Representing discourse referents in speech and gesture

Debreslioska, Sandra

2019

Document Version:
Publisher's PDF, also known as Version of record

Representing discourse referents in speech and gesture

SANDRA DEBRESLIOSKA
CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY

Representing discourse referents in speech and gesture

Sandra Debreslioska

LUND
UNIVERSITY

Cover photo by Carolina Larsson, Stefan Lindgren & Sandra Debreslioska

Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-88899-32-3 (print)
ISBN 978-91-88899-33-0 (digital)

Printed in Sweden by Media-Tryck, Lund University
Lund 2019

MADE IN SWEDEN

For my children

Table of Content

Acknowledgements 6

List of papers 8

1 Introduction 9

2 Background 11

2.1 Discourse reference in speech 11

2.1.1 Richness of expression 11

2.1.2 Nominal definiteness 13

2.1.3 Clause structure and grammatical role 15

2.1.4 Dimensions of information status/accessibility 17

2.1.5 Summary 18

2.2 Discourse reference in gesture 19

2.2.1 What are gestures? 19

2.2.2 Ways of classifying gestures 21

2.2.3 Gestures on the discourse level 27

2.3 The studies in this thesis 35

2.3.1 *When*? 35

2.3.2 *Where*? 35

2.3.3 *How*? 36

2.3.4 *What*? 36

3 Methods 37

3.1 Participants 37

3.1.1 Production studies 37

3.1.2 Perception study 37

3.2 Design 38

3.2.1 Production studies 38

3.2.2 Perception study 38

3.3 Stimulus materials 40

3.3.1 Production studies 40

3.3.2 Perception study 41

3.4 Procedures and tasks 43
3.4.1 Production studies 43
3.4.2 Perception study 44
3.5 Data treatment 45
3.5.1 Speech as a starting point for the examination of gesture 45
3.5.2 Annotation of speech and gestures in ELAN 46
3.5.3 Speech-gesture alignment 48
3.5.4 Further coding and reliability 49
4 Results 51
4.1 Paper I 51
4.2 Paper II 52
4.3 Paper III 52
4.4 Paper IV 54
5 Discussion 55
5.1 *When*? 55
5.2 *Where*? 58
5.3 *How*? 61
5.4 *What*? 63
6 Conclusion and future work 65
6.1 Some conclusions on the functions of gestures in discourse 65
6.2 Future work 69
References 75
Appendices 85
Appendix A: Story script production studies 85
Appendix B: Instructions production studies 89
Appendix C: Instructions perception study 91
Appendix D: Consent form 94
Papers I-IV 97

Acknowledgements

This dissertation would not have been possible without the guidance, support and encouragement from many people.

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Marianne Gullberg, for her continuous advice on doing research, writing papers and pursuing a career in academia. Her immense knowledge, scientific values and dedication have been a true inspiration, and I am more than thankful to have had her guiding me through this process.

I would also like to specially thank Joost van de Weijer for his collaboration on paper II and for his constant support with figuring out the statistics. My sincere thanks further go to Maria Graziano for providing very useful feedback on the final drafts of this dissertation, and Emanuela Campisi for her insightful comments and intriguing questions during my final seminar. I am also genuinely thankful to Mats Andrén, Elisabeth Engberg-Pedersen and Carita Paradis for accepting to be on my committee, and I feel very honored that Henriëtte Hendriks agreed to act as Faculty opponent.

The work in this dissertation has also benefited from the practical and technical assistance of several people. I thank Debora Strömberg for drawing the picture story used for the creation of the corpus, Nicole Weidinger for starring in the stimulus videos used for the perception experiments, Katrin Lindner and Judith Diamond for hosting my data collections in Germany, and Sabine Gosselke-Berthelsen, Wanda Jakobsen, Irene Lami and Nicolas Femia for their help in reliability coding. I am also very thankful to Carolina Larsson and Stefan Lindgren for helping me create a nice cover for the book.

My time as a PhD candidate in Lund was also particularly enjoyable because I was surrounded by many wonderful colleagues. I would like to thank Sabine Gosselke-Berthelsen and Frida Blomberg for their companionship and encouragement during the years, but in particular for their support during the final stages of my PhD journey. Sabine deserves special thanks for offering very useful comments on an earlier draft of this thesis. I extend my thanks and affection also to Susan Sayehli, Victoria Johansson and Annika Andersson for their many pep talks and invaluable advice on academic work and mom life. And of course, I also thank all the other PhD students and colleagues in general linguistics, phonetics and semiotics for being part of a stimulating work environment.

Since my journey leading to this PhD dissertation started before my time in Lund, I would also like to take this opportunity to acknowledge the positive influence from some of the special people that I have met during my time at Radboud University Nijmegen, the Max Planck Institute for Psycholinguistics and the University of Birmingham. I would like to thank Mandana Seyfeddinipur for supporting and

encouraging my decision to start an internship with Marianne Gullberg during my Master's studies. The internship was the starting point for my journey into the gesture world and academia. I am also deeply grateful to Asli Özyürek and Sotaro Kita for taking me on as their student, letting me be part of their labs and for their constant encouragement to go on. I further thank some of my fellow students and colleagues from the time, and in particular Giovanni Rossi, Anne-Therese Frederiksen, Reyhan Furman, Beyza Sumer and Katherine Mumford. The exchange of ideas with them have inspired me immensely.

I have also been lucky to meet Adam Kendon on multiple occasions during the past years. My work has greatly profited from his inspiring lectures and seminars, but also from the many stimulating discussions with him at different times.

Finally, my family and friends from home have been an indispensable source of support and encouragement throughout. I especially thank my parents for having always supported my many moves from country to country in order to follow my interests, as well as their never-ending love and trust along the way. My friends Judith and Desi have provided emotional support and positive thoughts whenever I needed them. I feel very lucky to have had them by my side for almost 25 years now.

Last but not least, I would like to thank the two most special people in my life: my partner Christian and my first child Valentin. Words cannot describe how happy I am to have met them during my time as a PhD candidate. Thank you for being my biggest fans. You are my world, and I cannot wait to start the next phase of my (and our) life with you!

List of papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals. The papers are appended at the end of the thesis.

I. *Gestures signal the difference between brand-new and inferable referents in discourse*

Debreslioska, S. & Gullberg, M. (submitted)

II. *Addressees are sensitive to the presence of gestures when tracking a single referent in discourse*

Debreslioska, S., van de Weijer, J. & Gullberg, M. (submitted)

III. *Discourse reference is bimodal: How information status in speech interacts with presence and viewpoint of gestures*

Debreslioska, S & Gullberg, M. (2019). *Discourse Processes, 56*, 41-60, DOI: 10.1080/0163853X.2017.1351909

(published online, 2017, August 24)

IV. *The semantic content of gestures varies with information status, definiteness and clause structure*

Debreslioska, S. & Gullberg, M. (submitted)

1 Introduction

The thesis examines the ways that speech and gestures are used to represent referents in connected discourse. Gestures are considered to be part of language and to form a tightly integrated system together with speech. Thus, when engaging in talk, speakers use a combination of speech and gestures to get their messages across. But while speech is *mostly* obligatory in order to communicate information to an addressee, gestures are not. Rather, during a certain stretch of discourse, there are moments in which gestures are produced and others when they are not. For instance, in the context of narrative discourse, if speakers want to introduce a new entity into the story, they will necessarily have to mention the entity in speech by using a referential expression denoting it[1]. If they do not, the addressee will have no representation of the entity in question. When it comes to gestures on the other hand, this obligatoriness does not apply in the same way. Speakers have the possibility to but do not necessarily always accompany each mention of a discourse entity with a gesture.

Furthermore, languages offer speakers different options for how to refer to discourse referents depending on the informational conditions in which they are mentioned. One of the central factors influencing these options is the accessibility of information in the preceding discourse. Previous research has shown that, depending on a referent's accessibility, speakers can vary the form of referential expressions, the clausal structures they are embedded in, and the grammatical roles they are instantiated in. For instance, speakers can choose between richer or leaner referential expressions to refer to an entity ('the bird' vs. 'it'), or between indefinite and definite expressions ('a bird' vs. 'the bird'). In addition, speakers can choose a clausal structure focusing on the existence of an entity or a structure that involves the referent in an event (e.g., 'there was a bird' vs. 'a/the bird came flying into the house'). Finally, speakers can vary the instantiation of entities as grammatical subjects or objects (e.g., 'she' vs. 'a bird' in 'she took a bird out of the cage').

Importantly, gestures too can vary along different dimensions for the representation of discourse referents. They vary in terms of *when* they are produced, *where* they are produced, *how* they are produced, and in terms of *what* information they express. For

[1] It is worth considering that in some pro drop languages, it might, under specific circumstances be possible to drop arguments even if they are new. This is especially the case for children (e.g., Allen, 2008).

instance, gestures can be used to represent referents at certain moments in the discourse, but not at others. Gestures can also be produced in specific locations in gesture space which can function as visual anaphora when they are reused by the speaker during the duration of the discourse. Furthermore, gestures can represent an entity from a character perspective, as when a speaker enacts a flapping motion of a bird by mapping the bird's wings onto her arms. Or they can represent an entity from an observer perspective, such as when a speaker draws a path through gesture space in order to represent the motion of a bird flying away, and thus looks onto the scene like an outside observer. Finally, gestures can provide information about the size, shape or location of an entity (e.g., a small, round bird sitting on the window sill). Whereas at other times gestures will represent actions or movements of an entity (e.g., a bird flapping its wings).

The studies in the current thesis examine the role that speech-associated gestures play in the production and perception of connected discourse by focusing on the representation of discourse referents. More specifically, the studies set out to examine how the variation in *when* gestures are produced, *where* they are produced, *how* they are produced, and *what* they express, patterns with variations in speech for the representation of discourse referents.

2 Background

2.1 Discourse reference in speech

Much of the linguistic work on discourse reference has shown that the way that speakers refer to discourse referents strongly relies on assumptions about the referents' accessibility or information status, that is, the process by which people focus their attention more on some discourse entities than on others (e.g., Ariel, 1988, 1991, 1996; Arnold, 1998, 2008, 2010; Chafe, 1994; Givón, 1983; Gundel, Hedberg & Zacharski, 1993; Prince, 1992). Speakers need to make assumptions about what their addressees know or are attending to at each point in the discourse and package the way they refer to discourse referents accordingly. This variation in the structuring of information can affect the form of a referential expression itself (on a 'local' level) and/or the packaging of the utterance that a referential expression is embedded in (on a 'global' level).

Reference to new or less accessible referents typically patterns differently than reference to given or more accessible referents on a range of different dimensions. These dimensions differ from language to language. In the current thesis, I focus on describing and analyzing German patterns, and thus I predominantly rely on previous research, which has considered discourse patterns in Western European languages (e.g., Chafe, 1987, 1994; Givón, 1983; Gullberg, 1998, 2003, 2006; Hickmann, Hendriks, Roland & Liang, 1996; Lambrecht, 1994). Accordingly, I will also provide German examples whenever it is appropriate throughout the thesis. The variations for discourse reference that are of particular interest in this thesis concern richness of expression and nominal definiteness on the word level, as well as the clause structure a referent is embedded in, and its grammatical role on the utterance level. Oftentimes, these different dimensions co-vary, but for reasons of clarity, I will discuss them separately.

2.1.1 Richness of expression

Richness of expression, as it is understood in this thesis, refers to the size of a referential expression which speakers vary with referent accessibility. Richness of expression has also been referred to as heaviness, weight/length or phonological size (e.g., Arnold, Losongco, Wasow & Ginstrom, 2000; Givón, 1983; Skopeteas, 2012). One typical pattern can be described as follows: When a discourse referent has not previously been

mentioned in the discourse, and therefore represents new information, or when it is not currently in the focus of attention of the addressee, and thus represents less accessible information, the speaker will typically use a richer, or more explicit, referential expression to refer to it. For instance, in (1), the referents *ein Mann* 'a man' (1a), *eine Kiste* 'a box' (1b), *ein Seil* 'a rope' (1c) and *ein anderer Mann* 'another man' (1e) are all mentioned for the first time in this piece of discourse and are all expressed by full lexical noun phrases (NPs). When a discourse referent has recently been mentioned, the speaker might assume it to be in the focus of attention, and they can then refer to it with leaner or reduced referential expressions, such as pronouns and zero anaphora (e.g., *der* 'he' and '∅' in 1b-d for the referent 'man'). When a referent is mentioned after a gap of absence, the speaker might assume that the referent is less accessible and can thus switch back to a richer, more explicit referential expression (e.g., *die Kiste* 'the box' in 1d after a gap of absence of one clause).

(1)

a *da ist ein Mann*$_1$

b *der*$_1$ *öffnet eine Kiste*$_2$

c *∅*$_1$ *holt ein Seil*$_3$ *heraus*

d *und ∅*$_1$ *schließt die Kiste*$_2$ *wieder*

e *dann kommt ein anderer Mann*$_4$ *die Treppe runter*

'**a** there is a man$_1$

b he$_1$ opens a box$_2$

c ∅$_1$ takes out a rope$_3$

d and ∅$_1$ closes the box$_2$ again

e then another man$_4$ comes down the stairs'

Referential expressions differing in richness can be ordered along a scale representing the degree of accessibility of referents (from low to high; e.g., Givón, 1983), as illustrated in (2).

(2) *lexical NP < pronoun < zero*

2.1.2 Nominal definiteness

Another variation of form on the word level, related to referent accessibility and information status, is nominal definiteness. Speakers of languages that encode definiteness tend to choose indefinite lexical NPs for first mentioned referents, which are assumed to be new to the addressee (e.g., the referent *ein Mann* 'a man' in 1a), and definite lexical NPs for already-mentioned referents, which are given but less accessible (e.g., the referent *die Kiste* 'the box' in 1d). Hence, indefinite lexical NPs typically refer to entities that have no explicit antecedent in the discourse context, whereas definite lexical NPs refer to entities that have an explicit antecedent (e.g., the referent *eine Kiste* 'a box' is the direct antecedent for the referent *die Kiste* 'the box' in 1).

An exception to this pattern are 'inferable' referents (Prince, 1981, 1992). Inferable referents do not have an explicit antecedent in the previous discourse but are nevertheless often represented with definite expressions. It has generally been agreed upon that this is due to a link between a first mentioned entity to a preceding 'trigger' entity by means of a contextual assumption, rendering it inferable (Gundel, 1996; see also Chafe, 1987, 1996; H. Clark, 1977; H. Clark & Haviland, 1977; Fillmore, 1982; Givón, 1995; Hawkins, 1984; Lambrecht, 1994; Prince, 1981, 1992). For instance, inferable referents often stand in a part/whole relationship to previous entities. An example would be body parts as illustrated in (3). The speaker mentions the referent *den Hals* 'the neck' (3d) for the first time in the discourse, and it thus represents new information to the addressee. However, the speaker refers to it with a definite lexical NP. It is likely that the previous mention of a trigger entity (in this case the referent 'man' in 3a-c) has rendered the concept of the referent 'neck' more accessible. The same principle applies to the referent *den Besenstiel* 'the broomstick' in (4d). The speaker mentions it for the first time in the discourse but uses a definite lexical NP to refer to it. This is presumably caused by the previous mention of the referent *Besen* 'broom' in (4b).

(3)

a *da ist ein Mann*$_1$

b *der*$_1$ *öffnet eine Kiste*

c *∅$_1$ holt ein Seil heraus*

d *und ∅$_1$ macht sich daraus einen Strick um den Hals$_2$*

'**a** there is a man$_1$

b he$_1$ opens a box

c ∅$_1$ takes out a rope

d and ∅$_1$ puts it as a cord around the neck$_2$'

(4)

a *dann versucht die Fee das Rutschen von der Torte aufzuhalten*

b *indem sie den Besen$_1$ dagegenstellt*

c *allerdings funktioniert das nicht*

d *weil die oberste Schicht der Torte dann den Besenstiel$_2$ runterrutscht*

'**a** then the fairy tries to stop the sliding of the cake

b by putting the broom$_1$ against it

c but it does not work

d because then the upper part of the cake is sliding down the broomstick$_2$'

In summary, indefinite lexical NPs are typically used for new (or least accessible) referents, whereas definite lexical NPs can be used for given, but less accessible referents on the one hand, and new, but somewhat accessible (inferable) referents on the other hand. Importantly, indefinite and definite lexical NPs both constitute rich referential expressions and therefore complement a scale of referential expressions representing referent accessibility (from low to high), as illustrated in (5).

(5) *indefinite lexical NP < definite lexical NP < pronoun < zero*

2.1.3 Clause structure and grammatical role

There are also clause level phenomena related to the accessibility or information status of discourse referents. When referents are new to the discourse, speakers are more likely to introduce them towards the end of the utterance (Chafe, 1994; H. Clark & Haviland, 1977; Hickmann et al., 1996). One way to achieve that is for speakers to use clause structures that are more specialized for referent introductions, such as locationals (i.e., existentials [6-7], locatives [8], and possessives [9]; E. Clark, 1978). These clause structures focus on the existence of a new referent, which is reflected in the verb semantics used (i.e., low content verbs, such as 'be' and 'have' or close variants), and/or in the use of locational elements (i.e., inanimate locations[2] as in *auf dem Tisch* 'on the table' in 8, or animate locations as in *die* 'she' in 9; E. Clark, 1978, see also Givón, 1983).

(6)

es gibt einen Tisch

'there is a table'

(7)

da sind drei Feen

'there are three fairies'

(8)

und auf einem Tisch steht eine riesen Torte

'and on a table is/stands a big cake'

(9)

und die hat ein Besen

'and she has a broom'

[2] Note that 'there' in 7 might in principle also constitute a location indication. However, in existential structures, it is not clear whether speakers and addressees process it as such.

More specialized clause structures for the introduction of referents can be contrasted with less specialized clause structures, which typically express events that entities are involved in (10-11). These can be either intransitive constructions, in which the new referent is the single argument/subject of the intransitive verb (*eine grüne Fee* 'a green fairy' in 10). Or transitive constructions, in which the new referent is typically instantiated as the transitive object (*einen Korb* 'a basket' in 11; Dixon, 1979; Du Bois, 1987). The contrast between more and less specialized clause structures is similar to the contrast between clauses in the descriptive versus narrative mode (Du Bois, 1980). Narrative (or less specialized) clauses are typically used to advance the story in contrast to descriptive (or more specialized) clauses which typically do not have this function, but are rather used to describe entities, their locations and/or their relationships to other discourse entities (see also McNeill & Levy, 1982 for a similar description).

(10)

dann kommt eine grüne Fee

'then comes a green fairy'

(11)

sie trägt einen Korb

'she carries a basket'

(12)

die Fee kommt wieder runter

'the fairy comes down again'

Most importantly, given/more accessible referents usually pattern differently from new/less accessible referents, in that they are more likely to be mentioned in less specialized or narrative clauses (*sie* 'she' in 11 and *die Fee* 'the fairy in 12). Furthermore, given/more accessible referents are more likely to take on the grammatical role of the subject than that of the object (e.g., Chafe, 1994; Givón, 1983; Du Bois, 1987). Specifically, in transitive clause structures, subjects are highly likely to be accessible and expressed with lean referential expressions (pronoun or zero) whereas objects tend to carry the new/less accessible information expressed by rich referential expressions (e.g., *sie* 'she' vs. *einen Korb* 'a basket' in 11; e.g., Du Bois, 1987; Kärkkainen, 1996; Schütze-Coburn, 1987 for German, cited in Du Bois, 1987).

2.1.4 Dimensions of information status/accessibility

2.1.4.1 First versus subsequent mentions

The main division into that which is new and that which is given concerns the difference between first and subsequent mentions. First mentions constitute introductions of new referents, whereas subsequent mentions maintain or track already-mentioned referents throughout the discourse. Both first and subsequent mentions can be further subdivided. First mentions can be divided into 'brand new' or 'inferable' (Prince, 1981, 1992), corresponding to less versus more accessible. Subsequent mentions can be divided into 'reintroduced' (after a gap of absence) versus 'maintained' (from the immediately preceding clause[s]), which also corresponds to less versus more accessible. A summary is given in Figure 1.

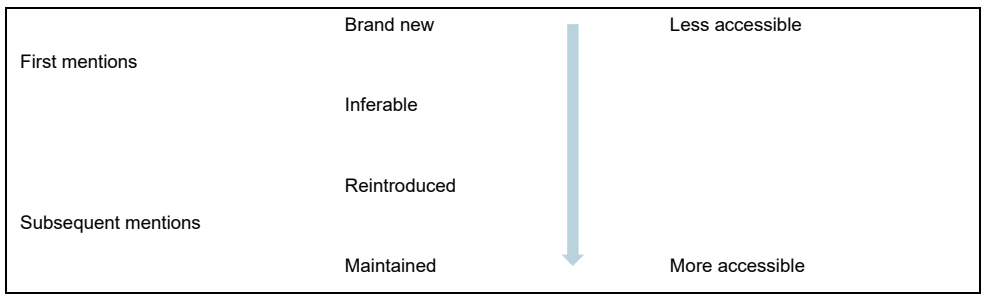


Figure 1: Information status/accessibility of referents in discourse

2.1.4.2 Referential distance

Another way of measuring information status or accessibility of referents in discourse is referential distance. Referential distance is a measurement that assesses the gap between a current mention of a referent and its previous occurrence in the discourse (Givón, 1983). When dealing with natural language production, this gap is typically expressed in terms of the number of clauses in between the two mentions (e.g., Arnold, 1998; Du Bois, 1987; Gullberg, 2006; Hickmann & Hendriks, 1999). The minimal value corresponds to one clause (i.e., when the current mention of a referent is coreferential with a referent in the immediately preceding clause), thereby indicating the highest level of accessibility. The maximal value is in principle infinite. Givón (1983) set an arbitrary boundary of 20 clauses as maximal value, considering everything above that boundary to be similarly low in accessibility (or new). Moreover, on the basis of the studies in Givón (1983), he defined an intermediate boundary spanning over three clauses, that is the 'immediately preceding register' (Givón, 1983: 14). This is to say, if a referent has been mentioned in the three clauses preceding its current mention, its status as a more accessible referent is typically kept. It is thus possible that the speaker is more likely to use zeros or pronouns for the expression of the referent in

this context. Conversely, if a referent has not been mentioned in the three clauses preceding its current mention, a lexical NP should be more likely. A special consideration is given to indefinite lexical NPs, which according to Givón (1983) do not need to be assessed in terms of referential distance. Rather, these forms can immediately be counted as new (or least accessible).

Importantly, a considerable number of studies examining different languages has found that referential distance correlates in important ways with referential form and/or grammatical role (e.g., Ariel, 1988; Arnold, 1998; Chafe, 1994; Clancy, 1980; Du Bois, 1987; Givón, 1984; Halliday & Hasan, 1976), which has also been supported by comprehension studies (e.g., H. Clark & Sengul, 1979; Duffy & Rayner, 1990; Ehrlich & Rayner, 1983; O'Brien, 1987). The pattern suggests that the further away the antecedent, the more likely it is that a rich referential expression is used and the more likely that the referent will be instantiated as intransitive subject or transitive object (e.g., Du Bois, 1987).

2.1.5 Summary

It is generally agreed upon that the way that speakers refer to discourse referents in speech depends on how accessible they are, and specifically, how accessible the speaker assumes them to be for the addressee. Two crucial variables that influence the assumptions about referent accessibility in discourse are inferability and referential distance. For referents that are mentioned for first time, the speaker must decide whether they represent brand-new information to the addressee, or whether the addressee is able to infer the existence of the referent by way of an inferential link to the previous discourse. For subsequent mentions, referential distance within the discourse, that is the length of the gap of absence between the current and the preceding mention of the referent, often plays an important role. In the light of these variables, the speaker will alter the way they refer to discourse referents on 'local' and more 'global' levels. I discussed four different dimensions, that is nominal definiteness, richness of expression, the structure of the clause in which the referent is mentioned, and the grammatical role it is instantiated in. Choosing the appropriate ways of referring to discourse referents along these dimensions is crucial for the creation of *cohesion*.

2.2 Discourse reference in gesture

The starting point for the consideration of gestures in discourse reference is that gestures are part of language and as such combine with speech not only on the word or sentence level, but also on the discourse level (McNeill, 1992). But while variations in information structure for discourse reference in speech are rather well described, we know comparatively little about the role that gestures play. In the following, I start by providing a definition of gestures, mainly following Kendon (1980, 1986, 2004) and McNeill (1992, 2005), and show how gestures can be classified. I will then present what is currently known about the discursive relationship between speech and gestures, and specifically when it comes to the representation of referents.

2.2.1 What are gestures?

Gestures are defined as visible actions of the hands and arms which speakers use while they are talking (Kendon, 1972; 1980; McNeill, 1992). Importantly, speakers in a communicative interaction perform many different bodily actions (i.e., self-adaptors, such as scratching their heads, adjusting their clothes, or other actions, such as drinking, cooking, etc.). But only those visible actions that are relevant to the talk in progress – or in other words, that are regarded as part of the speaker's total expression – are considered to be gestures (Kendon, 1980; 1986; but see Andrén, 2014, on how practical actions used by children can be considered 'gestural'). Kendon (1978) showed that, when asked to describe speakers' hand and arm movements, people were very good at recognizing which actions were part of what the speaker was trying to communicate and which ones were not. Recent neurocognitive evidence has further corroborated these observations by showing that the processing of speech-associated gestures differs in comparison to the processing of self-adaptors (Skipper, Goldin-Meadow, Nusbaum & Small, 2007) or other types of actions used while speaking (such as cutting, pouring water, etc.; Kelly, Healy, Özyürek & Holler, 2015).

Perhaps the most crucial feature that makes gestures recognizable as communicatively intended is their interplay with speech in terms of meaning and timing. In fact, gestures are semantically and temporally coordinated with speech such that they express closely related or complementary meaning at the same time (Kendon, 1986; McNeill, 1992). Figure 2 illustrates this interrelation between the modalities. The speaker is introducing the entity 'a mannequin' in the utterance *und die hat eine Puppe vor sich stehn* 'and she has a mannequin standing in front of her', by producing a gesture depicting the shape of the mannequin and by aligning the gesture exactly with the spoken referential expression (bold face indicates gesture alignment).

*und die hat ei**ne Puppe v**or sich stehn*

'and she has **a mannequin** standing in front of her'

Figure 2. Example of a gesture

This coordination in meaning and time is achieved despite the essential differences between the modalities with regard to their respective mode of expression. While speech has a standard of well-formedness and is linear/analytic, gesture has no standard of well-formedness and is global/synthetic/imagistic (McNeill, 1992). A consequence of this difference is that gestures can typically only be fully understood within the context of the spoken utterance that they co-occur with. The difference in mode of expression further entails that gestures can reveal non-redundant or different aspects of the meanings that the speaker is conveying in speech. For instance, gestures might express information about direction, size, shape or orientation (e.g., Beattie & Shovelton, 2007; Gullberg, 2011b; Kendon, 2004; Kita & Özyürek, 2003), even if this information is absent in speech. As shown in Figure 2, the speaker gesturally provides shape information about the entity 'mannequin' whereas she does not mention any aspects of its shape in speech.

The semantic coordination between speech and gesture is rarely a simple one-word-one-gesture mapping. Rather, gesture meaning parallels the meaning expressed by the phrasal or clausal context that the gesture appears in. In this case, a gesture is said to semantically coordinate with 'conceptual affiliates' (De Ruiter, 2000; but see also McNeill & Levy, 1982; McNeill, 1992). Because of gestures' imagistic nature, they can and do often express meanings that speech is not able to represent in one word. In Figure 3 the speaker is talking about candles on top of a cake while accompanying the referential expression 'candles' with a gesture drawing a (concave shaped) horizontal line. Previous to this utterance, the speaker had drawn the shape of a cake in front of her, extending from the height of her hips to the height of her chest. Thus, the gesture

in this example does not represent the candles as such, but rather reveals the location of the candles ('on top of the cake') and the fact that they are standing next to each other in a line. While synchronized with the referential expression 'candles', the gesture represents the concept that is represented by the whole spoken utterance.

*und auf der Torte sind **Kerzen** drauf*

'and on the cake are **candles** on top of it'

Figure 3. Example of a gesture

As illustrated by the examples, there is a clear parallelism on the word and clause level between meanings represented in speech and in gesture. But the coordination between the modalities goes beyond the word and sentence levels and further manifests itself on the discourse level. Before going into the details of this relationship, however, I will shortly discuss some classifications of gestures that will be relevant for the studies in this thesis.

2.2.2 Ways of classifying gestures

Gestures are typically divided into those gestures that are produced with speech and can only be understood in the presence of speech versus gestures that *can* be produced with speech, but that also have specific meanings when they are produced without speech. The latter ones typically have a standard of well-formedness and a well-defined meaning within a certain culture (e.g., the thumbs up gesture). They are often referred to as 'emblems' or 'quotable' gestures (Efron, 1941/1972; Ekman & Friesen, 1972; Kendon, 1995; Payrató, 1993). Emblems have traditionally been described as gestures that are autonomous from and can be used as substitutes for speech. However, they

often also occur with speech and interact with utterances' pragmatic meaning in important ways (Kendon, 1995).

Gestures that are used with speech, on the other hand, are typically described as spontaneous movements, which create meanings on the fly (McNeill, 2002). They have been variously referred to as 'gesticulations', 'co-speech gestures', 'speech-accompanying gestures', 'speech-associated gestures' or 'visible action as utterance'. Gestures are often further classified into referential versus pragmatic gestures on functional grounds (Kendon, 2004), or into representational gestures versus beat gestures on articulatory (or formal) and functional grounds (McNeill, 1992). Referential/representational gestures are used to represent entities, their properties, actions and movements or spatial relations to other entities by way of iconicity or deixis (Kita, 2000; see Figures 2 and 3, respectively). Gestures that represent entities via deixis have also been called 'pointing' gestures. Deictic or pointing gestures can either be concrete (indicating an object or person in the physical surrounding of speakers and addressees) or they can be abstract, in which case, the gestures are assigning locations in gesture space to discourse referents that are not physically present. Finally, pragmatic or beat gestures are mostly defined negatively as *not* having any semantic content and therefore no depictive functions (see for instance, McNeill, 1992, on the 'beat filter').

This thesis mainly considers referential/representational gestures, which can be divided further depending on the relevant research question. In paper II, we investigate congruent (or anaphoric) versus incongruent localizing gestures. In paper III, we use the division between Character versus Observer Viewpoint gestures (henceforth C-VPT and O-VPT). And in paper IV, we discuss 'entity' versus 'action' gestures. I give a short presentation of each of the divisions in turn. Further details are provided under 2.2.3 when discussing the background of each corresponding research question.

2.2.2.1 Localizing (anaphoric) gestures

The definition of localizing (anaphoric) gestures follows the work by Gullberg (1998, 2003, 2006). Speakers use localizing gestures to associate a referent with a certain location in space at their introduction and specifically in co-occurrence with the referential expression. Speakers can then refer back to the location and thus reactivate the referent at its reintroduction. The second localizing gesture that is produced in the same location for the same referent, and crucially also in co-occurrence with the referential expression, is called a localizing anaphoric gesture. Importantly, the definition is based on the spatial properties of a gesture (not function or semantics). Figures 4a-b illustrate the use of a localizing gesture followed by a localizing (anaphoric) gesture.

*und **der erste Mann** nimmt ein'n schwern Stein*

'and **the first man** takes a heavy stone'

Figure 4a: Example of a localizing gesture

*ähm der **Mann** hebt dann die Hand*

'uhm the **man** then raises his hand'

Figure 4b: Example of a localizing anaphoric gesture

2.2.2.2 Character and Observer Viewpoint gestures

The differentiation between C-VPT and O-VPT gestures follows the definition by McNeill (1992, 2005). According to McNeill (1992: 119), C-VPT gestures are those in which the speaker's body is incorporated into the gesture space, which is reflected by the speaker's hands representing the referent's hands. O-VPT gestures on the other

hand exclude the speaker from the gesture space. Rather it is as if the speaker was looking at the scene from the outside and their hand(s) represent(s) a referent as a whole. Figure 5-6 illustrate the difference between the two viewpoints. In Figure 5, the speaker is performing a sewing movement by pretending to hold a needle. In Figure 6, the speaker is representing the path of an egg yolk falling into a bowl with her left hand.

*und näht erst **das Ober**teil zusammen*

'and sews **the upper** part together first'

Figure 5: Example of a Character Viewpoint gesture

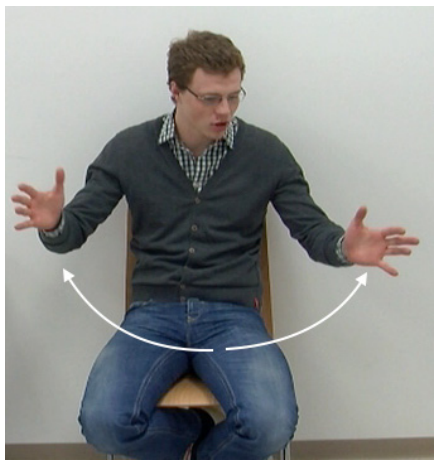
***das Eido**tter ist im Begriff in die Schüssel zu falln*

'**the egg y**olk is about to fall into the bowl'

Figure 6. Example of an Observer Viewpoint gesture

2.2.2.3 'Entity' and 'Action' gestures

The definition for the differentiation between 'entity' and 'action' gestures follows the work by Wilkin and Holler (2011). Gestures focusing on entity information are gestures that represent a referent itself, as in its shape, size or location (in relation to other referents). Gestures focusing on action information, on the other hand, are gestures that represent the action that a referent is involved in, whether the referent is the instigator of the action or the affected. Figures 7-9 show the difference between gestures focusing on entity information (shape and location in Figures 7-8 respectively) and gestures focusing on action information (Figure 9; see also Figure 6).



*und dann ist noch **n Korb d**a*

'and then there is **a basket**'

Figure 7. Example of a gesture focusing on entity information (drawing the shape of a basket)

*aber es ist dann irg**endwie n Koch**buch da*

'but there is some**how a cook**book there'

Figure 8. Example of a gesture focusing on entity information (indicating the location of a cook book)

*aber sie nimmt trot**zdem ein Stück Stoff** raus*

'but she takes out **a piece of cloth any**ways'

Figure 9. Example of a gesture focusing on action information (representing a person taking a piece of cloth out of a basket)

2.2.3 Gestures on the discourse level

The way that language users refer to entities in the flow of discourse is closely related to the information status of the referents and is thus crucial for the creation of cohesion (i.e., the connectedness of discourse). For speech, different strategies have been identified that speakers use to indicate whether a referent is new/less accessible or given/more accessible (see 2.1). The studies in this thesis take as their starting point these patterns and examine the way that speech-associated gestures are deployed in relation to them. The investigations can be considered along four main questions: *when*, *where*, *how* and *what*.

2.2.3.1 When are gestures used?

The question of *when* gestures are used refers to the incidence (or presence/absence) of gestures in relation to the different types of referential expressions that encode discourse referents. Some of the earliest studies on speech-associated gestures have examined this relationship and have taken the observed patterns as important evidence for the integrated nature of the two modalities, and specifically for the pragmatic/communicative function of gestures (Levy, 1984; Levy & McNeill, 1992; Marslen-Wilson, Levy & Tyler, 1982; see also Gullberg, 2003; McNeill, Levy & Pedelty, 1990).

Marslen-Wilson et al.'s (1982) study was the first to systematically examine the use of referential expressions in a narrative context by taking into consideration the contribution of gestures. The authors analyzed the spoken and gestural behavior of one subject who was retelling the content of a comic book story. During their retelling the subject had the comic book on their lap, which resulted in the production of exclusively concrete deictic gestures to the pictures of the two relevant characters in the story. The distribution of these gestures was not random. In fact, the speaker not only adjusted the form of their referential expressions according to referents' information status, but also their gestures. More specifically, the speaker accompanied the names and definite descriptions of protagonists with deictic gestures when they were first introduced in the narrative. Furthermore, the speaker accompanied the names of protagonists in reintroduction contexts, and most notably when a new episode started. But crucially, gestures never occurred with pronouns or zero anaphora, referential expressions that typically maintained referents from one clause to the next (see also Levy, 1984; McNeill et al., 1990). Marslen-Wilson et al. (1982) suggested that gestures have a reference fixing function. That is, they proposed that gestures function similarly to descriptions that accompany names, which indicate what the properties of a referent are.

Levy and McNeill (1992) further suggested that the combination of richer spoken expressions with accompanying gestures (in contrast to leaner spoken expressions without gestures) might reflect communicative dynamism (Firbas, 1971). Communicative dynamism is defined as the degree to which a piece of information

"pushes the communication forward" (Firbas, 1971: 136). Levy and McNeill proposed that communicative dynamism accumulates when a piece of information is new in relation to a previous stretch of discourse. This piece of new information should then be expressed by a more elaborate referential expression and accompanied by a gesture in order to reflect the higher level of communicative dynamism. Their examinations of three narratives by different speakers support this proposal (see also Levy & Fowler, 2000).

Gullberg (2003) also investigated the incidence of gestures in relation to referential context and the co-occurring spoken referential expression. The findings suggest that gestures might be sensitive to both referential context and the richness of the referential expression. In relation to referential context, she found that most gestures tended to accompany introductions of referents (25%), some gestures accompanied reintroductions (14%), but very few gestures accompanied maintained referents (2%; see also Yoshioka, 2008). In relation to spoken referential forms, she found that gestures predominantly occurred with lexical NPs (92%) and very few gestures accompanied pronouns (8%). Similarly, in Gullberg (2006), considering only subsequent mentions of referents, she found that lexical NPs were more likely to be accompanied by gestures than pronouns (23% vs. 0,5%; see also Perniss & Özyürek, 2015).

In summary, previous research on *when* gestures are used in order to represent discourse referents suggests that there is a strong relationship between the presence of gestures and the use of rich referential expressions, specifically in introduction and reintroduction contexts.

The current thesis adds to previous research by examining more closely the contexts of referent introductions and reintroductions. Paper I focuses on introductions of referents. Specifically, it takes as its starting point that while speakers generally tend to accompany newly introduced referents more than given/maintained ones, they still *do not* accompany *all* first mentions of referents with gestures (e.g., 39.8% in Foraker, 2011; 25% in Gullberg, 2003). Paper I addresses this gap by examining gesture incidence in relation to the information status of first mentioned referents (brand-new vs. inferable). Paper III (study 2) targets the question of whether gestures are used more often for introductions than for reintroductions of referents. Previous research has suggested that there is a qualitative difference between gestures in those two contexts (Gerwing & Bavelas, 2004; Wilkin & Holler, 2011), but there is little evidence for a potential variation in the incidence of gestures.

2.2.3.2 Where are gestures produced?

The question of *where* gestures are produced in gesture space refers to the potential cohesive use of space by gestures, a strategy that allows speakers to anaphorically track a referent in the visual modality. Just as speech uses anaphoric expressions in order to track a referent through discourse (e.g., 'a fairy in a red dress – the red one – she'),

gestures can fulfil that function as well, for instance by using a recurrent location in space. Production studies have shown that speakers make use of this strategy and a growing body of comprehension studies have provided evidence that addressees use spatial information from gestures (albeit in somewhat diffuse ways) when it comes to referent representation (Cassell, McNeill & McCullough, 1999; Goodrich Smith & Hudson Kam, 2012; Gunter & Weinbrenner, 2017; Gunter, Weinbrenner & Holle, 2015; Sekine & Kita, 2015, 2017).

Starting with the production studies, a number of studies has revealed the following pattern. When a referent is mentioned for the first time, a speaker can assign a specific location in space to that referent by using a localizing gesture in exact temporal alignment with the referential expression. When the speaker then introduces a second referent, they can choose another location in space for that referent in order to differentiate between the two referents spatially and in parallel to speech. Once assigned, the locations can be reused at any time and reactivate the referent in question. Importantly, however, speakers typically align a localizing anaphoric gesture with the referential expression only when a referent is reintroduced (typically with a lexical NP; Gullberg, 2003, 2006; McNeill, Cassell & Levy, 1993; McNeill & Levy, 1993; Perniss & Özyürek, 2015; So, Kita & Goldin-Meadow, 2009). The studies thus highlight that speech and gestures work in parallel when it comes to referent tracking. That is, when speakers use more marking material in speech (lexical NPs) to introduce or reintroduce a referent in discourse, they also use localizing gestures. But when speakers use less marking material in speech (pronouns) because they are maintaining a referent, they also tend *not* to use localizing gestures.

Beyond this pattern, Gullberg (2006) further sought to uncover the role that the addressee plays for the production of localizing gestures. She tested subjects in two conditions: full visual access (subjects sat across from each other at a table and had full visibility of each other's gestures) versus no visual access (a screen was placed in between the subjects in order to prohibit gesture visibility). The findings showed that the locations used for referents were more stable, and speakers kept locations apart more diligently in the full visual access condition than in the no visibility condition. This suggests that speakers design their gestures with the addressee in mind when it comes to localizations (see also Özyürek, 2002). Interestingly, Gullberg (1998, 2011a) also showed that in interactive stretches, addressees tended to point back to locations previously established for referents by the speakers. This in turn, provides evidence for the fact that addressees are picking up the information that spatial representations of referents create.

Turning to comprehension, a growing number of studies has aimed to support this view. Some studies have shown that localizing anaphoric gestures can facilitate processing in comparison to spatially incongruent gestures or speech alone (Cassell et al., 1999; Gunter & Weinbrenner, 2017; Gunter et al., 2015; Sekine & Kita, 2017).

For instance, in Cassell et al. (1999), participants watched taped retellings of a story by a person using congruent or incongruent localizing gestures. When asked to retell the stories, participants produced more retelling inaccuracies after the incongruent condition than after the congruent condition. In an ERP study, Gunter and Weinbrenner (2017) found that subjects who watched someone use localizing anaphoric gestures showed different activation patterns in the brain than when they watched someone using no gestures at all. This suggests that there is a neural underpinning for the facilitation effect in processing of anaphoric gestures in addition to speech (see also Gunter et al., 2015). Finally, Sekine and Kita (2017) showed that, in a reaction time experiment, subjects were significantly slower to respond in a condition with incongruent localizing gestures than in a no gesture condition. However, some of the same studies have also provided contradictory results. For instance, Gunter and Weinbrenner (2017) also examined brain responses in an experiment including three conditions, namely gesture congruent, gesture incongruent and no gesture, but found no difference between the conditions (see also Hudson Kam & Goodrich Smith, 2011, for similar results but with a different task). Similarly, Sekine and Kita (2017) found no facilitation effect of a gesture congruent condition in relation to a no gesture condition.

In summary, there seems to be a rather robust view in production studies that speech and gestures work in parallel, using space cohesively when introducing and reintroducing referents in discourse. Furthermore, speakers seem to qualitatively adjust their gestures with their addressees in mind. In perception studies, on the other hand, the findings diverge. Paper II discusses differences in research designs which could potentially explain the diverging results in previous studies and offers a new way of examining the sensitivity to localizing anaphoric gestures by addressees. Most notably, in contrast to previous studies, the design used in paper II reflects more closely the use of localizing (anaphoric) gestures in spontaneous communication and focuses on the tracking of a single referent instead of using a context of contrast/disambiguation, which has typically been used in previous studies on this topic.

2.2.3.3 How do gestures express meaning?

How gestures express meaning refers to differences in the techniques of representation in gesture in order to represent referents and/or their actions. Table 1 shows some techniques that have been identified by different scholars (Capirci, Cristilli, De Angelis & Graziano, 2011; Kendon, 2004; Marentette, Pettenati, Bello & Volterra, 2016; McNeill, 1992; Müller, 1998, 2014; see also Streeck, 2008).

Table 1: Techniques of representation

McNeill, 1992	Kendon, 2004	Müller, 2014	Capirci et al., 2011	Marentette et al., 2016
Observer Viewpoint (O-VPT)	Depicting	(Molding/Drawing)	Shape depiction, Delimitation	Size-and-shape
	Modeling	Representing	Hand becoming an object	Hand-as-object
Character Viewpoint (C-VPT)	Enactment	Acting	Mime, Manipulation	Hand-as-hand, Own-body

McNeill (1992) differentiates between O-VPT and C-VPT gestures. O-VPT gestures correspond to the techniques of representation that Kendon (2004) calls 'depiction' and 'modeling'. 'Depiction' refers to the hands molding or drawing the shape/size of an entity (e.g., drawing a square in the air to represent a box, or extending index finger and thumb to indicate the size of an object). 'Modeling' refers to a (or both) hand(s) representing an entity as a whole (e.g., stretched-out index finger for the referent 'needle'). Both 'depiction' and 'modeling' can further be used to represent the movements of an entity. For instance, a speaker can draw a line through gesture space in order to depict a path travelled by an entity. Similarly, a speaker can use their hand as a model for an entity and, at the same time, move it through space in order to represent the entity's path. C-VPT gestures, on the other hand, correspond to the technique of representation that Kendon calls 'enactment'. 'Enactment' refers to gestures in which a speaker is acting out an event from the perspective of a character. That is, the speaker's hands or body map onto an entity's hands or body (e.g., enacting someone sewing with a needle).

The relationship between techniques of representation in gesture and the accessibility of discourse referents has explicitly been formulated by McNeill (1992). He proposed that gestures can be put on a scale along which they progress in 'complexity'. The scale starts with no gestures, continues with beat and deictic gestures and ends with O-VPT and C-VPT gestures. McNeill further proposed that this progression is a reflection of communicative dynamism, whereby no gestures should be used in co-occurrence with the mention of a referent with very low communicative dynamism, and on the other end of the spectrum, C-VPT gestures should be used in co-occurrence with the mention of a referent with a very high degree of communicative dynamism. The variation between using an O-VPT versus C-VPT gesture should then, at least partly, depend on the accessibility or information status of the referent it represents.

One way of assessing the degree of communicative dynamism of a referent is by considering the form of the referential expression used to refer to it. In fact, McNeill

proposed to correlate his scale of gesture progression with Givón's (1983) scale of quantity for referential expressions (Figure 10). Based on a large range of cross-linguistic studies, which all examined the form of referential expressions in relation to the accessibility of referents, Givón formulated his scale of quantity ranking referential expressions according to their phonological size (or richness). Thus, one way of testing the validity of McNeill's proposition for the variation of O-VPT versus C-VPT gestures is to correlate the two scales directly and quantitatively (McNeill himself has only made qualitative observations).

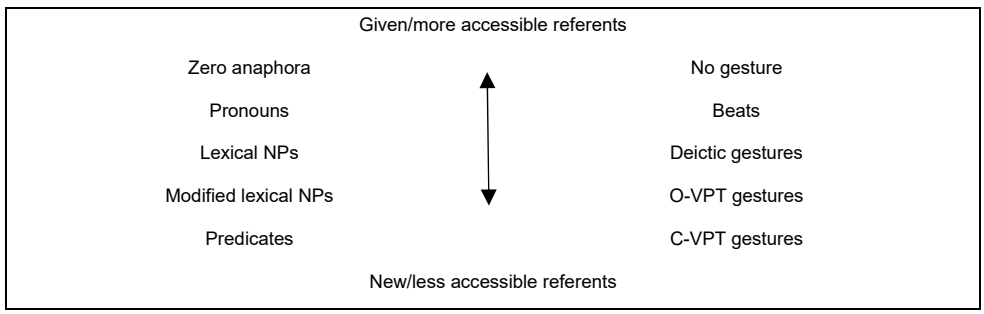


Figure 10: Alignment of scale of linguistic quantity and gesture progression (adapted from McNeill. 1992)

There are some indications in the literature that would support this proposition. Parrill (2012) conducted an experiment in which speakers retold a story to their addressees under two conditions: either the story was completely new to the addressee or the addressee was previously acquainted with the story. She found that speakers used more C-VPT gestures when addressees did not know the story, and conversely speakers used more O-VPT gestures when addressees already knew the story. Although it remains unclear which parts of speech the gestures were exactly aligned with, it is possible to assume that speakers used richer/indefinite referential expressions to mention referents in the first condition (because all referents were new to the addressee) whereas they used leaner/definite expressions to mention referents in the second condition (because the addressee already had knowledge of the referents). Therefore, Parrill's study provides indirect evidence for McNeill's proposition.

A study by Debreslioska, Özyürek, Gullberg and Perniss (2013) has provided more direct evidence that techniques of representation, and specifically gesture viewpoint, is sensitive to the information status of referents as reflected in the referential expressions representing them. The study found that gestures tended to be produced in O-VPT when representing discourse referents instantiated as intransitive subjects (typically less accessible), whereas they tended to be produced in C-VPT when representing discourse referents instantiated as transitive subjects (typically more accessible). In relation to McNeill's scale, this result seems to contradict the proposition that C-VPT gestures occur with less accessible referents. However, it is important to note that Debreslioska

et al. (2013)'s study was based on a clause level analysis, rather than the consideration of exact temporal alignment between speech and gestures. The latter, however, is the basis for the proposition made by McNeill (1992).

Paper III, study 1, sets out to test McNeill's (1992) proposition of a scale of gesture progression more directly by examining whether the differential use of gesture viewpoint can be linked to richness of expression. Paper III, study 2 goes beyond richness of expression (which McNeill has proposed as *one* possibility to test the scale) and further examines whether gesture viewpoint is sensitive to other indicators of a referent's information status, namely nominal definiteness and grammatical role. Contrary to Parrill (2012) and Debreslioska et al. (2013), the analysis of the relationship between gesture viewpoint, richness of expression, and nominal definiteness examines the exact temporal alignment between speech and gestures in order to link the results more directly to McNeill's scale. Furthermore, the analysis of the relationship between gesture viewpoint and grammatical role complements the study by Debreslioska et al. (2013) by specifically focusing on the variation of viewpoint with transitive subjects (typically more accessible) versus transitive objects (typically less accessible).

2.2.3.4 What meaning do gestures express?

The *what* question refers to the information that representational gestures express when they accompany discourse referents (i.e., their semantic content). A speaker can focus on different aspects concerning a referent in their gesture. For instance, when talking about a needle, a speaker could use a stretched-out index finger pointing downwards in order to provide information about the entity (and its orientation). Or she could enact the holding of a needle and do a sewing movement in order provide information about an action that the entity is involved in.

Much of the research showing what the semantic content of gestures is sensitive to has focused on gestures accompanying verbs representing events. One of the first studies in this domain (McNeill & Levy, 1982) examined gestures aligning with verbs and found that there were important correlations between some gesture features and some verb features. For instance, verbs implying a downward motion correlated with gestures that represented a downward path, while verbs implying a horizontal motion, correlated with gestures that represented a lateral movement from right to left.

Others examined verb semantics cross-linguistically and revealed that gestures parallel the information expressed in the verbs in a language-specific way (Brown, 2008; Brown & Chen, 2013; Brown & Gullberg, 2008; Choi & Lantolf, 2008; Hickmann, Hendriks & Gullberg, 2011; Kita & Özyürek, 2003; Stam, 2006). For instance, Gullberg (2011b) considered the domain of placement events. She showed that Dutch speakers preferred to use posture verbs, which are specific with regard to object properties (*zetten* 'sit/stand', *leggen* 'lay' and *hangen* 'hang'). In contrast, French speakers preferred to use

one neutral verb, which does not take into consideration any object properties (*mettre* 'put'). Interestingly, gestures reflected these patterns by incorporating object properties in the hand shape in Dutch, but not in French (see also Gullberg, 2009).

Other studies focused on the expression of manner and path in motion events and showed that the information that gestures express depends on sentence construction. For instance, in so-called satellite framed languages (such as English), speakers prefer to package path and manner in one clause (as in 'he rolls down'), whereas in so-called verb framed languages (such as Japanese), speakers prefer to package path and manner in two separate clauses (as in 'he descended while rolling'). Importantly, gestures are reflecting this choice by encoding both path and manner in one gesture when accompanying a one clause sentence (circling gesture moving downwards), whereas gestures encode path and manner in two separate gestures when accompanying a two-clause sentence (one downward moving gesture followed by a circling gesture; see also Fritz, Kita, Littlemore & Krott, 2019; Özyürek, Kita, Allen, Furman & Brown, 2005).

In summary, the literature has provided important findings showing how the semantic content of gestures can vary depending on the syntax used to represent events, and on the semantics of verbs. However, when it comes to what information gestures express when accompanying discourse referents, we know relatively little. There are only two previous studies suggesting that information status and nominal definiteness might play a role.

Foraker (2011) examined whether speakers varied semantic meaning in gestures when they co-occurred with first mentions of discourse referents versus subsequent mentions of referents. She tested four participants on a story retelling task and found that gestures differed in the information they expressed. Specifically, speakers tended to provide redundant information when the gesture accompanied first mentions (i.e., about the entity itself) whereas speakers tended to provide additional information when the gesture accompanied subsequent mentions (i.e., about the entity's action or another entity). Similarly, in a corpus study of 28 speakers retelling a story, Wilkin and Holler (2011) found that speakers' gestures differed in their semantic content depending on the definiteness of referential expressions. That is, for referents that were instantiated with indefinite nominals, speakers tended to use 'entity' gestures (i.e., representing the entity's shape, size or location), whereas when referents were instantiated with definite nominals, speakers tended to use 'action' gestures (i.e., representing actions or movements the entity was involved in).

In paper IV, the current thesis adds to previous research by examining the semantic content of gestures for discourse referents in a context of first mentions. The context of first mentions is particularly interesting for this endeavor because the expression of discourse referents in this context varies in three important ways, information status (brand-new vs. inferable), nominal definiteness (indefinite vs. definite) and clause structure in which the referent is embedded (more specialized for referent introduction,

or descriptive vs. less specialized, or narrative). Exploring the relationship with all three variables allows for more fine-grained insights into what it is that might drive the way that information is expressed in gesture.

2.3 The studies in this thesis

The studies this thesis focus on the way that referents are represented in speech as a function of discourse and explore the corresponding gesture patterns in production and perception. The examination of those patterns are approached from four different angles, namely by considering the questions *when?*, *where?*, *how?* and *what?*. *When* concerns the incidence of gestures in relation to the way referents are expressed in speech. *Where* concerns the affordance of gestures to use space cohesively, which is assumed to help the addressee. *How* concerns the way that gestures represent referents as a function of the way that referents are represented in speech. *What* concerns the semantic content that gestures express when representing referents, again in relation to how referents are referred to in speech.

2.3.1 *When*?

Paper I and part of paper III deal with the question of *when* gestures are used to accompany referents in discourse. Both studies take McNeill's (1992) theory of communicative dynamism and gestures as their starting point and examine how gestures pattern with referential expressions according to the accessibility of their referents and three markers of accessibility, namely nominal definiteness, richness of expression, and grammatical role. Paper I focuses on first mentioned discourse referents in particular, whereas Paper III examines the incidence of gestures in relation to both first *and* subsequent mentions of discourse referents, and specifically focuses on instantiations at the narrative level.

2.3.2 *Where*?

Paper II is concerned with *where* in space gestures are produced when accompanying discourse referents and turns its focus to the perception side. It takes as its starting point the results on visual anaphoricity in production (Gullberg, 1998, 2003, 2006; McNeill & Levy, 1993) on the one hand, and the suggestion that gestures in this context have an effect on the addressee (e.g., Gunter & Weinbrenner, 2017) on the other hand. In two experiments, paper II proposes a new way to test whether addressees are sensitive to localizing anaphoric gestures. More specifically, it focuses on tracking a single referent in narrative discourse, avoiding the use of contrast/disambiguation as it was

exclusively used in previous research. Importantly, it also reflects more closely the way that localizing (anaphoric) gestures are produced in spontaneous communication.

2.3.3 *How*?

Paper III deals with *how* gestures are produced when they accompany discourse referents with a specific focus on gesture viewpoint. Paper III also starts with McNeill's (1992) theory of communicative dynamism and gestures, which suggests that gesture viewpoint is sensitive to discourse organizational principles (see also Debreslioska et al. 2013; Parrill, 2010, 2012). Specifically, C-VPT gestures, in comparison to O-VPT gestures, are considered to be more complex and thus more likely to accompany less accessible information. Paper III explores this proposition by examining how gesture viewpoint relates to markers of referent accessibility, that is richness of expression, nominal definiteness and grammatical roles.

2.3.4 *What*?

Finally, paper IV examines *what* information gestures express when they accompany first mentioned discourse referents. The starting point is the observation that when speakers accompany the introduction of a new referent with a gesture, the gesture can either represent the entity itself (its shape, size or location) or an action that the entity is involved in (Foraker, 2011; Wilkin & Holler, 2011). On the basis of previous research, paper IV takes into consideration three variables in order to examine the variation in the semantic content of gestures, namely the information status of first mentions, their formal expression in terms of nominal definiteness, and the clause structures that referents are embedded in.

3 Methods

3.1 Participants

3.1.1 Production studies

Participants (n = 20, 16 females; mean age 26) for papers I, III and IV were recruited at Ludwig Maximilian University in Munich, Germany. Participants were native and monolingual speakers of German (most of them had basic knowledge of English and/or French from school). Early simultaneous bilinguals were excluded. If participants knew the experiment was about gestures, they were also excluded[3].

Participants were invited to come to the experiment with a German speaking friend, who would act as listener. Talking to a familiar person was assumed to encourage a less tense experimental environment and thus lead to more gesture production. However, if participants still did not gesture at all (or very little), they were excluded and a new participant pair was recruited instead until the limit of 20 participant pairs was reached. We provided detailed instructions for the addressees concerning their role (see procedure) to maintain some control over the situation.

3.1.2 Perception study

Participants for paper II, experiment 1, were recruited at DEKRA Hochschule in Berlin, Germany (n = 28, 19 females; mean age 23). For practical reasons, participants for paper II, experiment 2, were recruited at Lund University in Lund, Sweden (n = 29, 21 females; mean age 24). All participants were native speakers of German who had not grown up bilingually. The participants from study 2 had all arrived in Sweden no more than two months prior to taking part in the experiment.

[3] One participant recognized the name of one of the researchers and was familiar with their research focus.

3.2 Design

3.2.1 Production studies

All three production studies are based on a corpus of video recorded elicited narratives of a printed picture story. Narrative elicitation tasks (or semi-spontaneous narratives) are commonly used in linguistics and psycholinguistics for the collection of connected discourse in speech, but also for bimodal discourse including speech-associated gestures. The examination of narratives has proven very fruitful when it comes to phenomena that are best observed in a connected discourse context. This includes the study of discourse cohesion and information structure (Dimroth, 2012). In gesture studies, animated cartoons and printed picture stories are frequently used stimulus materials to elicit speech-accompanying gestures (e.g., Brown & Gullberg, 2008; Debreslioska et al. 2013; Gullberg, 1998, 2003, 2006; Gullberg, Hendriks & Hickmann, 2008; Marslen-Wilson et al., 1982; McNeill, 1992; So et al. 2009; Yoshioka, 2005).

The main advantages for using a narrative elicitation method are the control over the content of speech and the comparability between the uses of gesture. There are also specific advantages to using printed cartoons. For instance, the experimenter can more easily show participants shorter segments of the story and ask them to mention each picture in their retellings without taxing their memory. These two strategies were used in the creation of the corpus in order to guarantee as many data points as possible per participant and per item. Another more general advantage for using a narrative elicitation task is that it typically leads to more empirical material (besides the intended structures that the stimulus was designed for), and the corpus can often be used for further explorations. For instance, in the current thesis, the corpus was designed to elicit a similar amount of C-VPT and O-VPT gestures in events on the narrative level (see 3.3) but was then further explored for the semantic content of gestures in clauses in the narrative *and* descriptive modes.

3.2.2 Perception study

In paper II, we used reaction time experiments in order to test whether subjects were sensitive to localizing anaphoric gestures. Studies examining the process of referential access or anaphor resolution have used reaction time measures since the early 1980s (e.g., Chang, 1980; Cloitre & Bever, 1988; Gernsbacher, 1989; McKoon & Ratcliff, 1980). In these studies, experimenters have typically tested the effects of different types of spoken anaphora using probe recognition tasks (e.g., Gernsbacher, 1989; see also Emmorey, 1997, and Emmorey, Norman & O'Grady, 1991, for American Sign Language), or self-paced reading tasks (e.g., Gordon, Grosz & Gilliom, 1993). The studies assume that longer response latencies indicate more difficulty in matching an

anaphor to the target antecedent. This assumption is based on the more general idea in psycholinguistic studies that the time it takes a participant to carry out a task reflects the underlying complexity of the process under investigation (Garrod, 2006).

For the anaphoric use of localizing gestures in particular, only two previous studies have worked with tasks involving reaction time measurements (i.e., Nappa & Arnold, 2014; Sekine & Kita, 2017). However, the designs between those studies differ in important ways. Both studies work with narratives and contrast two gesturally tracked referents in space. But while Nappa and Arnold (2014) examine the use of concrete deictic gestures to indicate referents, Sekine and Kita (2017) examine the effects of abstract localizing gestures. Furthermore, in Nappa and Arnold (2014) a critical gesture is used in a compensatory fashion (i.e., in order to disambiguate a pronoun which could, in principle, refer to two preceding referents). Sekine and Kita (2017), on the other hand, did not use any gestures at all in the critical clauses. Rather participants saw black boxes with the protagonists' names in them. In both studies, participants answered questions about which referent was referred to during the critical clause. But while in Nappa and Arnold (2014), participants were asked a question after each critical clause, Sekine and Kita (2017) gave participants one general question for all experimental items at the outset of the experiment. Considering that these two cited studies are the only ones that use reaction times for the examination of anaphoric localizing gestures, it becomes clear that the subject is still under-researched and also that the existing studies vary considerably in terms of design.

The experiments in paper II are similar to these previous studies in that they also used a narrative context to test the sensitivity to gestural referent tracking. Furthermore, similar to Sekine and Kita (2017), we gave participants a question which was applicable to all items at the beginning of the experiment. However, the experiments in paper II also differ from previous studies in crucial ways. First, the gestures only tracked a single referent, thus avoiding contrast/disambiguation in gestures. Moreover, gestures temporally aligned with lexical NPs (and not with pronouns) in introduction and reintroduction contexts, which is in accordance with production studies (e.g., Gullberg, 2006; Levy & McNeill, 1992). We tested participants in three conditions: Gesture congruent (i.e., the speaker uses the same space to localize the referent at its introduction and reintroduction), gesture incongruent (i.e., the speaker uses a different space to localize the referent at its reintroduction), and no gesture.

3.3 Stimulus materials

3.3.1 Production studies

The corpus used for the production studies was based on retellings of a picture story created for the purposes of this thesis. The stimulus story is about three fairies each having to fulfil a task for which they decide to use magic. Figure 11 illustrate some of the scenes.

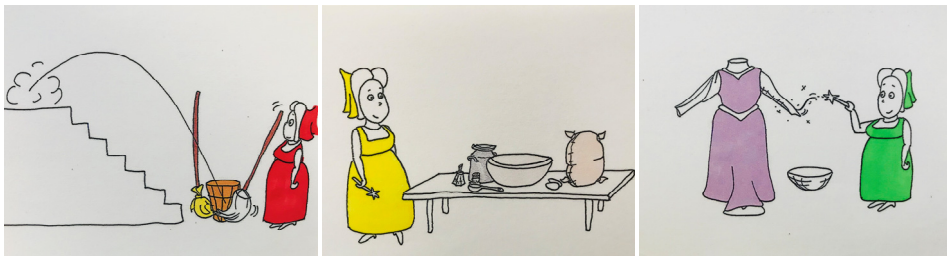


Figure 11: Examples of stimulus scenes in picture story

The story was built to include 36 intransitive and 36 transitive events, which were further controlled for agent animacy. Half of the (in)transitive events were carried out by animate entities and the other half was carried out by inanimate entities. The patients in transitive events were always inanimate. In addition, the events carried out by animate versus inanimate entities were always similar in structure (e.g., an event with an animate agent, such as 'a fairy takes out a wand', was always matched to a similar event with an inanimate agent, such as 'a needle takes out a bow'). The goal was to create items that would possibly generate both C-VPT and O-VPT gestures. C-VPT gestures typically co-occur with transitive events/clauses and O-VPT gestures with intransitive events/clauses (see e.g., Debreslioska et al. 2013; McNeill, 1992; Parrill, 2010). Furthermore, the variation in agent animacy controlled for the fact that C-VPT gestures are not solely produced when speakers are talking about and enacting animate/human agents. Finally, in order to make the story coherent, an additional 51 items (i.e., filler items) were added where needed.

In terms of information structure, the story was designed such that each agentive entity would be trackable. That is, each entity could be introduced and then referred back to at least two more times within the narrative (events) (see Appendix A for a script of the story).

3.3.2 Perception study

The stimulus material consisted of 50 videos of a female person telling short German narratives. She produced 20 narratives with spatially congruent and 20 narratives with spatially incongruent localizing gestures. In the remaining ten narratives, she produced no gestures. The speaker was trained to perform narratives and specifically the accompanying localizing gestures as naturally as possible. The narrator further kept the rest of her body as still as possible while speaking, kept the intonation of her speech as similar as possible, and spoke at a comparable speed across all narratives. While narrating, the speaker was sitting in a chair with no armrests in a room with plain, dark blue wall behind her.

All gestures were performed in central gesture space (coded as 'center right and left' in McNeill, 1992; cf. Gullberg & Kita, 2009), which corresponds to the culture-specific area for gesture production in German speakers (Müller, 1998). Figure 12 illustrates the gestures produced.

Figure 12: Examples of congruently and incongruently used localizing gestures in the stimulus material

The narratives had the same structure (13). The first utterance introduced the main protagonist with an indefinite lexical NP in an existential construction (e.g., 'There was a woman'). The second utterance was about a secondary character (e.g., 'husband'), who needs help carrying out a task. The third utterance then reintroduced the main protagonist with a lexical NP as grammatical subject (e.g., 'Then the woman…'), and it is explained how the protagonist intends to help the other character with the task. The fourth utterance stated whether the main protagonist *calls* or *writes* to someone for assistance. In experiment 1, paper II, participants had to respond to the action verb that was mentioned in the fourth utterance ('Did the protagonist call someone for help?'). In experiment 2, paper II, the participants had to press a button when they came across a mention of the protagonist, that is, at its introduction in the first utterance (e.g., 'a woman') and at its reintroduction in the third utterance (e.g., 'the woman'). A fifth and last utterance served as a wrap-up utterance.

(13)

Da war eine Frau$_1$*. Und ihr Mann konnte den Motor in seinem Auto nicht selbst reparieren. Also hat sich die Frau*$_2$ *dazu entschlossen, ihren Bruder anzurufen/anzuschreiben. Der soll ihm dann zur Hilfe kommen.*

'There was a woman$_1$. And her husband couldn't repair the engine of his car by himself. So, the woman$_2$ decided to call/write to her brother. He should come to help him out.'

$_1$ Gesture placed in right/left gesture space.

$_2$ Gesture placed in right/left gesture space.

Localizing gestures in the experimental items occurred in exact temporal alignment with the first and second referential expressions for the main protagonist. Gestures were performed with two hands (see Figure 12). In the gesture congruent condition, the first and second gestures were placed in the same location in space, half of the time to the right, the other half to the left. In the gesture incongruent condition, the second gesture was placed in the opposite locations in space, either left or right depending on the location of the first gesture.

3.4 Procedures and tasks

In all studies the participants were tested in a quiet room at their university.

For the production studies, participants came in pairs, whereby one participant acted as the speaker and the other as the addressee with roles distributed randomly. In cases where one of the participants was associated with the group hosting the data collection, that participant was automatically assigned the role of the addressee. Participants were offered light refreshments at the beginning and end of the session. The procedure lasted between 45- 90 minutes per participant pair.

For the perception studies, participants came alone and were compensated for their participation. In study 2 (which was carried out at Lund University), participants took part in a production experiment (not included in this thesis) after they finished the perception experiment. The procedure for the perception experiments lasted about 30 minutes in total.

In all studies, participants filled out a language and background information questionnaire (based on Gullberg & Indefrey, 2003), and provided written consent. They were debriefed orally.

3.4.1 Production studies

For the creation of the corpus, we used a guided elicitation task. Participants read their instructions on paper, but the experimenter repeated the most important points of the procedure orally. The speaker's task was to retell a picture story by answering the question 'what happened'. The speaker was further encouraged to say something about each picture. The experimenter always showed the speaker 4-6 pictures at once and the participant had as much time as they needed in order to memorize them. While the speaker was looking at the pictures, the addressee had to turn around and write down a summary of the part of the story that they had previously heard. When both participants were ready, they turned to face each other and the speaker retold the next part of the story. The speaker was filmed with a camera during the whole experiment. The camera captured the participant's head and torso. During the retelling, the addressee was not allowed to ask any questions, but backchanneling was not specifically discouraged and thus occurred naturally. The addressee was also asked to keep their arms on their upper thighs and not cross arms or legs. This procedure was chosen in order to minimize that speakers mirror a body position of the addressee that might be unfavorable for gesture production (Chartrand & Bargh, 1999; Kendon, 1973).

3.4.2 Perception study

For both experiments in paper II, the experimenter first orally introduced the experiment to the participant. The participant then read the instructions again in more specific form on paper.

3.4.2.1 Experiment 1

The clips were presented on a laptop running E-Prime software (version 2). The participants' task was to watch the videos of the narratives carefully and, for each narrative, respond to the question 'Did the main protagonist call someone for help?' as fast and accurately as possible by pressing the keys *j* for *ja* 'yes' in German or *f* for *falsch* 'no' in German on the keyboard. No explicit mention was made of the gesture information. The task implicitly probed the processing of information related to the referent. This task was chosen to avoid conscious and strategic processing of the gesture and its referent in speech (cf. Kelly, Creigh & Bartolotti, 2010). Participants were specifically encouraged to press the button as soon as they knew the answer and not to wait until the end of the video.

The correct answer was 'yes' for half of the narratives and 'no' for the other half (ending with the target verb 'write' instead of 'call', see Example 13). The instructions included an explanation that the main protagonist was always the first mentioned character, and that the narratives were about a problem that this protagonist had to solve. The instructions further contained three examples of narratives with corresponding correct responses and explanations.

3.4.2.2 Experiment 2

Participants carried out the experiment on a stationary computer in E-prime software (version 3). Before each clip, participants saw the target referent (e.g., 'girl', 'woman') written on the screen, indicating that this was the referent they had to track in the subsequent narrative. The instruction was to press the key *j* for *ja* 'yes' in German as fast as possible once they encountered the referent. We intentionally avoided using the word 'hear' in the instruction. For a third of the trials, a yes/no comprehension question appeared after the video clip. This question always related to details in the narratives. Participants responded to the questions by pressing the keys *j* for *ja* 'yes' in German or *f* for *falsch* 'no' in German on the computer keyboard. We added the comprehension questions to ensure that participants stayed focused on the content of the narratives.

3.5 Data treatment

In this section, I discuss the general methodological choices for data treatment. For the production data, I discuss the decision to use speech as a starting point to examine gesture and the implications this decision has for the initial annotation process. In a second step, I describe the annotation process of gestures in ELAN, in both perception and production studies. A special section is also dedicated to the detailed description of gesture phases in order to illustrate how gesture strokes and post-stroke holds are typically selected/identified. Finally, I will discuss speech-gesture alignment and why the current thesis works with exact temporal alignment between the modalities. The more specific coding of speech and gesture for each production study are detailed in the papers and will only be briefly touched upon here.

3.5.1 Speech as a starting point for the examination of gesture

The thesis asks the general question how gestures vary in terms of their incidence, form and content given a certain pattern used in speech. Therefore, the production studies consider speech as a starting point for the examination of gesture. From this point of view then, an important methodological choice concerned the delimitation of the search for relevant gestures depending on speech.

In papers I, III and IV, the focus is on gestures that accompany certain referential expressions in discourse. The first selection procedure was therefore to identify each relevant referential expression in the corpus independently of gesture and without access to the video. The clauses in which they appeared were annotated in video annotation software ELAN (Brugman & Russel, 2004; Sloetjes & Wittenburg, 2008), and the gestures that occurred within those clauses, and specifically co-occurred with the relevant referential expressions identified.

A note on the selection of the relevant clauses is in order because the procedure poses important challenges. First of all, speech production differs greatly from one person to another. This is the case, essentially, because language offers a multitude of ways to talk about the simplest events. The examples in (14) illustrate three ways of talking about 'a fairy going up the stairs'.

(14)

***a** und die gelbe Fee geht die Treppe hoch*

'and the yellow fairy goes up the stairs'

b *und währenddessen schleicht aber die Fee die Treppe hoch*

'and in the meantime, though, the fairy creeps up the stairs'

c *die sozusagen hochhüpft*

'who jumps up as it were'

Furthermore, speakers might repeat the event multiple times and in different ways, by using different verbs or constructions. Given these challenges, it is important to select target events in a principled way in order to retain objectivity in the data treatment and across all items and participants. I therefore selected the first complete mention of a target event.

3.5.2 Annotation of speech and gestures in ELAN

In all studies, I used ELAN software to analyze the alignment between speech and gestures. ELAN is a digital tool for creating annotations with different degrees of complexity on video or audio files (https://tla.mpi.nl/tools/tla-tools/elan/).

3.5.2.1 Production studies

For the production studies, speech was first transcribed verbatim in Microsoft Word and relevant speech segments were then transferred as annotations into ELAN. Gestures that were produced within these relevant stretches were also identified and further annotated in ELAN. During the gesture identification and annotation process, the sound of the speech was turned off. This procedure guaranteed an annotation process based on the articulatory properties of the gesture that is not influenced by or interpreted on the basis of speech. While this approach is not used by all gesture researchers and might not be relevant for all types of empirical questions, it is crucial for studies that examine the close relationship between speech and gestures, such as the ones in the current thesis.

3.5.2.2 Perception study

For the perception study, ELAN was used to annotate the stimulus material. That is, the two localizing gestures produced in each video clip were identified, and specifically their strokes phases (see section 3.5.2.3). The strokes phases had to align exactly with the relevant referential expression (e.g., 'a woman' and 'the woman', see 3.3.2) in order for an item to be used for the experiments.

In addition to the gesture coding, the onsets of the relevant action verbs (experiment 1; see 3.3.2) and the onsets of the anaphoric referential expressions (experiment 2) in each experimental item were also annotated. The onsets of the action verbs and the anaphoric referential expressions were used as reference points for the response time measures in experiment 1 and 2, respectively.

3.5.2.3 Gesture phases

Gesture movements have internal structure. A gesture is typically made up of a preparation phase, a stroke phase and a recovery (Kendon, 2004; McNeill, 1992). The stroke phase is defined as the most effortful part of the gesture, or the phase in which the 'expression' of the gesture is fully accomplished (Kendon, 2004: 112). The preparation phase leads up to the stroke phase, while the recovery phase, in which the hands are withdrawn or relax, typically follows the stroke phase. It is important to note that a gesture is only considered a gesture if there was a stroke phase. Preparation and recovery phases can sometimes be omitted. Kita (1990) has further identified an optional phase that may occur in between the stroke and recovery phases, which he calls the 'post-stroke hold'. The post-stroke hold refers to a brief suspension of the articulators at the end of a stroke, which seems to function as extension of the stroke's meaning (Kendon, 2004) and/or as a way to allow for the co-expressive part of the speech to be articulated before going into recovery (or to the next gesture). This is why Kendon (2004) considers the stroke and post-stroke hold to form the 'nucleus' of the gesture together (p. 112). In a similar fashion, a preparation phase can also be followed by a pre-stroke hold (Kita, van Gijn & van der Hulst, 1998). In this phase the hand is in stroke initial position but is held there for a short moment. Kita (1990) proposed that the function of such a hold is to allow for the stroke to achieve temporal co-ordination with the co-expressive part in speech. A schematic representation of the gesture phases is illustrated in Figure 13.

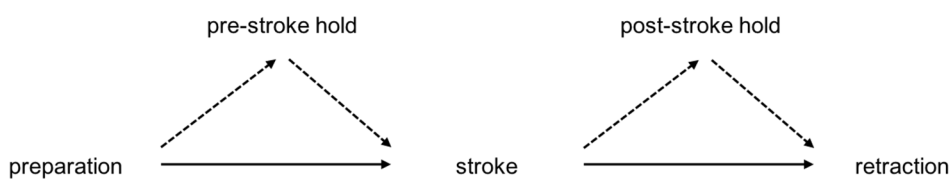


Figure 13. Schema of gesture phases (adapted from Seyfeddinipur, 2006)

Figure 14 illustrates a gesture stroke + post-stroke hold phase (i.e. the 'nucleus' of a gesture) in a concrete example. In picture (a) the speaker has placed his hands into a stroke initial position (i.e., during a preparation phase). His hands then move from (a) until (c), while (b) exemplifies a position within the stroke. The speaker is representing

the path of some egg shells landing on a table next to a bowl. After the stroke phase, the speaker performs a post-stroke hold. This is to say, he holds his hands in position (c) until the rest of the clause (in this case the expression *daneben* 'there besides') is uttered.

The beginning and end of gesture strokes can be identified by changes in shape or tension of the hand(s), direction of the movement and placement/orientation of the hands (aspects that are also considered to define signs in languages of the Deaf; i.a., Cormier & Fenlon, 2014; Engberg-Pedersen, 1993; Stokoe, 1960).

a b c

Die Schalen landen daneben

'the shells land there besides'

Figure 14. Example of gesture phases

3.5.3 Speech-gesture alignment

After the gestures were identified, the audio was turned back on and it was determined which gesture strokes temporally aligned with the relevant referential expressions. In the gesture literature, there are two ways in which the alignment of speech and gestures has been considered. While some studies choose to establish the exact temporal alignment between gestures and speech, others tend to examine clause level overlap (i.e., a gesture is annotated as co-occurring with a certain clause rather than the exact part[s] of word[s]). In studies of exact temporal alignment, it is typically assumed that the gesture is related to the stretch of speech that it temporally aligns with (e.g., Gullberg, 2006; McNeill, 1992; Stam, 2006). Whereas in studies of clausal co-occurrence, the gestures that are identified within a clause are typically further interpreted for their meaning in order to determine, which part of the clause the gesture is related to (e.g., So, Demir & Goldin-Meadow, 2010; So et al. 2009). In the current thesis, the first methodological approach was chosen in the interest of replicability, and objectivity. The gesture stroke had to span over at least one syllable of the referential

expression in order to be considered for analysis (Hickmann et al. 2011; Stam, 2006). Importantly, for each of the three production studies, a second coder identified 20-25% of all gestures that were used in the analyses to establish further reliability.

3.5.4 Further coding and reliability

3.5.4.1 Speech Coding

For the production studies, speech was coded for information structural dimensions, namely richness of expression ('the fairy' vs. 'she'), nominal definiteness ('a fairy' vs. 'the fairy'), grammatical role ('the fairy' in 'the fairy came in' vs. in 'she saw the fairy'), and clause structures that are more or less specialized for the introduction of referents ('there was a fairy' vs. 'a fairy came in'). In addition, all production studies included measures of accessibility, as in the difference between brand-new and inferable referents in papers I and IV, and referential distance in paper III. The categories are explained in detail in each of the papers accompanied by a variety of examples.

To assess the reliability of the coding, a second German native speaker recoded 20-25% of all referential expressions and for all dimensions considered in each of the papers. In cases of disagreement, the original coding by the author was kept.

3.5.4.2 Gesture Coding

Gestures were also further coded along the dimensions of gesture viewpoint, and semantic content in papers III and IV. More specifically, gestures were coded as C-VPT versus O-VPT gestures, and as gestures focusing on entity versus action (see Figures 5-9). Both of these coding procedures were performed with no access to sound in order to avoid any influence from the speech stream. However, for the coding of entity versus action gestures, the coder had the stimulus pictures at their disposal in order to reduce the semantic search space.

To assess the reliability of the coding, a second coder recoded 20-25% of all the gestures that went into the analysis. In cases of disagreement, the original coding by the author was kept.

4 Results

4.1 Paper I

Paper I examined *when* first mentions of discourse referents were accompanied by gestures. Previous research on the incidence of gestures for the representation of discourse referents has shown that speakers tend to use gestures with new referents rather than with given ones (Foraker, 2011; Gullberg, 1998, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill & Levy, 1993). This observation is supported by McNeill's (1992) theory of communicative dynamism and gestures, which states that the more a piece of information pushes the communication forward, the more likely a gesture will accompany it. Information status or accessibility is thus vital in that new or less accessible referents are more likely to occur with gestures than given or more accessible referents. However, some studies have shown that not all first mentions, representing new information, are in fact accompanied by gestures (Foraker, 2011; Gullberg, 2003). Paper I examined the variation in the incidence of gestures in the context of first mentions. Specifically, it hypothesized that the inferability of referents of first mentions (brand-new vs. inferable; Prince, 1981, 1992), paired with their expression as indefinite versus definite nominals may play a role. The predictions were that brand-new referents expressed as indefinite nominals would attract more gestures than inferable referents expressed as definite nominals.

A corpus of 20 bimodal narratives was analyzed. First mentions were identified and coded for information status and nominal definiteness. All gestures aligning with the referential expressions were identified. The data set consisted of 1,489 spoken referential expressions and 811 gestures. The speech results showed that, as expected, brand-new referents were more likely to be expressed by indefinite nominals and inferable referents were more likely to be expressed by definite nominals. The gesture results showed that speakers produced gestures for about half of all first mentions. Unexpectedly, the results further showed that inferable referents were more likely to be accompanied by gestures than brand-new referents.

4.2 Paper II

Paper II started from the observation in production studies that speakers can use gestures to create visual anaphoricity (Gullberg, 2003, 2006; McNeill & Levy, 1993). That is, speakers can produce localizing gestures to associate a referent with a certain location in space at their introduction and then gesturally refer back to the location to reactivate the referent at their reintroduction. Paper II asked whether addressees were sensitive to such localizing anaphoric gestures in a narrative context. In contrast to previous perception studies on this topic (e.g., Nappa & Arnold, 2014; Sekine & Kita, 2015), paper II used a design in which only one referent was gesturally tracked, and in which the typical alignment between speech and gestures found in spontaneous production of localizing (anaphoric) gestures was closely matched. Paper II presented two reaction time experiments testing whether addressees were faster to recognize or answer a question about a gesturally tracked referent when gestures were spatially congruent than when they were not. Three conditions were compared: gesture congruent (i.e., localizing gesture at introduction + congruent (i.e., anaphoric) localizing gesture at reintroduction), gesture incongruent (i.e., localizing gesture at introduction + incongruent localizing gesture at reintroduction) or no gesture. The hypothesis was that localizing anaphoric gestures would facilitate addressees' discourse processing. The predictions were that 1) participants would respond faster in the congruent condition than in the incongruent condition; 2) that participants would respond faster in the congruent condition than in the no gesture condition; and 3) that participants would be slower to respond in the incongruent condition than in the no gesture condition. The responses from 24 and 27 native German speakers were analyzed in experiment 1 and experiment 2, respectively.

The results showed that, contrary to predictions, participants responded faster in the gesture incongruent condition than in the gesture congruent condition (experiment 1), and also faster in the gesture incongruent than in the no gesture condition (experiment 2). This is to say that in both experiments, in the context of a single gesturally tracked referent, the presence of localizing anaphoric gestures did not speed up participants' performance. Rather, the results surprisingly suggest that there is an advantage in terms of processing speed in the incongruent conditions.

4.3 Paper III

Paper III addressed two main questions. It examined *when* gestures were produced with discourse referents, and if so, *how* gestures were produced. The paper took the proposed scale of gesture progression by McNeill (1992) as its point of departure. McNeill (1992) hypothesized that the incidence of gestures (*when*) paired with gesture viewpoint (*how*)

form a scale which is driven by communicative dynamism (see 2.2.3.3). One measure of communicative dynamism is the accessibility of discourse referents. McNeill (1992) therefore proposed to align his scale of gesture progression with Givón's (1983) scale of linguistic quantity for referential expressions in order to test its validity. Givón's (1983) scale of linguistic quantity orders referential expressions according to their phonological size (or richness of expression), which in turn signals referent accessibility. The first goal of paper III was to test whether there was any evidence for such alignment across scales, that is, 1) a progression from the use of pronouns to lexical NPs corresponding to a progression from the absence of gestures to the presence of gestures; and 2) a progression from the use of unmodified lexical NPs to the use of modified lexical NPs corresponding to a progression from the use of O-VPT gestures to C-VPT gestures. However, since the variation in richness of expression is only one aspect of the accessibility of referents, paper III also examined nominal definiteness and grammatical role assignment in relation to gesture incidence and viewpoint.

A corpus of 20 bimodal narratives was analyzed, targeting clauses which either referred to intransitive or transitive events (36 of each, 72 events in total). All gestures produced within those clauses and temporally aligned with a relevant referential expression were identified and annotated for gesture viewpoint (C-VPT vs. O-VPT). Only referential expressions produced as core arguments were further analyzed (i.e., [in]transitive subjects and transitive objects). The analyses were based on 1,237 referential expressions and 583 gestures.

The results in paper III can be summarized in three points for each question (i.e., *when* and *how*). Starting with the incidence of gestures (i.e. the question of *when*), we found that gesture incidence paralleled richness of expression. That is, gestures tended to accompany lexical NPs, but not pronouns. Second, gesture incidence paralleled nominal definiteness. That is, gestures were more likely to accompany indefinite than definite nominals. Third, gesture incidence paralleled grammatical roles. That is gestures were more likely to accompany direct objects than subjects (in transitive clauses).

Turning to gesture viewpoint (i.e. the question of *how*), we found that gesture viewpoint did not parallel richness of expression. That is, there was no evidence for a progression from O-VPT to C-VPT with unmodified to modified lexical NPs. Second, gesture viewpoint paralleled nominal definiteness. That is, C-VPT gestures tended to accompany indefinite rather than definite nominals. Third, gesture viewpoint paralleled grammatical roles. That is, C-VPT gestures were more likely to align with direct objects + verbs rather than subjects + verbs (in transitive clauses).

4.4 Paper IV

Paper IV investigated *what* semantic content gestures expressed when they accompanied first mentioned referents in discourse. Previous research has suggested that the structuring of information in discourse plays a key role for the variation in what information gestures represent, highlighting the influence of referents' information status (new vs. given), and nominal definiteness (definite vs. indefinite nominals; Debreslioska et al., 2013; Foraker, 2011; Gerwing & Bavelas, 2004; McNeill, 1992; McNeill & Levy, 1993; Parrill, 2012; Wilkin & Holler, 2011). However, most studies have focused on the referential expressions themselves without taking into account the clausal context in which they are embedded. Interestingly, other studies, not targeting discourse, have suggested a close connection between the structure of a clause and what information the gestures occurring in that clause represent (e.g., Özyürek et al. 2005). Therefore, paper IV examined the clausal context in which a referent was mentioned in addition to information status and nominal definiteness. The context of referent introductions is a good testing ground since first mentions vary in important ways for the three discursive aspects under consideration. First of all, they vary in information status, as in brand-new versus inferable (Prince, 1981, 1992; see also e.g., Chafe, 1994, 1996; Givón, 1995). Second, first mentions vary in nominal definiteness, such that brand-new referents are typically marked with indefinite lexical NPs, whereas inferable referents are marked with definite lexical NPs (e.g., H. Clark, 1977; Ward & Birner, 2001). Finally, the clauses in which first mentions are embedded vary in their syntactic and lexical structure, as in more specialized clause structures for the introduction of referents (e.g., *es gibt eine Fee/da ist eine Fee* 'there is a fairy') versus less specialized structures (e.g., *eine Fee* 'a fairy' or *ein Gürtel* 'a belt' in *dann kam eine Fee an* 'then came a fairy' and *und näht einen Gürtel* 'and sews on a belt').

Paper IV drew on a corpus of 20 bimodal narratives. All first mentions expressed as core arguments were identified, coded for their information status, and nominal definiteness. All gestures aligning with the referential expressions were identified and coded for semantic content (i.e., expressing mainly entity information or action information, see also 2.2.2.3). The analyses were performed on 462 gestures that aligned with first mentions.

The results showed that gestures focusing on entity information tended to accompany first mentions in more specialized clause structures, whereas gestures focusing on action tended to accompany first mentions in less specialized clause structures. Furthermore, entity gestures tended to co-occur with brand-new referents expressed by indefinite nominals, whereas action gestures tended to co-occur with inferable referents expressed by definite nominals.

5 Discussion

5.1 *When*?

The question of *when* speakers use gestures to accompany discourse referents was examined in papers I and III. The results in both papers suggest that the incidence of gestures varies with the accessibility of referents. However, they also differ in that they highlight two different functions for gestures, namely a complementary function in paper I, and a parallel function in paper III. The two papers are discussed in turn.

In the context of first mentions, Paper I showed that inferable referents were more likely to be accompanied by gestures than brand-new referents. Importantly, the results are generally in line with previous research since they show a link between the incidence of gestures and referent accessibility (Foraker, 2011; Gullberg, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill & Levy, 1993). However, they question the predictions derived from McNeill's (1992) theory of communicative dynamism and gestures.

According to McNeill's theory, the more a piece of information pushes the communication forward, the more likely it is to be accompanied by a gesture. McNeill (1992) has further suggested that referents' accessibility in discourse plays a crucial role for communicative dynamism (see also Firbas, 1971), such that less accessible referents would attract more gestures than more accessible referents. Thus, since brand-new referents are considered new to the hearer and new to the discourse, whereas inferable referents are considered to be new to the hearer but inferentially accessible through the discourse (Birner, 2013; Gundel, 1996), it seemed plausible to assume that brand-new referents would attract more gestures than inferable referents. However, this was not the case. Rather, the results suggest that gestures are used to indicate/enhance the accessibility of referents that need to be recovered inferentially.

This interpretation departs from previous research on this topic. It emphasizes a complementary rather than a parallel function of gestures to speech. Specifically, by highlighting the inferable pieces of information, gestures signal to the addressee that, even if the referents are marked by definite determiners, they should still add them as new information to the current discourse representation. Typically, inferable entities are linguistically encoded similarly to given information (by definite nominals), but by producing gestures more often with inferable referents, speakers can signal to the

addressee that the information is new rather than given, since there is not yet any active representation of the information in the discourse representation. That is, when speech does not provide an unambiguous clue as to whether information needs to be newly added to the current discourse representation (such as by indefinite nominals), gestures can do so instead.

This interpretation of a complementary function of gestures to speech is supported by McNeill's (1992) general view of gestures and speech as two dimensions of the same idea unit, whereby the two modalities do not always represent the same information. Rather, they come together to form a more complete representation. Kendon (2014) has also argued that speech and gestures together form a richer and more complex representation than if each of the modalities were considered alone.

Interestingly, the results can further be related to qualitative observations in children's speech production (Allen, 2008) and neurological studies on German discourse (Burkhardt, 2006). Allen (2008) examined the discourse-pragmatic influences on children's argument realization in Inuktitut, a pro-drop language. She found that 'newness' is one crucial factor determining whether children realize an argument overtly. However, she also observed that there are many arguments which are not expressed overtly even if new. A qualitative analysis of some of these cases showed that when referents represented new, but inferable information[4], children complemented the non-overt/dropped argument with gesture, and specifically with a deictic gesture pointing to the referent that the child wanted to refer to. This is similar to the findings in Paper I. Allen's (2008) examples also suggest that when new but inferable information is linguistically treated similarly to given information (i.e., non-overt or by a definite determiner), a gesture might indicate the referent's accessibility instead.

Moreover, in an ERP study comparing brain responses to new, given and inferable referents in German, Burkhardt (2006) found that references to these three types of referents are processed differently in the brain. Importantly, the property of inferable referents as lying in between new and given referents in terms of their accessibility, is reflected in brain responses. In their experiment, inferable referents first patterned similarly with given referents and then with new referents, forming a third type of pattern. In light of this, it seems plausible that, in language production, speakers would also treat inferable referents differently from brand-new referents.

In contrast to paper I, paper III highlights the parallel function of speech and gestures when representing discourse referents. The results suggest a relationship between three aspects of referent accessibility, namely richness of expressions, definiteness and grammatical role, and the incidence of gestures. The first finding was that lexical NPs (or richer referential expressions) tended to be accompanied by gestures significantly

[4] Allen (2008: 200) gives the example of the mentioning of food eaten with a fork when the fork was already mentioned.

more often than pronouns (or leaner referential expressions). This result replicates previous studies highlighting the importance of richness of expression in relation to the use of gestures (Gullberg, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson et al., 1982). The general interpretation of this finding is that the two modalities combine in their marking material in order to signal new/less accessible information in contrast to given/more accessible information. That is, rich referential expressions are typically used to express referents that are less accessible and need to be (re)activated. Leaner referential expressions are typically used to express referents that are more accessible, and which are assumed to be in the focus of attention at the moment of their mention. Gestures parallel this pattern by accompanying rich referential expression, adding marking material in order to contribute to the (re)activation process, but do not accompany lean referential expressions which signal that the referents are more easily recoverable. Levy and McNeill (1992) proposed that this is how speech and gesture together reflect moments of high or low communicative dynamism (see also Levy & Fowler, 2000).

The second finding concerning the incidence of gestures in Paper III was that indefinite nominals were more likely to occur with gestures than definite nominals. The result is commensurate with the view that the incidence of gestures reflects referent accessibility (Levy & McNeill, 1992; McNeill, 1992, 2005). That is, indefinite nominals are typically used to refer to new referents, whereas definite nominals are used to refer to given referents. The way that gestures pattern in relation to nominal definiteness reflects this difference by accompanying new referents more than given referents. Importantly, this finding further challenges results in the literature suggesting that richness of expression has an important influence on gesture when it is not linked to accessibility. For instance, previous research on second language learners has shown that over-specification in speech leads to over-specification in gesture (Gullberg, 2006). That is, second language speakers often use rich referential expressions in contexts where leaner expressions would have sufficed (i.e., for accessible/maintained referents). In such cases, speakers also produce gestures to accompany these expressions, suggesting that gestures might be sensitive to richness of expression even in the absence of a link to less accessible referents. However, in the case of indefinite versus definite nominals, both are rich referential expressions which vary with the incidence of gestures and therefore challenge these findings. Finally, the result stands in contrast to the only previous study that has also studied nominal definiteness and gesture production quantitatively. Wilkin and Holler (2011) did not find a difference between indefinite and definite nominals for the incidence of gestures. Possible reasons for the discrepancy between the results in paper III and Wilkin and Holler's study are the size of the data set, the kind of referential expressions taken into consideration (paper III only examined core arguments, whereas Wilkin and Holler included oblique arguments as well), or the language studied (German vs. English).

The last finding on the incidence of gestures in Paper III was that transitive objects were more likely to be accompanied by gestures than transitive subjects. Again, this result is in line with the view that the incidence of gestures reflects referent accessibility. Transitive objects typically represent less accessible referents and attract more gestures whereas transitive subjects typically represent more accessible referents and attract fewer gestures. Previous research has predominantly focused on the study of subject referents and their co-occurrence with gestures (e.g., Gullberg, 2006; Perniss & Özyürek, 2015; Yoshioka, 2008). Paper III thus adds to previous research by specifically considering other grammatical roles as well.

In summary, papers I and III shed new light on the question of *when* gestures are produced with discourse referents by showing that the incidence of gestures varies with (markers of) referent accessibility. Most importantly, they highlight the multifunctionality of gestures. While gestures are used in a complementary fashion in the context of first mentions (contrasting brand-new vs. inferable referents), they are used in parallel to speech in the context of first *and* subsequent mentions (contrasting new vs. given referents). This suggests that speakers can adapt their gestures with speech in a flexible way depending on the context and the corresponding communicative needs (Gullberg, 1998; Holler & Beattie, 2003; Kendon, 2004).

5.2 *Where*?

The question of *where* in space gestures are produced and whether the anaphoric use of space has any influence on the addressee was examined in paper II. The findings from two reaction time experiments showed that, contrary to expectations, there was an advantage of incongruent conditions. In other words, participants showed a processing advantage, as reflected in faster response times, in the incongruent condition in both experiments (compared to the congruent condition in experiment 1, and compared to the no gesture condition in experiment 2). Importantly, in experiment 2, faster response times in the incongruent condition compared to the no gesture baseline condition point to a facilitation effect. Therefore, addressees seem to be sensitive to or profit from the presence of localizing gestures, but not necessarily to or from their congruence. This finding stands in contrast to previous reaction time experiments on localizing gestures which found no difference between the incongruent and no gesture conditions (Nappa & Arnold, 2014) or that addressees were slower in the incongruent than in the no gesture condition (Sekine & Kita, 2017).

The suggestion made in paper II is that, in a context in which only one discourse referent is gesturally tracked, the presence of gestures is more important than their spatial congruence. Previous studies have predominantly examined the gestural tracking of two referents, each one localized on opposite sides in gesture space, thus creating a

contrast between referents and their associated locations (e.g., Nappa & Arnold, 2014; Sekine & Kita, 2017; Gunter & Weinbrenner, 2017). It is possible that in a contrastive context, the use of congruent localizing gestures leads to a facilitation effect (Nappa & Arnold, 2014), or that the use of incongruent localizing gestures leads to a hindrance effect (Sekine & Kita, 2017). However, in the context of a single gesturally tracked referent, (congruent) location information becomes less relevant, while the presence of gesture becomes more significant. Specifically, in previous designs with two tracked referents, the incongruent location for one referent typically represented the congruent location for the other referent. In contrast, in paper II the incongruent location was an unassigned location, and so presumably had no meaning. An incongruent location with no meaning might be easier for addressees to accept as a second location for a tracked referent than an incongruent and previously assigned location. This interpretation is supported by production studies, which have shown that speakers reuse an assigned location for a referent less than half of the time (35% in So et al. 2009, and 42% in Gullberg, 2006). Paper II (Figure 5) provides an example from spontaneous narrative discourse, in which a speaker introduces a referent by localizing it in their left gesture space and then uses an incongruent localizing gesture in their right gesture space for the referent at its next reintroduction. Since gestures by definition have to be produced somewhere in space, the use of spatially non-cohesive locations may be frequent and only become difficult to process when contrast or disambiguation come into play.

Finally, two secondary interpretations for the findings are proposed which open for testable predictions. First, gestures in the congruent condition may have been perceived as overexplicit. In speech, overexplicitness refers to the use of lexical NPs in contexts in which pronouns would have sufficed. Gordon et al. (1993) showed that such overexplicit use of lexical NPs leads to longer processing times in speech perception. This effect is called the *repeated noun phrase penalty*. We suggest that the use of localizing gestures in the congruent condition may also have been overexplicit. This would explain why participants needed more (or as much) processing time to integrate congruent gestures with spoken anaphoric expressions than incongruent gestures. Such a view receives some support from production studies. For instance, Marslen-Wilson et al. (1982) showed that localizing anaphoric gestures are typically used at the beginning of an episode (or at an episode boundary) rather than within single episodes. In the stimulus material in paper II, anaphoric gestures were used within an episode (with only one intervening utterance between the introduction and reintroduction of the referent in subject position). Thus, it is possible that participants did not expect a congruent/anaphoric gesture with the anaphoric expression within-episode, and therefore perceived them as overexplicit, causing longer processing times. This introduces a *repeated gesture penalty* hypothesis which can and should be tested.

The second alternative interpretation is that participants did not interpret the second gestures in the experiments anaphorically, as referring back to referents. Rather, they may have interpreted them as referring to a new event (e.g., 'the woman calling her

brother', see 3.1.2, Example 13). One reason for such an interpretation is that the second gesture appeared close to the discourse marker 'so', signalling an event shift. Alternatively, it is possible that addressees may not have tracked the exact temporal alignment of the second gesture with the mention of the main protagonist (e.g., 'the woman). Given the short time lag between the NPs 'the woman' and 'her brother' (three words in the German clauses), they may have assigned the gesture to the second NP (i.e., 'brother'). Further studies varying the alignment of gestures with different parts of the clauses are needed in order to establish the plausibility of such interpretations.

On a final note, it is important to acknowledge that the experiments were designed to test a parallel function of gestures, and specifically in the congruent condition. The rationale of the congruent condition as representing the parallel function of gestures can be explained with the notion of 'catchments' (McNeill & Levy, 1993; McNeill et al., 2001). A catchment is defined as two or more recurring gesture features that occur in at least two gestures (e.g., spatial location, handedness and form in paper II). In that case, the gestures are typically found to indicate the same 'image' or idea. This is similar to anaphoric expressions in speech that refer back to a previously mentioned referent, and accordingly also indicate the same idea (of a referent). However, since the congruent condition in paper II, which putatively represented the parallel function of gestures, did not reveal a facilitatory function, one possible conclusion is that the same functions are not equally important in gesture production as in gesture perception, at least in the context of discourse reference. More interestingly though, it also seems possible that the rationale behind a parallel function of gestures as defined by catchments is not applicable in every context. Rather, speakers might use both - the 'simple' incidence of gestures or certain aspects of their form - in order to create cohesion. It is thus possible that, in a context of a single tracked referent, the incidence of gestures is what constitutes the parallelism to speech and helps in the creation of cohesion. Importantly, it is evident that more ecologically valid perception studies are needed, specifically matching production processes in discourse, in order to gain a better understanding of the functions of gestures for addressees.

5.3 *How*?

The question of *how* discourse referents are represented in gesture was addressed by examining gesture viewpoint in paper III. The main findings suggest that there is a relationship between nominal definiteness and gesture viewpoint on the one hand, and between grammatical role and gesture viewpoint on the other hand.

Specifically, the results showed that indefinite nominals were more likely than definite nominals to be accompanied by C-VPT gestures. Conversely, definite nominals were more likely than indefinite nominals to be accompanied by O-VPT gestures. This finding supports McNeill's (1992) suggestion that gesture viewpoint is sensitive to the accessibility of referents. It is also compatible with McNeill's (1992) scale of gesture progression which places C-VPT gestures lower on the scale than O-VPT gestures, suggesting that C-VPT gestures represent less accessible information (see Figure 10 in 2.2.3.3). In other words, indefinite nominals typically refer to referents that are new whereas definite nominals typically refer to referents that are more accessible. The corresponding variation in gesture viewpoint suggests that C-VPT gestures link to new/less accessible information, and O-VPT gestures link to more accessible information.

More generally, the results are also in accordance with studies showing that *how* referents are represented in gesture is dependent on their discourse status (Gerwing & Bavelas, 2004; Parrill, 2012). Gerwing and Bavelas (2004) showed that speakers produced gestures that were bigger, more precise and combined more than one gestural movement when representing new referents than when representing given ones. Parrill (2012) reported that speakers produced significantly more C-VPT gestures when telling a story that their addressees had no previous knowledge of (thus, containing more new information). Conversely, speakers produced more O-VPT gestures when telling a story that addressees *had* previous knowledge about (thus, containing more given information). The result on the link between nominal definiteness and gesture viewpoint complements those findings.

Turning to grammatical role assignment, we found that C-VPT gestures tended to align with transitive objects + verbs, whereas O-VPT gestures tended to align with transitive subjects + verbs. This finding is also in line with McNeill's (1992) gesture scale. Transitive objects are typically less accessible and link to C-VPT gestures, whereas transitive subjects are typically more accessible and link to O-VPT gestures. While this result is very fine-grained and specific to the context of transitive clauses, it complements the findings in Debreslioska et al. (2013) who showed that O-VPT gestures are more frequent in intransitive events and C-VPT gestures in transitive events. However, it also raises important questions about speech-gesture alignment and semantic cross-modal coherence. In fact, Debreslioska et al. (2013) found that O-VPT gestures specifically link to intransitive subjects, whereas C-VPT gestures link to

transitive subjects. At first sight, this result seems to contradict the results in paper III (together with McNeill's scale of gesture progression), since transitive subjects are typically considered to be more accessible than both intransitive subjects *and* transitive objects (e.g., Du Bois, 1987). However, the two studies used different methodologies to analyze speech-gesture co-occurrence. While paper III examined exact temporal alignment between speech and gestures, Debreslioska et al. (2013) considered clause level overlap and added a meaning interpretation to each gesture (i.e., determining whether or not the gesture related to the referent in subject position and/or its actions/movements). Therefore, it is possible that the C-VPT gestures produced in transitive events did not necessarily align with the subjects but rather with other parts of the clause, such as the verb and the transitive object (just like in paper III). This interpretation is reinforced by the fact that most transitive subjects in Debreslioska et al. (2013) were expressed by zero anaphora (about 80%). Zero anaphora are empty slots and it is therefore not possible for a gesture to align with them. In light of this, the results between Debreslioska et al. (2013) and paper III do not necessarily contradict but may rather support each other.

Interestingly, the difference in methodological approach highlights an additional important issue when it comes to the interpretation of semantic coherence between speech and gesture. In studies on exact temporal alignment (to which paper III belongs), it is typically assumed that the gesture is related in meaning to the part of speech that it aligns with. However, it is also generally acknowledged that gestures can map to speech at a conceptual level. This means that gestures might express meanings on a more global level, as in representing the meaning of the whole clause rather than only the words that they exactly align with (e.g., De Ruiter, 2000; McNeill & Levy, 1982; cf. Paper IV). In studies on clause level overlap (e.g., Debreslioska et al. 2013), the gestures occurring in certain clauses are typically annotated for their semantic meaning in order to determine which semantic elements in the clause they are related to. However, this approach can also be problematic for the same reasons. If the goal is to decide whether a gesture represents a specific referent in the clause, but the gesture relates to the whole clause (encompassing more than one referent and the verb in its representation), it is difficult to draw the line between which element of the clause the gesture relates to 'most'. As shown through the comparison between paper III and Debreslioska et al's. (2013) study, this difference in approach can lead to different interpretations of the same phenomenon. It thus seems that the field of gesture studies would generally profit from studies that combine both approaches (cf. e.g., Hickmann et al., 2011).

5.4 *What*?

The question of *what* information gestures express about discourse referents was addressed in paper IV. Paper IV focused on the context of first mentions and examined how the semantic content of gestures related to three variables, namely the inferability of referents, nominal definiteness, and the clause structure in which referents were mentioned. The results showed that gestures focusing on entity information were strongly related to brand-new referents expressed as indefinite nominals. Conversely, gestures focusing on action information tended to be related to inferable referents expressed as definite nominals. Furthermore, gestures focusing on entity information tended to occur in clause structures that are more specialized for the introduction of referents. And conversely, gestures focusing on action information tended to occur in clause structures that are less specialized for the introduction of referents and more specialized for the expression of event information.

The findings suggest a link between the semantic content of gestures, the inferability of referents, and nominal definiteness. This is supported by previous research. Foraker (2011) similarly found that first mentions (or new referents) were likely to be accompanied by gestures focusing on entity information, whereas subsequent mentions (or given referents) were accompanied by gestures focusing on action information. Whereas paper IV focused on first mentions alone and used a more fine-grained measure of information status (brand-new vs. inferable), the results point in the same direction, suggesting that gestures focusing on entity information are linked to new(er) referents, whereas gestures focusing on action information are linked to (more) accessible referents. Similarly, Wilkin and Holler (2011) found that indefinite nominals were likely to be accompanied by gestures focusing on entity information, and definite nominals by gestures focusing on action information. The results in paper IV replicate these findings, but crucially extend them to the context of first mentions.

More importantly, paper IV adds new insight by showing that the clause structure in which a referential expression is embedded also influences what information gestures express about the referent. That is, when a more specialized clause structure focuses on the existence of a referent, then the gesture also focuses on entity information. Conversely, when a less specialized clause structure focuses on an event in which a referent is involved, then the gesture also focuses on action information. This contrast between more and less specialized clause structures is similar to the distinction between clauses that are in descriptive or in narrative mode, respectively (Du Bois, 1980). Clauses in the narrative mode advance the story whereas clauses in the descriptive mode typically rather describe entities, their location or relationships to other discourse entities (see also McNeill et al., 1993; McNeill & Levy, 1982; McNeill et al., 1990). Hence, representational gestures focusing on action (co-occurring with clauses in the narrative mode) seem to have the function of advancing the story. On the other hand,

representational gestures focusing on entity information (co-occurring with clauses in the descriptive mode) seem to have the function of describing characteristics of referents rather than advancing the story. This functional difference is compatible with the pattern that action gestures link to definite nominals/inferable referents, and entity gestures link to indefinite nominals/brand-new referents. That is, when referents are more easily inferable as indicated by definite nominals, the speaker can shift their focus to advancing the story by using an action gesture, whereas when referents are brand-new as indicated by indefinite nominals, the speaker is more likely to first indicate the properties of the referent before moving on to advancing the story.

More generally, these results tie in with studies suggesting that syntactic clause packaging and the semantics within a clause influence the information that gestures express (e.g., Fritz et al. 2019; Gullberg, 2011b; Özyürek et al. 2005).

Finally, the results complement the findings in paper III on gesture viewpoint. The analyses in paper III focused on the narrative level and showed that C-VPT gestures are related to new referents, and O-VPT gestures to given referents. Paper IV adds the important insight that those gestures (produced in C-VPT or O-VPT) are more likely to represent dynamic information about actions and movements, rather than static information about referent shape or location.

6 Conclusion and future work

6.1 Some conclusions on the functions of gestures in discourse

The thesis set out to examine the role of speech-associated gestures in the production and perception of connected discourse with a specific focus on referent representation. Four studies assessed the role that gestures play in the representation of discourse referents from different perspectives: *when* gestures are produced, *where* they are produced, *how* they are produced, and *what* they express. Together, the studies contribute to a deeper understanding of the close relationship between speech and gestures. Specifically, the results have highlighted that gestures co-construct connected discourse together with speech on different levels. Two main functions of gestures in the construction of discourse are identified, a parallel and a complementary function. For gesture perception, a facilitatory function is suggested.

The parallel function of speech and gestures is revealed by the incidence of gestures in relation to richness of expression. Speech and gestures work in parallel in terms of marking material. That is, the more marking material is used in speech to refer to a (new) discourse referent, the more likely a gesture is to occur as well. Similarly, the parallel function is revealed by gesture viewpoint or gesture 'complexity'. That is, the more 'complex' a gesture is, as in C-VPT instead of O-VPT, the more likely that a new(er) discourse referent is signaled. Finally, speech and gesture also work in parallel in terms of semantic focus. Gestures expressing entity information occur with brand-new referents expressed as indefinite nominals in clause structures specialized for the introduction of referents. Conversely, gestures expressing action information occur with inferable referents expressed as definite nominals in clause structures representing events. The parallel focus on entity versus action in speech and gesture add up to either describing referents or to moving the narrative forward.

These findings support previous research highlighting the parallel function of gestures in discourse (e.g., Gullberg, 2003, 2006; Levy, 1984; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill et al., 1990). It is worth mentioning that the studies in this thesis draw on a bigger data set than previous work. For instance, qualitative studies such as Marslen-Wilson et al. (1982) analyzed the narrative of one single speaker, and even quantitative studies have examined only four to ten speakers. In contrast, the

production studies in this thesis draw on narratives from 20 participants and analyzed 811, 538 and 462 gestures in papers I, III and IV. Therefore, the thesis adds substantial grounding to previous studies suggesting a parallel function of gestures. It should also be noted that studies on gesture are generally smaller in scope than studies that examine speech alone.

More importantly, most previous research that has highlighted the parallel function of gestures in discourse has particularly focused on gesture's relation to richness of expression in different discursive contexts (Gullberg, 2003, 2006; Levy, 1984; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill et al., 1990; Perniss & Özyürek, 2015). While part of the studies in this thesis replicate this focus (paper III, study 1), all production studies further add to this previous research in important ways. First, they take into consideration the link between gesture production and nominal definiteness, a relationship that has largely been neglected so far. Some previous studies have indirectly considered this relationship, as in the examination of the incidence of gestures with introduced versus reintroduced referents (e.g., Gullberg, 2003; Marslen-Wilson et al., 1982). Only one study has directly compared indefinite and definite nominals, but only in terms of gesture incidence and found no link between the variables (Wilkin & Holler, 2011). The thesis extends the findings from previous studies by showing that speakers *can* vary the incidence of gestures between introduced and reintroduced referents as measured by nominal definiteness. In addition, it shows that speakers also vary gesture viewpoint in parallel to the variation in nominal definiteness. It should be noted however, that the interpretation of a parallelism in the context of viewpoint is specifically tied to McNeill's (1992) proposition of a scale of gesture progression.

Furthermore, the studies in this thesis demonstrate that the parallel function of gestures can also come to light in the semantic content that gestures express. While this aspect has previously been examined in relation to discourse (Foraker, 2011; Wilkin & Holler, 2011), this thesis provides important novel insights. The results on the semantic focus of gestures are new as they highlight that discursive factors on the word level need to be complemented with discursive factors on the clause level to understand the variation in gesture. Only when both levels were taken into consideration was a parallelism between speech and gestures revealed. In fact, previous research has failed to provide a compelling explanation for the variation in semantic focus when referential expressions were considered in isolation (Foraker, 2011; Wilkin & Holler, 2011).

The complementary function of gestures is revealed by the finding that gestures are more likely to occur with (definite) referents that are inferable from the previous discourse than with (indefinite) referents that are brand-new. The results suggest a sophisticated mechanism whereby gestures can highlight actual information status when speech does not provide an unambiguous clue about the information status of a given referent. The finding presents a departure from previous studies which have

mostly suggested a parallel function of gestures in discourse reference. However, those studies do not contradict the current findings. Rather, together they highlight the multifunctionality of gestures in the context of discourse. In fact, gestures are a flexible and adaptable resource and can relate to speech in different ways. Kendon (1986) has suggested that gestures can be used as complements or supplements, sometimes even as substitutes or alternatives, to spoken expressions, as long as they are in accordance with the underlying communicative effort or intent. Similarly, McNeill (1992) has suggested that gestures and speech are two dimensions of the same idea unit, whereby gestures do not always represent the same information as speech. Rather speech and gesture together form a more complete representation. A complementary function of gestures is thus compatible with the descriptions of the general nature of the gestural modality. In the context of discourse, however, the gesture complementarity view is rather novel (but see So et al., 2010).

Finally, the facilitatory function of gestures in perception is revealed by the study of processing speed. Contrary to predictions, the results showed that the mere presence of localizing gestures facilitated discourse processing and referent tracking in particular, but that spatial anaphoricity did not matter. This finding stands in opposition to previous studies on anaphoric gestures in language comprehension, most of which show that spatial anaphoricity either plays an important role (e.g., Goodrich Smith & Hudson Kam, 2012; Nappa & Arnold, 2014; Sekine & Kita, 2017) or that the presence of gesture is irrelevant (e.g., experiment 1 in Gunter & Weinbrenner, 2017; Hudson Kam & Goodrich Smith, 2011). It must be noted, however, that the methodologies and designs used in previous research differs greatly from one study to the other, and importantly also in comparison to the study carried out for this thesis. The experiments in this thesis took into consideration the exact temporal alignment between referential expressions and localizing gestures in accordance with production studies on this topic. In addition, the current experiments focused on the tracking of one single referent whereas all previous comprehension experiments have worked with contrast/disambiguation between two referents. These differences in design suggest that the findings in this thesis do not necessarily contradict previous research, but rather complement it in an important way. Namely, it is possible that gestural information is used differently in contexts in which gestures track two different referents (Sekine & Kita, 2017), or in which gestures' function was specifically designed to disambiguate referents (Goodrich Smith & Hudson Kam, 2012; Nappa & Arnold, 2014). Whereas in such contexts spatial information might be relevant (but see experiment 1 in Gunter & Weinbrenner, 2017; Hudson Kam & Goodrich Smith, 2011), it becomes less significant in contexts in which a single referent is tracked. Rather, in the latter case, the presence of a gesture, independently of any spatial congruence with a previous gesture for the same referent, matters more. This interpretation is supported by evidence from production studies showing that in less than half of the time speakers reuse a location previously assigned to a referent at its next mention (Gullberg, 2006;

So et al. 2009). The interpretation is further in line with Gullberg's (2006) suggestion that the use of cohesive space might be related to speech production, planning and cognitive load.

Overall, the findings in this thesis have provided new evidence that speech and gestures work together to build a piece of connected discourse. They do so in a parallel but also in a complementary fashion. Furthermore, addressees seem to be sensitive to gestures in discourse perception. This suggests that gestures are a constitutive part of language in production and in perception and should therefore be taken into consideration in linguistic theories about discourse. However, very few theoretical proposals dealing with discourse and discourse referents mention gestures or the potential role they could play (but see e.g., Ariel's, 1991, hierarchy of accessibility markers, which mentions the possibility of a gesture combining with a stressed pronoun). The studies in this thesis suggest that studying gestures in combination with speech offers another source of evidence to shed new light on (discourse) patterns that have been found in speech. Importantly, the study of gestures can unveil how cohesion is achieved, even in cases where speech seems to offer limited resources.

In relation to gesture studies, the findings in this thesis highlight the importance of considering the structure of discourse when examining gestures in any context. Since gestures are *always* produced in accompaniment with a certain stretch of discourse, whether it is one clause or multiple connected clauses, the way that information is structured in that very stretch of discourse will necessarily play a role for *when*, *where*, and *how* gestures are produced, and for *what* information they express. For instance, if a study sets out to examine how a certain motion event is encoded in gesture, the gestures might differ depending on whether the verb(s), particles or other aspects of the clause constitute new or given information. This consideration becomes particularly essential for cross-linguistic comparisons (cf. Choi & Lantolf, 2008; Gullberg 2011b; Hickmann et al., 2011). Only an equivalent discourse context will allow for an assessment of the 'real' differences in the speech-gesture relationship between languages. In addition, the thesis emphasizes the relevance of examining different levels of coordination between speech and gestures. The studies considered exact temporal alignment, as well as word, clause and discourse level co-expressions, and showed that only through combining these different degrees of granularity, do we get a fuller picture of gesture functions, and of how the modalities work together. In summary, these observations suggest that gestures should always be examined together with their context in speech, and ideally by considering multiple levels of possible coordination with speech.

6.2 Future work

The studies in this thesis open new possible lines of inquiry. Some concern specific follow-ups of the current studies, others are more generally connected to the topic of the thesis.

Follow-up work from the production studies mainly concern more detailed considerations of referential expressions and different ways of assessing referent accessibility. First of all, for the referents being referred to, it would be useful to add measures of how conducive they are to gesture production in general, and to a certain gesture form or content in particular. Aspects that have been suggested to matter are afforded action (Chu & Kita, 2016), familiarity (Campisi & Özyürek, 2013), or the event structure referents are mentioned in (Parrill, 2010). For example, Chu and Kita (2016) showed that objects with a smooth surface are more likely to be gestured about than objects with a spiky surface. Campisi and Özyürek (2013) showed that speakers tend to use more, bigger and more informative gestures when teaching addressees new knowledge about an object that they are not or less familiar with (in this case children). Finally, Parrill (2010) has suggested that (in English) an event involving handling, an emotional state or the torso of an animate referent is typically represented in C-VPT rather than O-VPT. It is worth examining possible interactions between these aspects and the discursive variables considered in this thesis. Additionally, the referential expressions themselves could be considered from new perspectives. For instance, it might be useful to compare referential expressions taking on different types of semantic roles in order to check whether some of them might attract more gestures or gestures with a specific semantic content. Previous research has suggested that gestures vary in the way they represent referents depending on their semantic role, specifically contrasting agent versus patient roles (McNeill & Levy, 1982). However, we know little about whether other semantic roles might attract gestures using specific techniques of representation or with specific semantic content over another.

Turning to referent accessibility, we need new ways of assessing accessibility, specifically concerning the difference between brand-new and inferable referents. We have used a rather conservative measure in the studies of this thesis (i.e., a part/whole or content/container relationship between the first mention of a referent and the previously mentioned 'trigger' referent). However, there are much broader measures that could be applied. For instance, in the broader sense of a frame or script (e.g., Fillmore, 1982; Schank & Abelson, 1977), the inferability of the current mention of a referent could be triggered by a larger piece of preceding discourse describing a certain situation. It is worth exploring such relations as well. In addition, it would be useful to test participants on how inferable they perceive some referents to be in certain contexts before they carry out a narrative production task involving those referents. This would potentially allow us to test more directly how speakers treat inferable referents in speech

and in gesture. Conversely, perception studies should test whether and if so how addressees profit from gestures that indicate the information status of inferable referents in a complementary fashion to speech. This would clarify whether the increased use of gestures for inferable referents is specifically meant to enhance accessibility of a referent for the addressee.

Moving beyond the context of brand-new versus inferable referents which are low(er) in accessibility, we should also turn our attention to the other end of the accessibility scale and consider the variation of referring forms expressing higher accessibility, such as different types of pronouns. While it is the case that pronouns are less likely to be accompanied by a gesture than lexical NPs, it is also generally acknowledged that demonstrative pronouns, marking lower accessibility than personal pronouns (Ariel, 1991), are often accompanied by gestures (Cooperrider, 2011; Kendon, 2004; Kita, 2003; Wilkins, 2003). Furthermore, in some pro-drop languages overt pronouns seem to be accompanied by gestures more often than in non-pro-drop languages, possibly because they are more likely to mark information structural contexts, such as contrast (see Azar, Backus & Özyürek, 2018 for Turkish as an example and a discussion on this topic). Investigations into the use of gestures with pronouns will further clarify whether gestures follow a parallel or a complementary pattern in the context of more accessible information and/or whether they interact with other information structural dimensions, such as emphasis and highlighting.

Follow-ups to the perception experiment include four main avenues. First, the most obvious next step would be to compare the gestural tracking of one referent to the gestural tracking of two or more referents (in the same naturalistic setting and with similar tasks) in order to establish whether there is indeed a difference between the contexts of one versus two or more gesturally tracked referents. Another avenue is to test how closely addressees track gestures by varying the alignment with referential expressions versus with verbs versus with other parts of the utterances. Some previous research has shown that addressees are rather flexible when it comes to the start of gestures in relation to the onset of the congruent speech (Kirchhof, 2017). Therefore, it would be useful to combine an experiment in which the alignment of a referent tracking gesture is varied with a test of what spoken element addressees perceive the gesture to be semantically congruent with. A third extension is to test whether there is a difference when space is used in an abstract fashion versus when it is used topographically (i.e., when locations in gesture space are used as counterparts to physical locations in the [imagined] world). In fact, Emmorey, Corina and Bellugi (1995) found that American Sign Language users treat topographic locations differently from what in Sign Languages is called syntactic locations. Syntactic locations can be compared to the abstract use of space as we have designed it in the perception experiment in this thesis. It is possible that the function of localizing gestures might differ in these two contexts as well. Finally, we proposed the *gesture penalty hypothesis*, which will need to be further explored. According to this hypothesis, addressees'

language processing will be slowed down if gestures are used overexplicitly. However, in order to identify the contexts in which localizing anaphoric gestures might be perceived as overexplicit (independently of speech), more (sophisticated) production studies on the use of space in gestures are needed.

Finally, some more general lines of inquiry that should be explored include the study of gesture complexity, the study of cross-linguistic differences concerning the speech-gesture relationship in discourse, and the role of gestures in other strategies used to create cohesion in speech.

The study of gesture complexity is related to McNeill's (1992) proposition of a scale of gesture progression. McNeill suggested that C-VPT gestures are linked to less accessible referents in comparison to O-VPT gestures. Importantly, he further related this variation to gesture 'complexity'. By gesture 'complexity' McNeill (1992) seemed to understand the representation of more or less differentiated parts (of an event) in a gesture (p. 125). As example for a less complex O-VPT gesture, McNeill showed how a speaker was flexing their hand backwards in order to represent a character rising upwards, while saying 'he tries climbing up the rain barrel' (p. 108). As an example for a more complex C-VPT gesture, he showed how a speaker gripped something in the space right in front of their head, and then pulled it back down to their shoulder, while saying 'and he bends it way back' (p.12). McNeill (1992) explained that, while the O-VPT gesture in this example only represented the direction of the movement, the C-VPT gesture incorporated many more elements, such as the character's hand, the character's body, the shape of the hand, the trajectory that the hand followed, etc. (p. 125). In addition, McNeill (1992) suggested that gesture 'complexity' can also be defined by the physical properties of the gestural movements. Such physical properties could involve the use of one versus two hands, the additional movement of the fingers, a change in hand shape during the stroke phase, etc. (p.126). Importantly, however, the observations showing that C-VPT gestures are generally more 'complex' than O-VPT gestures are rather limited. They rest on qualitative examinations of a rather restricted data set. Besides those observations, there is no other systematic and/or quantitative evidence for a difference in complexity between C-VPT and O-VPT gestures. In fact, there are even proposals in the literature conflicting with McNeill's view. For instance, Debreslioska et al. (2013) proposed that O-VPT gestures can be considered more complex than C-VPT gestures. Particularly, that is because C-VPT gestures do not incorporate a ready means for referent differentiation whereas O-VPT gestures do. That is, while C-VPT gestures necessarily need to map each referent onto the same body in the same location (i.e., the speaker's), O-VPT gestures can use the gesture space in front of the speaker to differentiate between referents. Speakers can assign different locations in space to different referents and refer back to them at any given time. The contrast between the proposals of McNeill (1992) and Debreslioska et al. (2013) identifies a gap in our knowledge that needs to be addressed in the future. More studies are needed developing measures of complexity in gesture (for a possible

approach, see Hogrefe, Ziegler & Goldenberg, 2011). Moreover, larger corpora containing many varieties of viewpoint gestures need to be analyzed in order to get a better understanding of how the two main dimensions (C-VPT and O-VPT) might differ.

Turning to cross-linguistic differences, it is unclear whether discourse patterns for the production of gestures interact with any language specific patterns. Previous research has suggested that there are cross-linguistic differences when it comes to the way that gestures represent information. For instance, languages differ in the way that semantic content about motion events is expressed in gestures and also in the way that they align with the corresponding semantic content expressed in speech (e.g., Choi & Lantolf, 2008; Hickmann et al., 2011). However, only a few studies have examined discursive patterns in speech and gestures from a cross-linguistic perspective (but see Duncan, 1996; Gullberg, 1998; Yoshioka, 2005). Duncan (1996) revealed interesting differences between English and Chinese patterns of gesture use in the construction of discourse, and specifically as regards to *how* and *what* information gestures represent. First, she showed that English and Chinese speakers differ in preferences for C-VPT (English) and O-VPT (Chinese) gestures, and in their focus on action (English) or entity (Chinese). In relation to the question of *when* gestures are used, Duncan also showed a difference in semantic synchronization across the two languages. That is, English speakers tend to align gestures with the semantically coherent information in speech, whereas Chinese speakers instead tend to align gestures with information that precedes the semantically coherent element. These differences highlight that language specific tendencies may interact with general discourse patterns. More cross-linguistic studies are needed, specifically targeting the patterns found in this thesis, in order to evaluate the results further.

Finally, there are other possible elements that hold discourse together. For instance, on a more global level, the way that time and space are represented, introduced and maintained throughout a piece of discourse plays an important role for cohesion and coherence (Hendriks, 1993). On a local level, elements such as connectives are important for holding a piece of discourse together (Halliday & Hasan, 1976). No study has directly considered the dimensions of time and space in discourse from a gesture perspective so far. Also, very few studies have examined gestures' function in relation to connectives (but see Graziano, 2009). More studies moving away from the representation of discourse referents are needed to complement previous research and thus further our understanding of the role that gestures play in the construction of connected discourse.

In conclusion, the thesis contributes to our understanding of the close relationship between speech and gestures. While there is general agreement that the two modalities are integrated in language production and comprehension, the specifics of *when*, *where* and *how* gestures are used and *what* they express when they are used, are less well

understood. The studies in this thesis provide new insights into these questions in the context of the representation of referents in connected discourse. They suggest that gestures contribute to discourse cohesion on different levels by facilitating language processing in discourse comprehension, and more importantly by paralleling and complementing speech in discourse production.

References

Allen, S.E.M. (2008). Interacting pragmatic influences on children's argument realization. In M. Bowerman & P. Brown (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 191-210). Mahwah, NJ: Lawrence Erlbaum Associates.

Andrén, M. (2014). On the lower limit of gesture. In M. Seyfeddinipur and M. Gullberg (Eds.), *From gesture in conversation to visible action as utterance: Essays in honor of Adam Kendon* (pp. 153-174). Amsterdam, The Netherlands: John Benjamins.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24, 65–87.

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of pragmatics*, 16, 443-463.

Ariel, M. (1996). Referring Expressions and the+/-Coreference Distinction. In T. Fretheim and J. Gundel (Eds.), *Reference and Referent Accessibility* (pp. 13-33). Amsterdam, The Netherlands: John Benjamins.

Arnold, J. E. (1998). *Reference form and discourse patterns* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and cognitive processes*, 23, 495-527.

Arnold, J.E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4, 187-203

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28-55.

Azar, Z., Backus, A. & Özyürek, A. (2018). General- and Language-Specific Factors Influence Reference Tracking in Speech and Gesture in Discourse. *Discourse Processes*, DOI: 10.1080/0163853X.2018.1519368

Beattie, G., & Shovelton, H. (2007). The role of iconic gesture in semantic communication and its theoretical and practical implications. In S. Duncan, J. Cassell & E. T. Levy (Eds.) *Gesture and the Dynamic Dimension of Language. Essays in honor of David McNeill* (pp. 221-241). Amsterdam, The Netherlands: John Benjamins.

Birner, B. (2013). Discourse functions at the periphery: Noncanonical word order in English. In B. Shaer, W. Frey, & C. Maienborn (Eds.), *Proceedings of the Dislocated Elements Workshop (ZAS Papers in Linguistics 35)* (pp. 41–62). Berlin, Germany: ZAS.

Brown, A. (2008). Gesture viewpoint in Japanese and English: Cross-linguistic interactions between two languages in one speaker. *Gesture*, 8, 256-276.

Brown, A., & Chen, J. (2013). Construal of Manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics*, 24, 605-631.

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, 225-251.

Brugman, H., & Russel, A. (2004). Annotating Multimedia/ Multi-modal resources with ELAN. *In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*. Lisbon, Portugal.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98, 159-168.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, 47, 14-27.

Capirci, O., Cristilli, C., De Angelis, V., & Graziano, M. (2011). Learning to use gesture in narratives. In G. Stam, M. Ishino & R. Ashley (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 187–200). Amsterdam, The Netherlands: John Benjamins.

Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & cognition*, 7, 1-34.

Chafe, W. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse: Typological studies in language* (pp. 21-51). Amsterdam, The Netherlands: John Benjamins.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: University of Chicago Press.

Chafe, W. (1996). Inferring identifiability and accessibility. In T. Fretheim and J. Gundel (Eds.), *Reference and Referent Accessibility* (pp. 37-46). Amsterdam, The Netherlands: John Benjamins.

Chang, F. R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory & Cognition*, 8, 58-64.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910.

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, 191-224.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 257.

Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. In W. L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 127-201). Norwood, NJ: Ablex.

Clark, Eve V. (1978). Locationals: Existential, locative, and possessive constructions. In J.H. Greenberg (Ed.), *Universals of human language*, 4, 85-126. Stanford, CA: Stanford University Press.

Clark, H. H. (1977) Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 243-263). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. Discourse production and comprehension. *Discourse processes: Advances in research and theory*, 1, 1-40.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition*, 7, 35-41.

Cloitre, M., & Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and cognitive processes*, 3, 293-322.

Cooperrider, K. A. (2011). *Reference in action: Links between pointing and language* (Doctoral dissertation). University of California, San Diego, San Diego, CA. Retrieved from: https://escholarship.org/uc/item/52g063b5

Cormier, K. & Fenlon, J. (2014). *BSL Corpus Annotation Guidelines, v. 1*. Cognition and Language Research Centre, University College London. URL: http://bslcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines_23October2014.pdf

Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, 50, 431–456.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284-311). Cambridge, UK: Cambridge University Press.

Dimroth, C. (2012). Videoclips zur Elizitation von Erzählungen: Methodische Überlegungen und einige Ergebnisse am Beispiel der "Finite Story". In B. Ahrenholz (Ed.), *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren* (pp. 77-98). Berlin, Germany: De Gruyter.

Dixon, R. M. (1979). Ergativity. *Language*, 59-138.

Du Bois, J. W. (1980). Beyond definiteness: The trace of identity in discourse. In W.L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 203-274). Norwood, NJ: Ablex.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, 805-855.

Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, 33, 103-119.

Duncan, S. (1996). *Grammatical form and "thinking for speaking" in Mandarin Chinese and English: An analysis based on speech-accompanying gestures* (Doctoral dissertation). University of Chicago, Chicago, IL. Retrieved from: http://mcneilllab.uchicago.edu/pdfs/susan_duncan/S.Duncan_dissertation.PDF

Efron, D. (1941/1972). *Gesture, race and culture; a tentative study of the spatiotemporal and "linguistic" aspects of the gestural behavior of eastern Jews and southern Italians in New York City, living under similar as well as different environmental conditions.* The Hague, The Netherlands: Mouton.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of communication*, 22, 353-374.

Emmorey, K. (1997). Non-antecedent suppression in American Sign Language. *Language and Cognitive Processes*, 12, 103-120.

Emmorey, K., Corina, D., & Bellugi, U. (1995). Differential processing of topographic and syntactic functions of space. In K. Emmorey & J. Reilly (Eds.), *Language, gesture and space* (pp. 43-62). Hillsdale, NJ: Lawrence Erlbaum Associates.

Emmorey, K., Norman, F., & O'Grady, L. (1991). The activation of spatial antecedents from overt pronouns in American Sign Language. *Language and Cognitive Processes*, 6, 207-228.

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. Hamburg, Germany: Signum-Verlag.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of verbal learning and verbal behavior*, 22, 75-87.

Fillmore, C. J. (1982). Frame semantics. In D. Geeraerts (Ed.), *Cognitive linguistics: Basic readings* (pp. 373-400). Berlin, Germany: De Gruyter Mouton

Firbas, J. (1971). On the concept of communicative dynamism in the theory of functional sentence perspective. *Philologica Pragensia*, 8, 135-144.

Foraker, S. (2011). Gesture and discourse: how we use our hands to introduce and refer back. In G. Stam, M. Ishino & R. Ashley (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 279–292). Amsterdam, The Netherlands: John Benjamins.

Fritz, I., Kita, S., Littlemore, J., & Krott, A. (2019). Information packaging in speech shapes information packaging in gesture: The role of speech planning units in the coordination of speech-gesture production. *Journal of Memory and Language*, 104, 56-69.

Garrod, S. (2006). Psycholinguistic Research Methods. In: K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 251–257). Amsterdam, The Netherlands: Elsevier.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99-156.

Gerwing, J. & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, 4, 157-195.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1–42). Amsterdam, The Netherlands: John Benjamins.

Givón, T. (1984). Universals of discourse structure and second language acquisition. In W. E. Rutherford (Ed.), *Language universals and second language acquisition* (pp. 109-136). Amsterdam, The Netherlands: John Benjamins.

Givón, T. (1995). Coherence in text vs. coherence in mind. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 59-115). Amsterdam, The Netherlands: John Benjamins.

Goodrich Smith, W., & Hudson Kam, C. K. (2012). Knowing 'who she is' based on 'where she is': The effect of co-speech gesture on pronoun comprehension. *Language and Cognition*, 4, 75–98.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science*, 17, 311-347.

Graziano, M. (2009). *Rapporto fra lo sviluppo della competenza verbale e gestuale nella costruzione di un testo narrativo in bambini dai 4 ai 10 anni* (Unpublished doctoral dissertation). Università degli Studi 'Suor Orsola Benincasa', Napoli, Italy & Université Stendhal Grenoble 3, Grenoble, France.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish*. Lund, Sweden: Lund University Press.

Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. In C. Dimroth & M. Starren (Eds.), *Information structure, linguistic structure and the dynamics of language acquisition* (pp. 311-328). Amsterdam, The Netherlands: John Benjamins.

Gullberg, M. (2006). Handling discourse: Gestures, reference, tracking, and communication strategies in early L2. *Language Learning*, 56, 155–196.

Gullberg, M. (2009). Reconstructing verb meaning in a second language. How English speakers of L2 Dutch talk and gesture about placement. *Annual Review of Cognitive Linguistics*, 7, 222-245.

Gullberg, M. (2011a). Multilingual multimodality: Communicative difficulties and their solutions in second language use. In Goodwin, C., LeBaron, C. & Streeck, J. (Eds.) (pp. 137-151). Cambridge, UK: Cambridge University Press.

Gullberg, M. (2011b). Language-specific encoding of placement events in gestures. In Bohnemeyer, J. & Pederson, E. (Eds.), *Event representation in language and cognition* (pp. 166-188). Cambridge, UK: Cambridge University Press.

Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language*, 28, 200-236.

Gullberg, M. & Indefrey, P. (2003). *Language background questionnaire.* The Dynamics of Multilingual Processing. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics. URL: http://www.mpi.nl/research/projects/Multilingualism/Lang_Hist_Quest_Engl.pdf

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, 33, 251-277.

Gundel, J. K. (1996). Relevance theory meets the givenness hierarchy: an account of inferrables. In T. Fretheim and J. Gundel (Eds.), *Reference and Referent Accessibility* (pp. 141-153). Amsterdam, The Netherlands: John Benjamins.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.

Gunter, T. C., & Weinbrenner, J. E. D. (2017). When to take a gesture seriously: On how we use and prioritize communicative cues. *Journal of Cognitive Neuroscience*, 29, 1355-1367.

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, 1–10.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: Routledge.

Hawkins, J. A. (1984). A note on referent identifiability and co-presence. *Journal of Pragmatics*, 8, 649-659.

Hendriks, H. (1993). *Motion and location in children's narrative discourse. A developmental study of Chinese and Dutch* (Unpublished doctoral dissertation). Leiden University, Leiden, The Netherlands.

Hickmann, M., & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language*, 26, 419-452.

Hickmann, M., Hendriks, H., & Gullberg, M. (2011). Developmental perspectives on the expression of motion in speech and gesture: A comparison of French and English. *Language, Interaction and Acquisition*, 2, 129-156.

Hickmann, M., Hendriks, H., Roland, F., & Liang, J. (1996). The marking of new information in children's narratives: a comparison of English, French, German and Mandarin Chinese. *Journal of Child Language*, 23, 591-619.

Hogrefe, K., Ziegler, W., & Goldenberg, G. (2011). Measuring the formal diversity of hand gestures by their hamming distance. *Integrating Gestures: The interdisciplinary nature of gesture* (pp. 293–308). Amsterdam, The Netherlands: John Benjamins.

Holler, J., & Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process?. Semiotica, 146, 81-116.

Hudson Kam, C. L., & Goodrich Smith, W. (2011). The problem of conventionality in the development of creole morphological systems. *The Canadian Journal of Linguistics*, 56, 109–124.

Kelly, S., Creigh, P., & Bartolotti, J. (2010). Integrating Speech and Iconic Gestures in a Stroop-like Task: Evidence for Automatic Processing. *Journal of Cognitive Neuroscience*, 22, 683–694.

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic bulletin & review*, 22, 517-523.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Seigman & B. Pope (Eds.), *Studies in Dyadic Communication* (pp.177-216). Elmsford, New York: Pergamon Press.

Kendon, A. (1973). The role of visible behaviour in the organization of face-to-face interaction. In M. Von Cranach & I. Vine (Eds.), *Social communication and movement: studies of interaction and expression in man and chimpanzee* (pp. 29–74). London, UK: Academic Press.

Kendon, A. (1978). Differential perception and attentional frame in face-to-face interaction: Two problems for investigation. *Semiotica*, 24, 305-316.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207-227). The Hague, The Netherlands: De Gruyter Mouton.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica*, 62, 3-28.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of pragmatics*, 23, 247-279.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kendon, A. (2014). The 'poly-modalic' nature of utterances and its implication. In D. Dor, C. Knight, & D. Lewis (Eds.), *The social origins of language* (pp. 67-76). Oxford, UK: Oxford University Press.

Kirchhof, C. (2017). *The shrink point: audiovisual integration of speech-gesture synchrony* (Doctoral dissertation). Bielefeld University, Bielefeld, Germany. Retrieved from: https://pub.uni-bielefeld.de/record/2908762

Kita, S. (1990). *The temporal relationship between gesture and speech: A study of Japanese-English bilinguals* (Unpublished master's thesis). University of Chicago, Chicago, IL.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge, UK: Cambridge University Press.

Kita, S. (2003). Pointing: A foundational building block of human communication. In S. Kita (Ed.), *Pointing* (pp. 9-16). Mahwah, NJ: Lawrence Erlbaum Associates.

Kita, S., Van Gijn, I., & Van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer interaction* (International Gesture Workshop, Bielefeld, 17-19 September, 1997; pp. 23-35). Berlin/Heidelberg: Springer.

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, 16-32.

Kärkkäinen, E. (1996). Preferred argument structure and subject role in American English conversational discourse. *Journal of Pragmatics*, 25, 675-701.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge Studies in Linguistics 71). Cambridge, UK: Cambridge University Press.

Levy, E. T. (1984). Communicating thematic structure in narrative discourse: The use of referring terms and gestures (Unpublished doctoral dissertation). University of Chicago, Chicago, IL.

Levy, E. T., & Fowler, C. A. (2000). The role of gestures and other graded language forms in the grounding of reference. In D. McNeill (Ed.), *Language and gesture* (pp. 215–234). Cambridge, UK: Cambridge University Press.

Levy, E. T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Processes*, 15, 277–301.

Marentette, P., Pettenati, P., Bello, A. & Volterra, V. (2016). Gesture and Symbolic Representation in Italian and English-Speaking Canadian 2-Year-Olds. *Child Development*, 87, 944-961.

Marslen-Wilson, W. D., Levy, E., & Tyler, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R. J. Jarvella & W. Klein (Eds.), *Language, place, and action: Studies in deixis and related topics* (pp. 339–378). Chichester, UK: Wiley.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, 19, 668-682.

Nappa, R. & Arnold, J. (2014). The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology*, 70, 58-81.

McNeill, D. (1992). *Hand and mind*. Chicago, IL: University of Chicago Press.

McNeill, D. (2002). Gesture and language dialectic. *Acta Linguistica Hafniensia*, 34, 7-37.

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

McNeill, D., Cassell, J., & Levy, E. T. (1993). Abstract deixis. *Semiotica*, 95, 5-20.

McNeill, D. & Levy, E.T. (1982). Conceptual representations in language activity and gesture. In R. Javella & W. Klein (Eds.), *Language, place, and action. Studies in deixis and related topics* (pp. 271 -295). Chichester, UK: Wiley.

McNeill, D., & Levy, E. T. (1993). Cohesion and gesture. *Discourse Processes*, 16, 363–386.

McNeill, D., Levy, E. T. & Pedelty, L. L. (1990). Speech and Gesture. In G. R. Hammond (Ed.), *Cerebral control of speech and limb movements* (p. 203-256). Amsterdam, The Netherlands: North Holland.

McNeill, D., Quek, F., McCullough, K. E., Duncan, S. D., Furuyama, N., Bryll, R., & Ansari, R. (2001). Catchments, prosody and discourse. *Gesture*, 1, 9-33.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz GmbH.

Müller, C. (2014). Gestural Modes of Representation as techniques of depiction. In: C. Müller, A, Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Eds.), *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction*. (pp. 1687-1702). Berlin, Germany: De Gruyter Mouton.

O'Brien, E. J. (1987). Antecedent search processes and the structure of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 13, 278-290.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688-704.

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, 219-240.

Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25, 650-668.

Parrill, F. (2012). Interactions between discourse status and viewpoint in co-speech gesture. In B. Dancygier & E. Sweetser (Eds.), *Viewpoint in language: A multimodal perspective* (pp. 97-112).

Payrató, L. (1993). A pragmatic view on autonomous gestures: A first repertoire of Catalan emblems. *Journal of Pragmatics*, 20, 193-216.

Perniss, P. & Özyürek, A. (2015). Visible cohesion: A comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science*, 7, 36-60.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In Peter Cole (Ed.), *Radical pragmatics* (p. 223-56). New York, NY: Academic Press.

Prince, E. F. (1992). The ZPG Letter: Subjects, definiteness and information status. In S. Thompson & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295-325). Amsterdam, The Netherlands: John Benjamins.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schütze-Coburn, S. (1987). Topic management and the lexicon: A discourse profile of three-argument verbs in German (Unpublished master's thesis). University of California, Los Angeles, Los Angeles, CA.

Sekine, K., & Kita, S. (2015). Development of multimodal discourse comprehension: Cohesive use of space in gesture. *Language, Cognition, and Neuroscience*, 30, 1245–1258.

Sekine, K. & Kita, S. (2017). The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures. *Acta Psychologica*, 179, 89-95.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture* (MPI Series in Psycholinguistics) (Unpublished doctoral dissertation). Radboud University Nijmegen, Nijmegen, The Netherlands.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101, 260-277.

Skopeteas, S. (2012). The empirical investigation of information structure. In M. Krifka & R. Musan (Eds.), *The expression of information structure* (pp.217-248). Berlin, Germany: De Gruyter Mouton.

Sloetjes, H., & Wittenburg, P. (2008, May/June). Annotation by category – ELAN and ISO DCR. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco.

So, W. C., Demir, Ö. E., & Goldin-Meadow, S. (2010). When speech is ambiguous, gesture steps in: Sensitivity to discourse-pragmatic principles in early childhood. *Applied Psycholinguistics*, 31, 209-224.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, 33, 115-125.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *International Review of Applied Linguistics*, 44, 143-169.

Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication system of the American deaf. *Studies in Linguistics. Occasional papers*, 8.

Streeck, J. (2008). Depicting by gesture. *Gesture*, 8, 285-301.

Ward, G. & Birner, B. (2001). Discourse and Information Structure. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *Handbook of discourse analysis* (119-137). Oxford, UK: Basil Blackwell.

Wilkin, K., & Holler, J. (2011). Speakers' use of 'action'and 'entity'gestures with definite and indefinite references. In G. Stam & M. Ishino (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 293–308). Amsterdam, The Netherlands: John Benjamins.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing* (pp. 9-16). Mahwah, NJ: Lawrence Erlbaum Associates.

Yoshioka, K. (2005). Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse (Doctoral dissertation). University of Groningen, Groningen, The Netherlands. Retrieved from: https://www.rug.nl/research/portal/en/publications/linguistic-and-gestural-introduction-and-tracking-of-referents-in-l1-and-l2-discourse(ed3d66a8-1714-48c5-a48f-1bffce7a38f5).html

Yoshioka, K. (2008). Gesture and information structure in first and second language. *Gestur*e, 8, 236–255.

Appendices

Appendix A: Story script production studies

(For the pictures, see: https://portal.research.lu.se/portal/en/publications/-(252b43cf-652d-45d6-8a0c-c64f96aa9305).html)

1. Three fairies (in green, yellow and red dresses) are standing in a room. One fairy stands next to a table with a cake. The second fairy stands next to a mannequin. The third fairy has a bucket, broom and mop.
2. Fairy 1 lights the candles on the cake.
3. Cake is falling.
4. Fairy tries to save falling cake by pushing a broom against it.
5. The upper part of the cake slides down the broom.
6. Fairy 2 is sewing a dress.
7. Fairy 2 is cutting out a belt.
8. Fairy 2 puts the belt around the dress.
9. Fairy 2 is cutting little triangles out of a piece of cloth.
10. Dress falls apart.
11. Fairy 2 is sad, Fairy 1 looks at Fairy 2.
12. Fairy 1 goes to Fairy 2.
13. Fairy 1 and Fairy 2 both look sad.
14. Fairy 3 is sweeping the floor with a broom.
15. A blast of wind comes in. Leaves and dust are flying around.
16. Fairy 3 is sweeping the floor with a mop.
17. Fairy 3 slips and falls down.
18. Fairy 3 sits on the floor.
19. Fairy 2 goes to Fairy 3.
20. Fairy 1 goes to Fairy 3.
21. All fairies look sad.
22. Fairy 3 has an idea.
23. Fairy 3 goes up the stairs.
24. Fairy 3 comes down the stairs with a box with 3 magic wands.
25. Fairy 1 takes a wand out of the box.

26. Fairy 2 takes a wand out of the box.
27. Fairy 3 takes a wand out of the box.
28. All fairies go back to their places in the room.
29. Fairy 1 stands next to her table. On top of the table are a bag of flour, two eggs, a bowl, a milk can, a spoon, a sugar bowl and a saltshaker.
30. Fairy 1 enchants all the objects on her table.
31. The objects become alive.
32. Fairy 1 shows the objects a recipe in a book.
33. The bowl moves to the middle of the table.
34. The flour bag moves next to the bowl.
35. The milk can moves next to the bowl.
36. A spoon moves next to the bowl.
37. The spoon takes some flour out of flour bag.
38. The spoon puts the flour into the bowl.
39. The eggs fly over the bowl.
40. The milk can pours milk into the bowl.
41. The eggs break in the air, the yolk and white fall into bowl.
42. The spoon stirs the dough.
43. Fairy 1 comes to the table.
44. A sugar bowl flies above the bowl.
45. A salt shaker flies above the bowl.
46. Fairy 1 takes a pinch of salt with her hand.
47. Fairy 1 adds the salt to the bowl.
48. Salt shaker falls into the dough.
49. Fairy 1 takes saltshaker out of the bowl.
50. The sugar bowl pours sugar into spoon.
51. The spoon pours sugar into the bowl.
52. Fairy 1 is stirring the dough with her wand.
53. Bowl pours dough into a cake.
54. Fairy 2 is next to her mannequin and enchants a basket with purple cloth, scissors, a needle and thread in it.
55. A first piece of cloth puts itself around upper body of the mannequin.
56. A second piece of cloth puts itself around the lower body of the mannequin.
57. Fairy 2 looks at her dress.
58. A first needle with thread sews the top.
59. A second needle with thread sews the skirt.
60. A white piece of cloth and scissors fly into the air.
61. The scissors cut the cloth into two pieces.
62. The two pieces put themselves on the dress as sleeves.

63. Fairy 2 sews one sleeve with her wand.
64. Fairy 2 flies over the dress to the other side.
65. Fairy 2 sews the other sleeve with her wand.
66. Fairy 3 is at the bottom of the stairs. The bucket, mop and broom are at the top. Fairy 3 enchants the objects.
67. Bucket comes down the stairs.
68. Broom comes down the stairs.
69. Mop comes down the stairs.
70. Fairy 3 flies above the buck.et
71. Fairy 3 pours cleaning agent into bucket.
72. Broom sweeps the floor.
73. Flour bag jumps into the air.
74. Fairy 1 jumps into the air.
75. Fairy 3 jumps into the air.
76. Mop goes into water bucket.
77. Mop sweeps floor.
78. Milk can jumps into the air.
79. Fairy 2 jumps into the air.
80. Fairy 2 is still in the air, the bucket is under her on the floor.
81. Fairy 2 falls into the bucket.
82. Fairy 3 comes to help Fairy 2 out of the bucket.
83. An icing bag is putting sugar dots onto cake.
84. Fairy 3 flies to cake.
85. Fairy 3 is decorating the cake with some sprinkles.
86. Fairy 1 goes up the stairs.
87. Fairy 1 comes down the stairs with a box full of candles.
88. Fairy 1 flies next to the cake.
89. Fairy 1 puts the candles onto the cake.
90. A match comes flying to cake.
91. Icing bag is putting sugar hearts onto cake.
92. Match lights the candles.
93. Cake jumps into the air.
94. Fairy 1 and 3 look over to Fairy 2.
95. Fairy 2 goes up the stairs.
96. Fairy 2 comes down the stairs with a box.
97. Fairy 2 takes out a cloth.
98. Scissors cut the cloth into a collar.
99. One needle sews the collar onto dress. Two more needles are flying next to the box.

100. Second needle takes a bow out of box.
101. Second needle sews the bow onto the dress.
102. Third needle takes a bow out of box.
103. Third needle sews the bow onto the dress.
104. First needle takes a bow out of box
105. First needle sews the bow onto the dress.
106. Fairy 3 looks over to the dress.
107. Fairy 3 tiptoes to the bucket.
108. Fairy 3 enchants herself. She shrinks.
109. Fairy 3 jumps into water bucket to hide.
110. Fairy 3 uses magic to change the color of the dress.
111. Fairy 2 changes back the color of the dress.
112. Fairy 3 changes the color of the dress again.
113. Fairy 2 goes to the bucket and tells Fairy 3 off.
114. Fairy 2 tells Fairy 3 to go back to her work.
115. Fairy 3 enchants herself again. She grows back to her normal size.
116. Fairy 3 is back with her broom, bucket and mop.
117. Broom goes up the stairs.
118. Mop goes up the stairs.
119. Bucket goes up the stairs.
120. Fairy 3 changes the color of the dress again.
121. Fairy 2 changes back the color of the dress.
122. Fairy 1 tells Fairy 2 and Fairy 3 to stop.
123. Fairy 1 gets the box for the wands.
124. Fairy 2 puts her wand back into the box.
125. Fairy 3 puts her wand back into the box.
126. Fairy 1 changes the color of the dress (i.e., it has two colors now).
127. All fairies are happy.

Appendix B: Instructions production studies

German original

Was ist passiert?

Ihre Aufgabe ist es eine Bildergeschichte über drei Feen nachzuerzählen. Die Bilder werden Ihnen Stück für Stück von der Forscherin gezeigt (auf dem Tisch neben Ihnen). Es werden immer 4-6 Bilder auf einmal sein. Insgesamt sind es 127 Bilder.

Die leitende Frage lautet: Was ist passiert in den Abschnitten, die Ihnen gezeigt werden?

Achten Sie bitte auf Folgendes, wenn Sie die Abschnitte Ihrem Gegenüber nacherzählen:

1. *Bitte erwähnen Sie immer jedes einzelne Bild.*
2. *Nehmen Sie sich all die Zeit, die Sie brauchen, um sich den Ablauf, und was in jedem Bild passiert, einzuprägen.*
3. *Drehen Sie sich dann bitte zu Ihrem/Ihrer Zuhörer(in) und erzählen Sie ihm/ihr, was passiert ist.*
4. *Seien Sie so genau wie möglich bei der Nacherzählung, denn Ihr(e) Zuhörer(in) wird nach jedem Abschnitt auf der Basis ihrer Nacherzählung, Fragen zur Geschichte beantworten müssen.*
5. *Ihr(e) Zuhörer(in) kennt die Geschichte nicht und hat die Bilder nie gesehen.*
6. *In dieser Studie geht es um Alltagssprache, also sprechen Sie gerne so, wie Sie es sonst auch tun mit Ihren Freunden.*

Vielen Dank für Ihre Teilnahme und viel Spaß!

English translation

What happened?

Your task is to retell a story about three fairies on the basis of a picture story. The experimenter will show you the pictures bit by bit (on the table next to you). You will always see 4-6 pictures at once. In total, there are 127 pictures.

The leading question is: What happened in the section you have just seen?

Please, mind the following aspects when retelling the sections to the person you are addressing:

1. Please, always mention every picture.
2. Take all the time you need in order to memorize the sequence of events, and what happens in every picture.

3. Then turn around to your addressee and tell him/her what happened.
4. Be as precise as possible since your addressee will have to answer questions about each section on the basis of your retelling.
5. Your addressee does not know the story and has never seen the pictures.
6. In this study, the focus is on everyday language. Therefore, you are welcome to speak in a way that you typically do with your friends.

Thank you for your participation!

Appendix C: Instructions perception study

Experiment 1: German original

Es werden Ihnen Videos vorgespielt, in denen eine Frau kurze Geschichten erzählt. In jeder Geschichte wird zuerst die Hauptperson vorgestellt. Es folgen Informationen über ein Problem, das aufkommt, welches die Hauptperson lösen will. Ihre Aufgabe ist es genau zuzuhören und auf die folgende Frage zu antworten:

Hat die Hauptperson jemanden ANGERUFEN, um das Problem zu lösen/um Hilfe zu bekommen?

Wenn das der Fall ist, drücken Sie auf die Taste „j" (=ja)

Wenn das nicht der Fall ist, drücken Sie auf die Taste „f" (=falsch/nein)

Hier sind drei Beispiele:

1. *Da war eine Dame (Hauptperson). Und ihr Hund wurde von einem Auto angefahren (Problem). Da hat die Dame beschlossen, den Notruf anzurufen (Lösung). Der sollte ihm dann zur Hilfe kommen.*
   → Die richtige Antwort ist „ja", also Taste „j", weil die Hauptperson, jemanden angerufen hat, um Hilfe zu bekommen und um somit das Problem zu lösen.
2. *Ein Junge (Hauptperson) schlenderte gerade auf der Straße entlang als ihn eine Frau ansprach. Sie hatte sich aus ihrer Wohnung ausgeschlossen und kein Telefon zur Hand (Problem). Also rief der Junge den Schlüsseldienst von seinem Handy aus an (Lösung).*
   → Die richtige Antwort ist „ja", also Taste „j", weil die Hauptperson, jemanden angerufen hat, um Hilfe zu bekommen und um somit das Problem zu lösen.
3. *Es ging um einen Mann (Hauptperson). Und seine Tochter wurde in der Schule ständig gehänselt (Problem). Also beschloss der Mann, ihre Lehrerin anzuschreiben (Lösung). Sie würde ihm sicher bei der Lösung des Problems helfen.*
   → Die richtige Antwort ist „falsch/nein", also Taste „f", weil die Hauptperson, NICHT angerufen, sondern jemanden angeschrieben hat, um das Problem zu lösen.

Bitte drücken Sie die Tasten so genau und so schnell, wie möglich (d.h. Sie sollen nicht warten bis das Video vorbei ist). Halten Sie, während des gesamten Experiments, den rechten Zeigefinger auf die Taste „j" und den linken Zeigefinger auf die Taste „f" gelegt. Versuchen Sie bitte keine anderen Tasten zu drücken. Bitte schauen sie während der gesamten Dauer des Experiments auf den Bildschirm! Das Experiment dauert ca. 15 min.

Vielen Dank für Ihre Teilnahme und viel Spaß!

Experiment 1: English translation

You will see videos of a woman telling different short stories. In each story, a main character is introduced. Then follows information about a problem that the main character wants to solve. Your task is to listen carefully and answer the following question:

Did the main character CALL someone to solve the problem/to get help?

If that is the case, then press the key 'j' (= *ja* 'yes').

If that is not the case, then press the key 'f' (= *falsch/nei*n 'false/no').

Here are some examples:

1. There was a lady (main character). And her dog was hit by a car (problem). So, the lady decided to make an emergency call (solution). They would come to help him out.
   → The correct answer is 'yes', and therefore key 'j' because the main character made a call to solve the problem.
2. A boy (main character) was strolling around the streets when a woman approached him. She had locked herself out of her apartment and did not have a cell phone at hand (problem). So, the boy called the locksmith from his phone (solution).
   → The correct answer is 'yes', and therefore key 'j' because the main character called someone to solve the problem.
3. There was a man (main character). And his daughter was constantly being teased at school (problem). So, the man decided to write to her teacher (solution). She would certainly help to solve the problem.
   → The correct answer is 'no', and therefore key 'f' because the main character wrote to someone for help.

Please press the keys as fast as possible (i.e. do not wait until the video has finished). Keep your right index finger on the key 'j' and your left index finger on the key 'f' throughout the whole experiment. Please, try not to press any other keys. Please look at the screen during the whole duration of the experiment. The experiment lasts about 15 minutes.

Thank you for your participation!

Experiment 2: German original

Es werden Ihnen Videos vorgespielt, in denen immer die gleiche Frau verschiedene kurze Geschichten erzählt. Vor jedem Video wird ein Wort (= Bezeichnung einer Person) eingeblendet. Ihre Aufgabe ist es jedes Mal, wenn Sie dem Wort in der Geschichte begegnen, die Taste „j" zu drücken. Tun Sie dies bitte so schnell wie möglich, da es uns um die Geschwindigkeit Ihrer Antwort geht. Nach der Erzählung wird dann manchmal eine Frage zu der Geschichte eingeblendet. Beantworten Sie diese Frage bitte auch so schnell wie möglich. Drücken Sie auf die Taste „j" (= ja) oder „f" (= falsch/nein).

Halten Sie, während des ganzen Experiments, den rechten Zeigefinger auf die Taste „j" und den linken Zeigefinger auf die Taste „f" gelegt. Versuchen Sie bitte keine anderen Tasten zu drücken. Schauen Sie während der gesamten Dauer des Experiments bitte auf den Bildschirm und machen sie keine Pausen zwischendurch.

Das Experiment dauert ca. 15 min.

Vielen Danke für Ihre Teilnahme und viel Spaß!

Experiment 2: English translation

You will see videos, in which the same woman is telling different short stories. Before a video starts, you will see a word on the screen (= the designation of a person). Your task is to press the key 'j' each time you come across that word during the story. Please, press the key as fast as possible since we are interested in the speed of your response. After some of the stories, you will be shown a question about the preceding story. Please, answer that question as fast as possible. Press the key 'j' (= *ja* 'yes') or 'f' (= *falsch/nein* 'false/no').

Keep your right index finger on the key 'j' and your left index finger on the key 'f' throughout the whole experiment. Please, try not to press any other keys. Please look at the screen during the whole duration of the experiment and do not take any breaks. The experiment lasts about 15 minutes.

Thank you for your participation!

Appendix D: Consent form

Production studies: German original

Einverständniserklärung

Ich gebe hiermit Sandra Debreslioska, Lund Universität, die Genehmigung die heutigen Aufnahmen (Audio und Video) für die folgenden Zwecke zu nutzen:

(Bitte, kreuzen Sie die entsprechende Box an,"□", wenn Sie die Genehmigung dafür erteilen.)

□*1. Für wissenschaftliche Analysen zu Forschungszwecken;*

□*2. Als Illustrationen der Analysen/Ergebnisse bei wissenschaftlichen Konferenzen, Vorlesungen und wissenschaftlichen Veröffentlichungen;*

□*3. Als Illustrationen auf der Webseite der Universität Lund.*

Meine Anonymität ist voll und ganz garantiert. Unter keinen Umständen wird meine Identität offengelegt werden. Niemand, außer der oben genannten Forscherin, wird über meine Identität Bescheid wissen (d.h., zum Beispiel, dass keine Namen in Präsentationen der Ergebnisse benutzt werden).

Production studies: English translation

Consent form

I herewith give Sandra Debreslioska, Lund University, the permission to use today's recordings (audio and video) for the following purposes:

(Please, cross the box, '□', if you give permission.)

□1. For scientific analyses for research purposes;

□2. As illustrations of the analyses/results at scientific conferences, lectures or in scientific publications;

□3. As illustrations on Lund University's website.

My anonymity is guaranteed. Under no circumstances will my identity be disclosed to anyone else than the above-mentioned researcher (e.g., no names will be used in presentations about the results or recordings).

Perception experiments: German original

Einverständniserklärung

Ich gebe hiermit Sandra Debreslioska, Lund Universität, die Genehmigung die heutigen Aufnahmen (d.h., Antworten per Tastatur) für wissenschaftliche Analysen zu Forschungszwecken zu nutzen. Das bedeutet auch, dass die Ergebnisse der Analysen in Form von Graphen (oder ähnlichem) bei wissenschaftlichen Konferenzen, in Vorlesungen und in wissenschaftlichen Veröffentlichungen gezeigt werden.

Meine Anonymität ist voll und ganz garantiert. Unter keinen Umständen wird meine Identität offengelegt werden. Niemand außer der oben genannten Forscherin, wird über meine Identität Bescheid wissen (d.h., zum Beispiel, dass keine Namen in Präsentationen der Ergebnisse benutzt werden).

Perception experiments: English translation

Consent form

Herewith I give Sandra Debreslioska, Lund Unversity, the permission to use today's recordings (i.e., key press answers) for scientific analyses for research purposes. This also means that the analyses/results will be shown in the form of graphs (or similar) at scientific conferences, in lectures or in scientific publications.

My anonymity is guaranteed. Under no circumstances will my identity be disclosed to anyone else than the above-mentioned researcher (e.g., no names will be used in presentations about the results).

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

Representing discourse referents in speech and gesture

Gestures are part of language. When speakers produce discourse, they use speech but also gestures, and addressees reliably recognize such gestures as communicatively meaningful. This thesis examines the details of how speech and gestures work together in discourse production, and how addressees use gesture information in discourse perception. The focus is on discourse referents (entities talked about), and on how they are represented in the two modalities. Speakers refer to referents in speech differently as a function of discourse, for example depending on whether they are new to discourse or already mentioned. The thesis takes such variations in speech as their starting point and examines the way that gestures pattern accordingly. In four studies, the thesis investigates *when* gestures are produced for the representation of discourse referents, *where* they are produced, *how* they are produced, and *what* they express. The findings highlight the multifunctionality of gestures, showing that gestures can have a parallel or complementary function to speech depending on the context. In discourse perception, gestures further seem to have a facilitatory function. The studies in this thesis contribute to our understanding of the close relationship between speech and gestures, and advocate that gestures be considered in linguistic studies on discourse, and that connected discourse be considered in gesture studies.

LUND UNIVERSITY

LUND UNIVERSITY
The Faculties of Humanities and Theology
Centre for Languages and Literature
ISBN 978-91-88899-32-3