how speech synchronises with representational gestures (i.e., iconic gestures and deictic gestures) was proposed: The Attraction Point Hypothesis of gesture-speech synchronisation. The main idea

of this hypothesis is that semantic affiliates of a gesture function as “attraction points” which have a pulling effect on the gesture. This hypothesis can explain 1. why gestures do not necessarily overlap with their semantic affiliates (Gesture Onset Rule), 2. why gestures are lengthened in case of an early onset (Gesture Offset Rule) and 3. why some affiliates might have a stronger pulling effect (Semantic Overlap Rule).

7.2.3. Gesture-speech Production

In the first chapter of this thesis the close semantic and temporal relationship between gesture and speech has been outlined and the importance of this relationship for gesture-speech production models has been highlighted. In the planning unit chapter (Chapter 3), for the first time, direct evidence was found that the amount of information a speaker packages into a planning cycle in speech translates to gestural planning units. In the synchronisation chapter (Chapter 4), the proposed Attraction Point Hypothesis combines the gesture onset and gesture duration data in order to answer questions about how gesture and speech interact during production. Taken both studies together, they are an important step towards a better understanding of the coordination of gesture and speech during production and during the gesture’s execution phase within a sentential context. The findings from both speech-gesture production studies support the idea that gesture and speech are generated interactively and that this feedback channel is open from early production stages (i.e., from pre-linguistic stages) until the gesture has been produced.

7.2.4. Discourse and Gesture Processing

Previous studies highlighted that gesture ambiguity and gesture synchronisation play an important role in how gestures are processed within a speech context (e.g., Habets et al., 2011; Obermeier et al., 2010). The results from Chapter 6 (behavioural and ERP) suggest that discourse is a further factor that influences gesture processing. More specifically, the studies

presented showed that discourse information preceding the gesture constrains gesture interpretation and thus also impacts gesture processing when gesture-speech synchrony is not present. Thus, the study demonstrated that gesture processing is discourse-dependent (similar to discourse processing in speech (Van Berkum, 2008)).

The most important contribution of the ERP study presented in Chapter 6 is the distinction between three different gesture integration processes within a discourse context. The integration processes are reflected by different ERP components: Nref, N400, P600. More specifically these three components were linked to the following processes: search for a referent in preceding discourse (Nref), context driven meaning construction/semantic lexical retrieval (N400), post-semantic integration into a discourse model (P600). The interpretation of the N400 and P600 was based on the Retrieval Integration Account (Brouwer et al., 2012). The Nref has been observed only in a speech setting prior to this study (written and spoken) (Van Berkum et al., 2007).

Moreover, the literature review of previous ERP studies on gesture-speech processing suggests, that the distinction between these three ERP components could also be used to explain findings on gesture integration of the following studies: Cornejo et al. (2009); Habets et al. (2011); Holle and Gunter (2007); Ibáñez et al. (2011); Kelly et al. (2004); Obermeier and Gunter (2014); Obermeier et al. (2010); Özyürek et al. (2007).

7.2.5. Gesture-speech Synchrony

In terms of gesture-speech synchrony, the ERP study in Chapter 6 showed that integration processes of gestures are more flexible than often argued (cf. Habets et al., 2011; McNeill, 1992). More specifically, the results from the ERP study demonstrated that gestures could be integrated regardless of the discourse information provided and despite the gesture not synchronising with its semantic affiliate. However, the integration processes of asynchronous

gestures seem to differ from gestures that do synchronise with their semantic affiliates, i.e., evident in the absence of an N400 gesture mismatch effect time-locked to the semantic affiliate.

These findings of the role of synchrony on gesture integration further suggest that asynchronous gestures found in previous production studies (e.g., chapter 4; Morrel-Samuels & Krauss, 1992; Schegloff, 1984), would have had a communicative value. In other words, if a gesture precedes its semantic affiliate, the listener is still extracting meaning from the gesture and also integrates the gesture's meaning into the discourse model. Thus, the ERP study in Chapter 6 suggests that the gestures from the gesture-speech production experiments (i.e., Chapter 3 and 4) that did not synchronise (i.e., overlap) with their semantic affiliates would have still become part of a listener's discourse model.

7.3. Open Questions & Future Directions

The findings on gesture-speech synchronisation and the findings from the comprehension part of this thesis open new research questions for future studies. Based on the findings of this thesis, two strands of possible research directions were identified: (1) further investigations of the Attraction Point Hypothesis, (2) the addressing of further questions about the influence of discourse information on gesture-speech processing/interpretation.

7.3.1. Attraction Point Hypothesis

The Attraction Point Hypothesis proposed in chapter 4 suggests the surface occurrence of semantic affiliates influences gesture onset and gesture duration because the gesture's semantic affiliates function as attractors which pull the gesture. Whether this hypothesis also holds for gesture-speech synchronisation in other languages and/or testing domains, needs to be investigated in future studies. Suggestions for future studies are listed below.

Motion Events

For the synchronisation study presented in Chapter 4, only Path Gestures and Conflated Gestures were included in the analysis but not Manner Gestures. This was due to a low proportion of responses of Manner Only Gestures that did not allow a sound analysis (i.e., 9 %). Although a low proportion of manner gestures is a common finding in studies of motion event gestures (e.g., Choi & Lantolf, 2008; Kita & Özyürek, 2003; Kita et al., 2007; Wessel-Tolvig & Paggio, 2016), it did not allow us to provide a full picture of motion event synchronisation patterns, i.e., path gestures, manner gestures and conflated gestures. Gathering a large enough dataset for an analysis of manner gestures would be very informative regarding the pulling effect of the gesture's semantic affiliates. For manner gestures, the Attraction Point Hypothesis would predict that gesture onset and gesture duration of manner gestures would be exclusively linked to the manner verb because the manner verb is the only semantic affiliate that would function as an attractor (following Talmy, 2000).

Furthermore, future studies could test whether the pulling effect of semantic affiliates also holds for particle verbs describing motion events in narratives. Testing the Attraction Point Hypothesis in a discourse setting would enable the researcher to establish whether synchronisation patterns found in motion event descriptions or narratives (e.g., Chui, 2009; Duncan, 2006; Kellerman & van Hoof, 2003) are discourse driven (e.g., gesture synchronises with the speech element that the speaker puts the focus on) or whether the same pulling effect of the manner verb and path particle found in this thesis applies. This could be done in either German or in Dutch where particle verbs are also separated on the surface structure (e.g., Kellerman & van Hoof, 2003).

Finally, future studies should also test whether an early gesture offset in relation to its semantic affiliate results in a post-stroke hold (or lengthens the post-stroke hold) in order to balance out gesture-speech asynchronies. In particular, it would be informative to compare post-

stroke holds between repetitive (i.e., Conflated) and non-repetitive (i.e., Path) gestures. Based on Kita's (1990) findings that non-repetitive gestures are more often followed by a post-stroke hold compared to repetitive-gestures, we would expect fewer and shorter post-stroke holds for Conflated gestures compared to Path gestures because Conflated gestures are easier to prolong until the gesture's semantic affiliate is encoded.

Testing Domain

The Attraction Point Hypothesis was proposed as a general theory of how gesture and speech are coordinated on a temporal level. Thus, it would be informative to directly test this hypothesis within different testing domains. It would need to be a domain where the gesture's semantic affiliate is separated on the surface structure. These constructions are for example found in German giving of directions (e.g., "*Biegen Sie am Ende der Straße links ab*", literal translation: "*Turn you at the end of the street left to*"; English translation: "Turn left at the end of the street").

Another testing domain suitable for testing the Attraction Point Hypothesis is the domain of placement events which have already been investigated from a co-speech gesture perspective (e.g., Gullberg, 2011). Similar to motion events, placement events, in German and English for example, usually have more than one semantic affiliate in speech (e.g., "put the glass on the table").

Finally, previous studies also suggested that prosodic features (de Ruiter, 1998; Nobe, 1996) influence gesture-speech synchronisation. Thus, this aspect should also be taken into account in future studies that test the Attraction Point Hypothesis.

7.3.2. *Gesture and Discourse*

Based on the outcomes of the comprehension part of this thesis, Future research questions were identified which could be investigated with the stimuli created for the experiments presented in Chapter 6.

First, the inclusion of a no gesture condition in the study design presented in Chapter 6 could be used as a baseline in order to further investigate the integration processes of gestures that do not synchronise with their semantic affiliates. In particular, a baseline condition would be informative in terms of how the semantic affiliate of the gesture is processed without a gesture. This would further help to untangle the effect of the gesture on the processing of the semantic affiliate.

Second, the stimuli set could also be used for further behavioural experiments. For example, it could be tested how ambiguous the gestures used in the stimuli videos are without any speech. This would allow a comparison between gesture interpretation within a discourse context (i.e., related vs. unrelated discourse) and the interpretation of gestures without any speech context. Thus, it could be further investigated how discourse information provided before the occurrence of a gesture influences/guides gesture interpretation.

List of References

- Akhavan, N., Nozari, N., & Göksun, T. (2017). Expression of motion events in Farsi. *Language, Cognition and Neuroscience*, 32(6), 792-804. doi:10.1080/23273798.2016.1276607
- Argyriou, P., Mohr, C., & Kita, S. (2017). Hand Matters: Left-Hand Gestures Enhance Metaphor Explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 874-886. doi:10.1037/xlm0000337
- Bates, D. (2009). Is it right to specify a random slope for the dummy variables. Retrieved from <https://stat.ethz.ch/pipermail/r-sig-mixed-models/2009q1/001736.html>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *2015*, 67(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. *R package version*, 1(7), 1-23.
- Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, 99(3-4), 239-272. doi:10.1515/semi.1994.99.3-4.239
- Beattie, G., & Shovelton, H. (1999). Mapping the Range of Information Contained in the Iconic Hand Gestures that Accompany Spontaneous Speech. *Journal of Language and Social Psychology*, 18(4), 438-462. doi:doi:10.1177/0261927X99018004005
- Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *Br J Psychol*, 93(Pt 2), 179-192.
- Bock, J. K. (1982). Toward a Cognitive-Psychology of Syntax - Information-Processing Contributions to Sentence Formulation. *Psychological Review*, 89(1), 1-47. doi:Doi 10.1037/0033-295x.89.1.1

- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain research reviews*, *59*(1), 55-73.
- Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., . . . Swaab, T. Y. (2015). Sensitivity to Referential Ambiguity in Discourse: The Role of Attention, Working Memory, and Verbal Ability. *Journal of Cognitive Neuroscience*, *27*(12), 2309-2323.
doi:10.1162/jocn_a_00837
- Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539-550. doi:10.1037/0096-3445.136.4.539
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*, *41*, 1318-1352.
doi:10.1111/cogs.12461
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res*, *1446*, 127-143.
doi:10.1016/j.brainres.2012.01.055
- Brouwer, H., & Hoeks, J. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*(758).
doi:10.3389/fnhum.2013.00758
- Brown, A. (2008). Gesture viewpoint in Japanese and English: Cross-linguistic interactions between two languages in one speaker. *Gesture*, *8*(2), 256-276. doi:10.1075/gest.8.2.08bro
- Brown, A. (2015). Universal development and L1-L2 convergence in bilingual construal of manner in speech and gesture in Mandarin, Japanese, and English. *Modern Language Journal*, *99*(S1), 66-82. doi:10.1111/j.1540-4781.2015.12179.x
- Brown, A., & Chen, J. (2013). Construal of manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics*, *24*(4), 605-631. doi:10.1515/cog-2013-0021

- Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition*, 30(02), 225-251. doi:10.1017/S0272263108080327
- Brown, A., & Gullberg, M. (2010). Bidirectional cross-linguistic influence in event conceptualization? Expressions of Path among Japanese learners of English. *Bilingualism: Language and Cognition*, 14(01), 79-94. doi:10.1017/s1366728910000064
- Casey, S., & Emmorey, K. (2008). Co-speech gesture in bimodal bilinguals. *Language and Cognitive Processes*, 24(2), 290-312. doi:10.1080/01690960801916188
- Casey, S., Emmorey, K., & Larrabee, H. (2012). The effects of learning American Sign Language on co-speech gesture(). *Bilingualism (Cambridge, England)*, 15(4), 677-686. doi:10.1017/S1366728911000575
- Cassell, J., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition*, 7(1), 1-34.
- Choi, S., & Lantolf, J. P. (2008). Representation and embodiment of meaning in L2 communication: Motion events in the speech and gesture of advanced L2 Korean and L2 English speakers. *Studies in Second Language Acquisition*, 30(2), 191-224. doi:10.1017/S0272263108080315
- Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, 143(4), 1726-1741. doi:10.1037/a0036281
- Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140(1), 102-116. doi:10.1037/a0021790
- Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, 37(6), 871-887. doi:10.1016/j.pragma.2004.10.016
- Chui, K. (2009). Linguistic and imagistic representations of motion events. *Journal of Pragmatics*, 41(9), 1767-1777. doi:http://dx.doi.org/10.1016/j.pragma.2009.04.006

- Chui, K. (2012). Cross-linguistic comparison of representations of motion in language and gesture. *Gesture*, 12(1), 40-61. doi:10.1075/gest.12.1.03chu
- Church, R. B., Kelly, S. D., & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience*, 29(3), 345-354. doi:10.1080/01690965.2013.857783
- Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition*, 70(1), 42-52. doi:http://dx.doi.org/10.1016/j.bandc.2008.12.005
- Coulson, S. (2007). Electrifying results: ERP data and cognitive linguistics. *Methods in cognitive linguistics*, 18, 400–423.
- Croft, W. A., Barðdal, J., Hollmann, W., Sotirova, V., & Taoka, C. (2010). Revising Talmy's typological classification of complex event constructions In H. C. Boas (Ed.), *Contrastive Studies in Construction Grammar* (pp. 201-236). Amsterdam/Philadelphia: John Benjamins.
- de Ruiter, J. P. (1998). *Gesture and Speech Production*. (PhD), Radboud University, Nijmegen.
- de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 248-311). Cambridge: Cambridge University Press.
- Dipper, L., Pritchard, M., Morgan, G., & Cocks, N. (2015). The language–gesture connection: Evidence from aphasia. *Clinical Linguistics & Phonetics*, 29(8-10), 748-763. doi:10.3109/02699206.2015.1036462
- Duncan, S. (2002). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture*, 2(2), 183-206. doi:10.1075/gest.2.2.04dun
- Duncan, S. (2006). Co-expressivity of speech and gesture: Manner of motion in Spanish, English, and Chinese. In *Proceedings of the 27th Berkeley Linguistic Society Annual Meeting* (pp. 353-370). Berkeley, CA: : Berkeley University Press.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505.

- 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:[http://dx.doi.org/10.1016/0749-596X\(91\)90004-4](http://dx.doi.org/10.1016/0749-596X(91)90004-4)
- Feyereisen, P. (1997). The Competition between Gesture and Speech Production in Dual-Task Paradigms. *Journal of Memory and Language*, 36(1), 13-33.
doi:<http://dx.doi.org/10.1006/jmla.1995.2458>
- Feyereisen, P., Vandewiele, M., & Dubois, F. (1988). The Meaning of Gestures - What Can Be Understood Without Speech. *Cahiers de Psychologie Cognitive*, 8(1), 3-25.
- Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, 6(1), 35-53. doi:[http://dx.doi.org/10.1016/0010-0277\(78\)90008-2](http://dx.doi.org/10.1016/0010-0277(78)90008-2)
- Freleng, F. (Writer). (1950). Canary Row. Film, animated cartoon. In. New York: Time Warner.
- Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *J Cogn Neurosci*, 8(2), 89-106.
doi:10.1162/jocn.1996.8.2.89
- Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Belknap of Harvard UP.
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257-283. doi:10.1146/annurev-psych-113011-143802
- Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing Knowledge Through Gesture: Using Children's Hands to Read Their Minds. *Cognition and Instruction*, 9(3), 201-219.
doi:10.1207/s1532690xci0903_2
- Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology*, 10(2), 96-106. doi:10.1080/17470215808416261
- Goldman-Eisler, F. (1972). Pauses, Clauses, Sentences. *Language and Speech*, 15(2), 103-113.
doi:doi:10.1177/002383097201500201

- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95-112. doi:10.1007/bf02289823
- Gullberg, M. (2010). Methodological Reflections on Gesture Analysis in Second Language Acquisition and Bilingualism Research. *Second Language Research*, 26(1), 75-102.
- Gullberg, M. (2011). Language-specific encoding of placement events in gesture. In J. Bohnemeyer & E. Pederson (Eds.), *Event Representation in Language and Cognition* (pp. 166-188). Cambridge: Cambridge University Press.
- Gullberg, M., & Narasimhan, B. (2010). What gestures reveal about how semantic distinctions develop in Dutch children's placement verbs. *Cognitive Linguistics*, 21(2), 239-262. doi:10.1515/COGL.2010.009
- Gunter, T. C., & Weinbrenner, J. E. D. (2017). When to Take a Gesture Seriously: On How We Use and Prioritize Communicative Cues. *J Cogn Neurosci*, 1-12. doi:10.1162/jocn_a_01125
- Gunter, T. C., Weinbrenner, J. E. D., & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology*, 6(80). doi:10.3389/fpsyg.2015.00080
- Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neurosciences*, 23, 1845-1854.
- Hadar, U., & Pinchas-Zamir, L. (2004). The Semantic Specificity of Gesture: Implications for Gesture Classification and Function. *Journal of Language and Social Psychology*, 23(2), 204-214. doi:10.1177/0261927x04263825
- Hagoort, P. (2003). Interplay between Syntax and Semantics during Sentence Comprehension: ERP Effects of Combining Syntactic and Semantic Violations. *Journal of Cognitive Neuroscience*, 15(6), 883-899. doi:10.1162/089892903322370807
- Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos Trans R Soc Lond B Biol Sci*, 363(1493), 1055-1069. doi:10.1098/rstb.2007.2159

- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439-483.
- Hagoort, P., & Van Berkum, J. J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 801-811. doi:10.1098/rstb.2007.2089
- Hauk, O., Davis, M. H., Ford, M., Pulvermuller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, 30(4), 1383-1400. doi:10.1016/j.neuroimage.2005.11.048
- Hauk, O., & Pulvermuller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clin Neurophysiol*, 115(5), 1090-1103. doi:10.1016/j.clinph.2003.12.020
- Hoeks, J. C., Brouwer, H., & Holtgraves, T. (2014). Electrophysiological research on conversation and discourse. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 365-386). Oxford: Oxford University Press.
- Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Brain Res Cogn Brain Res*, 19(1), 59-73. doi:10.1016/j.cogbrainres.2003.10.022
- Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19, 1175-1192.
- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol Bull*, 137(2), 297-315. doi:10.1037/a0022128
- Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., & Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters*, 471(1), 48-52. doi:https://doi.org/10.1016/j.neulet.2010.01.009
- Ibáñez, A., Toro, P., Cornejo, C., Urquina, H., Manes, F., Weisbrod, M., & Schroder, J. (2011). High contextual sensitivity of metaphorical expressions and gesture blending: A video event-related potential design. *Psychiatry Res*, 191(1), 68-75. doi:10.1016/j.psychresns.2010.08.008

- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446.
doi:<https://doi.org/10.1016/j.jml.2007.11.007>
- Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591.
- Kaan, E., & Swaab, T. Y. (2003). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, *17*(3), 621-635.
- Kellerman, E., & van Hoof, A.-M. (2003). Manual Accents. *IRAL*, *41*(3), 251-269.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*(1), 253-260. doi:[http://dx.doi.org/10.1016/S0093-934X\(03\)00335-3](http://dx.doi.org/10.1016/S0093-934X(03)00335-3)
- Kelly, S. D., Manning, S. M., & Rodak, S. (2008). Gesture Gives a Hand to Language and Learning: Perspectives from Cognitive Neuroscience, Developmental Psychology and Education. *Language and Linguistics Compass*, *2*(4), 569-588. doi:[10.1111/j.1749-818X.2008.00067.x](https://doi.org/10.1111/j.1749-818X.2008.00067.x)
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological Science*, *21*(2), 260-267.
doi:[10.1177/0956797609357327](https://doi.org/10.1177/0956797609357327)
- Kendon, A. (1980). Gesture and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *Nonverbal Communication and Language* (pp. 207-227). The Hague: Mouton.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Key, A. P., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: an ERP primer. *Dev Neuropsychol*, *27*(2), 183-215. doi:[10.1207/s15326942dn2702_1](https://doi.org/10.1207/s15326942dn2702_1)
- Kircher, T. T. J., Brammer, M. J., Levelt, W. J. M., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: engagement of left temporal cortex during pauses in speech. *Neuroimage*, *21*(1), 84-90. doi:<http://dx.doi.org/10.1016/j.neuroimage.2003.09.041>

- Kita, S. (1990). *The Temporal Relationship between Gesture and Speech: A Study of Japanese-English Bilinguals*. (MA Thesis), University of Chicago,
- Kita, S. (1993). *Language and thought interface: a study of spontaneous gestures and Japanese mimetics*. (PhD), University of Chicago, Chicago.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge: Cambridge University Press.
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, 24(2), 145-167. doi:10.1080/01690960802586188
- Kita, S. (2014). Production of Speech-Accompanying Gesture. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *Oxford Handbook of Language Production* (pp. 451-459). USA: Oxford University Press.
- Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol Rev*, 124(3), 245-266.
doi:10.1037/rev0000059
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16-32.
doi:http://dx.doi.org/10.1016/S0749-596X(02)00505-3
- Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, 22(8), 1212-1236.
doi:10.1080/01690960701461426
- Kita, S., van Gijn, I., & van der Hulst, H. (1997). *Movement Phases in signs and co-speech gestures, and their transcription by human coders*. Paper presented at the International Gesture Workshop Bielefeld, Germany, Bielefeld.

- Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1-36.
- Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal Behavior and Nonverbal Communication: What do Conversational Hand Gestures Tell Us? *Advances in Experimental Social Psychology*, *28*, 389-450. doi:[http://dx.doi.org/10.1016/S0065-2601\(08\)60241-5](http://dx.doi.org/10.1016/S0065-2601(08)60241-5)
- Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The Communicative Value of Conversational Hand Gesture. *Journal of Experimental Social Psychology*, *31*(6), 533-552. doi:<http://dx.doi.org/10.1006/jesp.1995.1024>
- Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, *61*(5), 743-754. doi:<http://dx.doi.org/10.1037/0022-3514.61.5.743>
- Kuperberg, G. R. (2008). Electroencephalography, event-related potentials, and magnetoencephalography. *Essentials of neuroimaging for clinical practice*, 117-127.
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain Lang*, *100*(3), 223-237. doi:10.1016/j.bandl.2005.12.006
- Kuperberg, G. R., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117-129.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463-470.
- Kutas, M., & Federmeier, K. D. (2007). Event-related brain potential (ERP) studies of sentence processing. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 385-406). Oxford: Oxford University Press.

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647. doi:10.1146/annurev.psych.093008.131123
- Kutas, M., Federmeier, K. D., Staab, J., & Kluender, R. (2007). Language. In G. Berntson, J. T. Cacioppo, & L. G. Tassinary (Eds.), *Handbook of Psychophysiology* (3 ed., pp. 555-580). Cambridge: Cambridge University Press.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, *2*(0).
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933.
- Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, *41*(3), 841-849. doi:10.3758/brm.41.3.841
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104.
- Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*(2), 133-164. doi:http://doi.org/10.1016/0749-596X(85)90021-X
- Luck, S. J. (2005a). *An introduction to the event-related potential technique*. Cambridge, MA: MIT press.
- Luck, S. J. (2005b). Ten simple rules for designing ERP experiments In *Event-related potentials: A methods handbook* (pp. 17-32). Cambridge, MA: The MIT Press.
- Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*(1467). doi:10.3389/fpsyg.2014.01467
- Macedonia, M., & Knösche, T. R. (2011). Body in Mind: How Gestures Empower Foreign Language Learning. *Mind, Brain, and Education*, *5*(4), 196-211. doi:10.1111/j.1751-228X.2011.01129.x

- Mayer, M. (1969). *Frog, Where Are You?*. NY: Dial Press.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92, 350–371.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: The University of Chicago Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: The University of Chicago Press.
- McNeill, D. (2009). Imagery for Speaking. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & Ö. S. (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 517-530). London: Taylor & Francis.
- McNeill, D. (2015). Gesture in Linguistics In *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (Vol. 10, pp. 109-120). Oxford: Elsevier.
- McNeill, D., & Duncan, S. (2000). Growth Points in Thinking-for-Speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 141-161). Cambridge: Cambridge University Press.
- Mol, L., & Kita, S. (2012). *Gesture structure affects syntactic structure in speech*. Paper presented at the 34th Annual Conference of the Cognitive Science Society, Austin, TX.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 615-622. doi:10.1037/0278-7393.18.3.615
- Müller, C. (2002). Eine kleine Kulturgeschichte der Gestenbetrachtung. *Psychotherapie und Sozialforschung*, 4(1), 3-29.
- Neguera, E., Lantolf, J. P., Jordan, S. R., & Gelabert, J. (2004). The “private function” of gesture in second language speaking activity: a study of motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics*, 14(1), 113-147. doi:10.1111/j.1473-4192.2004.00056.x
- Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *J Cogn Neurosci*, 19(2), 228-236. doi:10.1162/jocn.2007.19.2.228

- Nieuwland, M. S., & Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Brain Res Cogn Brain Res*, *24*(3), 691-701. doi:10.1016/j.cogbrainres.2005.04.003
- Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155-167. doi:http://dx.doi.org/10.1016/j.brainres.2006.08.022
- Nieuwland, M. S., & Van Berkum, J. J. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, *106*(2), 119-131. doi:http://dx.doi.org/10.1016/j.bandl.2008.05.001
- Nobe, S. (1996). *Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/threshold model of gesture production*. (unpublished PhD dissertation), University of Chicago, Chicago.
- Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and Gesture* (pp. 186-198). Cambridge Cambridge University Press.
- Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From Action to Abstraction: Using the Hands to Learn Math. *Psychological Science*, *25*(4), 903-910. doi:10.1177/0956797613518351
- Nunez, M. D., Nunez, P. L., & Srinivasan, R. (2016). Electroencephalography (EEG): Neurophysics, Experimental Methods, and Signal Processing. *Handbook of Statistical Methods for Brain Signals and Images*, 175-197.
- Obermeier, C., Dolk, T., & Gunter, T. C. (2012). The benefit of gestures during communication: evidence from hearing and hearing-impaired individuals. *Cortex*, *48*(7), 857-870.
- Obermeier, C., & Gunter, T. C. (2014). Multisensory Integration: The Case of a Time Window of Gesture–Speech Integration. *Journal of Cognitive Neuroscience*, *27*(2), 292-307. doi:10.1162/jocn_a_00688

- Obermeier, C., Holle, H., & Gunter, T. C. (2010). What Iconic Gesture Fragments Reveal about Gesture–Speech Integration: When Synchrony Is Lost, Memory Can Help. *Journal of Cognitive Neuroscience*, *23*(7), 1648-1663. doi:10.1162/jocn.2010.21498
- Obermeier, C., Kelly, S. D., & Gunter, T. C. (2015). A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *Soc Cogn Affect Neurosci*, *10*(9), 1236-1243. doi:10.1093/scan/nsv011
- Osterhout, L. (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain Lang*, *59*(3), 494-522. doi:10.1006/brln.1997.1793
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806.
- Otten, M., & Van Berkum, J. J. (2008). Discourse-Based Word Anticipation During Language Processing: Prediction or Priming? *Discourse Processes*, *45*(6), 464-496. doi:10.1080/01638530802356463
- Özçalışkan, S. (2016). Do gestures follow speech in bilinguals' description of motion? *Bilingualism*, *19*(3), 644-653. doi:10.1017/S1366728915000796
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition*, *148*, 10-18. doi:http://dx.doi.org/10.1016/j.cognition.2015.12.001
- Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philos Trans R Soc Lond B Biol Sci*, *369*(1651), 20130296. doi:10.1098/rstb.2013.0296
- Özyürek, A., Kita, S., & Allen, S. (2001). Tomato Man movies: Stimulus kit designed to elicit manner, path and causal constructions in motion events with regard to speech and gestures. Nijmegen, the Netherlands: Max Planck Institute for Psycholinguistics, Language and Cognition group.
- Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology*, *44*(4), 1040-1054. doi:10.1037/0012-1649.44.4.1040

- Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture*, 5(1-2), 219-240.
- Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line Integration of Semantic Information from Speech and Gesture: Insights from Event-related Brain Potentials. *Journal of Cognitive Neuroscience*, 19(4), 605-616. doi:10.1162/jocn.2007.19.4.605
- Pawley, A. (1987). Encoding events in Kalam and English: different logics for reporting experience. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 329-360). Amsterdam: Benjamins.
- Pawley, A. (2010). Event representation in serial verb constructions. In J. Bohnemeyer & E. Pederson (Eds.), *Event Representation in Language and Cognition*: (pp. 13-42). Cambridge: Cambridge University Press.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, 191, 191-225.
- Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, 459(3), 142-146.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413-425. doi:<https://doi.org/10.1016/j.jml.2008.02.002>
- R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Schegloff, E. A. (1984). On some gesture's relation to talk. In M. A. J. Heritage (Ed.), *In Structures of Social Action: Studies in Conversation Analysis* (pp. 266-296). Cambridge: Cambridge University Press.
- Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, 20(11), 2037-2057.

- Slobin, D. I. (1987). Thinking for Speaking. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 435-445.
- Slobin, D. I. (2000). A Dynamic Approach To Linguistic Relativity And Determinism. In S. Niermeier & R. Dirven (Eds.), *Evidence for Linguistic Relativity* (pp. 107-138). Amsterdam/Philadelphia: John Benjamin Publishing Company.
- Slobin, D. I. (2003). Language and Thought Online: Cognitive Consequences of Linguistics Relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language In Mind* (pp. 157-192). Massachusetts: The MIT Press.
- Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59-81). Amsterdam/Philadelphia: John Benjamins.
- Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73(3), 205-246. doi:[http://dx.doi.org/10.1016/S0010-0277\(99\)00053-0](http://dx.doi.org/10.1016/S0010-0277(99)00053-0)
- Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *IRAL*, 44(2), 145-171.
- Stam, G. (2015). Changes in thinking for speaking: A longitudinal case study. *Modern Language Journal*, 99(S1), 83-99. doi:10.1111/j.1540-4781.2015.12180.x
- Supalla, T., Newport, E., Singleton, J., Supalla, S., Metlay, D., & Coulter, G. (n.d.). *The test battery for American Sign Language morphology and syntax*. University of Rochester. New York.
- Talmy, L. (2000). *Toward a Cognitive Semantics, Vol II: Typology and Process in Concept Structuring*. Cambridge: Massachusetts Instit Technology Press.
- Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3), 382-392. doi:<http://dx.doi.org/10.1016/j.ijpsycho.2011.12.007>
- Van Berkum, J. J. (2008). Understanding Sentences in Context. *Current Directions in Psychological Science*, 17(6), 376-380. doi:10.1111/j.1467-8721.2008.00609.x

- Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276-316): Palgrave Macmillan.
- Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158-171. doi:<http://dx.doi.org/10.1016/j.brainres.2006.06.091>
- Van Berkum, J. J., Zwitserlood, P., Bastiaansen, M., Brown, C. M., & Hagoort, P. (2004). *So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension*. Paper presented at the Annual meeting of the Cognitive Neuroscience Society (CNS-2004), San Francisco.
- van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Brain Res Cogn Brain Res, 17*(3), 701-718.
- van den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 364-372.
- van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brainpotentials. *Memory & Cognition, 18*(4), 380-393.
- van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.
- Vigliocco, G., & Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes, 21*(7-8), 790-816. doi:10.1080/016909600824070
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209-232. doi:<http://dx.doi.org/10.1016/j.specom.2013.09.008>
- 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. *J Exp Psychol Learn Mem Cogn, 36*(2), 423-440. doi:10.1037/a0018619

- Ward, J. (2015). *The student's guide to cognitive neuroscience*: Psychology Press.
- Wessel-Tolvig, B., & Paggio, P. (2016). Revisiting the thinking-for-speaking hypothesis: Speech and gesture representation of motion in Danish and Italian. *Journal of Pragmatics*, *99*, 39-61.
doi:10.1016/j.pragma.2016.05.004
- Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, (J. B. Carroll Ed.). Cambridge, MA: MIT Press
- Willems, R. M., van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat Rev Neurosci*, *15*(3), 193-201.
doi:10.1038/nrn3679
- Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41-62). Chichester: John Wiley.

APPENDIX 1

Particle Verbs from Experiment 1

German and English particle verbs given to describe the motion events depicted in the cartoon clips in Experiment 1. The particle verb “bore through” was given to describe two different video stimuli.

English	German
<i>Example Item</i> climb up	hinaufklettern
<i>Practice Item</i> ride around	herumfahren
<i>Items</i> bore through climb up crawl out dance around drill down float down jump around jump into jump over roll into slide down spin up	durchbohren hinaufklettern hervorkriechen herumtanzen hineindrehen hinunterschweben herumhüpfen hineinspringen drüberspringen hineinrollen hinunterrutschen hinaufdrehen

Particle Verbs from Experiment 2

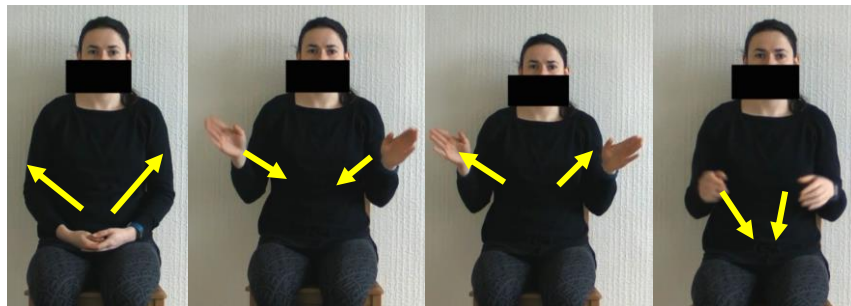
German particle verbs given to describe the motion events depicted in the cartoon clips in Experiment 2. The particle verbs “bore through” and “jump around” were both given to describe two different video stimuli.

German	English Translation
<i>Example Item</i> hinunterrutschen	slide down
<i>Practice trial 1</i> herumfahren	ride around
<i>Practice Item 2</i> hinaufklettern	climb up
<i>Items</i> durchbohren herumhüpfen herumtanzen hinaufdrehen hinaufhüpfen hinaufklettern hinaufrollen hinaufsteigen hineindrehen hineinrollen hinunterhüpfen hinunterrollen hinunterschweben	bore through jump around dance around spin up jump up climb up roll up walk up drill down roll into jump down roll down float down

APPENDIX 3

Example Stimuli Gesture-speech Comprehension Study

Example Stimulus 1: Gestures depicting the verbs “flying” and “swimming”



flying: preparation stroke stroke retraction to resting position

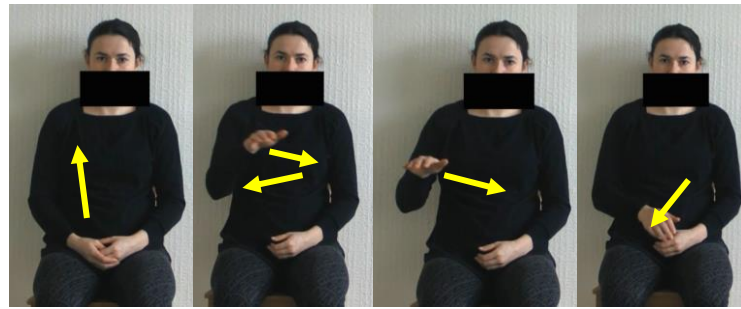


swimming: preparation stroke stroke retraction back to resting position

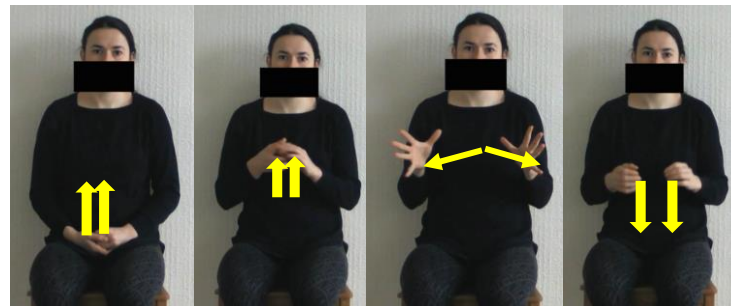
<i>Condition</i>	<i>Introductory Sentence</i>	<i>Target Sentence</i>	<i>Gesture (match/mismatch)</i>
Stimuli Set 1			
Unrelated Discourse Condition	This Thursday I finally took a day off.	I could see from the park bench that something white was flying around.	Match (flying gesture)
	This Thursday I finally took a day off.	I could see from the park bench that something white was flying around.	Mismatch (swimming gesture)
Related Discourse Condition	I went to watch the ducks by the river.	I could see from the park bench that some creatures were flying around.	Match (flying gesture)
	I went to watch the ducks by the river.	I could see from the park bench that some creatures were flying around.	Mismatch (swimming gesture)

Stimuli Set 2			
Unrelated Discourse Condition	This Thursday I finally took a day off.	I could see from the park bench that some creatures were flying around.	Match (flying gesture)
	This Thursday I finally took a day off.	I could see from the park bench that some creatures were flying around.	Mismatch (swimming gesture)
Related Discourse Condition	I went to watch the ducks by the river.	I could see from the park bench that something white was flying around.	Match (flying gesture)
	I went to watch the ducks by the river.	I could see from the park bench that something white was flying around.	Mismatch (swimming gesture)
Stimuli Set 3			
Unrelated Discourse Condition	This Thursday I finally took a day off.	I could see from the park bench that something white was swimming around in the pond.	Match (flying gesture)
	This Thursday I finally took a day off.	I could see from the park bench that something white was swimming around in the pond.	Mismatch (swimming gesture)
Related Discourse Condition	I went to watch the ducks by the river.	I could see from the park bench that some creatures were swimming around in the pond.	Match (flying gesture)
	I went to watch the ducks by the river.	I could see from the park bench that some creatures were swimming around in the pond.	Mismatch (swimming gesture)
Stimuli Set 4			
Unrelated Discourse Condition	This Thursday I finally took a day off.	I could see from the park bench that some creatures were swimming around in the pond.	Match (flying gesture)
	This Thursday I finally took a day off.	I could see from the park bench that some creatures were swimming around in the pond.	Mismatch (swimming gesture)
Related Discourse Condition	I went to watch the ducks by the river.	I could see from the park bench that something white was swimming around in the pond.	Match (flying gesture)
	I went to watch the ducks by the river.	I could see from the park bench that something white was swimming around in the pond.	Mismatch (swimming gesture)

Example Stimulus 2: Gestures depicting the verbs “floating” and “bursting”



floating: preparation stroke stroke retraction to resting position



bursting: preparation preparation stroke retraction to resting position

<i>Condition</i>	<i>Introductory Sentence</i>	<i>Target Sentence</i>	<i>Gesture (match/mismatch)</i>
Stimuli Set 1			
Unrelated Discourse Condition	Tonight we had a great BBQ in our garden.	I could see from our terrace that something colourful was floating towards the sunset	Match (floating gesture)
	Tonight we had a great BBQ in our garden.	I could see from our terrace that something colourful was floating towards the sunset	Mismatch (bursting gesture)
Related Discourse Condition	There were a lot of balloons at my sister’s birthday party today.	I could see from our terrace that something above me was floating towards the sunset.	Match (floating gesture)
	There were a lot of balloons at my sister’s birthday party today.	I could see from our terrace that something above me was floating towards the sunset.	Mismatch (bursting gesture)
Stimuli Set 2			
Unrelated Discourse Condition	Tonight we had a great BBQ in our garden.	I could see from our terrace that something above me was floating towards the sunset.	Match (floating gesture)
	Tonight we had a great BBQ in our garden.	I could see from our terrace that something above me was floating towards the sunset.	Mismatch (bursting gesture)
Related Discourse Condition	There were a lot of balloons at my sister’s birthday party today.	I could see from our terrace that something colourful was floating towards the sunset.	Match (floating gesture)
	There were a lot of balloons at my sister’s birthday party today.	I could see from our terrace that something colourful was floating towards the sunset.	Mismatch (bursting gesture)

Stimuli Set 3			
Unrelated Discourse Condition	Tonight we had a great BBQ in our garden.	I could see from our terrace that something colourful was bursting into pieces.	Match (bursting gesture)
	Tonight we had a great BBQ in our garden.	I could see from our terrace that something colourful was bursting into pieces.	Mismatch (floating gesture)
Related Discourse Condition	There were a lot of balloons at my sister's birthday party today.	I could see from our terrace that something above me was bursting into pieces.	Match (bursting gesture)
	There were a lot of balloons at my sister's birthday party today.	I could see from our terrace that something above me was bursting into pieces.	Mismatch (floating gesture)
Stimuli Set 4			
Unrelated Discourse Condition	Tonight we had a great BBQ in our garden.	I could see from our terrace that something above me was bursting into pieces.	Match (bursting gesture)
	Tonight we had a great BBQ in our garden.	I could see from our terrace that something above me was bursting into pieces.	Mismatch (floating gesture)
Related Discourse Condition	There were a lot of balloons at my sister's birthday party today.	I could see from our terrace that something colourful was bursting into pieces.	Match (bursting gesture)
	There were a lot of balloons at my sister's birthday party today.	I could see from our terrace that something colourful was bursting into pieces.	Mismatch (floating gesture)