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.
Representing discourse referents in speech and gesture

Sandra Debreslioska

LUND UNIVERSITY
Cover photo by Carolina Larsson, Stefan Lindgren & Sandra Debreslioska

Copyright pp 1-96 Sandra Debreslioska
Paper I © by the Authors (submitted)
Paper II © by the Authors (submitted)
Paper III © Taylor & Francis
Paper IV © by the Authors (submitted)

Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-88899-33-0 (digital)

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

Media-Tryck is an environmentally certified and ISO 14001 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se
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*
    (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

*b* der1 öffnet eine Kiste

*c* ∅1 holt ein Seil3 heraus

*d* und ∅1 schließt die Kiste2 wieder

*e* dann kommt ein anderer Mann4 die Treppe runter

‘a there is a man

*b* he1 opens a box

*c* ∅1 takes out a rope

*d* and ∅1 closes the box again

*e* then another man4 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

b der öffnet eine Kiste
c $\emptyset_1$ holt ein Seil heraus

d und $\emptyset_1$ macht sich daraus einen Strick um den Hals.

‘a there is a man

b he opens a box

c $\emptyset_1$ takes out a rope

d and $\emptyset_1$ puts it as a cord around the neck’

(4)

a dann versucht die Fee das Rutschen von der Torte aufzuhalten

b indem sie den Besen, dagegenstellt

c allerdings funktioniert das nicht

d weil die oberste Schicht der Torte dann den Besenstiel runterrutscht

‘a then the fairy tries to stop the sliding of the cake

b by putting the broom against it

c but it does not work

d because then the upper part of the cake is sliding down the broomstick’

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.

<table>
<thead>
<tr>
<th>First mentions</th>
<th>Less accessible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand new</td>
<td>Inferable</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent mentions</td>
<td>More accessible</td>
</tr>
<tr>
<td>Reintroduced</td>
<td>Maintained</td>
</tr>
</tbody>
</table>

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 eine Puppe vor 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.

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

---

*Figure 3. Example of a gesture*

*und auf der Torte sind Kerzen drauf*

‘and on the cake are **candles** on top of it’
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.

Figure 5: Example of a Character Viewpoint gesture

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).

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

`und dann ist noch n Korb da`

‘and then there is a basket’
aber es ist dann irgendwie *n Kochbuch* da
‘but there is somehow a *cook*book there’

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

aber sie nimmt trotzdem *ein Stück Stoff* raus
‘but she takes out *a piece of cloth* anyways’

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

27
“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).
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).

<table>
<thead>
<tr>
<th>Given/more accessible referents</th>
<th>New/less accessible referents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero anaphora</td>
<td>No gesture</td>
</tr>
<tr>
<td>Pronouns</td>
<td>Beats</td>
</tr>
<tr>
<td>Lexical NPs</td>
<td>Deictic gestures</td>
</tr>
<tr>
<td>Modified lexical NPs</td>
<td>O-VPT gestures</td>
</tr>
<tr>
<td>Predicates</td>
<td>C-VPT gestures</td>
</tr>
</tbody>
</table>

**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 Debreslowska, 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 Debreslowska
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.

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., Debréslioska 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](image)
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 Frau1. Und ihr Mann konnte den Motor in seinem Auto nicht selbst reparieren. Also hat sich die Frau2 dazu entschlossen, ihren Bruder anzurufen/anzuschreiben. Der soll ihm dann zur Hilfe kommen.

‘There was a woman1. And her husband couldn’t repair the engine of his car by himself. So, the woman2 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’
‘and in the meantime, though, the fairy creeps up the stairs’

‘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).

**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; Debresloska 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 ‘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 information4, 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
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 Debresljoska 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, Debresljoska 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


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.


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 bucket.
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.
Second needle takes a bow out of box.
Second needle sews the bow onto the dress.
Third needle takes a bow out of box.
Third needle sews the bow onto the dress.
First needle takes a bow out of box.
First needle sews the bow onto the dress.
Fairy 3 looks over to the dress.
Fairy 3 tiptoes to the bucket.
Fairy 3 enchants herself. She shrinks.
Fairy 3 jumps into water bucket to hide.
Fairy 3 uses magic to change the color of the dress.
Fairy 2 changes back the color of the dress.
Fairy 3 changes the color of the dress again.
Fairy 2 goes to the bucket and tells Fairy 3 off.
Fairy 2 tells Fairy 3 to go back to her work.
Fairy 3 enchants herself again. She grows back to her normal size.
Fairy 3 is back with her broom, bucket and mop.
Broom goes up the stairs.
Mop goes up the stairs.
Bucket goes up the stairs.
Fairy 3 changes the color of the dress again.
Fairy 2 changes back the color of the dress.
Fairy 1 tells Fairy 2 and Fairy 3 to stop.
Fairy 1 gets the box for the wands.
Fairy 2 puts her wand back into the box.
Fairy 3 puts her wand back into the box.
Fairy 1 changes the color of the dress (i.e., it has two colors now).
All fairies are happy.
Appendix B: Instructions production studies

German original

Was ist passiert?


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 Alltagsprache, 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 Schlüsselfeldienst 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.

   → 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/nein ‘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


*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.


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.


Perception experiments: English translation

Consent form

Herewith I give Sandra Debreslioska, Lund University, 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).
Papers I-IV
Paper I: Gestures signal the difference between brand-new and inferable referents in discourse

Abstract

The literature on bimodal discourse reference has shown that gestures are sensitive to referents’ information status in discourse. Gestures occur more often with new referents/first mentions than with given referents/subsequent mentions. However, not all new entities at first mention occur with gestures. The current study therefore examines whether gestures are sensitive to a more fine-grained difference in information status between brand-new and inferable referents (i.e., referents that are ‘truly’ new vs. referents that are inferentially linked to an already-mentioned referent), and/or to a variation in referential form, as in nominal definiteness. Unexpectedly, the results show that gestures are more frequent with inferable referents than with brand-new referents. The findings reveal new aspects of the relationship between gestures and speech in discourse, specifically suggesting a complementary (disambiguating) function for gestures in the context of first mentioned discourse entities. The results thus highlight the multi-functionality of gestures in relation to speech.

1 Introduction

When producing a stretch of discourse, speakers can use speech and speech-associated gestures to indicate to whom or what they are referring. For instance, when mentioning the entity ‘broom’ for the first time, a speaker can accompany the referential expression with a gesture in which she pretends to hold the broom (Figure 1).
Bimodal referring is a widely acknowledged phenomenon, but the mechanism explaining why gestures occur at specific moments when speakers mention entities in discourse is less well understood. McNeill (1992, 2005) proposes that communicative dynamism (CD) – the degree to which a piece of information “pushes the communication forward” (Firbas, 1971: 136) – determines the presence versus absence of gesture. McNeill takes information status, one of three factors influencing CD (Firbas, 1971) as a starting point and shows that the less accessible the information, the more likely a gesture is to occur. Conversely, the more accessible the information, the less likely a gesture is to occur. This would suggest that new entities in discourse are more likely to occur with gestures than already mentioned ones, an observation that is well supported in the literature (Foraker, 2011; Gullberg, 1998, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson, Levy & Tyler, 1982; McNeill & Levy, 1993).

However, there is evidence that not all entities which are mentioned for the first time in discourse, representing the lowest degree of accessibility (or highest degree of newness), are accompanied by gestures (Foraker, 2011; Gullberg, 2003). Hitherto, this variation has gone unmentioned. The current study therefore examines the variation in the incidence of gesture with entities mentioned for the first time, and specifically probes the possibility that gesture production may be related to entities’ information status (brand-new vs. inferable; Prince, 1981; see also Chafe, 1994; Clark, 1977; Fillmore, 1982; Givón, 1995; Gundel, 1996), which in turn may interact with nominal definiteness (definite vs. indefinite NPs).
1.1 Speech-associated gestures

When speakers engage in talk, bodily action is always mobilized, which goes beyond the use of the anatomical apparatus needed for speaking (Kendon, 2014). This bodily action can involve the face and eyes, the neck and head, the upper body and trunk, and, importantly, the hands and arms. A large body of research shows that the hand and arm movements speakers perform while speaking (also called gesticulations, co-speech gestures, speech-associated gestures, manual gestures, or simply gestures) are organized as patterns of movement that are rhythmically coordinated with speech production (Kendon 1972, 1980). At the same time, they are also considered to be meaningful, specifically in how they relate to the meanings in the speech they accompany (McNeill, 1992; Kendon 2004). For instance, speakers may use gestures to provide iconic representations of what is being talked about, or they may use them to point to or locate entities (see Figures 2-3). In Figure 2, the speaker mentions the entity Ärmel ’sleeve’ for the first time within a sewing event. In exact temporal co-occurrence with this mention, she uses a gesture to represent the sewing action performed on the sleeve by moving her right hand in a circular fashion along her left arm producing an iconic depiction. In Figure 3, the speaker mentions the existence of the entity Tisch ’table’. She raises both hands in parallel to chest level, with flat hands and palms facing each other, in order to indicate the shape/size of the table. This tight coordination in meaning and timing of two modalities is at the basis of the consideration that gestures and speech are conceptually linked (Kendon, 2004).

![Image of hand gesture]

wie sie zuerst auf der Seite, auf der die Fee steht, den Ärmel zunäht

‘how she sews the sleeve on the side, on which the fairy is standing’

Figure 2: Iconic representation of ‘sewing a sleeve’
1.2 Speech-associated gestures and the information status of entities

The relationship between speech and gestures extends from the local level of one composite expression to more global interactions of the two modalities, as is the case for the organization of connected discourse. Gestures and speech vary in a coordinated fashion in the way they are deployed depending on the unfolding of information in discourse. For example, for the tracking of referents in discourse, a growing number of studies demonstrate a close link between gestures and speech, emphasizing the role played by the information status of entities. When entities are new or less accessible, they are typically expressed with richer referential expressions in speech (as in lexical NPs) and are accompanied by gestures. In contrast, when entities are given or more accessible, they are expressed with leaner referential expressions in speech (as in pronouns) and are typically not accompanied by gestures (e.g., Debreslioska & Gullberg, 2017; Debreslioska, Özyürek, Gullberg, Perniss, 2013; Gullberg, 1998, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill, 1992, 2005; McNeill & Levy, 1993; Parrill, 2012; Perniss & Özyürek, 2015; Wilkin & Holler, 2011; Yoshioka, 2008). This pattern reflects Givón’s so-called principle of quantity (Givón, 1983), which predicts more marking material for less accessible information and less marking material for more accessible information. More importantly, the pattern is also at the heart of McNeill’s theory of communicative dynamism (CD) and gestures, which posits that the more a piece of information “pushes the communication forward” (Firbas, 1971:136), the more likely it is that a gesture co-occurs with it. The information status (or how accessible a referent is) is one
important factor influencing the CD of an expression (Firbas, 1971). Thus, findings showing the parallelism between speech and gesture to signal new information (richer referential expressions and gestures) versus given information (leaner referential expressions and few/no gestures) are considered to provide evidence for McNeill’s theory.

An example of this pattern is illustrated in Figure 4, which is taken from the current data set. In order to signal that referents are new, indefinite lexical NPs are used in speech for the referents Kerzen ‘candles’ in the first utterance (4a), and Fee ‘fairy’ in the second utterance (4b). When the referent ‘candles’ is mentioned for the second time in the second utterance (4b), the speaker uses a pronoun to refer back to it (die ‘they’). In gesture, this alternation between richer/leaner expressions is reflected in a variation in gesture incidence. Both first mentions are accompanied by gestures (i.e., the referents ‘candles’ are represented by a gesture locating them on top of a cake, which in turn had previously been placed underneath the gesture indicating the candles, and the ‘fairy’ is represented by a localizing gesture to the right of the cake), however the subsequent mention of the referent ‘candles’ by the pronoun die ‘they’ is not accompanied by a gesture.
Although the literature shows that new referents are more likely to occur with gestures than old/given ones, it also shows that not all first mentions are accompanied by gesture (e.g., 39.8% in Foraker, 2011; 25% in Gullberg, 2003). This observation, in turn, seems to challenge predictions derived from McNeill (1992, 2005). Since a referent
mentioned for the first time should always push the communication forward (or carry higher CD), we might expect every first mention to be accompanied by gesture. But it is not. It remains unclear why this should be the case.

One possibility is that a more fine-grained notion of information status is needed to account for the incidence of gestures. Specifically, in the context of new information and first mentions, entities could be divided into those that are brand-new and those that are inferable from the preceding context. Prince (1981, 1992) defines brand-new entities as being new to the preceding discourse and also new to the addressee. Inferable entities, on the other hand, are new to the preceding discourse, but their existence can be inferred by the addressee. A referent is typically rendered inferable by virtue of a trigger entity which has previously been mentioned in the discourse (Prince, 1981, 1992). For instance, inferable referents are entities that stand in a part/whole relationship or in a content/container relationship to already-mentioned entities. For example, if the referent Besen ‘broom’ has already been mentioned in a particular stretch of discourse, then a current mention of the referent Stiel ‘broomstick’ can be considered inferable. Similarly, if the referent Salzstreuer ‘saltshaker’ has already been mentioned, then a current mention of Salz ‘salt’ can be considered inferable information. Note that these kinds of relationships that give rise to inferables hold true even if in some cases a particular referent does not have a certain part or content (e.g., an empty saltshaker). It is considered sufficient that the relationship typically holds true (Birner, 2013). The relationship between an inferable referent and its trigger entity has been described and discussed in terms of ‘frames’ or ‘schemas’ by other functional and cognitive linguists (e.g., Chafe, 1987, 1994; Fillmore, 1982; Givón, 1995; Hawkins, 1984; Lambrecht, 1994). In a narrow sense of Fillmore (1982)’s frame semantics, we could say that the mentioning of a trigger entity by the speaker evokes a frame (of the trigger entity) in which the inferable entity plays a role. The inferable entity is thus brought into consciousness because the whole frame with all its parts is always activated when one part of the frame is being referred to (Fillmore, 1982). Chafe (1987, 1994) calls the referents that are not explicitly mentioned but which are indirectly activated by the frame ‘accessible’. However, since the term ‘accessible’ is typically also used for referents that have explicit antecedents in the text (e.g., Chafe, 1987; Givón, 1995; Lambrecht, 1994), we will, for the purposes of this paper, stick to Prince’s (1981) term ‘inferable’ in order to make it clear that we do not examine referents which have previously been explicitly mentioned in the discourse (i.e., evoked but not currently active), but rather those referents that have not previously been mentioned but that can be inferentially accessed (i.e., inferables).

The variation in information status between brand-new versus inferable referents can further be signaled in speech by a formal variation in nominal definiteness. Speakers are likely to refer to inferable entities with definite lexical NPs (also called bridging expressions, as in e.g., ‘the broomstick’) more often than to brand-new entities (e.g., ‘a broom’; Clark, H., 1975, 1977). In principle, however, inferable entities can be
represented by both indefinite and definite lexical NPs (Gundel, 1996; Prince, 1992), as illustrated in (1-2), taken from the current data set. In each case, the speaker has already introduced a broom as a whole into the discourse. At a later point, one speaker mentions the referent ‘broomstick’ by using an indefinite lexical NP (1), whereas the other speaker chooses a definite lexical NP (2)\(^1\). In order to avoid circularity (i.e., by assuming that each definite nominal used for a first mention automatically represents an inferable entity, and vice versa), we will keep the formal marking of nominal definiteness separate from information status while still assuming that the two measures will likely co-vary, such that inferables will be referred to with definites more often.

(1)

der hat nen braunen Stiel und gelbe Borsten

‘it has a brown broomstick and yellow bristles’

(2)

der Besenstiel ist braun

‘the broomstick is brown’

McNeill’s (1992, 2005) theory of communicative dynamism and gesture, but also most other previous research on gestures in discourse, would predict that brand-new referents – which are ‘truly’ new since they have never been mentioned and cannot be inferred from previously mentioned referents – should attract more gestures than inferable referents. Furthermore, if it is the case that indefinite lexical NPs signal brand-new referents more than definite lexical NPs, then they should also attract gestures more than definite lexical NPs (Debreslioskata & Gullberg, 2017, but see Wilkin & Holler, 2011).

The current study seeks to test these predictions in order to further our understanding of when first mentions attract gestures or not.

\(^1\) Note that there is a difference in utterance structure between the two examples, which interacts with the definiteness marking (i.e., indefinite markers are less likely to be used for referents instantiated as subject/topic of an utterance). However, for the purposes of the current study, this is not directly relevant, since the goal is to establish whether the frequency of gestures interacts with a referent’s information status, irrespective of its position in the utterance.
1.3 The current study

The current study examines when discourse entities that are mentioned for the first time co-occur with gestures and when they do not. Particularly, it explores two variables, information status (brand-new vs. inferable reference) and nominal definiteness (definite vs. indefinite nominals) to test whether these two factors are related to the incidence of gestures (presence vs. absence) in bimodal discourse.

For speech, we predict that a) brand-new entities are more likely to be mentioned with indefinite nominals, and conversely, that inferable entities are more likely to be represented with definite nominals. For gesture, we predict that, if information status and definiteness have an effect on the incidence of gestures, b) brand-new referents will co-occur with gestures more than inferable referents, and c) indefinite lexical NPs will co-occur with gestures more than definite lexical NPs.

2 Method

2.1 Participants

We invited 20 native German speakers (16 female, mean age 26, range 20-39) to participate in the study at Ludwig-Maximilian University, Munich, Germany. All participants came with a native German-speaking friend who acted as listener. Everyone provided written consent.

2.2 Materials/Design

We used a picture story to elicit narrative speech and gestures. The story consisted of 127 pictures about three fairies, each having to fulfill a task (baking a cake, sewing a dress, cleaning the floor), which they fail at, and consequently use magic to achieve (see Figure 5 for example stimulus pictures). References to the three fairies and a range of inanimate entities were considered (see Appendix B for a full list).
2.3 Procedure

Participants sat across from each other and only the speaker was captured by a video camera, focusing on head and torso. Participants read instructions on paper, and the experimenter further repeated the main points orally to them. Speakers had to retell the picture story by answering the question ‘what happened?’. They always saw four to six pictures at once and had unlimited time to memorize them. Speakers were encouraged to say something about each picture. The listener was not supposed to ask any questions, but to write down a short summary of each part of the story just heard. While only the speaker was of interested for the current study, this was not disclosed to the participants. The listener was further instructed not to cross legs or arms in order to avoid mirroring by the speaker, which could be unfavorable for gesture production (e.g., Chartrand & Bargh, 1999; Kendon, 1973). The roles of speaker and listener were assigned randomly². A session lasted between 45-90 mins. The produced narratives were 20 mins long on average. All participants were debriefed orally at the end of the experiment and were offered refreshments as compensation. Furthermore, all participants signed consent forms, while speakers also filled out a more detailed (language) background questionnaire based on Gullberg and Indefrey (2003).

2.4 Speech Coding

A native speaker of German transcribed speech of all narratives using German standard orthography, also taking note of filled pauses, word truncation, repetitions, etc. We then identified all referential expressions mentioning an entity for the first time. For the purposes of this study, we only selected references to concrete animate (i.e., the fairies) and inanimate entities (e.g., cake, broom, needle; see Appendix B for a full list of entities) that played a role in the story. We excluded references to referents representing general knowledge (such as die Luft ‘the air’) and to all abstract/non-

² However, if one of the participants knew that the experimenter researched gestures (e.g., if a research assistant from the local university working on the topic of gestures came with a friend), then she was automatically assigned as listener.
spatial/immaterial objects (3). We also excluded references to ‘non-referential referents’ (Chafe, 1994). Non-referential referents do not factually exist at the moment of mention, and speakers typically mention them in an irrealis context or present their existence as ‘hypothesized, predicted, or denied’ (Chafe, 1994; Example 4). Importantly, non-referential referents are not trackable, and thus represent a different category of referents than those that are to be explored in the present study. Finally, we also excluded references to the pictures themselves (5).

(3)

sie hat eine Idee

‘she has an idea’

(4)

das soll vielleicht so ein Mehlsack sein

‘it should perhaps be a bag of flour’

(5)

die grüne äh steht in der Mitte des Bildes

‘the green fairy stands in the middle of the picture’

Entities were either mentioned as core arguments (subjects and direct objects) in presentative utterances (such as existentials or locatives), transitive or intransitive clauses (corresponding to 60% of all referential expressions; 6-12; relevant expressions are underlined). In all three of these utterance types, the starting point is typically either an inanimate or animate locational element (e.g., 6 vs. 9), the dummy subject es ‘it’ or the adverbial da ‘there’, so that the first mentioned entities are placed toward the end of the utterance. In intransitive utterances, the speakers can further use subject-verb inversion in order to place the first mentioned entity toward the end of the utterance (12). Placing the referents in utterance final (focal) position is typical in the context of first mentions. The rest of the entities were instantiated as either oblique arguments (29% of all referential expressions; 13), or in verbless utterances (11% of all referential expressions; 14).
(6) und in dieser Schüssel sind drei Zauberstäbe
‘and in the bowl are three wands’

(7) es gibt einen Tisch
‘there is a table’

(8) da sind drei Feen
‘there are three fairies’

(9) die hat n Eimer
‘she has a bucket’

(10) sie holt ein kleines Kästchen
‘she goes to get a little box’

(11) da kommen Funken raus
‘there are coming out sparks’

(12) dann fliegt ein Streichholz herbei
‘then flies by a match’
2.4.1 Information status

For each referential expression, we determined whether it referred to a brand-new or inferable entity. A brand-new entity was a ‘truly’ new entity, which had never been mentioned before, and was not inferentially linked to a previous entity in the discourse. Conversely, an inferable entity corresponded to an entity that was mentioned for the first time, but that was linked to a previous ‘trigger’ entity in the discourse via an inferential link (following Prince, 1981, 1992). In the current data set, two different links connected first mentions to previous entities, namely part/whole (e.g., sleeve – dress, egg shells – eggs), and content/container relationships (e.g., milk – milk can, sugar – sugar bowl; see Appendix B for a full list).

In relation to the way that entities were embedded in different utterance types, we observed that brand-new entities were introduced as core arguments in 67% of the cases, as oblique objects in 21% of the cases, and in verbless utterances in 12% of the cases. Inferable entities were mentioned as core arguments in 41% of the cases, as oblique objects in 50% of the cases and in verbless utterances in 9% of the cases.
2.4.2 Noun phrase definiteness

We considered lexical NPs to be indefinite if they were mentioned as bare nouns, marked by indefinite determiners or numerals (Milch/ein Besen ‘milk/a broom’; drei Feen ‘three fairies’). We considered them to be definite when they were marked by definite determiners, such as definite articles, demonstrative and possessive pronouns (die/diese Fee ‘the/that fairy’; ihr Kleid ‘her dress’).

2.5 Gesture Coding

We used frame-by-frame analysis of digital video in the software ELAN (Sloetjes & Wittenburg, 2008) to annotate manual gestures. We identified the most meaningful part of the gestural movements, the stroke phase (Kendon, 2004; McNeill, 1992), with sound turned off. We consider it necessary to turn the sound off during the annotation of gesture phases to provide an objective and replicable annotation, which is based on physical features of the hand/arm movements alone. We determined the onset and offset of a stroke when there were changes in the trajectory or movement of the hand(s), as well as when there were changes in the tension, shape or placement of the hand(s) (see Kendon, 2004; Seyfeddinipur, 2006 for more detailed descriptions/instructions). In the case of deictic gestures, we counted the accelerated movement toward the end configuration together with the momentary stop in the end configuration as the stroke. For all other gestures, we also included post stroke holds, defined as movement cessations of the hand(s) at the end of a gesture stroke, as meaningful parts of the gesture. One of the functions of post stroke holds is to allow for the rest of the co-expressive speech to be uttered before the hand(s) go(es) into retraction or the next gesture (Kita, 1990; McNeill, 1992). They are therefore relevant for our analysis. Since the goal of the current examination is to find out when gestures are aligned with new referents in discourse, it is crucial to take into consideration the full chunk of speech that the meaningful part of the gesture is related to. In a last step, we identified which gestures co-occurred temporally with at least one syllable of the relevant referential expressions (following Gullberg, Hendriks & Hickmann, 2008, and Stam, 2006), and only took those gestures into account for the analyses.

2.6 Reliability Coding

A second German native speaker recoded speech for information status (brand-new vs. inferable) and nominal definiteness (indefinite vs. definite) for the referential expressions of four participants, corresponding to about 20% of the total amount of referential expressions used in the analyses. The agreement between coders was 90% for the coding of information status. Interrater reliability was computed using Cohen’s kappa (Kappa = 0.796, SE of kappa = 0.035). The agreement between coders was 98%
for nominal definiteness coding. The interrater reliability was also measured using Cohen’s kappa (Kappa = 0.979, SE of kappa = 0.012).

A second coder recoded gestures for the same four participants in our data set, identifying gestures in the target clauses (i.e., those containing first mentions). The target gestures in those clauses constitute about 20% of the total amount of gestures that went into the analysis. Agreement was reached when the gesture that coder 2 identified aligned with the same referential expression as the one that coder 1 identified. The agreement between coders was 96%.

2.7 Analyses

The analyses focus on first mentions of referents, brand-new or inferable, encoded by definite or indefinite nominals, and produced with or without gestures. The data set consisted of 1,489 spoken referential expressions, and 811 gestures.

We used linear mixed effects models with the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017) in RStudio (RStudio Team, 2016) for all analyses. Since the dependent variables were categorical and binary, the analyses were run with variants of generalized linear models with binomial error structure. The models analyze how likely a category from the independent variable is to combine with a category from the dependent variable. Crucially, they allow for the use of the raw data instead of using transformed proportions as was the norm in older statistical analyses of categorical data (for a discussion of the problems arising from transforming categorical data into proportions and using ANOVAs for categorical variables in general, see e.g., Jaeger, 2008).

Table 1 summarizes the two main analyses. Analysis 1 concerns speech alone, examining the relationship between the information status of referents and their formal representation in speech as definite versus indefinite nominals. Analysis 2 then examines whether the presence of gesture is modulated by these variations in information status and definiteness.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Dependent variable</th>
<th>Levels</th>
<th>Predictor variable</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Definiteness</td>
<td>Indefinite/Definite</td>
<td>Information status</td>
<td>Brand-New/Inferable</td>
</tr>
<tr>
<td>2</td>
<td>Presence of gesture</td>
<td>yes/no</td>
<td>Information status</td>
<td>Brand-New/Inferable</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Definiteness</td>
<td>Indefinite/Definite</td>
</tr>
</tbody>
</table>

Table 1. Variables and levels
3 Results

3.1 Speech

In a first step, we explored the relationship between information status and definiteness in speech alone (table 1, analysis 1). Figure 6 presents the observed distribution of indefinite nominals across brand-new (82%) and inferable referents (27%). The analyses revealed that, as expected, brand-new referents were significantly more likely to be expressed as indefinite nominal expressions than inferable referents. Conversely, inferable referents were significantly more likely to be encoded with definite than with indefinite nominal expressions ($EST = -5.83$, $SE = 0.32$, $z$-value = -18.46, $p = .000$).

![Figure 6](image)

**Figure 6.** Mean proportions of indefinite nominals representing brand-new referents (0.82; SE = .02) versus inferable referents (0.27; SE = .02)

3.2 Gesture

Next, we examined the relationship between the incidence of gestures and first mentions. We found that speakers produced gestures for 55% ($SD = 23\%$) of all first mentioned entities (mirroring 60% in Foraker, 2011). We tested whether the incidence of gesture is modulated by two independent variables, namely information status operationalized as brand-new versus inferable, and referents’ representation in speech as indefinite versus definite nominals (table 1, analysis 2). Figure 7 presents the observed distribution of (mean proportions of) gestures across inferable (65%) versus brand-new referents (52%). Figure 8 presents the observed distribution of (mean proportions of) gestures across definite (56%) versus indefinite (54%) nominals.
We ran 5 different models in order to determine the model that fit the data best. The first model included no predictor variables. The second and third models each included only one predictor variable, information status and definiteness, respectively. Finally, the fourth and fifth models included both predictor variables, but one was a simple model and the other an interaction model. All models included ‘subject’ as a random predictor variable. We compared AIC values (the Akaike information criterion) between all models in order to determine the model which represents the data set best. More specifically, the AIC is an estimate of predictive accuracy, which measures how well a regression model will fit when applied to a new sample (see Long, 2012 for a
detailed description). Lower AIC values correspond to better fit (see Appendix A for a full list of models ranked according to their AIC values).

The model comparisons showed that the simple model including the two predictor variables, information status and definiteness, best explained the present data. The analysis revealed that there was a significant effect of information status on the incidence of gestures, but in the opposite direction from the prediction. Inferable referents were significantly more likely to occur with gestures than brand-new referents ($EST = -0.73$, $SE = 0.16$, $z$-value = -4.51, $p < .000$). There was no significant effect of definiteness ($EST = -0.25$, $SE = 0.15$, $z$-value = -1.68, $p = .092$).

4 Discussion

The existing literature on discourse reference and gestures has shown that gestures are sensitive to referents’ information status in discourse such that they occur more often with new referents/first mentions than with given referents/subsequent mentions. However, because not all new entities are gestured about at their introduction, the current study set out to examine when first mentions of discourse entities are accompanied by gestures and when they are not. In particular, we considered the possible connection between gesture production and a more fine-grained difference in information status between brand-new and inferable entities, as well as the variation in linguistic encoding between indefinite and definite nominals, reflecting this difference in speech.

The results can be summarized in two points. First, the speech results showed that, as predicted, brand-new referents tend to be expressed by indefinite nominals (e.g., ‘a broom’), whereas inferable referents tend to be expressed by definite nominals (e.g., ‘the broomstick’). These findings are in line with previous research on this topic (e.g., Clark, 1975, 1977; Fraurud, 1990; Prince, 1981, 1992; see also Hickmann, Hendriks, Roland & Liang, 1996 for marking of newness in German), showing that referential form is sensitive to the inferability of referents mentioned for the first time.

Second, the gesture results revealed a link between gesture production and the brand-new/inferable distinction. Contrary to prediction, however, inferable referents were significantly more likely to be accompanied by gestures than brand-new ones. To give an example, the brand-new referent ‘dust pile’ is introduced in a presentative utterance in *man sieht da vorn dran sonen kleinen Haufen* ‘one sees there in front a little pile’, where no gesture co-occurs with the first mention. Compare this to the first mention of the inferable referent ‘egg yolk’ in the presentative und *man sieht jetzt das Eigelb* ‘and one sees now the egg yolk’, in which a gesture localizing the egg yolk above a bowl accompanies the referential expression denoting the inferable entity (Figure 9).
This result poses a challenge to McNeill’s (1992, 2005) theory of communicative dynamism and gestures, which posits that the more a piece of information pushes the communication forward, the more likely it is to co-occur with a gesture. It seemed plausible to assume that brand-new referents, which mark the lowest degree of accessibility of referents in discourse, push communication forward more than inferable referents, and would thus be accompanied by gestures more often. However, the current results do not support this assumption. Rather, the results suggest that gestures are used to indicate or enhance the accessibility of entities that need to be recovered inferentially.

Returning to the main question of the current study, we asked whether information status plays a role for the incidence of gestures with first mentioned entities in discourse. The results show that this is the case: gestures are significantly more likely to occur with inferable than with brand-new referents. Although these results go in unanticipated directions, they still suggest that gesture production is sensitive to the subtle distinction in information status suggested by the difference between brand-new and inferable referents. The findings therefore generally support previous research on the relationship between information status and gesture production in discourse (e.g., Foraker, 2011; Gullberg, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill & Levy, 1993).

A more important question arises, namely why gestures should be more strongly linked to inferable than to brand-new entities. We propose that gestures in the context of first mentioned entities are used to indicate or enhance the accessibility of inferable referents. As discussed above, inferable entities are considered ‘discourse old’ (Birner,
2013) in the sense that they are inferentially accessible in the discourse, but ‘hearer new’ (Birner, 2013) in the sense that they are not active in the current discourse space (Chafe, 1994; Fillmore, 1982; Lambrecht 1994; Langacker, 2001). We suggest that speakers use gestures to highlight the inferable pieces of information in order to signal to the addressee that, even if the referents are marked by definite determiners, they are still to add them as new information to the current discourse space. In other words, since inferable entities are linguistically encoded similarly to given information (by definite nominals), speakers may produce gestures more often with them to signal to the addressee that the information is not in fact given, but new since there is not yet any active representation of the information in the discourse space. By this account, gestures and speech in this particular context work together in a complementary rather than a parallel fashion. That is, when speech does not provide an unambiguous clue as to whether information needs to be newly added to the current discourse space (such as by indefinite nominals), gestures can do so instead.

In relation to the view on discourse in cognitive linguistics, there is little evidence to suggest that inferable referents are added to a current discourse space differently than given referents. In fact, Langacker (2001) proposes that the definite article alone achieves the function of signaling to the addressee that the inferable referent is available in the ‘minus frame’ (i.e., previous discourse) and needs to be added to the current discourse space (Langacker, 2001: 168). The results in the current study challenge this view and suggest that gestures play a role as well. Specifically, this means that while the definite article might be sufficient to make a given referent accessible to the addressee, it needs support from a gesture when it comes to making an inferable referent accessible to the addressee. This view does not preclude that given referents are at times also accompanied by gestures. In fact, there are circumstances under which already-mentioned referents move from an ‘active’ to a ‘semi-active’ state (e.g., when they have not been mentioned for a long time; Chafe, 1994; Givón, 1983), in which case they might, similarly to inferable referents, also need enhancement by a gesture in order make them accessible. This is also supported by quantitative studies on gesture production in referent maintenance contexts (Gullberg, 2006).

The interpretation of a complementary function of gestures in the context of discourse that we provide here is something of a departure from previous studies which have mainly emphasized that the two modalities work in parallel. However, the interpretation is commensurate with McNeill’s (1992) view on gestures and speech as two dimensions of the same idea unit, where gestures do not always represent the same information as speech. The suggestion is that together, speech and gesture form a more complete representation. Similarly, Kendon (2014) suggests that gestures and speech together form a richer and more complex expression than if words or gestures are considered alone. In order to form such a complex expression, gestures can be used in flexible ways, as complements or supplements, sometimes even as substitutes or alternatives, to spoken expressions, always in accordance with the underlying
communicative effort or intent (Kendon, 1986). The two modalities can thus be seen as adaptable resources allowing speakers to vary how they coordinate them depending on the communicative needs in different types of situations (Gullberg, 1998; Holler & Beattie, 2003; Kendon, 2004).

4.1 Limitations and Future research

A first important limitation is that although referent inferability explains a considerable part of the data, we still find inferable referents that are not accompanied by gestures (36%), as well as brand-new referents that do co-occur with gestures (52%). This means that there must be other aspects (possibly related to information status) which affect the presence of gestures in general, and with first mentions in particular. One aspect concerns the operationalization of inferability. In the current study we only considered inferential relations between first mentioned and already-mentioned trigger entities in the narrow sense of a frame. However, previous research suggests that a first mentioned entity can also be inferentially related to previously mentioned frames or scripts in a much broader sense (e.g., Fillmore, 1982; Schank & Abelson, 1977; see also Ward & Birner, 2001; Ward & Hirschberg, 1985; Ward & Prince, 1991). For instance, after having mentioned a baking situation, a speaker might refer to the referent ‘spoon’ with a definite nominal because she considers it inferable given that people often use spoons when baking. It is worth considering such relations in future studies.

A further aspect is more linguistic in nature. Firbas (1971), in his original work on communicative dynamism (CD) in discourse, suggests that the amount of CD a speech unit carries (whether it is a referential expression, a verb or any other unit of meaning) does not solely depend on information status, but also on the semantics and the word order used in a given utterance. It is therefore possible to complement an analysis of information status of first mentions with, for instance, the semantics of the verbs used to introduce an entity into discourse, or the position of the referent in the utterance. It is already known that semantics plays an important role in the way that gestures represent information (e.g., Gullberg, 2009, 2011; Gullberg et al., 2008; Kendon, 2004; Kita & Özyürek, 2003; McNeill, 1992, 2005). However, it is rather unclear whether and if so how the semantics of a referential expression and/or the verb used to introduce a referent would also affect the incidence of gestures. Other studies suggest a relationship between the way speakers package information morpho-syntactically and the way that gestures represent information (e.g., Gullberg et al., 2008; Kita & Özyürek, 2003; Kita et al., 2007; Özyürek, Kita, Allen, Furman & Brown, 2005). However, also for these studies, it is unclear how morpho-syntactic packaging would influence the incidence of gesture rather than the mode of representation in gesture. Hence, examining the interplay between semantics, word order, and information status in discourse might provide further useful insights into why some entities occur with
gestures and others do not, and on the relationship between gestures and speech on the discourse level more generally.

Finally, there are other non-discursive aspects to consider. For instance, some entity properties may be particularly conducive to gesture production. Different objects afford action on them to different degrees, which in turn may affect how likely people are to gesture about them. For example, Chu and Kita (2016) found that speakers produced speech-associated gestures more often when the stimulus objects they saw afforded action (i.e., objects with a smooth surface) than when they did not (i.e., objects with a spiky surface). Another issue is familiarity. For instance, if someone is not, or supposes the addressee is not, familiar with a certain entity or action, such as decorating a cake with an icing bag, they might be more likely to gesture about it (cf. Campisi & Özyürek, 2013). Lastly, of course, it is also possible that the specific task in this study might have influenced why speakers did or did not gesture about entities. For instance, we encouraged speakers to say something about each picture, which might have led them to talk about aspects of the stories that they would have left out otherwise. When speakers leave out information in a narrative context, it is typically because the information is not relevant to the story at hand, or because the information is old/given. It is therefore possible that this is the reason why some speakers refrained from gesturing about certain entities they talked about. These suggestions will have to be explored in future studies.

4.2 Conclusion

The current study examined when gestures are produced in co-occurrence with first mentions of entities in discourse. Specifically, it tested whether there is a relationship between the incidence of a gesture and an entity’s information status as brand-new versus inferable, as well as the referent’s formal expression in terms of nominal definiteness. The literature on bimodal reference has suggested that gestures tend to be used with new rather than given information, suggesting that the two modalities mirror each other in the construction of discourse (i.e., the newer the referent, the fuller the referential form, and the more likely a gesture). The current results differ from previous studies by showing that gestures are in fact more likely to co-occur with inferable referents (which are considered more easily accessible because they have previously/indirectly been activated by a frame) than with brand-new referents (which are considered to be inactive at the moment of utterance). Therefore, in the context of first mentions in discourse, gestures seem to fulfill a complementary function to speech. Future work will need to explore wider measures of the notion of frame (Fillmore, 1982) and its interaction with the incidence of gestures, since the current study has applied a rather narrow sense of the term (as in Prince, 1981, 1992). Furthermore, more studies are needed that explore the specific utterances in which first mentions are embedded in order to test whether they interact with the incidence of gestures. Finally,
testing the affordances of and speakers’ familiarity with certain objects might also shed light on the nature of discourse level phenomena.

In conclusion, the study has provided new evidence for the claim that the incidence of gestures in discourse is related to the referential status of entities. The focus on first mentions in relation to gesture is novel and, unlike previous studies on this topic suggesting a parallel link between the modalities, this study reveals a complementary function of speech and gestures in discourse. We propose that speakers use gestures to signal that inferable referents, despite their inferential link to the previous discourse, are new to the hearer and that, consequently, addressees need to add them as new to the current discourse representation. Gestures may help them do this. The findings are in line with the view that gestures and speech work together to build a coherent piece of discourse, but further highlight the many and flexible functions that gestures can fulfill in relation to speech in general and in bimodal discourse reference in particular.

5 References


Appendix A: Model selection on the basis of AIC (Long, 2012)

<table>
<thead>
<tr>
<th>Specification</th>
<th>Predictors</th>
<th>AIC</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Gesture ~ 1 + (1</td>
<td>Subject)</td>
<td>none</td>
<td>1795.1</td>
</tr>
<tr>
<td>Model 2: Gesture ~ information status + (1</td>
<td>Subject)</td>
<td>1</td>
<td>1778.4</td>
</tr>
<tr>
<td>Model 3: Gesture ~ definiteness + (1</td>
<td>Subject)</td>
<td>1</td>
<td>1796.4</td>
</tr>
<tr>
<td>Model 4: Gesture ~ information status + definiteness + (1</td>
<td>Subject)</td>
<td>2 (simple)</td>
<td>1777.5</td>
</tr>
<tr>
<td>Model 5: Gesture ~ information status * definiteness + (1</td>
<td>Subject)</td>
<td>2 (interaction)</td>
<td>1779.4</td>
</tr>
</tbody>
</table>

Appendix B: List of entities

Fairy 1, Fairy 2, Fairy 3, broom, bucket, mop, pile of dust, wand(s), hat, fairy’s dress, leaves, stairs, stars/sparks, pot, bowl, table, cake, candles, flame, cleaning agent, water, match, cookbook, milk can, saltshaker, sugar bowl, bowl, spoon, flour bag, eggs, icing bag, dough, sugar hearts, sugar dots, sprinkles, shoes, mannequin, dress, needle(s), thread, basket, belt, scissors, pieces of cloth (different colors), collar, box, cloth triangle, bow(s)

Parts/content of entities:

broomstick, bristle, border of bucket, hand, arm, head, eye, foot, buttocks, face, shoulders, finger, mouth, parts of the dress (upper, lower, skirt part, edge, neckline, sleeves), parts of mannequin (hip, upper/lower body, arms, waist, shoulder, belly), parts of cloth, parts of stairs (stair head, steps), corner/part of book, egg yolk, egg white, egg shells, salt, sugar, flour, milk, parts of table (edge, middle)
Paper II: Addressees are sensitive to the presence of gesture when tracking a single referent in discourse

Abstract

Production studies show that anaphoric reference is bimodal. Speakers can introduce a referent in speech by also using a localizing gesture, assigning a specific locus in space to it. Referring back to that referent, speakers then often accompany a spoken anaphor with a localizing anaphoric gesture (i.e., indicating the same locus). Speakers thus create visual anaphoricity in parallel to the anaphoric process in speech. In the current perception study, we examine whether addressees are sensitive to localizing anaphoric gestures and specifically to the (mis)match between recurrent use of space and spoken anaphora. The results of two reaction time experiments show that, when a single referent is gesturally tracked, addressees are sensitive to the presence of localizing gestures, but not to their spatial congruence. Addressees thus seem to integrate gestural information when processing bimodal anaphora, but their use of locational information in gestures is not obligatory in every discourse context.

1 Introduction

Discourse needs to be cohesive for addressees to understand it. They have to know at all times who is doing what to whom. Therefore, speakers need to manage reference to discourse entities constantly and consistently, a process known as anaphoric reference. Most entities are mentioned multiple times throughout discourse. When a speaker mentions a referent for the first time, she will typically use a rich referential expression (e.g., an indefinite lexical noun phrase, ‘a girl’). Once the referent is introduced, the speaker has a choice of different anaphoric expressions with which to refer back to it. Depending on the referential context, that is whether the referent is maintained from one clause to the next or reintroduced after a gap, the speaker will either choose a lean expression (e.g., a pronoun, ‘she’) or a rich one (e.g., a definite lexical noun phrase, ‘the
Beyond that, speakers can also realize visual anaphoric reference through speech-accompanying gestures (Kendon, 2004; McNeill, 1992). Various studies have described localizing gestures as playing an important role in this endeavor (e.g., Gullberg, 1998, 2003, 2006; Kendon & Versante, 2003; McNeill & Levy, 1993; So, Kita & Goldin-Meadow, 2009). Speakers tend to assign a location in gesture space to a referent at its introduction by way of a localizing gesture (Figure 1a). They can then reuse this location when the referent is referred back to later in discourse (Figure 1b). The second gesture is what we call a localizing anaphoric gesture. Importantly, the use of localizing anaphoric gestures depends on the discourse context. When speakers maintain a referent, they are less likely to align a gesture with the spoken referential expression (often a pronoun). But after a gap, when speakers need to reintroduce a referent (often using a richer nominal form), they frequently also accompany the mention of the referent with a localizing anaphoric gesture (Gullberg, 2006). Thus, gestures reflect the information status of a referent in parallel with speech. In production, less marking material is used for highly accessible referents (pronominal forms + absence of gesture), and more marking material is used for the reactivation of referents (nominal forms + localizing gestures) across both modalities (i.e., speech and gesture; Debreslioska & Gullberg, 2017; Gullberg, 1998, 2003, 2006; Levy & McNeill, 1992; Marslen-Wilson, Levy & Tyler, 1982; McNeill, 1992; McNeill & Levy, 1993; Perniss & Özyürek, 2015; Yoshioka, 2008). Interestingly, some production studies also suggest that parts of the gestural reactivation process are meant for the listener. For instance, Gullberg (2006) showed that speakers adhered more consistently to locations set up by localizing gestures when addressees could see them than when they could not (i.e., when speakers and addressees were separated by a screen preventing eye contact and gesture visibility). Furthermore, Gullberg (1998, 2011) showed that in interactive stretches, some addressees even pointed back to locations previously established for referents by the speakers. This is interpreted as evidence that addressees understand when spatial representations of referents were created.
Figure 1. Example of a localizing gesture at a referent's introduction (1a) and a localizing anaphoric gesture at a referent's reintroduction after a gap of absence (1b). The stroke phases of the gestures pictured in this Figure align with the words in bold face. The first picture (1a) shows a gesture which indicates a spatial area (indicated by the dot in white) above the speaker's right knee for the referent at its introduction. The second picture (1b) shows a gesture which places the referent exactly in that area at its reintroduction after a gap of absence (the white arrow indicates the gestural movement toward the spatial area).
A wealth of perception studies on cross-modal information integration support the fact that addressees integrate information from gestures with the meaning in speech. Evidence for this view comes from behavioral studies (e.g., Beattie & Shovelton, 1999; Graham & Argyle, 1975; Kelly, Barr, Church & Lynch, 1999; Kelly, Özyürek & Maris, 2010; Riseborough, 1981; Thompson & Massaro, 1986, 1994), ERP studies (e.g., Kelly, Ward, Creigh & Bartolotti 2007; Özyürek, Willems, Kita & Hagoort, 2007; Sheehan, Namy & Mills, 2007; Wu & Coulson, 2007); and fMRI studies (e.g., Dick, Mok, Beharelle, Goldin-Meadow & Small, 2014; Holle, Gunter, Rüschemeyer, Hennenlotter & Iacoboni, 2008; Skipper, Goldin-Meadow, Nusbaum & Small, 2007; see also Hostetter, 2011, and Kendon, 1994, on the communicative function of speech-accompanying gestures).

Perception studies specifically testing the processing of localizing anaphoric gestures also generally support this view. Some studies suggest that localizing gestures that are congruent with previous gestures can facilitate processing in comparison to incongruent localizing gestures (Cassell, McNeill & McCullough, 1999), while others suggest that congruent localizing gestures facilitate processing in comparison to speech alone (Gunter & Weinbrenner, 2017; Gunter, Weinbrenner & Holle, 2015). Finally, anaphoric localizing gestures are shown to reinforce expectations about referent resolution in speech (Goodrich Smith & Hudson Kam, 2012; Nappa & Arnold, 2014), and to help identify referents (Sekine & Kita, 2015).

However, there are many inconsistencies within and across those studies, which also suggest addressee insensitivity. For instance, in Cassell et al. (1999) participants retold taped narratives in which they had seen a speaker either use localizing gestures congruently or incongruently. They produced more retelling inaccuracies in the incongruent than in the congruent condition. Interestingly, however, only 32% of all incongruencies resulted in retelling inaccuracies. Thus, although an effect was observed in comparison to the congruent condition, participants were also very likely not to be influenced by the incongruent information provided in gesture (68% of the time).

Similarly, Hudson Kam and Goodrich Smith (2011) found that addressees were insensitive to gesturally established locations for entities in narratives. They showed participants taped narratives in which a speaker used (multiple) congruent localizing gestures for each of two entities, placed left and right. They found that participants did not adopt a consistent perspective when asked to choose one of two pictorial representations of the story. The pictures were always mirror images of each other showing one entity on the right and the other on the left. In another study, Goodrich Smith and Hudson Kam (2012) used similar taped narratives, but in a critical clause, the speaker used an ‘ambiguous’ pronoun (i.e., a pronoun that could refer to either of the two same gendered preceding referents) with a localizing gesture that either matched the first or second protagonist. They found that participants preferred the first protagonist as referent for the pronoun when no gestures were used, replicating the
order-of-mention effect, a well-established cue for pronoun resolution in many spoken languages (e.g., Gernsbacher & Hargreaves, 1988). The presence of gestures indicating the second protagonist changed this pattern, and participants chose the second participant more often (38%). Importantly, however, they still chose the first mentioned protagonist even more often (44% of the time; and 18% of the time, participants did not choose either of the two relevant referents).

Reaction time experiments using comparable designs further show diverging results. On the one hand, Nappa and Arnold (2014) found that addressees profit from gestures that reinforce expectations coming from speech (as in order-of-mention for pronoun resolution), leading to faster responses. But they also showed that addressees were not influenced by gestures that went against expectations drawn from speech, that is, their performance was not slowed down by incongruently used gestures. In contrast, Sekine and Kita (2017) found the opposite. In comparison to speech alone, addressees were slowed down in the incongruent condition, but they were not faster to respond in the congruent gesture condition.

Gunter and Weinbrenner (2017), examined event-related brain responses in participants who watched videos of a person talking about topics of a dualistic nature, introducing and referring back to each topic multiple times by gesturally placing them left and right in gesture space. The results suggested a difference in activation patterns when brain responses to critical expressions accompanied by congruent gestures were compared to those with no gesture, but showed no difference when congruent, incongruent and no gesture conditions were compared.

The contradictory results in these studies may be due to the underlying assumption about the function of anaphoric gestures and to certain methodological choices. We here would like to discuss five important points. First, the natural alignment of speech and gesture in the context of referent tracking is not always taken into account. Production studies show that speakers tend to produce gestures in alignment with nominal forms in reintroduction contexts (e.g., Gullberg, 2006). In discourse, localizing anaphoric gestures thus typically do not have a disambiguating function when it comes to referent identification (e.g., Gullberg, 2006; So et al. 2009), and they typically do not occur with pronouns. In many experiments, however, this is how gestures are used in the stimulus materials with localizing gestures co-occurring with (ambiguous) personal pronouns (e.g., Goodrich Smith & Hudson Kam, 2012; Nappa & Arnold, 2014; Sekine & Kita, 2017).

Second, there is an overemphasis on contrast. Most studies work with contrast between two referents located in two opposite spaces. This choice means that an incongruent localizing gesture for one referent is always produced in the space previously assigned to the other referent. The underlying assumption seems to be that localizing gestures that are incongruent with a referent locus should always be produced in a space that
has already had a meaning assigned to it. It is therefore unclear how a gesture produced in an unassigned location may affect comprehension.

Third, there is a confound of handedness. In all studies, narrators use their right and left hands to locate referents to the right and left in gesture space, respectively. This experimental choice means that it is hard to disentangle which gestural level of representation is crucial for addressees’ processing difficulties or enhancement, that is, whether it is handedness, location, or possibly both. Location and handedness are generally considered two different dimensions of gestural representation processes in discourse (e.g., McNeill & Levy, 1993). While Cassell et al. (1999) attribute the effect in their study to handedness, suggesting that addressees associate each hand with a different referent, all other studies assume location in space to be the determining factor (Sekine & Kita, 2015, 2017; Gunter et al. 2015, Gunter & Weinbrenner, 2017; Goodrich Smith & Hudson Kam, 2012). However, none of the studies provide decisive evidence either way.

Fourth, there is potentially altered allocation of attention to gesture. Some studies have chosen to blur (Gunter et al. 2015, Gunter & Weinbrenner, 2017), or cover the narrators’ faces with masks (Sekine & Kita, 2015, 2017). While this technique might control important aspects of an experiment (e.g., being able to use the same audio for different videos), it also means that participants’ attention to the gestures may be increased. There is evidence that addressees typically focus their gaze on the speaker’s face and only process gesture input in peripheral vision (Gullberg & Holmqvist, 1999, 2006). However, with the face masked or blurred, attention allocation is likely to be altered towards gestures. In addition, in some of the same studies (Gunter et al. 2015; Sekine & Kita, 2015, 2017), the gestures were produced at shoulder height, which also draws more attention to them, considering that this is a rather marked area for gesture production (McNeill, 1992 for coding scheme of gesture space; Müller, 1998). In Sekine and Kita (2015, 2017) the narrators further used marked resting positions for the hands after they had performed the gestures. That is, when narrators had gesturally introduced referents by locating them, they held their hands in those spaces (at shoulder height) for the rest of the narrative. This might lead to over-specification since the locations were kept active throughout the narrative.

Finally, there is a lack of distractors and control of possible learning effects. Only two studies report using distractor items or items with gestures fulfilling other functions in relation to speech (Cassell et al. 1999; Goodrich Smith & Hudson Kam, 2012). By not including distractors, studies may have increased participants’ awareness of the topic being studied, that is, drawn attention to gestures with a referential function more generally, and possibly even to location/handedness of gesture in particular. This is especially important in experiments in which only a congruent condition was compared to a no gesture condition (Gunter & Weinbrenner, 2017, experiment 2). Such a design might have led participants to learn over the course of the experiment that all gestures
reliably have the same function because they always provided the same information, and thus that the gestures have to be useful for the task at hand.

In contrast to previous studies, the current experiments focus on a more naturalistic setting, in which the speaker’s face can be seen and gestures are produced in central gesture space (i.e., between chest and hip height and relatively close to the body on the left and right), while still controlling for handedness and learning effects. Moreover, the current study goes beyond previous research by testing participants’ sensitivity to anaphoric gestures in the context of a single gesturally tracked referent. The study therefore takes a first step towards addressing the potential methodological confounds discussed above.

2 The current study

The present study examines whether addressees are sensitive to the use of localizing anaphoric gestures. We conducted two reaction time experiments with differing tasks comparing performance in three conditions: gesture congruent, gesture incongruent, and no gesture. The same stimulus narratives were used in both experiments. In the gesture congruent condition, a referent is introduced with a localizing gesture in utterance 1 and reintroduced after an intervening utterance by a localizing anaphoric gesture in utterance 3. In the incongruent condition, the referent is reintroduced by a localizing gesture in a different, previously unassigned location in space (note that an incongruent gesture is not technically an anaphoric gesture). We also added a no gesture condition as a baseline condition in both experiments. The results from the comparisons between the gesture and no gesture conditions need to be considered very cautiously though, because they are not perfectly comparable. In the gesture conditions, the spoken referential expressions are aligned with the stroke phase (or the most meaningful part) of the localizing gestures. However, gestures have a preparatory phase that precedes the stroke, and therefore typically start before the referential expressions are uttered. Thus, gestures might provide information to addressees before the referential expressions are even produced. This is, of course, not the case in the no gesture condition.

In experiment 1, the task for participants was to answer a question about an action performed by the tracked referent in a fourth and critical utterance. The assumption was that a preceding congruent localizing gesture should facilitate responses to the content question, whereas a preceding incongruent gesture could render decisions regarding the referent more difficult (cf. Sekine & Kita, 2017 for a similar task). In experiment 2, the participants saw the referent to be tracked in written form before the start of the narratives. Their task was to press a key as fast as possible every time they encountered the referent during the subsequent narrative. The assumption was that
congruent gestures speed up the recognition of a bimodal anaphoric expression, whereas incongruent gestures slow it down.

In addition, the present study differs theoretically and methodologically to the existing literature on bimodal anaphor perception in a number of ways. First, the narrator gesturally tracks only one referent rather than two. The assumption is that if addressees indeed associate a certain location with a discourse referent, then that will be the case even if there is no contrast between that referent and another. Second, the referent is located twice in the narrative, once at its introduction and once at its reintroduction, respecting the discourse context in which localizing anaphoric gestures are typically found in production. The assumption is that addressees can create a spatial representation of a referent in a minimal context, even after only two instances of localization (cf. Sekine & Kita, 2017). Third, the narrator always uses two hands to locate a referent rather than one in order to exclude handedness as a potential confound for referent assignment. Fourth, the narrator’s face is visible and the gestures are produced in central gesture space (cf. Gullberg & Kita, 2009). Finally, we added distractor items with gestures fulfilling other functions to obscure the goal of the study.

3 Experiment 1

We test the hypothesis that addressees are sensitive to the use of localizing anaphoric gestures. We predict a) that participants perform faster in the congruent condition than in the incongruent condition. In relation to the no gesture condition, we predict b) that participants will perform faster in the gesture congruent condition, and more slowly in the gesture incongruent condition.

3.1 Method

3.1.1 Participants

Twenty-eight students enrolled at DEKRA Hochschule, Berlin, Germany, participated in the study (19 female; mean age 23). All participants were native monolingual speakers of German. We recruited participants through notices at the school, and word of mouth. They received a small fee for their participation in the study.
3.1.2 Stimuli/Materials/Design

The experimental stimuli were 50 video-taped narratives told by a female native speaker of German. She produced ten narratives without gestures, 20 with congruent gestures, and 20 with incongruent gestures. The 20 narratives in the congruent condition were the same as the 20 narratives in the incongruent condition. The speaker was trained to perform narratives and accompanying localizing gestures as naturally as possible. She was also trained to keep the rest of her body as still as possible, keep the intonation of her speech as similar as possible, and to speak at a comparable speed across all narratives. The speaker was recorded sitting in a chair with no armrests against a plain, dark blue background. She performed all gestures in central gesture space (coded as ‘center right and left’ in McNeill, 1992; cf. Gullberg & Kita, 2009) because this corresponds to the typical culture-specific area for German speakers (Müller, 1998).

All narratives consisted of 30-35 words, lasted between 8.7-11 seconds, and had the same utterance structure (Example 1). In the first utterance, the main protagonist is introduced with an existential construction and an indefinite lexical NP as grammatical subject (e.g., ‘There was a woman.’). The second utterance is about a secondary character (e.g., ‘husband’), who does not manage to carry out a certain task. In the third utterance, the main protagonist is reintroduced with a lexical NP as grammatical subject (e.g., ‘Then the woman...’), and it is explained how she intends to help the other character with the task. In the fourth utterance, the main protagonist either calls or writes to someone for assistance. This action corresponds to the relevant action verb that participants need to respond to (henceforth called ‘target verb’). The fifth and last utterance served as a wrap-up utterance. There were always 11 syllables between the anaphoric expression and the target verb. We measured the time (in ms) between the onset of the anaphoric expression and the onset of the target verb. The average time was 2,203 ms ($SD = 176$) in the congruent condition, 2,126 ms ($SD = 136$) in the incongruent condition, and 2,001 ms ($SD = 37$) in the no gesture condition. The time difference between onset of the anaphoric expression and the onset of the target verb was added as an additional predictor variable into our models for analysis in order to control for this variation (see analyses in 3.1.5).
Da war eine Frau\textsubscript{1}. Und ihr Mann konnte den Motor in seinem Auto nicht selbst reparieren. Also hat sich die Frau\textsubscript{2} dazu entschlossen, ihren Bruder anzurufen/anzuschreiben. Der soll ihm dann zur Hilfe kommen.

‘There was a woman\textsubscript{1}. And her husband couldn’t repair the engine of his car by himself. So, the woman\textsubscript{2} decided to call/write to her brother. He should come to help him out.’

\textsuperscript{1} Gesture placed in right/left gesture space.
\textsuperscript{2} Gesture placed in right/left gesture space.

In the experimental items, localizing gestures occurred in exact temporal alignment with the first and second referential expressions for the main protagonist. All gestures were performed with two hands (Figure 2-3). 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 gestures were placed in the opposite locations in space, either left or right depending on where the first gesture was placed.
Gesture preparations started between 200 and 680 ms before the onset of the spoken anaphoric expression (see Figure 4). Gesture preparations started slightly earlier in the gesture incongruent ($M = 542$ ms, $SD = 103$ ms) than in the gesture congruent condition ($M = 408$ ms, $SD = 118$ ms). The time difference between onset of the gesture preparation in relation to the onset of the spoken anaphoric expression was added as an additional predictor variable into our models for analysis in order to control for this variation (see analyses in 3.1.5). Referential expressions used for the main protagonists were the common nouns Mädchen ‘girl’ and Frau ‘woman’ for the gesture conditions, and Junge ‘boy’ and Mann ‘man’ for the no gesture condition.

We also created 30 distractor narratives that differed from the experimental narratives in various ways. Half of the distractor items differed in the spoken clausal structure for the introduction and reintroduction of protagonists; the other half retained the structure of the experimental items. In half of the distractor items, the introductions of referents were accompanied by two-handed localizing gestures as in the experimental items, whereas in the other half the introductions of referents were not accompanied by gestures. In all distractor items the narrator also naturally performed other gestures, mostly depicting actions or simple beats, that were aligned with (and thus highlighted) parts of speech other than the relevant referential expressions. Distractor items also varied in length, and crucially, differed in terms of where the target verb was mentioned. This was done in order to ensure that participants stayed attentive to the content of the speech at all times.

We created two versions of the experiment each with 30 experimental items (10 gesture congruent, 10 gesture incongruent, 10 no gesture items) mixed with 30 filler items. Each participant saw an experimental narrative in only one version, congruent or incongruent. Between participants the versions were counter-balanced. In the experiment, experimental trials alternated with distractor trials. Otherwise, the order of the trials was randomized.
3.1.3 Post-processing and stimulus selection

We used a Canon Legria High Definition16E consumer camera to tape the narratives. The recording format was AVCHD. The videos were transformed into .mpg files with a frame rate of 25 frames per second and a resolution of 1920 × 1080 and edited in Adobe Premier Pro video editing software (cropping, cutting beginning and end of videos, color adjustment for normalization purposes).

Each narrative was videotaped 10-15 times to allow the actor to practice and perform as naturally as possible. The criteria for selecting the best instance of each item were that spoken referential expressions overlapped in time with the localizing gestures, and that gesture handshape and location in space corresponded between the first and second gestures (Figures 2-3). We analyzed the recordings in the video annotation software ELAN (Sloetjes & Wittenburg, 2008) and identified the gesture stroke, defined as the expressive and meaningful part of the gesture movement, to determine whether or not the stroke phase temporally aligned with the corresponding spoken referential expression. The narrative was excluded if this was not the case. At least one syllable of a relevant referential expression had to be temporally aligned with the time it took the speaker to perform the stroke phase (cf. McNeill, 1992). Other parameters, such as intonation, blinking, head position of the speaker or movement of other body parts were carefully observed, and those narratives that matched each other as closely as possible on all parameters were selected as stimuli.

3.1.4 Procedure

The experiment was carried out in a quiet room at the university. The clips were presented on a laptop running E-Prime version 2. The room was darkened (blinds down at all times) in order to avoid differences in lightning during the day and possible reflections on the screen. The experimenter first orally introduced the experiment. Participants then read specific instructions on paper. Their 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 ‘yes’ (ja) or f for ‘no’ (falsch) 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 write instead of call, see Example 1). 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, mirroring the difference between experimental items and two kinds of distractor items.

The experiment lasted 10-15 minutes, after which participants filled out a consent form, and a language and background information questionnaire. The experimenter debriefed participants verbally.

We had to exclude data from 4 participants because more than a third of their responses were incorrect or given after the narratives had ended. The analyses were performed on the remaining 24 participants.

3.1.5 Analyses

We fitted linear mixed effects models with the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017) in RStudio (RStudio Team, 2016) to the participants’ response times. Response times were time locked to the onset of the relevant part of the target verb (rufen ‘call’ or schreiben ‘write’; see Example 1). We excluded 16 incorrect trials (4 congruent, 6 incongruent, 6 no gesture) from the analysis (i.e., when participants responded incorrectly to the question). Furthermore, we excluded three responses that were given earlier than 100 ms after the onset of the target verb, and 24 responses given after the narrative had ended, corresponding to a total of 5.9% of the data.

The predictor variables were (1) experimental condition (congruent, incongruent, no gesture), (2) the time difference between onset of gesture preparation and onset of spoken anaphoric expression, (3) the time lag between onset of spoken anaphoric expression and onset of target verb, and (4) trial number. We also added random intercepts for each subject. Note that we also ran models with random intercepts for each experimental item, but since there was no difference between the models, we report only on the simpler ones here (but see Appendix A for information about the additional models).

It is important to control for (2) the difference between the onset of the gesture preparation in relation to the onset of the spoken anaphoric expression since it is well-known that gestures usually start before the onset of the expression to which they are semantically related (e.g., Kendon, 1972; Schegloff, 1984). Thus, an anaphoric gesture might provide information about which entity will be mentioned before the spoken expression itself has been produced and before the gesture stroke has begun. Moreover, since there is natural variation between the different items in our material, it is important to take that into account. There is also natural variation in terms of (3) the temporal distance between the anaphoric referential expression and the target verb which also needs to be controlled for.

Since (2) the time difference between onset of gesture preparation and onset of spoken anaphoric expression, only applies in the two gesture conditions, we ran two analyses.
In the first analysis, we compare the two gesture conditions (congruent vs. incongruent) including (2). In the second analysis, we compare the two gesture conditions to the no gesture condition by excluding variable (2) the time difference between onset of gesture preparation and onset of spoken anaphoric expression. In this analysis, the no gesture condition is coded as the intercept in the model. We report the estimates derived from the analyses in tables.

3.2 Results

3.2.1 Comparison between the gesture congruent and gesture incongruent conditions

First, we examined the response times in the two gesture conditions. Table 1 shows the estimated response times per condition derived from the analysis. The results suggest that participants were faster to respond in the gesture incongruent than in the gesture congruent condition ($EST = -106.4$, $SE = 37.37$, $t$-value $= -2.85$, $p = .005$). There was no significant effect of (2) the time difference between onset of gesture preparation and spoken anaphoric expression ($EST = -0.04$, $SE = 0.16$, $t$-value $= -0.27$, $p = 0.789$), suggesting that it did not matter when the gesture preparation started in relation to the spoken anaphoric expression for the response times. There was also no significant effect of (3) the time lag between onset of anaphoric referential expression and onset of target verb ($EST = -0.01$, $SE = 0.11$, $t$-value $= 0.07$, $p = .949$), suggesting that the variation in distance between the anaphoric expression and target verb also had no influence on participants’ response times. There was a marginal effect of trial number ($EST = -3.65$, $SE = 1.89$, $t$-value $= -1.93$, $p = .054$), suggesting that participants’ response time decreased over the course of the experiment.

Table 1. Response time estimates with 95% confidence intervals derived from models in experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Congruent gesture (95% CI limits)</th>
<th>Incongruent gesture (95% CI limits)</th>
<th>No gesture (95% CI limits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis 1: RT in ms</td>
<td>861 (313-1409)</td>
<td>755 (201-1308)</td>
<td>-</td>
</tr>
<tr>
<td>Analysis 2: RT in ms</td>
<td>804 (346-1263)</td>
<td>693 (254-1133)</td>
<td>742 (326-1158)</td>
</tr>
</tbody>
</table>

3.2.2 Comparison between the gesture (in) congruent and no gesture conditions

Next, we compared the gesture conditions to the no gesture condition. Table 1 shows the estimated response times per condition derived from the analysis. The results suggest that there was no difference between the gesture congruent and the no gesture condition ($EST = 62.34$, $SE = 37.64$, $t$-value $= 1.66$, $p = .098$). There was also no difference between the gesture incongruent condition and the no gesture condition.
Further, there was no effect of the time lag between onset of anaphoric referential expression and onset of target verb (\(EST = 0.02, SE = 0.10, t\)-value = 0.20, \(p = 0.844\)), again suggesting that this variation had no influence on participants’ response times. Finally, there was an effect of trial number (\(EST = -4.86, SE = 1.50, t\)-value = -3.25, \(p = .001\)), suggesting that participants responded significantly faster at their last trial than at their first.

### 3.3 Discussion

The results from experiment 1 show that, contrary to predictions, participants were faster to respond in the gesture incongruent than in the gesture congruent condition. There were no significant differences between any of the two gesture conditions and the no gesture condition. These results seem to suggest that incongruent localizing gestures might facilitate processing speed in comparison to congruently used (i.e., anaphoric) localizing gestures. No previous studies have reported an advantage of the incongruent condition in comparison to the congruent condition. We can therefore only speculate as to the reasons for this result and will provide possible interpretations in the general discussion.

But first, we explore a possible explanation that is related to the design in experiment 1. Since there is a relatively long temporal distance between the (in)congruent localizing gestures and the target verbs in the stimulus narratives, it is possible that the effect of the (in)congruent gesture had subsided by the time participants came across the target verb. This could explain why there was no difference between the gesture and no gesture conditions. To probe this possibility, we conducted a second experiment with the same set of stimuli, but with a different task. Participants saw a referential expression in written form on the screen before a narrative started and were instructed to track the given referent by pressing a key each time they encountered it during the narrative. This task allows us to measure processing of spoken anaphoric expression \pm (in)congruent gestures more directly, by examining how quickly participants recognize a (bimodal) anaphoric expression.

### 4 Experiment 2

For experiment 2, we make the same predictions as in Experiment 1, namely that a) participants perform faster in the congruent than in the incongruent condition; and b) in comparison to the no gesture condition, participants perform faster in the gesture congruent condition, and more slowly in the incongruent condition.
4.1 Method

4.1.1 Participants

Twenty-nine native German speaking students enrolled as exchange students at Lund University, Lund, Sweden participated in the study (21 female; mean age 24). All participants were monolingual native speakers of German, born and raised in Germany. All of them were international exchange students. They were recruited through social media groups for international students at the university, and by word of mouth. Participants received a voucher for their participation in the study.

4.1.2 Procedure

We used the same stimuli as in Experiment 1. Participants carried out the experiment on a stationary computer in E-prime software (version 3) at in the lab facilities of the university. 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 \textit{j} for ‘yes’ (\textit{ja}) as fast as possible once they encountered the referent. We intentionally avoided using the word \textit{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 \textit{j} for ‘yes’ (\textit{ja}) or \textit{f} for ‘no’ (\textit{falsch}) on the computer keyboard. We added the comprehension questions to ensure that participants stayed focused on the content of the narratives. The experiment lasted approximately 15 minutes. After the experimental session, participants filled out a consent form and a language and background information questionnaire. The experimenter debriefed participants verbally.

We excluded data from two participants. One participant had answered more than a third of the comprehension questions incorrectly, the other one provided only 1 out of 10 responses in the no gesture condition. The analyses were performed on the remaining 27 participants.

4.1.3 Analyses

As in experiment 1, we fitted linear mixed effects models with \textit{lmerTest} package (Kuznetsova et al., 2017) to participants’ response times. We time locked response times to the onset of the spoken anaphoric expression. If participants provided a keypress after they had encountered an anaphor, we assumed that they had recognized the anaphor, and thus used that data point in our analysis. We excluded 14 responses from the no gesture condition because they were given before, or within 100 ms after, the onset of the spoken anaphoric expression (corresponding to 1.7% of the data). Participants further failed to detect (i.e., did not press a key) the anaphoric expression 44 times in total (7 in congruent, 14 in incongruent and 23 in no gesture), which corresponds to another 5.4% of the total data set.
As predictor variables we used (1) experimental condition (congruent, incongruent, no gesture), (2) the time difference between onset of gesture preparation and onset of spoken anaphoric expression, and (3) trial number. We also added random intercepts for each subject. As in experiment 1, we ran two analyses. The first one compared the gesture conditions in order to include (2) the time difference between onset of gesture preparation and onset of spoken anaphoric expression, and the second analysis compared the gesture conditions to the no gesture condition excluding (2). Again, we ran the models with random intercepts for each experimental item, but since there was no difference between the models, we report only on the simpler ones here (but see Appendix B for information about the additional models).

4.2 Results

First, we analyzed response times for anaphor recognition in the two gesture conditions. Table 2 shows the estimated response times derived from the analysis. The analysis revealed no difference between the gesture conditions, meaning that participants were equally fast at recognizing anaphoric expression + gesture in both the gesture congruent and gesture incongruent conditions (EST = -3.36, SE = 26.84, t-value = -0.13, p = .901). Furthermore, there was a significant effect of (2) time difference between onset of gesture preparation and onset of spoken anaphoric expression (EST = -0.22, SE = 0.11, t-value = -2.03, p = .043), suggesting that gesture preparations that started earlier than others in relation to the anaphoric expression provided an advantage for anaphor recognition. There was also an effect of trial number (EST = -4.42, SE = 1.32, t-value = -3.35, p = .001), suggesting that participants’ response time significantly decreased over the course of the experiment.

Finally, we compared response times in the gesture conditions to the no gesture condition (see Table 2 for estimated response times derived from the analysis). The results showed that participants were significantly faster to respond in the gesture incongruent condition than in the no gesture condition (EST = -51.51, SE = 23.48, t-value = -2.19, p = .029), but there was no significant difference between the gesture congruent and no gesture conditions (EST = -18.92, SE = 23.26, t-value = -0.81, p = .416). As in the previous analysis, there was also an effect of trial number (EST = -4.39, SE = 1.09, t-value = -4.04, p = .000), suggesting that participants responded significantly faster at their last trial than at their first.
Table 2. Response time estimates with 95% confidence intervals derived from models in experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gesture congruent (95% CI limits)</th>
<th>Gesture incongruent (95% CI limits)</th>
<th>No gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis 1: RT in ms</td>
<td>651 (548-754)</td>
<td>647 (518-776)</td>
<td>-</td>
</tr>
<tr>
<td>(95% CI limits)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 2: RT in ms</td>
<td>563 (511-614)</td>
<td>530 (471-583)</td>
<td>582 (530-633)</td>
</tr>
<tr>
<td>(95% CI limits)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3 Discussion

The analyses in experiment 2 revealed that gesture congruency did not affect recognition speed of anaphoric expressions; participants were equally fast to recognize anaphoric referential expressions accompanied by congruent or incongruent gestures. Importantly, predictor (2), that is the time difference between onset of gesture preparation and onset of anaphoric expression, indicated that the earlier the preparation phase of the gesture started, the faster participants responded, however any possible location information provided by the gesture before the spoken anaphoric expression started did not matter.

In contrast to the no gesture condition, we found that participants performed significantly faster in the incongruent condition than in the no gesture condition, but there was no difference between the no gesture and the gesture congruent condition. This result suggests that the presence of gesture matters more than their congruence. Previous reaction time studies have either reported no difference between the incongruent and no gesture condition (Nappa & Arnold, 2014) or slower reaction times in the incongruent condition (Sekine & Kita, 2017). These studies, however, have worked with disambiguation and contrast, respectively, whereas in the present study only one referent was gesturally tracked. The incongruent location was previously unassigned and therefore arguably had no meaning. It is therefore difficult to directly compare the results of all three studies. Rather, the current experiment adds to the understanding of the phenomenon by showing that, in the context of gesturally tracking one referent, addressees’ processing seems to be enhanced by the presence of a gesture regardless of its spatial congruence in relation to a previous one.

5 General discussion

The aim of the current study was to examine whether addressees are sensitive to and/or profit from the use of localizing anaphoric gestures (i.e., the congruent condition) when processing a stretch of connected discourse. The results suggest that addressees are indeed sensitive to the use of localizing gestures, but in unexpected ways. Both
experiments showed a lack of processing benefit of congruent gestures over incongruent gestures or a no gesture baseline condition. Instead, the results show that the incongruent condition speeds up performance in comparison to the congruent condition (experiment 1), and to the no gesture condition (experiment 2). The results from both experiments suggest similar interpretations, namely that the presence of gestures matters more than their spatial congruence in relation to a previous gesture in contexts of tracking a single referent and in the absence of ambiguity and contrast.

This interpretation is supported by patterns found in spontaneous speech-gesture production. The natural input for addressees in face-to-face interactions appears to be rather imprecise and/or incongruent when it comes to localizing gestures. Production studies have convincingly shown that speakers reuse a congruent location for a referent previously assigned to a location in space less than half of the time (35% in So et al. 2009, and 42% in Gullberg, 2006). Thus, it is possible that addressees have a high level of acceptance for imprecise and/or incongruently used locations. We therefore assume that the incongruence manipulation in the current study was not perceived as such by addressees. We show a qualitative example from a data set of elicited narrative production to illustrate this point.
5.1 Different location for same referent at introduction and reintroduction

5a
dann läuft er auf diesen Steg zu
‘then he goes onto this bridge’

5b
legt davor seine Sachen aufn Steg
‘puts his things onto the bridge’

Figure 5. Example of an incongruent localizing gesture in spontaneous production
Figure 5 shows an example of an incongruent localizing gesture in spontaneous narrative speech-gesture production. In 5a the speaker produces a gesture which indicates a spatial area to the speaker’s left for the referent ‘bridge’ at its introduction (i.e., first mention). In 5b the speaker reintroduces the referent ‘bridge’ after a gap of 3 utterances and uses a gesture which places the referent to the right side of the speaker.

A further interpretation of the current results is that gestures were perceived as overexplicit in the congruent condition. This would also explain why participants needed more or at least just as much processing capacity to integrate congruent gestures with spoken anaphoric expressions as incongruent gestures. Overexplicitness in speech refers to re-mentions of referents by the use of a noun when a pronoun would have sufficed. In speech perception studies, the repeated noun penalty effect (Gordon, Grosz & Gilliom, 1993) predicts increased processing times for such overexplicit information (see also Vonk, Hustinx & Simons, 1992). We suggest that the use of localizing gestures in the congruent condition in the present study may also have been overexplicit. In fact, the two congruent gestures were used in the context of a rather short piece of discourse with minimal requirements for referent reintroduction in speech (i.e., one intervening utterance containing one intervening new referent as grammatical subject). However, the minimal context justifying a lexical noun phrase to reintroduce a referent in speech may not also be the minimal context for the use of an anaphoric gesture. Some qualitative studies on gesture production in discourse show that (localizing) anaphoric gestures are not only sensitive to the local information status of a referent, but also to bigger units, such as episode boundaries (Marslen-Wilson et al. 1982; Levy & McNeill, 1992; McNeill & Levy, 1993). Those studies indicate that an episode boundary might even be a stronger predictor for the occurrence of an anaphoric gesture. They suggest that more anaphoric gestures are used at the beginning of an episode (or at an episode boundary) than within episodes. Since in our stimulus material anaphoric gestures were used within an episode (with only one intervening utterance), it is possible that participants did not expect a congruent/anaphoric gesture to co-occur with the anaphoric expression and thus, perceived them as overexplicit. The longer processing times in the congruent condition could then reflect a repeated gesture penalty. Further gesture perception research is required to examine the effects of overexplicit gestures on comprehension to support a repeated gesture penalty hypothesis.

Finally, it may be possible that addressees did not interpret the second gestures in the experiments anaphorically as referring back to referents, but rather as referring to the new event (e.g., ‘the woman calling her brother’), since the second gesture occurs close to a discourse marker signaling an event shift (‘so’). Alternatively, addressees may also have interpreted the second gesture as referring to the new referent introduced in the third clause (e.g., ‘her brother’). This could be the case if addressees did not closely track the onset of a gesture (i.e., the second gesture could potentially refer to the new referent ‘brother’ if the time lag between the mentioning of the two referents, ‘the woman’ and ‘her brother’, is rather short; three words in the current material). Both
these options are, in principle, conceivable for gestures with no iconic relationship to the referents with which they align, and/or for second gestures that use the opposite location (as in the incongruent condition). These explanations could potentially also explain the results in Gunter and Weinbrenner (2017). Future studies should test these possibilities by varying the alignment of gestures with referential expressions versus verbs versus other parts of the utterance.

The discrepancy between the results of the current study and previous research on this topic is mainly due to difference in design. In the present study, we used stimulus stories in which we matched production processes very closely, and we used only one referent that was gesturally localized and tracked in space. There was also no contrast or disambiguation worked into the narratives. Thus, our design is different from all previous studies on this topic. Therefore, we conclude that in a context, in which there is a contrast or in which a mismatch needs to be resolved, we can expect the congruent condition to enhance, and the incongruent condition to rather slow down processing (but see Gunter & Weinbrenner, 2017, experiment 1). However, in a context in which location information is used as a means to map discourse onto space without any added disambiguating or contrasting function, the same expectation does not apply. Rather, the presence of a gesture seems to matter more than its spatial congruence, at least at its second appearance.

To test this assumption, we must directly compare the gestural tracking of referents in an ambiguous/contrastive context versus a non-ambiguous/non-contrastive context for different numbers of referents. Finally, evidence about how precise gestural location information actually is in production is rather sparse (but see Gullberg, 2006). Further research should explore how consistent speakers typically are when tracking referents in different contexts. This type of enquiry would greatly deepen our understanding of the phenomenon and bridge the gap between production and perception studies.

5.2 Conclusion

The results from the current study suggest that, in a context of a single gesturally tracked discourse referent, the mere presence of a gesture is more useful to addressees than its spatial congruence. This interpretation is supported by speech-gesture patterns found in spontaneous production, which show that approximate/incongruent locations are rather common when it comes to gesturally tracking a referent. We also suggest that the relatively slow processing of congruent localizing gestures in the current and previous studies on this topic may be due to an overexplicitness of such repeated gestures in the tested contexts (the repeated gesture penalty hypothesis). This proposal will need further supporting research. Most importantly, the study highlights the importance of the context in which localizing anaphoric gestures are examined. The current results stand in contrast to previous studies that have mainly examined contexts
in which anaphoric gestures fulfill a disambiguating or contrastive function. We conclude that gestures can be used to make discourse more coherent for addressees by paralleling referent tracking in speech but that the way gestures are deployed and integrated differs by context and number of referents.

6 References


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.


7 Appendices

Appendix A: Models and additional models Experiment 1

(a)

Summary of mixed effects model 1: comparison of gesture congruent and gesture incongruent conditions

\[ lm.RT \sim c + Onset.RE2 + Onset.target.word + Onset.prep.onset.RE2 + trial + (1|Subject), data \]

Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>27676</td>
<td>166.4</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE2 – target word</td>
<td>-0.00707</td>
<td>0.10960</td>
<td>-0.065</td>
<td>0.94860</td>
</tr>
<tr>
<td>G prep – RE2</td>
<td>-0.04196</td>
<td>0.15658</td>
<td>-0.268</td>
<td>0.78886</td>
</tr>
<tr>
<td>Trial</td>
<td>-3.64696</td>
<td>1.89098</td>
<td>-1.929</td>
<td>0.05446</td>
</tr>
<tr>
<td>Condition:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesture congruent vs. Gesture incongruent</td>
<td>-106.40180</td>
<td>37.37686</td>
<td>-2.847</td>
<td>0.00464</td>
</tr>
</tbody>
</table>

(b)

Summary of additional analysis 1: comparison between gesture congruent and gesture incongruent conditions, including 'item' as random factor

\[ lm(RT \sim c + Onset.RE2 + Onset.target.word + Onset.prep.onset.RE2 + trial + (1|Subject) + (1|Item), data) \]

Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>28000</td>
<td>167.33</td>
</tr>
<tr>
<td>Item</td>
<td>9775</td>
<td>98.87</td>
</tr>
</tbody>
</table>
### Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE2 – target word</td>
<td>-0.007388</td>
<td>0.1524</td>
<td>-0.048</td>
<td>0.9616</td>
</tr>
<tr>
<td>G prep – RE2</td>
<td>-0.04059</td>
<td>0.2162</td>
<td>-0.188</td>
<td>0.8521</td>
</tr>
<tr>
<td>Trial</td>
<td>-3.711</td>
<td>1.843</td>
<td>-2.014</td>
<td>0.0447</td>
</tr>
<tr>
<td><strong>Condition:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesture congruent vs. Gesture incongruent</td>
<td>-106.6</td>
<td>51.42</td>
<td>-2.073</td>
<td>0.0450</td>
</tr>
</tbody>
</table>

(c)

Summary of mixed effects model 2: comparison between gesture (in)congruent and no gesture conditions

\[ \text{lmer(}RT\text{-condition+Onset\_RE2\_Onset\_target\_word+trial+(1|Subject),data)\]}

### Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>27260</td>
<td>165.1</td>
</tr>
</tbody>
</table>

### Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE2 – target word</td>
<td>0.01999</td>
<td>0.10174</td>
<td>0.196</td>
<td>0.844307</td>
</tr>
<tr>
<td>Trial</td>
<td>-4.86310</td>
<td>1.49886</td>
<td>-3.245</td>
<td>0.001238</td>
</tr>
<tr>
<td><strong>Condition:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech vs. Gesture congruent</td>
<td>62.33991</td>
<td>37.63903</td>
<td>1.656</td>
<td>0.098158</td>
</tr>
<tr>
<td>Speech vs. Gesture incongruent</td>
<td>-48.64506</td>
<td>33.45167</td>
<td>-1.454</td>
<td>0.146383</td>
</tr>
</tbody>
</table>

(d)

Summary of additional analysis 2: comparison between gesture (in)congruent and no gesture conditions, including ‘item’ as a random factor

\[ \text{lmer(}RT\text{-condition+Onset\_RE2\_Onset\_target\_word+trial+(1|Subject)+(1|Item),data)\]}

### Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>27230</td>
<td>165.02</td>
</tr>
<tr>
<td>Item</td>
<td>5632</td>
<td>75.05</td>
</tr>
</tbody>
</table>

### Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE2 – target word</td>
<td>0.01993</td>
<td>0.12621</td>
<td>0.158</td>
<td>0.87514</td>
</tr>
</tbody>
</table>
Appendix B: Models and additional models Experiment 2

(a)

Summary of mixed effects model 1: comparison between gesture congruent and gesture incongruent conditions

\[ rt.lmer = \text{lmer}(RT\text{-condition+Onset_prep_onset_RE2+trial+}(1|Subject), data) \]

Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>4758</td>
<td>68.98</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>G prep – RE2</td>
<td>-0.2158</td>
<td>0.1061</td>
<td>-2.033</td>
<td>0.04256</td>
</tr>
<tr>
<td>Trial</td>
<td>-4.4216</td>
<td>1.3199</td>
<td>-3.350</td>
<td>0.00087</td>
</tr>
<tr>
<td>Condition: Gesture congruent vs. Gesture incongruent</td>
<td>-3.3575</td>
<td>26.8363</td>
<td>-0.125</td>
<td>0.90049</td>
</tr>
</tbody>
</table>

(b)

Additional analysis 1: comparison between gesture congruent and gesture incongruent conditions, including ‘item’ as a random factor

\[ lmer(RT\text{-condition+Onset_prep_onset_RE2+trial+}(1|Subject)+(1|Item), data) \]

Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>4758</td>
<td>68.98</td>
</tr>
<tr>
<td>Item</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>G prep – RE2</td>
<td>-0.2158</td>
<td>0.1061</td>
<td>-2.033</td>
<td>0.042552</td>
</tr>
<tr>
<td>Trial</td>
<td>-4.4216</td>
<td>1.3199</td>
<td>-3.350</td>
<td>0.000869</td>
</tr>
</tbody>
</table>
Condition:
Gesture congruent vs. Gesture incongruent

<table>
<thead>
<tr>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.3575</td>
<td>26.8363</td>
<td>-0.125</td>
<td>0.900486</td>
</tr>
</tbody>
</table>

(c)

Summary of mixed effects model 2: comparison between gesture (in)congruent and no gesture conditions

\[ \text{lmer}(RT \sim \text{condition} + \text{trial} + (1|\text{Subject}), \text{data}) \]

Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>3448</td>
<td>58.72</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>-4.391</td>
<td>1.088</td>
<td>-4.037</td>
<td>0.0000598</td>
</tr>
<tr>
<td>Condition:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech vs. Gesture congruent</td>
<td>-18.921</td>
<td>23.259</td>
<td>-0.813</td>
<td>0.4162</td>
</tr>
<tr>
<td>Speech vs. Gesture incongruent</td>
<td>-51.505</td>
<td>23.481</td>
<td>-2.193</td>
<td>0.0286</td>
</tr>
</tbody>
</table>

(d)

Additional analysis 2: comparison between gesture (in)congruent and no gesture conditions, including ‘item’ as a random factor

\[ \text{lmer}(RT \sim \text{condition} + \text{trial} + (1|\text{Subject}) + (1|\text{Item}), \text{data}) \]

Random Effects

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>3430.9</td>
<td>58.57</td>
</tr>
<tr>
<td>Item</td>
<td>230.6</td>
<td>15.19</td>
</tr>
</tbody>
</table>

Fixed Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff. Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>-4.391</td>
<td>1.088</td>
<td>-4.038</td>
<td>0.0000598</td>
</tr>
<tr>
<td>Condition:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech vs. Gesture congruent</td>
<td>-19.063</td>
<td>23.965</td>
<td>-0.795</td>
<td>0.4318</td>
</tr>
<tr>
<td>Speech vs. Gesture incongruent</td>
<td>-51.721</td>
<td>24.180</td>
<td>-2.139</td>
<td>0.0394</td>
</tr>
</tbody>
</table>
Discourse Reference Is Bimodal: How Information Status in Speech Interacts with Presence and Viewpoint of Gestures

Sandra Debreslioska and Marianne Gullberg

ABSTRACT
Speakers use speech and gestures to represent referents in discourse. Depending on referents’ information status, in speech speakers will vary richness of expression (e.g., lexical noun phrase [NP]/pronoun), nominal definiteness (indefinite/definite), and grammatical role (subject/object). This study tested whether these three linguistic markers of information status interact with presence of gestures and gestural viewpoint (observer/character). The results show that gestures are more frequent with less accessible referents expressed with richer spoken forms but that richness of expression does not interact with viewpoint. In contrast, nominal definiteness and grammatical role interact with both presence and viewpoint of gestures. Gestures occur mainly with indefinite lexical NPs and objects. Character viewpoint gestures occur mainly with indefinite lexical NPs and objects plus predicates. The results shed light on when and how speakers use gestures in connected discourse and specifically highlight the discursive function of gestural viewpoint.

KEYWORDS
gestures; mode of representation; referential expressions; information status; bimodal discourse

Introduction
Discourse reference is a bimodal endeavor. Speakers combine spoken referential expressions with precisely timed gestures to represent referents in the flow of information in discourse. Depending on the referents’ information status, speakers vary different properties of the referential expressions and the number or mode of representation of gestures. For instance, if a referent is new to the discourse, speakers tend to use richer and indefinite referential expressions and (in transitive constructions) prefer to introduce the referent as an object (e.g., ‘she saw a fairy with a red dress’). If a referent is given, speakers tend to choose definite and/or leaner referential expressions and maintain the referent as a subject (e.g., ‘the fairy/she was baking a cake’). In gesture, speakers vary the presence versus absence of gestures according to the information status of referents, reflecting richness of spoken expression (Gullberg, 2006; Levy & McNeill, 1992; Marslen-Wilson, Levy, & Tyler, 1982). They produce more gestures with new/less accessible referents expressed with lexical noun phrases (NPs) (i.e., richer expressions) and fewer/no gestures with given/more accessible referents expressed with pronouns (Prons) (i.e., leaner expressions). Furthermore, speakers can vary mode of representation in gesture as a function of information status. Specifically, they may vary gesture viewpoint or the means whereby an entity can be depicted from the perspective of an observer (observer viewpoint [OVPT]) or enacted from the perspective of a character (character viewpoint [CVPT]) to indicate whether a referent is new/less accessible or given/more accessible (Debreslioska, Özyürek, Gullberg, & Perniss, 2013; McNeill, 1992). In fact, McNeill (1992) proposes that the presence/
absence of gestures and gesture viewpoint together form a scale of gesture progression, which is
driven by information status.

In this study we explore this idea by examining the interaction between three linguistic markers of
information status—richness of expression, definiteness, and grammatical role—and the presence/
absence and viewpoint of gestures. The aim is to specify a scale of progression for gesture by
examining different levels of information status. First, we explore richness of expression and its link
to presence/absence of gestures and gesture viewpoint. In his suggested scale, McNeill (1992)
predicts a connection between the progression from Pron to lexical NPs and lexical NPs of differing
size (unmodified vs. modified) and a progression from absence to presence of gesture and then from
OVPT to CVPT gestures. In a first study we explore this hypothesis.

In a second study we examine nominal definiteness and grammatical role and their link to
presence/absence and viewpoint of gestures. Some studies suggest that speakers can vary mode of
representation in gestures, for instance their size, preciseness, and content, according to referents’
information status more generally (as in newly mentioned vs. already given) but also more specif-
ically when expressed in terms of nominal definiteness (as in indefinite vs. definite lexical NPs)
(Gerwing & Bavelas, 2004; Wilkin & Holler, 2010). It remains unexplored whether the presence and
viewpoint of gestures are also sensitive to nominal definiteness. Other studies suggest that gesture
viewpoint varies with grammatical role (Debreslioska et al., 2013), specifically marking the difference
between transitive/intransitive subject referents, which are known to vary in information status in
speech (DuBois, 1987). We examine here whether speakers also use the presence or viewpoint of
gestures to mark the difference between other grammatical roles known to differ in information
status, such as transitive subjects versus objects.

Together, the two studies allow us to evaluate a possible scale of gesture progression. More
generally, they also allow us to shed more light on when and how speakers produce gestures to refer
to discourse referents and to construct a connected bimodal stretch of discourse.

Background

Referential gestures

In the field of gesture studies, gestures are generally defined as expressive and deliberate movements
or actions of the hands and arms that are produced during the act of speaking (Kendon, 2004;
McNeill, 1992, 2005). Speakers use gestures in intricate ways to communicate their messages, and,
importantly, addressees also reliably recognize them as meaningful and communicative (e.g., Beattie
Therefore, gestures are seen as an integral part of the process of utterance production and compre-
hension (Kendon, 1986). Gestures’ systematic and intricate links to language and speech are revealed
in their synchrony with the co-expressive part in speech, such that gestures and speech express the
same underlying idea unit at the same time (Kendon, 2004; McNeill, 2005).

The gestures of interest in the present study are so-called referential (Kendon, 2004) or repre-
sentational gestures (McNeill, 1992). Their main function is to refer to and represent discourse
referents, such as their properties (e.g., shape, size, spatial characteristics, etc.), spatial relations to
other referents, actions, and movements. Referential gestures do this by representing these aspects
through iconicity and deixis (Kita, 2000). McNeill (1992) further suggests that referential gestures
can vary in mode of representation. The two main distinctions are CVPT and OVPT gestures.

In CVPT gestures the speaker’s hand(s) represents the character’s hand(s). In OVPT gestures the
speaker’s hand(s) instead represents one or more characters or entities as wholes. McNeill (1992,
2005) suggests that CVPT gestures are “first-person point of view” gestures, since they correspond to
enactments during which the speaker’s body seems to become part of the gesture space (e.g., the
speaker is pretending to hold an apple and moving it toward her mouth to represent “someone
eating an apple”). In contrast, OVPT gestures are “third-person point of view” gestures, because they
correspond to depictions during which the speaker seems to be looking onto the scene from the outside (e.g., the hand is drawing an entity’s path from left to right through gesture space in front of the speaker).

Importantly, these two ways of representing discourse referents cannot be understood as entirely dichotomous, although this study treats them as such for the purpose of quantitative analysis. Rather, they should be thought of as “dimensions,” since most gestures are “multifaceted” (McNeill, 2005, p. 38). In such cases gestures are often said to have mixed or dual viewpoints (McNeill, 1992; Parrill, 2009; Perniss & Özyürek, 2008; for other ways of distinguishing different modes of representation in gesture, see, e.g., Müller, 2014).

**Constructing bimodal discourse**

Referential gestures play an important role in the construction of discourse, including for discourse reference. Speakers use spoken referential expressions and accompany them with precisely timed gestures to represent referents bimodally in the flow of information in discourse. Moreover, speakers vary properties of both referential expressions and gestures in parallel to reflect referents’ information status.

**Richness of expression and gesture**

Richness of expression is a linguistic marker of information status that interacts with presence/absence of gesture. New referents are typically expressed with richer referential expressions (e.g., ‘a fairy’), whereas given referents are expressed with leaner ones (e.g., ‘she’; e.g., Ariel, 1988; Chafe, 1994; Gundel, Hedberg, & Zacharski, 1993; Lambrecht, 1996; Prince, 1981). Seminal studies on gesture’s role in cohesive discourse show that gestures reflect this pattern by accompanying rich expressions more than lean referential forms. Thus, presence/absence of gesture and richness of expression in speech work together to create cohesion in discourse (Gullberg, 1998, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1992; Marslen-Wilson et al., 1982; McNeill, Cassell, & Levy, 1993).

McNeill (1992) further hypothesizes that gesture viewpoint is sensitive to information status and should therefore also interact with richness of expression. He proposes a scale of gesture progression, which he suggests is parallel to Givón’s (1983) scale of phonological size for referential expressions (Fig. 1). Both scales are driven by information status.

Lean referential expressions (e.g., zero anaphora and Prons) should be linked to “no gesture,” whereas rich referential expressions of different length (i.e., unmodified vs. modified lexical NPs) should be linked to referential gestures of different “complexity” (i.e., OVPT vs. CVPT gestures; McNeill, 1992, pp. 206–212). McNeill also adds predicates to the linguistic scale to account for the

![Figure 1](image)

*Figure 1.* Alignment between scales of phonological size and gesture progression (adapted from McNeill, 1992).
fact that gestures often align with the verbal elements, usually containing new information, in a clause. Thus, CVPT gestures should be more likely to occur with predicates than OVPT gestures.

**Nominal definiteness and gesture**

Newer studies also point toward a link between mode of representation in gesture and other linguistic markers of information status, such as nominal definiteness (e.g., Wilkin & Holler, 2010). Nominal definiteness is typically used to mark the difference between a new entity and an already given one. In languages such as German or English speakers can use determiners to express a referent’s information status (eine Fee ‘a fairy’ vs. die/diese Fee ‘the/that fairy’; cf., Givón, 1995). Referential gestures also appear to be sensitive to this variation in definiteness. In a corpus study, Wilkin and Holler (2010) found that speakers varied the way in which they depicted referents expressed by indefinite versus definite lexical NPs. New referents were associated with “entity” gestures (depicting the properties of the referents), whereas given referents were associated with “action” gestures (depicting referents’ actions or movements). Gerwing and Bavelas (2004) further found that speakers’ gestures were larger and more precise for new information but smaller and less precise (“sloppier”) for given information. In terms of viewpoint, Parrill (2012) showed that speakers were more likely to produce OVPT gestures in retellings of a story that listeners were already familiar with. Conversely, they were more likely to use CVPT gestures when retelling a story that was unknown/new to their listeners. In discourse terms this result suggests that new information in discourse (typically expressed by indefinite lexical NPs) is more likely to be accompanied by CVPT gestures, whereas given information (including definite lexical NPs) is more likely to be accompanied by OVPT gestures.

**Grammatical role and gesture**

In addition, mode of representation in gesture seems to be linked to a third linguistic marker of information status, namely grammatical role. Different grammatical roles expressed by referential expressions are of crucial importance in the structuring of information in discourse. For instance, speakers are more likely to introduce new referents as objects and maintain given referents as subjects in transitive clauses. The tendency to vary grammatical role according to a referent’s information status (most often operationalized by referential distance) is well established (Allen & Schröder, 2003; Clancy, 2003; DuBois, 1987; Kärkkäinen, 1996). In terms of the relationship between grammatical role and gesture viewpoint, Debreslioska et al. (2013) found that gesture viewpoint varies with transitive versus intransitive subjects, which differ in information status (cf. DuBois, 1987). More specifically, they found that speakers produced more OVPT gestures to represent less accessible intransitive subject referents, and CVPT gestures to represent more accessible transitive subject referents. This result suggests that gesture varying in viewpoint might also mark the difference between other grammatical roles with different information status, such as transitive subject versus transitive object.

**Current study**

The aim of the current study is to examine the details of a possible scale of gesture progression driven by information status, as proposed by McNeill (1992). We examine three linguistic markers of information status and their interactions with presence and viewpoint of gestures to determine whether, and if so how, gesture parallels speech in expressing information status. The study focuses on when and how speakers gesture about discourse referents with different information status. It thus also reveals more about the cohesive role of gesture in discourse reference (also known as reference tracking; Gullberg, 2006; Yoshioka, 2008).

We present two studies. Study 1 examines the parallelism between McNeill’s (1992) hypothesized scale of gesture progression and Givón’s (1983) scale of phonological size for referential expressions, as used for discourse referents. This comparison investigates whether richness of expression is related to
the presence and viewpoint of gestures. Study 2 examines nominal definiteness and grammatical role and investigates whether these are related to the presence and viewpoint of gestures.

**Study 1**

Study 1 tests whether the presence and viewpoint of gestures interact with richness of expression in speech. First, assuming that lexical NPs are less accessible than Prons, and modified lexical NPs are less accessible than unmodified lexical NPs, we make the following predictions for gesture: (1) Lexical NPs will align with gestures more than Prons; (2) modified lexical NPs will align with CVPT gestures more than unmodified lexical NPs; and (3) predicates will align with CVPT gestures more than lexical NPs. Note that since the evidence on a progression from unmodified to modified lexical NPs according to information status is not as robust as for the progression from Prons to lexical NPs (e.g., Engelhardt, Bailey, & Ferreira, 2006), we also examine speech.

**Methods**

**Participants**

Twenty participants (16 women, mean age 26) took part in the study. All participants were native speakers of German. We recruited participants at Ludwig-Maximilian University, Munich, Germany through flyers and e-mails and by word of mouth. Participants were asked to come with a native German-speaking friend who acted as listener. All participants were offered snacks for their participation, and all provided written consent.

**Materials and design**

We used a picture story about three fairies to elicit narrative speech and gestures. The story consisted of 76 experimental items (+51 fillers) (see Appendix). The events presented in the experimental items all depicted actions and were controlled for agent animacy and event transitivity. Half of the events depicted actions carried out by animate entities, and the other half depicted actions carried out by inanimate entities. Each action either constituted a transitive event (agent + inanimate patient) or an intransitive event (only agent).

**Procedure**

Participants were assigned roles as speaker or listener randomly. Speaker and listener sat across from each other. A camera captured the speaker’s head and torso. Participants read their instructions and were also given an oral repetition of the main points in the instructions by the experimenter to ensure they understood the procedure correctly. The speaker’s task was to retell a picture story answering the question “what happened?” The experimenter showed the speaker four to six pictures at once. The participant had unlimited time to memorize them. During this time the listener turned around and wrote down a short summary of the part of the story just narrated. The listener was not allowed to ask any questions, but occasional back-channeling naturally occurred. The listener was asked not to cross legs or arms but to put his or her arms on the upper thighs (which was not disclosed to the speaker). This was done to avoid the speaker mirroring the interlocutor’s body position, which might be unfavorable for gesture production (e.g., Chartrand & Bargh, 1999; Kendon, 1973).

**Speech coding**

We transcribed speech using German standard orthography, also taking note of filled pauses, word truncation, repetitions, and so on.
Event identification and clause annotation

For each experimental item we annotated one target clause (i.e., a unit containing a unified predicate; Berman & Slobin, 1994, p. 660). The unit of analysis was always the first complete mention of an event. If more than one clause described the same event, all mentions after the first complete mention were excluded (example 1; bracketed clause was excluded).

(1)

und dann stürzt aber leider der Salzstreuer ab (first complete mention)
[und fällt auch in die Schüssel] (second complete mention)
‘And then the salt shaker unfortunately crashes down
[and also falls into the bowl]’

Inserted main or metanarrative clauses as well as unrestrictive (i.e., descriptive) relative clauses were excluded from analysis.

Core arguments

Our analyses were limited to the referential expressions for the core arguments of each target clause. In German, intransitive clauses have one core argument, the subject. Transitive clauses have two core arguments, the subject and the direct object.

Richness of expression

Each referential expression was coded for form: zero anaphora (Ø), Pron, unmodified lexical NPs, and modified lexical NPs. A zero anaphor is an omitted argument (as in und Ø kommt zu ihr ‘and Ø comes to her’). Pron were personal (sie ‘she’), demonstrative (die ‘that’), relative (die ‘who’), indefinite (alles ‘everything’), or clitic (s ‘she’).

For unmodified lexical NPs, we coded bare nouns (example 2), lexical NPs + determiners (such as (in)definite articles and demonstrative, indefinite or possessive Pron) (example 3) and elliptical constructions (example 4).

(2)

und mit der rechten Hand steckt sie Kerzen in den Kuchen
‘and with the right hand she puts candles on the cake’

(3)

und dann kommt von der Seite ein/das/dieses/irgendein/ihr Streichholz herangeflogen
‘and then a/the/that/some/her match comes flying in from the side’

(4)

dann kommt die andere/die grüne (Fee)
‘then comes the other/the green (fairy)’

For modified lexical NPs, we coded all lexical NPs that had a larger phonological size than unmodified lexical NPs: lexical NPs modified by one or more adjectives (e.g., der hölzerne Eimer ‘the wooden bucket’), by a prepositional phrase (e.g., die Hexe mit dem gelben Kleid ‘the witch with the yellow dress’), or by a restrictive relative clause (e.g., die Nadel, die vorher den Kragen angenäht hat ‘the needle that earlier sewed the collar’).

Information status

We determined the information status of each referential expression by measuring referential distance to its previous mention. A referent was coded as less accessible/new if it was mentioned for the first time in the narrative or if it was not mentioned in the three clauses preceding its current mention. A referent was coded as more accessible/given if it had been mentioned in the three clauses preceding its current mention, following Givón (1983) (but see also Clark & Sengul, 1979; Du Bois, 1987; Hickmann & Hendriks, 1999 for similar approaches).
Excluded referential expressions

Referential expressions were excluded if they were uttered with a disfluency, defined as a repetition, repair and/or interruption of a referential expression, or referential expressions immediately preceded by filled pauses. Pauses were only counted as disfluencies if they were produced intraclausally. Filled pauses at clause boundaries are seen as planning behavior and were therefore not considered to be disfluencies (Graziano & Gullberg, 2013).

Gesture coding

To annotate gestures, we used frame-by-frame analysis of digital video in ELAN (Sloetjes & Wittenburg, 2008). In all target clauses gesture strokes and poststroke holds were identified with sound turned off. A gesture stroke is defined as the meaningful and most effortful part of the gestural movement (Kendon, 2004; McNeill, 1992). To determine onset and offset of gesture strokes, we considered changes in the parameters of shape, placement, and tension of the hands as well as the trajectory or direction of the hands’ motion (Kendon, 2004; Seyfeddinipur, 2006). A poststroke hold is defined as a movement cessation of the hand at the end of a gesture stroke. One function of a hold is to allow for the rest of the co-expressive speech to be uttered before the hand goes into retraction or the next gesture (Kita, 1990; McNeill, 1992). Since we are interested in examining what the meaningful part of the gesture is aligned with in speech, we also took poststroke holds into account.

Speech–gesture alignment

Speech that was exactly aligned with the gesture strokes ± poststroke holds was annotated. If the gesture covered a vocalized syllable of a referential expression, it was counted as co-occurring with the full speech element (Gullberg, Hendriks, & Hickmann, 2008; McNeill, 1992; Stam, 2006) and taken into consideration for the analysis. For instance, if the stroke co-occurred with Kr in Kragen ‘collar’, which is not a full syllable, we did not consider the referential expression to be aligned with the gesture, and it was thus excluded from analysis. But if the stroke co-occurred with Kra, a full syllable, we took it into account for the analysis. A syllable is a good proxy for the referential expression it belongs to, because it is a clearly identifiable phonological unit of a word. If a gesture was aligned with more than one referential expression, the gesture and corresponding referential expression(s) were excluded.¹

Furthermore, we considered gestures that were aligned solely with a predicate for one part of the analyses. The alignment with predicates included alignment with the lexical verb, the auxiliary verb, and the particle of a verb. Gestures that were aligned with more than one predicate were excluded.

Excluded gestures

We excluded all beats/pragmatic gestures (on the so-called beat filter, see McNeill, 1992; on pragmatic gestures, see Kendon, 2004) and highly conventionalized gestures (e.g., listing gestures, i.e., when a speaker is counting her fingers to introduce new characters), together with the referential expressions they were aligned with. We also excluded all gestures that were produced during speech disfluencies (see Excluded referential expressions, above). If gestures were aligned with one or more disfluencies or with the word immediately after a disfluency, they were excluded from analysis. In addition, gestures were excluded if they were produced during the lengthening of a syllable and long unfilled pauses (>200msec) (see Seyfeddinipur, 2006), which could signal a covert disfluency. In total 176 gestures were excluded.

¹Note that, in contrast to some previous studies, we do not examine zero anaphora. This is because the current study examines exact temporal alignment between gestures and speech. Since zero anaphora are empty “slots,” meaning they have no linguistic material, there is nothing for a gesture to align with. This is not to say that zero anaphora do not play an important role in reference, but a study of gesture alignment with spoken referential expressions cannot take them into account.
Gesture viewpoint

All remaining gestures were coded as OVPT or CVPT representations. A gesture was classified as CVPT if the speaker’s hand(s) could be considered to enact a person doing the movement or action gestured about. A gesture was classified as OVPT if the speaker’s hands depicted an entity as a whole or if the speaker’s hand shape did not represent enactment. Consider a scene in which an icing bag is decorating a cake by itself, putting white dots on the cake. In Figure 2, the speaker chose to represent the icing bag by acting as if holding the icing bag in her hand and performing an up and down movement to depict the dotting action. This is a CVPT representation. In Figure 3, the speaker chose to represent the icing bag by extending her index finger and perform the same movement. This is an OVPT representation because the whole hand represents the entity in question.

A speaker can also indicate an entity’s location in gesture space by using an OVPT hand shape (representing the entity as a whole and placing it, for instance, to the right in gesture space) or CVPT hand shape (enacting the entity’s hands and placing it to the left in gesture space). A “floppy” or relaxed hand shape (i.e., a hand shape with no iconic relationship to the properties of the entity) was always considered an OVPT representation. Finally, an OVPT representation can also correspond to modeling or drawing the shape and/or size of an entity.

Inter-rater reliability

A native German speaker recoded 25% of all speech variables for inter-rater reliability (produced by all participants). A second coder annotated 25% of all gesture strokes (produced by 18 participants). For the inter-rater reliability of gesture we first determined whether there was an overlap between the gesture strokes as annotated by coder 1 and coder 2. We then further qualified the overlap by considering speech. If the gestures coded by coder 1 and 2 aligned with the same speech element (i.e., the same referential expression or verb), they were considered to be in agreement. However, if a gesture annotated by coder 2 aligned with an additional

Figure 2. CVPT entity + action.
referential expression or verb than that of coder 1, then they were not considered to be in agreement. This procedure was chosen because these speech–gesture packages are of particular importance in this study (see Speech–gesture alignment, above). The second coder also recoded 25% of all analyzed gestures for viewpoint (CVPT/OVPT). The kappa statistic was performed to determine consistency among raters. Table 1 summarizes all inter-rater reliability measures.

**Analyses**

The analyses are based on 1,237 referential expressions and 583 referential gestures. We analyzed the data using generalized logistic mixed effect regression (glmer function from the lme4 package in the software R, version 3.1.3). All analyses made use of variants of the generalized linear model with binomial error structure because the dependent variables were binary. The models are used to predict how likely a category from the dependent variable is to combine with different values of the independent variables, while controlling for random variation with the events used in the material (see Appendix for specifications and output of all models).

<table>
<thead>
<tr>
<th>Identification of:</th>
<th>Cohen’s Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech</strong></td>
<td></td>
</tr>
<tr>
<td>Lexical NPs, Prons, Ø</td>
<td>.95</td>
</tr>
<tr>
<td>Unmodified vs. modified lexical NPs</td>
<td>.89</td>
</tr>
<tr>
<td>Information status:</td>
<td></td>
</tr>
<tr>
<td>more vs. less accessible</td>
<td>.81</td>
</tr>
<tr>
<td><strong>Gesture</strong></td>
<td></td>
</tr>
<tr>
<td>Gesture strokes ± poststroke holds</td>
<td>.79</td>
</tr>
<tr>
<td>Gesture viewpoint:</td>
<td></td>
</tr>
<tr>
<td>CVPT vs. OVPT</td>
<td>.82</td>
</tr>
</tbody>
</table>
Results

Richness of expression and gesture progression

Speech
First, we examined whether less accessible referents are more likely to be expressed by lexical NPs than Pron and whether less accessible referents are more likely to be expressed by modified than unmodified lexical NPs. The dependent variable for both analyses was information status (more/less accessible referent) and the independent variable was richness of expression (lexical NP/Pron and modified/unmodified lexical NP, respectively). The analyses revealed that, as expected, lexical NPs are significantly more likely to encode less accessible referents than Pron ($\beta = 2.83$, $SE = .22$, $z$-value = 13.01, $p = .000$). However, contrary to predictions, unmodified and modified lexical NPs were equally likely to encode less accessible referents ($\beta = .004$, $SE = .21$, $z$-value = .02, $p = .985$).

Gesture
We next tested whether lexical NPs are more likely to align with gesture than Pron. The dependent variable was presence of gesture (yes/no), and the independent variable richness of expression (lexical NP/Pron). The analysis revealed that, as expected, lexical NPs are significantly more likely to align with gesture than Pron ($\beta = .81$, $SE = .17$, $z$-value = 4.65, $p = .000$).

Next, we tested whether modified lexical NPs are more likely to align with a CVPT gesture than unmodified lexical NPs. The dependent variable was gesture viewpoint (CVPT/OVPT), and the independent variable richness of expression (modified/unmodified lexical NP). Contrary to predictions, the analysis revealed that unmodified rather than modified lexical NPs are significantly more likely to align with CVPT gestures ($\beta = 1.12$, $SE = .34$, $z$-value = 3.24, $p < .001$). When a gesture aligns with a modified lexical NP, it is more likely to be produced in OVPT.

Finally, we examined whether predicates are more likely than lexical NPs to align with CVPT gestures. The dependent variable was gesture viewpoint (CVPT/OVPT), and the independent variable was clause constituent (predicate/referential expression). For this analysis we excluded gestures that aligned with referential expressions + part of a predicate (leaving us with 233 referential expressions and 147 predicates). Contrary to predictions, the analysis revealed that lexical NPs and predicates are equally likely to align with CVPT gestures ($\beta = -.42$, $SE = .3$, $z$-value = −1.4, $p = .16$).

Discussion

Study 1 tested the link between richness of referential expressions and presence and viewpoint of gestures to examine the suggested parallelism between Givón’s (1983) scale of phonological size and McNeill’s (1992) hypothesized scale of gesture progression. The results show a clear relationship between richness of expression and presence of gesture such that richer referential expressions (i.e., lexical NPs) align with gestures, whereas leaner referential expressions (i.e., Pron) do so less, replicating previous findings (e.g., Gullberg, 2006).

In contrast, the predictions for variation in viewpoint were not borne out. Following McNeill (1992), we expected modified lexical NPs to align with CVPT gestures more than unmodified lexical NPs, since they were assumed to encode less accessible referents. However, the results instead show that speakers often produced modified lexical NPs for more accessible referents when simpler forms would have sufficed, meaning they were over-specific (Deutsch & Pechmann, 1982; Engelhardt et al., 2006). We assume this result is a task effect in that speakers in our study used modified lexical NPs as proper names. That is, they described referents in a certain fashion at the beginning (e.g., die rote Fee ‘the red fairy’) and kept using those expressions as names irrespective of the information structural context that the referents were in (e.g., Garrod & Anderson, 1987).

For gestures, modified lexical NPs tended to align with OVPT gestures. This unexpected result might also be a task effect and reflect the naming strategy in speech. In fact, one study suggests that
gestures associated with names are more likely to correspond to an OVPT representation (Marslen-Wilson et al., 1982).

Finally, there was no difference in viewpoint when comparing gestures aligned with lexical NPs versus predicates. This is perhaps not surprising, since lexical NPs and predicates are very different constituents in the production of discourse (lexical NPs refer to referents; predicates refer to referents’ actions/movements). Also, given that the assumption of a progression of gesture viewpoint is based on the information status of a referent, it makes more sense to focus on combinations of referential expressions + predicates to further explore the variation in mode of representation for discourse referents.

In summary, the results support assumptions about a parallelism between richness of expression in speech and the presence/absence of gesture. However, task effects meant that we found no relationship between richness of expression and gesture viewpoint. Importantly, however, richness of expression is only one linguistic marker of information status. Some studies have suggested a link between modes of representation in gesture and other important markers of information status, such as nominal definiteness (Gerwing & Bavelas, 2004; Wilkin & Holler, 2010), and grammatical role (Debreslioska et al., 2013). In study 2 we therefore test whether nominal definiteness and/or grammatical role modulate presence/absence of gesture and gesture viewpoint.

**Study 2**

Study 2 tests whether presence and viewpoint of gestures interact with nominal definiteness and/or grammatical role. First, assuming that nominal definiteness is determined by referents’ information status, we predict that indefinite lexical NPs are more likely to be aligned with gestures than definite lexical NPs and specifically with CVPT gestures. Further, assuming that in transitive clauses direct objects are less accessible than subjects (e.g., Du Bois, 1987), we make the following predictions for gesture: Direct objects are more likely to be aligned with gestures than subjects and specifically with CVPT gestures. Based on the results of Study 1, we also took into account gesture alignment with referential expressions + predicates and predict that CVPT gestures are aligned with direct objects + predicates more than with subjects + predicates. This latter prediction also corresponds to McNeill’s (2005) suggestion that CVPT gestures may be related to whole verb phrases (see also Cassell, Stone, Douville, Prevost, & Achorn, 1994).

**Methods**

**Further speech coding**

**Definiteness**

We coded all lexical NPs for definiteness (indefinite/definite lexical NPs). Indefinite lexical NPs comprise all bare nouns (Kerzen ‘candles’), lexical NPs with an indefinite article (eine Kerze ‘a candle’), or with an indefinite Pron (ein paar Kerzen ‘some candles’). Definite lexical NPs comprise all lexical NPs with a definite article (die Kerze ‘the candle’) or a demonstrative Pron (diese Kerze ‘that candle’). Possessive NPs were excluded because they refer to two entities, which differ highly in their discourse characteristics (Ariel, 2002).

**Grammatical role**

We coded all referential expressions in transitive clauses for grammatical role (subject/direct object). We identified the transitive subject, which is the argument whose referent is most likely to be relevant for the success of the activity being identified (the initiator, controller of the activity), and the transitive object, which is the argument whose referent is most likely to be saliently affected by the activity (Dixon, 2009). An intransitive clause only contains one core argument, a (intransitive) subject.
Inter-rater reliability

A native German speaker recoded 25% of the two speech variables for inter-rater reliability (Table 2). The gestures are the same as used in Study 1.

Analyses

We analyzed the data using generalized logistic mixed effect regression with binomial error structure because, again, all dependent variables are binomial (using glmer function from the package lme4 in R, version 3.1.3; see Appendix for model specifications and output).

Results

Nominal definiteness and gesture

First, we examined whether indefinite lexical NPs are more likely to align with gesture than definite lexical NPs. The dependent variable was presence of gesture (yes/no) and the independent variable nominal definiteness (definite vs. indefinite lexical NP). The analysis revealed that, as expected, indefinite lexical NPs are significantly more likely to align with gesture than definite lexical NPs ($\beta = -0.37, SE = 0.18, z$-value = $-2.1, p < .05$).

Second, we probed whether indefinite lexical NPs are more likely to align with CVPT gestures than definite lexical NPs. The dependent variable was gesture viewpoint (CVPT/OVPT) and the independent variable nominal definiteness (definite vs. indefinite lexical NP). The analysis revealed that, as expected, indefinite lexical NPs are significantly more likely to align with CVPT gestures than definite lexical NPs ($\beta = -0.9, SE = 0.35, z$-value = $-2.58, p < .01$). Conversely, when gesture strokes align with definite lexical NPs, they are more likely to be produced in OVPT.

Figure 4 exemplifies the result and shows a participant producing a CVPT hand shape and performing an arc movement with her right hand to enact how a needle picked up a bow from a basket by itself. The gesture stroke aligns with the indefinite lexical NP (eine Schleife ‘a bow’).

Grammatical role and gesture

We further tested whether referential expressions encoded as direct objects are more likely to align with gesture than referential expressions encoded as subjects. The dependent variable was presence of gesture (yes/no) and the independent variable grammatical role (subject vs. direct object). As expected, direct objects are more likely to align with gesture than subjects ($\beta = -1.04, SE = 0.17, z$-value = $-6.22, p < .000$).

Next, we examined whether direct objects are more likely to align with a CVPT gesture than subjects. The dependent variable was gesture viewpoint (CVPT/OVPT) and the independent variable grammatical role (subject vs. direct object). Contrary to predictions, subjects and direct objects are equally likely to align with a CVPT gesture ($\beta = -0.14, SE = 0.33, z$-value = $-0.45, p = .66$).

Finally, the last analysis had gesture viewpoint as dependent variable (CVPT/OVPT) and grammatical role + predicate as the independent variable (subjects + predicate/object + predicate) and revealed, as expected, that CVPT gestures are significantly more likely to align with direct objects + predicate than with subjects + predicate ($\beta = -1.19, SE = 0.57, z$-value = $-2.1, p < .05$) (see Fig. 4, in Table 2. Inter-Rater Reliability Coding Study 2.

<table>
<thead>
<tr>
<th>Identification of:</th>
<th>Cohen’s Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject vs. direct object</td>
<td>0.96</td>
</tr>
<tr>
<td>Definite vs. indefinite lexical NPs</td>
<td>0.96</td>
</tr>
</tbody>
</table>
which the gesture aligns with the direct object and part of the predicate). Thus, when gesture strokes align with subjects + predicates, they are more likely to be produced in OVPT.

**Discussion**

The results from Study 2 suggest a firm relationship between nominal definiteness and both presence of gesture and gesture viewpoint. Indefinite lexical NPs are more likely to align with gestures overall, and specifically with CVPT gestures, than definite lexical NPs. These results support McNeill’s (1992) idea that the progression from presence to absence, and from CVPT to OVPT reflects a progression in information status from new to given. Furthermore, there is a connection between grammatical role and the presence of gesture. Objects in transitive constructions are less accessible and align with gestures more than more accessible subjects. This also replicates the idea that gesture use increases with an increase in novelty of information (see also McNeill’s idea on growth points, defined as “the element in thought that stands out in the context and may be the point of greatest relevance” [1992, p. 220] and thus should be associated with gesture more than points of lesser relevance).

In contrast, the relationship between grammatical role and gesture viewpoint is less firm. Transitive subjects and objects alone are equally likely to align with CVPT gestures. However, a detailed analysis of gestures aligning with subjects + predicates versus direct objects + predicates revealed that direct objects + predicates are significantly more likely to be accompanied by CVPT gestures. Again, this result supports McNeill’s (1992) suggestion that CVPT gestures are related to whole verb phrases, which usually express the new information in the clause (see also Chafe, 1994 on how verb + object combinations express new information together).
**Presence and viewpoint of gestures interact with definiteness**

The results on nominal definiteness showed effects on both presence and viewpoint of gestures. In relation to gesture presence, the results contradict previous findings by Wilkin and Holler (2010), who found no increase in the number of gestures for indefinite versus definite lexical NPs. There are several possible explanations for this discrepancy. First, Wilkin and Holler examined an English corpus, whereas we worked with German data. Second, our data set was considerably larger. Finally, we restricted our analysis to core arguments of transitive and intransitive clauses, whereas Wilkin and Holler considered all kinds of referential expressions in their corpus (including prepositional objects).

In relation to gesture viewpoint, on the other hand, the current results support previous studies showing a variation in the mode of representation (or physical form of the gesture) for less versus more accessible information (Gerwing & Bavelas, 2004; Parrill, 2012; Wilkin & Holler, 2010).

**Presence and viewpoint of gestures interact with grammatical role**

The findings on grammatical role also show an effect on presence of gesture. Gesture use increases with direct objects, which usually express less accessible information, and decreases with (transitive) subjects, which usually express more accessible information. This result is in line with previous studies on this topic, drawing a connection between the presence of gestures and more or less accessible referents. These studies have, however, mostly focused on clausal subjects and their information status (e.g., Gullberg, 2006; Perniss & Özyürek, 2014; Yoshioka, 2008). The current study widens the scope by specifically examining nonsubjects. It thus opens up the area of investigation for new questions. For instance, it seems important to examine whether gesture use differs for prepositional phrases (+ predicates), covering semantic roles such as location or instrument, in comparison with direct objects. Such a comparison could explain some of the differences found in relation to previous studies (i.e., Wilkin & Holler, 2010) but also uncover whether gesture is sensitive to the difference between core and optional arguments, which are presumably less accessible in the flow of discourse (Hovav & Levin, 2007).

Furthermore, information status is not only a major factor for assignment to grammatical role or referential form but also to position in the clause. The factors usually correlate: Highly accessible entities tend to be pronominalized, realized as subjects, and occur early in the utterance (e.g., Chafe, 1994). The current study has provided evidence for gesture interacting with two of these factors, referential form (richness of expression and definiteness) and grammatical role. Further studies will need to examine whether gesture also interacts with position in the clause. The current study indirectly touches on this issue by showing that there is variation in viewpoint for subjects + predicates (associated with OVPT) versus direct objects + predicates (associated with CVPT). It is thus plausible that gesture is sensitive to position in the clause as well.

Finally, we still know very little about the role that each individual factor plays independently for the variation in the number or mode of representation in gesture. For instance, it would be relevant to examine whether the difference in the number of gestures between subjects and direct objects correlates with richness of expression and/or whether it holds when both grammatical roles are expressed with rich referential expressions (i.e., lexical NPs). In fact, previous research in second language acquisition has shown that over-specification in speech is accompanied by over-specification in gesture (Gullberg, 1998, 2003, 2006; see also So, Kita, & Goldin-Meadow, 2009 for native speakers). L2 speakers often use rich referential expressions in contexts where leaner forms would have sufficed and also accompany those expressions with gestures, suggesting that gesture production is potentially more—or at least just as—sensitive to richness of expression (when it is not used as a marker of information status) as to referential context. The results on definiteness in the present study clearly challenge the view that richness of expression, when it is not used as a marker of information, might have a more important influence on gesture use than when it is linked to the
referential context. That is, the results clearly show that for native speakers, gesture use varies with indefinite versus definite lexical NPs, which both constitute rich referential expressions, but differ in information status. Another study on bilingual children also shows that in their less dominant language, children do not necessarily gesture about given referents even if richer forms were used to express them in speech (So, Lim & Tam, 2014). It will therefore be important to design studies specifically to disentangle to what degree and in which circumstances gesture is more or less sensitive to richness of expression and referential context or information status.

General discussion

The current study examined the role of gesture in discourse reference and more specifically the interaction of referential gestures with spoken referential expressions reflecting referents’ information status. We investigated three linguistic markers of information status—richness of expression, nominal definiteness, and grammatical role—and found that all three markers vary with the presence/absence of gesture, but only nominal definiteness and grammatical role also vary with gesture viewpoint. The results indicate progressions from absence to presence of gestures and from OVPT to CVPT gestures as a function of information status, similar to what was proposed by McNeill’s (1992) gesture scale. Figure 5 summarizes the findings.

Example 5 illustrates the intricate ways in which the three linguistic markers of information status work together with the presence/absence and viewpoint of gestures in order to construct connected discourse (square brackets indicate a co-occurring gesture).

(5)
1 die grüne Fee kommt zurück
2 Ø trägt jetzt [ein Kästchen] in der Hand (CVPT gesture)
3 Ø [läuft] die Treppe runter
4 Ø stellt [das Kästchen] erstmal ab (OVPT gesture)
5 und Ø holt da [n Stück weisse]n Stoff wieder hera (CVPT gesture)
6 der Stoff [fliegt] in die Luft
‘1 the green fairy comes back
2 Ø carries a box in her hand (CVPT gesture)
3 Ø [goes] down the stairs
4 Ø puts [the box] down (OVPT gesture)
5 and Ø takes out [a white piece] of cloth (CVPT gesture)
6 the cloth [flies] into the air’

In line 2, the speaker introduces the referent ‘ein Kästchen’ as a direct object with an indefinite lexical NP and accompanies it with a CVPT gesture. She uses both hands to depict how the fairy is holding the object, thus also representing the shape of the box. In line 4 she then uses a definite
lexical NP, again encoded as a direct object, for the same referent (‘das Kästchen’) and produces an OVPT gesture. She uses both hands again (closely held together), but this time her hands no longer depict holding the box but rather the direction of its placement. She raises her hands during the preparation phase of the gesture and then produces a downward movement (the stroke phase) by also extending her fingers to depict the path of the box being put down. The fact that she repeats some of the features of the previous gesture for the same referent suggests that speakers try to create a connection between different mentions of the same referent in gesture (Gerwing & Bavelas, 2004; McNeill, 1992). Importantly, however, the speaker varies viewpoint for these two mentions of the referent (from CVPT to OVPT) in parallel with the change in information status from new to given (here ‘introduced’ to ‘reintroduced’). Finally, in line 5 she introduces yet another referent as a direct object, using an indefinite lexical NP and a CVPT gesture (‘n Stück weissen Stoff’). That referent is then maintained as subject with a definite lexical NP in line 6 without any co-occurring gesture. Thus, in this instance, the speaker varies the presence of gestures with a progression in information status from new to given (here ‘introduced’ to ‘immediately maintained’).

Importantly, this study did not find evidence for a linear progression from absence to presence of gesture and then from OVPT to CVPT, as suggested by McNeill’s scale. Instead, there seem to be two levels—presence and viewpoint of gestures—that speakers can use both when indicating the difference between new/less accessible or given/more accessible entities. Thus, speakers can highlight newness or relevance simply through the presence of gesture or vary mode of representation. The context in which the relevant piece of information is expressed may matter for the choice. It would be particularly interesting to examine for nominal definiteness since both presence and viewpoint of gestures co-varied with this linguistic marker of information status. Grammatical role also seems interesting to probe given that gestures co-occurring with objects + predicates are associated with CVPT, whereas gestures co-occurring with objects alone are equally likely produced in OVPT and CVPT. With more detailed analyses like these, a scale of gesture progression combining the two separate scales proposed in this paper might be possible.

Finally, the findings raise important questions about the co-expression of meaning between speech and gesture and cross-linguistic differences. The method of analyzing speech–gesture alignment in this study leads us to ask whether gestures always align with the elements in speech that they also semantically cohere with. For instance, it remains unresolved whether gestures aligning with objects represent the referents of these nominal expressions or rather the actions that are performed on them by the agent in subject role (or both). It is important to resolve this issue given contradictory findings in the literature. For example, Debreslioska et al. (2013) found that more accessible subjects in transitive clauses were associated with CVPT gestures, whereas in the current study actions on less accessible objects in transitive clauses are associated with CVPT gestures. The different findings are motivated by a difference in methodology. Debreslioska et al. (2013) examined speech–gesture co-expressivity at the clause level, not exact temporal alignment, and considered only those gestures that depicted the (actions of the) subject referent. In contrast, in the current study we examined exact temporal alignment and did not consider semantic overlap between referential expression and gestural depiction. Thus, the findings from the studies do not contradict each other. In fact, both studies suggest that speakers tend to use OVPT gestures with less accessible entities (subjects) and CVPT gestures with actions on new/less accessible entities (objects). Both studies demonstrate that the way we look at speech–gesture co-expression matters more than it is typically acknowledged in the literature. For instance, a CVPT gesture aligning with action on a less accessible object in the current study is very likely to be depicting some property of its referent (e.g., ‘the box’ in (5)). But the gesture might equally well be depicting the more accessible subject via its action (‘the fairy’ in (5)). Thus, it is important to resolve the question of how we consider the timing of a gesture in relation to its semantic content, and also how to best assess the semantic content of a gesture in general.

These questions become even more challenging when cross-linguistic differences are taken into account. For instance, Duncan’s (1996) study of English and Chinese narratives reveals gestural...
differences between the two languages. In English there is a tendency for a tight semantic synchron-
ization (i.e., gestures align tightly with semantically coherent pieces of information in speech). How-
ever, in Chinese gestures tend to align with speech elements that precede the semantically 
coherent piece of information. Gestures then often align with given elements in speech while 
simultaneously representing the new information (Duncan, 1996: 180–182). It remains to be 
explored what such (cross-linguistically driven) differences in patterns of speech-gesture alignment 
and semantic integration can tell us about the speech–gesture relationship in general. We clearly 
need more cross-linguistic studies examining the variation of gesture use at the discourse level, 
taking into consideration different speech phenomena such as how new versus given information is 
expressed, how referents are tracked in discourse, etc., and different methods of investigating gesture 
alignment or co-expression.

**Conclusion**

Although there is widespread agreement that speech and gestures are tightly integrated in the 
production and comprehension of language, the questions of why and how speakers deploy gestures 
at particular moments in the flow of discourse remain a matter of debate. The present study suggests 
that information status plays a crucial role in the way speakers organize speech and gesture in 
connected discourse. The presence and/or viewpoint of gestures co-vary with three linguistic 
markers of a referent’s information status: the richness of expression, nominal definiteness, and 
grammatical role. This suggests that gestures are orchestrated as a cohesive device in combination 
with speech on different levels. The production of gesture can therefore partly be explained by the 
relevance and position of a piece of information in discourse. Discourse context and exact temporal 
alignment between speech and gesture become vital to furthering our understanding of the role of 
gesture in language use.

**Acknowledgements**

We gratefully acknowledge financial support for part of this study from the IDO foundation for language research in memory of Hellmut Röhnisch. We thank Debra Svensson for her help in creating the stimuli, and Sabine Gosselke-Berthelsen and Wanda Jakobsen for reliability coding. We also express our thanks to Prof. Katrin Lindner and Dr. Nicole Weidinger, Ludwig-Maximilians-University Munich, for hosting us and supporting us during the data collection in Munich.

**ORCID**

Sandra Debreslioska [http://orcid.org/0000-0002-1186-853X](http://orcid.org/0000-0002-1186-853X)
Marianne Gullberg [http://orcid.org/0000-0002-3589-9605](http://orcid.org/0000-0002-3589-9605)

**References**


Appendix

Experimental items (descriptions)

1. F1 lights candles on cake
2. Cake falls apart
3. F2 is sewing dress
4. F2 is cutting out belt
5. F2 is cutting triangles into belt
6. F1 goes to F2
7. F3 is sweeping the floor with broom
8. F3 is sweeping the floor with mop
9. F3 falls
10. F2 goes to F3
11. F1 goes to F3
12. F3 goes up stairs
13. F3 goes down stairs with box
14. F1 takes wand out of box
15. F2 takes wand out of box
16. F3 takes wand out of box
17. Bowl moves to the middle of table
18. Flour bag moves next to bowl
19. Milk can moves next to bowl
20. Spoon moves next to bowl
21. Spoon takes flour out of flour bag
22. Milk can pours milk into bowl
23. Spoon stirs dough
24. Salt shaker flies above bowl
25. Sugar bowl flies above bowl
26. F1 takes pinch of salt
27. F1 adds salt to bowl
28. Salt shaker falls into dough
29. Sugar bowl puts sugar into spoon
30. F1 stirs dough with wand
31. Bowl pours dough into a cake
32. Needle1 sews top
33. Needle2 sews skirt
34. Scissors cut cloth
35. F2 sews sleeve with wand
36. F2 flies over dress
37. F2 sews other sleeve with wand
38. Broom comes down stairs
39. Mop comes down stairs
40. Bucket comes down stairs
41. F3 flies above bucket
42. F3 pours cleaning agent into bucket
43. Broom sweeps floor
44. Flour bag jumps into the air
45. F1 jumps into the air
46. Mop goes into water bucket
47. Mop sweeps floor
48. F2 jumps up into the air
49. F2 falls into bucket
50. Icing bag puts sugar drops on cake
51. F3 flies to cake
52. F3 decorates cake with glitter
53. F1 goes up stairs
54. F1 comes down stairs with box of candles
55. F1 flies to cake
56. F1 puts candles on cake
57. Match flies to cake
58. Icing bag puts sugar hearts on cake
59. Match lights candles
60. Cake jumps up into the air
61. F2 goes up stairs
62. F2 brings scissors and 3 needles
63. Scissors cut cloth into a collar
64. Needle1 sews collar onto dress
65. Needle2 takes bow out of box
66. Needle3 takes bow out of box
67. Needle1 takes bow out of box
68. F3 jumps into water bucket
69. Broom goes upstairs
70. Mop goes upstairs
71. Bucket goes upstairs
72. F3 packs wand into box

Specification and Output of Mixed Effects Models

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Random Factor</th>
<th>Estimate</th>
<th>SE</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information status (new/given)</td>
<td>Lexical NP/Pron</td>
<td>Event*</td>
<td>2.83</td>
<td>.22</td>
<td>13.01</td>
<td>&lt;.000</td>
</tr>
<tr>
<td>Information status (new/given)</td>
<td>Unmod/Mod Lexical NP</td>
<td>Event</td>
<td>.004</td>
<td>.21</td>
<td>0.02</td>
<td>.985</td>
</tr>
<tr>
<td>Gesture presence (yes/no)</td>
<td>Lexical NP/Pron</td>
<td>Event</td>
<td>.81</td>
<td>.17</td>
<td>4.65</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gesture viewpoint (cvpt/ovpt)</td>
<td>Unmod/Mod Lexical NP</td>
<td>Event</td>
<td>1.12</td>
<td>.34</td>
<td>3.24</td>
<td>&lt;.005</td>
</tr>
<tr>
<td>Gesture viewpoint (cvpt/ovpt)</td>
<td>Lexical NP/Predicate</td>
<td>Event</td>
<td>−.42</td>
<td>.3</td>
<td>−1.4</td>
<td>.16</td>
</tr>
<tr>
<td>Gesture presence (yes/no)</td>
<td>Indef/Def Lexical NP</td>
<td>Event</td>
<td>−.37</td>
<td>.18</td>
<td>−2.1</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Gesture viewpoint (cvpt/ovpt)</td>
<td>Indef/Def Lexical NP</td>
<td>Event</td>
<td>−.9</td>
<td>.35</td>
<td>−2.58</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Gesture viewpoint (cvpt/ovpt)</td>
<td>Subject/Object</td>
<td>Event</td>
<td>−1.04</td>
<td>.17</td>
<td>−6.22</td>
<td>&lt;.000</td>
</tr>
<tr>
<td>Gesture viewpoint (cvpt/ovpt)</td>
<td>Subject/Object</td>
<td>Event</td>
<td>−.14</td>
<td>.33</td>
<td>−.45</td>
<td>.66</td>
</tr>
<tr>
<td>Gesture viewpoint (cvpt/ovpt)</td>
<td>Subject+verb/Object+verb</td>
<td>Event</td>
<td>−1.19</td>
<td>.57</td>
<td>−2.1</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

*We only used “event” as a random factor, because the variance component for “participant” was practically zero, suggesting a need to simplify the structure of the model. Therefore, “participant” was removed as random factor. This removal did not affect the outcome of the fixed effects, i.e., the same fixed effects were significant before and after the removal.
Paper IV
Paper IV: The semantic content of gestures varies with definiteness, information status and clause structure

Abstract

When speakers mention referents for the first time in discourse, they can also accompany them with gestures. This study explores whether the semantic content expressed by such gestures, that is whether gestures focus on the entity itself or on an action that the referent is involved in, is related to discourse properties. We consider the potential effects of a referent’s information status (brand-new/inferable), definiteness of the referring expression (indefinite/definite), and the clause structure in which it occurs (more/less specialized for referent introduction). Results suggest that ‘entity’ gestures specifically accompany brand-new referents expressed by indefinite nominals (e.g., indicating the shape of ‘a broom’). Furthermore, ‘entity’ gestures co-occur with first mentions in clause structures specialized for the introduction of referents (e.g., ‘there was a broom’). ‘Action’ gestures, on the other hand, are more likely to occur with inferable referents expressed by definite nominals (e.g., indicating the path of some egg shells falling into a bowl), and first mentions in less specialized clause structures focusing on events (e.g., ‘the egg shells fell into the bowl’). The study highlights the importance of considering discourse variations on the word and clause level for understanding the semantic content of gestures, and the interplay between gestures and speech to create cohesive discourse.

1 Introduction

When speakers introduce new entities in discourse for the first time, they often produce gestures with the spoken referential expressions. In such gestures, speakers can either focus on the entity itself or on an action involving the entity. For instance, when introducing the referent ‘ball’ in clauses like ‘there was a ball on the table/she put a ball on the table’, a speaker can focus on the entity, the ball, by shaping the form of the ball
by drawing a circle in the air with her finger. But the speaker can also focus on the action, the holding/moving of the ball, by enacting such an event (e.g., by pretending to hold the ball). In this study, we examine when speakers produce representational gestures with a focus on entity versus a focus on action in the context of entities that are new/mentioned for the first time in discourse. More specifically, we investigate how this variation in what we will call the semantic content of the gesture (entity/action) is related to the information status of the referent, the form of the referring expression, and its clausal embedding.

1.1 First mentions in speech

When speakers mention referents for the first time, they need to mark their status as new using the possibilities that their language offers. Languages typically provide ‘local’ markings on the noun phrase, as in an indefinite article, and more ‘global’ markings affecting the entire clause structure, such as placing the new referent toward the end of the clause (Hickmann, Hendriks, Roland & Liang, 1996). In this study, German is the example language and we will thus present the German local and global marking patterns.

1.1.2 Local markings - definiteness and information status

In languages like German or English, local markings for new referents consist of indefinite determiners on the noun phrase (1). It is typically assumed that when speakers use an indefinite lexical noun phrase (e.g., eine Fee ‘a fairy’), the referent represents information that is new to the discourse and new to the addressee. This information status is also called ‘brand-new’ (Prince, 1981, 1992). However, speakers can also use definite lexical NPs when mentioning a referent for the first time (e.g., der Besenstiel ‘the broomstick’ underlined in 2). In this case, the referent is considered to be ‘inferable’ (Chafe, 1972, 1994; H. Clark, 1975; Givón, 1995; Gundel, 1996; Prince, 1981, 1992). Inferable referents are typically inferentially linked to an already-mentioned referent (e.g., der Besen ‘the broom’ in 2), which triggers the definite reference, but they are still new since they do not exist in the addressee’s knowledge prior to their mention (Birner, 2013).

(1)

dann kommt eine Fee herein

‘then a fairy comes in’
(2)

dann nimmt sie den Besen

um den Kuchen abzustützen

und dis läuft aber so ein bisschen den Besenstiel runter

‘then she takes the broom

in order to stabilize the cake

but it runs down the broomstick a little bit’

1.1.3 Global markings - clause structure

In addition to local markings, speakers can also choose to mark newness more globally on the clause level. One general principle is that new information is preferably placed towards the end of a clause (e.g., Chafe, 1994; H. Clark & Haviland, 1977; Hickmann et al., 1996). Speakers can use different types of clause structures to achieve this. In German or English, speakers can use more and less specialized clause structures (Hickmann et al., 1996).

More specialized (or ‘presentational’) clause structures include existentials, locatives, and possessives (3-6 respectively). These clause structures focus on the existence of a new referent, which is typically reflected in the verb semantics used (‘be’ and ‘have’ or close variants as in 3), and the use of locational elements (E. Clark, 1978). Locational elements can be inanimate, as is typical for locatives (Tisch ‘table’ in 5), or animate, as is typical in possessive constructions (die ‘she’ in 6 or sie ‘she’ in 8). Furthermore, speakers often use the dummy subject es ‘it’ (3), the adverbial da ‘there’ (4), or a location indication (5-6) at the beginning of the clause, allowing them to place the new referent towards the end. The close relationship between these types of clause structures is well-established (e.g., Lyons, 1967), and documented cross-linguistically (E. Clark, 1978).

In contrast, less specialized clause structures for referent introduction often focus on events in which new referents are involved. These are typically intransitive constructions, with the new referent instantiated as the grammatical subject and as agent (the instigator of an action), or transitive constructions, with the new referent

---

1 Note that existentials, locatives and possessives do not necessarily need to have a presentational function (see also Prince, 1992; Ward & Birner, 1995).
typically instantiated as the grammatical object and as patient (the entity affected by an action) (7-8). In this case, speakers often choose other information ordering strategies, which allow them to place the new referent towards the end of the clause. For instance, speakers might use subject-verb inversion in an intransitive construction (7) or an already-mentioned referent as a starting point in a transitive construction (8)².

(3)

es gibt einen Tisch

‘there is a table’

(4)
da sind drei Feen

‘there are three fairies’

(5)
und auf einem Tisch steht eine riesen Torte

‘and on a table is/stands a big cake’

(6)
und die hat ein Besen

‘and she has a broom’

(7)
dann kommt eine grüne Fee

‘then comes a green fairy’

² Note here that transitive constructions are similar to possessives (6) in that they have a subject-verb-object structure. The difference between these two types of clause structures crucially lies in their semantics, whereby possessives express existence of an entity in an animate place whereas transitives involve an action by an agent onto a patient.
In summary, referents that are mentioned for the first time in discourse can vary in information status as brand-new versus inferable, typically also marked locally by a variation in definiteness on the noun phrase (indefinite/definite). Furthermore, new referents can be introduced in more or less specialized clause structures focusing either on the existence of the new discourse referent or on an action that the referent is involved in, respectively.

1.2 Gestures accompanying discourse referents

There is accumulating evidence that gestures are also sensitive to referents' discourse status and to discourse related variations in speech (e.g., Debreslioska & Gullberg, 2017; Foraker, 2011; Gerwing & Bavelas, 2004; Gullberg, 2003, 2006; McNeill, 1992; Levy & McNeill, 1992; Wilkin & Holler, 2011). This body of research is taken as important evidence for the notion that gestures form an integrated whole with speech and language, as suggested by Kendon (1972, 1983, 2004), McNeill and Levy (1982), and McNeill (1992, 2005).

1.2.1 Representational gestures and information status

The focus in this paper are representational gestures, that is gestures that represent entities and actions via iconicity and deixis (Kita, 2000). More specifically, there seems to be a relationship between ‘how’ and ‘what’ meaning is expressed in representational gestures and the information status of the referential expressions they accompany.

1.2.1.1 ‘How’ and ‘what’ meaning is expressed in representational gestures

The issues of ‘how’ and ‘what’ meaning is expressed in gestures is often not clearly differentiated in the gesture literature. This may be because, as McNeill (1992) states, “[if] we explain the meaning of a gesture we explain the form” (p. 23), which suggests that there is a very close connection between the meaning (‘what’) and the form (‘how’) of a gesture. However, these two levels still represent partially different ways of examining gestures. Table 1 gives an overview of the techniques of representation (i.e., ‘how’ gestures represent) described in the literature with different terminology, and shows their relationship to the semantic content of gestures (i.e., ‘what’ meaning is expressed in gesture).
‘Depiction’ in Kendon’s (2004) terms corresponds to a speaker molding or drawing the shape or size of an entity (e.g., drawing a circle in the air to represent a ball). ‘Modeling’ corresponds to a speaker’s hand being used to represent an entity as a whole (e.g., a stretched-out index finger to represent a match). In McNeill’s (1992) terms, these gestures are O-VPT representations. This is essentially because the speaker seems to assume an observer perspective, looking onto the scene from the outside. ‘Enactment’ in Kendon’s classification corresponds to a speaker acting out an event from the perspective of the character herself (assuming an inside perspective or C-VPT in McNeill’s terms). The speaker’s hands or body then map onto the entity’s hands or body (e.g., enacting someone holding or throwing a ball). In the current study, the techniques of ‘depiction’ and ‘modeling’, in the way just described, correspond to gestures expressing entity information (see Figure 1). The technique of ‘enactment’ corresponds to gestures expressing action information (see Figure 2). However, speakers can also use ‘depiction’ and ‘modeling’ to express action information. For instance, a speaker can draw a line in order to represent a path travelled by an entity. Similarly, they can use a hand as a model for an entity (e.g., index finger for match), but their hand may simultaneously move through space in order to represent the entity’s movement. In this case, the gesture is also about action. It is therefore important to note that the techniques of representation are not necessarily predictive of the meanings that the gestures will express.
Figure 1. Example of an ‘entity’ gesture. The speaker depicts a basket’s shape with his hands/arms. The gesture stroke starts in picture a) and ends in the configuration shown in picture b). He has drawn an upward facing semi-circle in order to represent the basket.

Figure 2. Example of an ‘action’ gesture. The speaker represents how a fairy takes a piece of cloth out of a basket. He maps the fairy’s grasping hand onto his own hand, and also depicts the upward path of the cloth as it is lifted out of the basket.
1.2.1.2 Meaning in gesture and information status

Returning to the relationships between ‘how’ and ‘what’ meaning is expressed in gestures and referents’ information status, the following patterns have been described. Gerwing and Bavelas (2004) find that gestures are bigger, more precise and combine more than one gestural movement when representing new referents than when representing given referents. In a similar vein, Debreslioska and Gullberg (2017) find that speakers are more likely to align so-called C-VPT gestures with new information and O-VPT gestures with given information. This result is commensurate with McNeill’s (1992) suggested gestural scale of progression in which C-VPT gestures represent more complexity (than O-VPT gestures) in their physical forms. It is also in line with Gerwing and Bavelas’s (2004) suggestion that more complexity corresponds to new information (but see Debreslioska, Özyürek, Gullberg & Perniss, 2013, for a different proposal on the difference of complexity between O-VPT and C-VPT gestures).

Foraker (2011) examines ‘what’ meaning gestures express when accompanying first versus subsequent mentions of referents. She finds that gestures are more likely to express entity information (e.g., shape or size) when they co-occur with first mentions, and more likely to express action information when they co-occur with subsequent mentions. Similarly, Wilkin and Holler (2011) find that gestures tend to express entity information when they co-occur with referents encoded with indefinite lexical NPs, and action information when they co-occur with referents encoded with definite lexical NPs. Thus, there appears to be a semantic shift in gesture from entity to action representation as a function of information status from new to given (operationalized as first/subsequent mention, and indefinite/definite nominals).

1.2.1.3 Summary: how gestures represent entities in spoken discourse

Previous research suggests that there is a close connection between ‘how’ and ‘what’ information gestures express and discourse organizational principles. It is, at least partly, possible to tease apart the two levels in gesture representation (the ‘how’ and the ‘what’) and studies have focused on both levels in order to show that gestures reflect discourse. For instance, shifts from more to less complex gestures, from C-VPT to O-VPT, and from ‘entity’ to ‘action’ gestures have been found to correlate with a shift from new to given referents in discourse. But since all these studies either focus on the difference between first versus subsequent mentions (Foraker, 2011; Gerwing & Bavelas, 2004) or a combination of these (Debreslioska & Gullberg, 2017; Wilkin & Holler, 2011; see also others having only considered different kinds of subsequent mentions, i.e., Azar, Backus & Özyürek, 2018; Debreslioska et al., 2013; Gullberg, 2006; Perniss & Özyürek, 2015), it remains unclear whether and if so how gestures vary in their representations of first mentioned referents in discourse. Given the previous findings, it seems plausible that the information status of first mentions (brand-new/inferable) and their encoding as definite/indefinite nominals might interact with the semantic
content of the gestures co-occurring with them. Specifically, there seems to be a link between gestures with an entity focus and ‘newer’ referents (i.e., brand-new rather than inferable) and/or referents encoded as indefinite nominals. Conversely, there is a suggested link between gestures with an action focus and ‘older’ referents (i.e., inferable rather than brand-new) and/or referents encoded as definite nominals (Foraker, 2011; Wilkin & Holler, 2011).

1.2.2 Meaning expressed in gestures, clause structures and first mentions

The gesture literature also highlights an important relationship between ‘what’ meaning gestures express and the way that information is packaged at the clause level in speech, specifically the way that speech packages and distributes information lexically and syntactically. The domain of motion has been investigated in this regard. Languages offer their speakers different possibilities to encode motion information (Aske, 1989; Berthele, 2004; Engberg-Pedersen & Trondhjem, 2004; Ragnasdóttir & Strömqvist, 2004; Slobin, 1996; 2004; Talmy, 1985, 1991 inter al.). For instance, for events involving manner and path of motion, speakers of some languages (so-called ‘satellite-framed’ languages, such as English and German) prefer to package both elements within a one-clause sentence (e.g., ‘he rolls down’), whereas speakers of other languages (so-called ‘verb-framed’ languages, such as Spanish or Japanese) prefer to package manner and path into separate clauses (e.g., ‘he descended while rolling’). Studies examining the accompanying gestures find that gestures reflect these encoding patterns functionally, even if the details vary somewhat across studies. On the whole, speakers of satellite-framed languages are more likely to encode both manner and path in one gesture, reflecting the preferred one-clause sentence (e.g., a ‘circling’ gesture moving downwards for ‘he rolls down’), than speakers of verb-framed languages who are comparatively more likely to encode manner and path in separate gestures, reflecting the two-clause sentence structure (e.g., a downward moving gesture, representing path, followed by a second ‘circling’ gesture, representing manner for ‘he descended while rolling’; e.g., Brown & Gullberg, 2008; Gullberg, Hendriks & Hickmann, 2008; Hickmann, Hendriks & Gullberg, 2011; Kita & Özyürek, 2003; Kita et al., 2007; Özyürek, Kita, Allen, Furman & Brown, 2005).

Turning to the way in which entities are mentioned for the first time in discourse, recall that speakers can use two main types of clause structures, more/less specialized. More specialized clause structures focus on the existence of referents, describing their location, appearance or relation to other referents. Less specialized clause structures, instead introduce referents within an event, in which they are the subject/agent (in an intransitive construction) or the object/patient (in a transitive construction). The studies on the speech-gesture parallelisms in motion events suggest that the semantic content of gestures may similarly reflect this variation in clause structure. For instance, it is possible that ‘entity’ gestures are used in contexts where the existence of a referent
is in focus, whereas ‘action’ gestures are used in contexts where the action a referent is involved in is in focus.

1.3 The present study

The present study examines what semantic content gestures express when they co-occur with first mentions of discourse referents, and specifically, whether they focus on entity or on action. Previous studies suggest that information status and nominal definiteness play an important role for this difference. However, those studies ignore the clausal context in which the referents are mentioned. Therefore, the current study also examines the clause structure in which referents are first mentioned.

The focus on first mentions adds to previous research in two important ways. First, previous studies mainly deal with the difference between first mentions (new referents) and subsequent mentions (given referents), or between different kinds of subsequent mentions. Second, first mentions are a good testing ground for how gestures relate to the three discursive domains of interest since they vary in information status (brand-new/ inferable), nominal definiteness (indefinite/definite), but also in how they are embedded in different clause structures (more/less specialized).

We hypothesize that the semantic content of gestures is sensitive to information status and nominal definiteness. We predict that a) gestures expressing entity information will align with brand-new referents and/or indefinite nominals, whereas gestures expressing action information will align with inferable referents and/or definite nominals. Furthermore, we hypothesize that the semantic content of gestures is also sensitive to variations in the clause structure used for first mentions. We therefore predict that b) gestures expressing entity information will align with referents in more specialized clause structures, whereas gestures expressing action information will align with referents in less specialized clause structures.

2 Method

2.1 Participants

Twenty native German speaking students from Ludwig-Maximilian University, Munich, Germany participated in the study (16 female, mean age 26). Participants were tested at their University, each of them together with a native German-speaking
friend who acted as listener/addressee. The role as speaker/addressee was randomly assigned. All participants provided written consent.

2.2 Materials and Task

A picture story about three fairies involved in baking, sewing and cleaning was used to elicit narrative speech and gestures. The story also contained a range of inanimate entities and consisted of 127 pictures in total (see example pictures in Figures 3-5 and Appendix B for a full list of entities that went into the analysis). The participants’ task was to retell the story as seen on the pictures to addressees who had not seen the pictures by answering the question ‘what happened?’.

Fairy 1 is involved in baking activities
Figure 3. Stimulus example 1

Fairy 2 is involved in sewing activities
Figure 4. Stimulus example 2

Fairy 3 is involved in cleaning activities
Figure 5. Stimulus example 3

2.3 Procedure

Speaker and addressee came to a quiet room at their University. The experimenter gave both participants instructions on paper but repeated the procedure again orally before starting the experiment. Participants sat across from each other. Only the speaker was filmed by a high-speed video camera capturing head and torso. The speaker was always given 4-6 pictures to memorize at once, and then retold them to the addressee by answering the question ‘what happened?’.

The addressee was asked to listen carefully and write a summary of each piece of retelling while the speaker was given the next set of pictures. The addressee was not allowed to ask any questions, but natural back channeling behavior was not explicitly discouraged. Addressees were also told not to cross arms or legs, but rather to put their arms on the upper thighs, in order to avoid

---

3 If it was the case though that a participant knew the topic of research by the experimenter (e.g., because she was someone from the local university who was involved in the same kind of research but wanted to help out by bringing along a friend), then she was automatically assigned as listener.
that speakers mirrored arm crossing behavior and thus, potentially stopped gesturing (see Chartrand & Bargh, 1999; Kendon, 1973). If speakers asked whether they were allowed to gesture, the experimenter responded that they should tell the story as was most comfortable for them. The experiment lasted between 45 – 90 minutes per session. Participants talked for an average of 20 minutes in total. At the end of the session, both participants filled out a consent form and speakers additionally filled out a (language) background questionnaire, based on Gullberg and Indefrey (2003).

2.3 Speech Coding

All narratives were transcribed verbatim. We identified all referential expressions mentioning a referent for the first time. We selected references to concrete entities (animate and inanimate) and excluded references to abstract and/or immaterial objects (9) and non-referential referents (10), since they are not trackable (cf. Chafe, 1994). References to the pictures themselves were also excluded.

(9)
sie hat eine Idee
‘she has an idea’

(10)
ein, das soll vielleicht so ein Mehl sack sein?
‘a, it might perhaps be a flour bag?’

We analyzed only core arguments for the purposes of the current study. These included subjects (11), complements in clauses using dummy subjects (12), and direct objects (13).

(11)
dann ist noch eine grüne Fee da
‘then is also a green fairy there’
2.3.1 Information status

Each expression was further coded as either brand-new or inferable. An inferable referent either stood in a part/whole or content/container relationship to a previously mentioned referent (e.g., sleeve/dress, egg shells/eggs, milk/milk can, sugar/sugar bowl), whereas there was no such relationship for brand-new referents. Rather, brand-new referents were ‘truly’ new in the sense that they did not inferentially link to the previous discourse (e.g., Birner, 2013; H. Clark, 1975; Fraurud, 1990; Prince, 1981, 1992). We found no other inferential links between a first mentioned and a previously mentioned referent. In some cases, a referent might have been inferentially linked to a previous activity or event. However, we chose to follow a conservative approach and only consider the relationship between current and trigger referent in the interest of replicability.

2.3.2 Noun phrase form

Each referential expression was categorized as either indefinite or definite. Indefinite nominals were those marked by no or indefinite determiners, as well as numerals (Nadel/ein Besen ‘needle/a broom’; drei Feen ‘three fairies’). Definite nominals were marked by definite articles, demonstratives and possessive pronouns (die/diese Fee ‘the/that fairy’; ihr Kleid ‘her dress’).

2.3.4 Clause structure

We categorized the clauses in which each referential expression occurred into more specialized clause structures for the introduction of referents expressing existence, or into less specialized clause structures (for the introduction of referents) expressing events (i.e., [in]transitive constructions).

More specialized clause structures included existential, locative and possessive constructions (14-18; E. Clark, 1978; Lyons, 1968). Furthermore, we included clauses in which a referent is presented in an external perspective (19; Hickmann et al., 1996). Note that in German existentials/locatives can also appear with posture verbs like stehen.
‘stand’ or *liegen* ‘lay’ (16-17). The following is a ranked list of all verbs used in more specialized clause structures by frequency of appearance: *sein* ‘to be’ (39%), *haben* ‘to have’ (26%), *stehen* ‘stand’ (10%), *man sieht* ‘one sees’ (8%), *liegen* ‘lay’ (6%), *schweben* ‘hover’ (6%), *es gibt* ‘there exists’ (3%), *es geht um* ‘it is about’ (2%).

(14)

*es gibt einen Tisch*

‘there is a table’

(15)

*da sind drei Feen*

‘there are three fairies’

(16)

*und es liegt außerdem noch keine Kanne am Boden*

‘and there lies also a can on the floor’

(17)

*und auf einem Tisch steht eine riesige Torte*

‘and on a table is/stands a big cake’

(18)

*und die dritte hat eine ziemlich hohe Torte*

‘and the third has a fairly tall cake’

(19)

*und man sieht noch zwei Nadeln*

‘and one also sees two needles’
For the less specialized clause structures, it is notable that intransitive constructions often contained subject-verb inversion and the first mentioned referent was typically instantiated as a semantic agent (20), whereas transitive constructions typically mentioned new referents as grammatical objects, taking on the semantic role of the patient, and following a known referent in subject position (21).

(20)

und dann fliegt ein angezündetes Streichholz herbei

‘and then a burning match flies by’

(21)

und füllt etwas Putzmittel in den Eimer rein

‘and fills some cleaning agent into the bucket

2.4 Gesture annotation

We identified gestures with the sound turned off in order to avoid circularity in the identification process. We identified manual gesture strokes (i.e., the most effortful parts of gestural movements) together with possible post-stroke holds (i.e., movement cessations of the hand[s] at the end of a gesture stroke) in frame-by-frame analysis of digital video in the software ELAN (Sloetjes & Wittenburg, 2008). Gesture strokes and possible post-stroke holds together represent the most meaningful parts of the gestural movements (Kendon, 2004; Kita, 1990; McNeill, 1992). Each gesture stroke ± post-stroke hold that aligned with at least one syllable of a first mentioned referent in speech was further coded for the meaning it expressed (cf. Hickmann et al. 2011; McNeill, 1992; Stam, 2006).

2.5 Semantic content coding of gestures

We divided gestures into ‘entity’ and ‘action’ gestures following Wilkin and Holler (2011)’s definitions, and coded them with the sound turned off. The coders had the story pictures at their disposal in order to know to which scenes the speakers were referring. Crucially, however, the coding was done without direct access to speech in order to avoid circularity. When the aim is to examine the relationship and possible parallelism between speech and gesture, it is vital not to code the semantic content of gestures relying on speech. However, access to the pictures provides the coder with a semantic search space for entities and actions.
‘Entity’ gestures express information about referents’ shape, size or location. For instance, speakers can model the shape of a multi-level cake by using both hands or by using a finger to draw it (cf. Figure 1). Furthermore, speakers can indicate a referent’s location by pointing to or placing a gesture in a certain location in gesture space. In Figure 6, after having talked about a cake standing on a table, the speaker indicates the location of some candles on top of the cake by using a localizing gesture, drawing a little circle with her extended index finger. The techniques of representation correspond to what Kendon (2004) calls depiction and modeling (Table 1).

![Figure 6: ‘Entity’ gesture indicating location](image)

‘Action’ gestures primarily express information about referents’ actions and movements. A speaker might use a hand shape that represents holding a broom and move her hand from side to side in order to enact a sweeping action. Or she might draw an upward path with her finger in order to depict how an entity went up some stairs (cf. Figure 2). These gestures might also include information about referents, such as when a speaker represents the shape of a match by using a flat hand shape while at the same time moving the hand around to represent its path. In comparison to ‘entity’ gestures, however, ‘action’ gestures seem to foreground the action whereas ‘entity’ gestures solely encode the properties of referents themselves (Wilkin & Holler, 2011). The techniques of representation correspond to what Kendon (2004) calls enactment and depiction (Table 1).

2.6. Interrater reliability

A second German native speaker recoded 20% of the total amount of referential expressions for information status (brand-new/inferable), nominal definiteness
(indefinite/definite) and clause structure (more/less specialized). A third coder annotated 20% of all gestures used in the analysis. They identified the stroke phases of gestures in the target clauses (containing first mentions). Agreement between coders for the identification of gestures was reached when both coders identified a gesture that co-occurred with the same first mention (a match). If one coder identified a gesture co-occurring with a first mention, but the other coder did not (and instead identified a gesture that aligned with a neighboring word, for example), there was no match. Finally, a fourth coder categorized 20% of all gestures as ‘entity’ versus ‘action’ gestures. Agreement between coders and interrater reliability computed using Cohen’s kappa (where applicable) are given in Table 2.

Table 2. Interrater agreement and reliability

<table>
<thead>
<tr>
<th>Modality</th>
<th>Variable</th>
<th>Agreement</th>
<th>Cohen’s kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td>Information status</td>
<td>81%</td>
<td><em>Kappa</em> =.799, SE =.049</td>
</tr>
<tr>
<td>Speech</td>
<td>Definiteness</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Speech</td>
<td>Clause structure</td>
<td>99%</td>
<td><em>Kappa</em> =.968, SE =.032</td>
</tr>
<tr>
<td>Gesture</td>
<td>Identification</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Gesture</td>
<td>Semantic content</td>
<td>86%</td>
<td><em>Kappa</em> =.726, SE =.069</td>
</tr>
</tbody>
</table>

2.6 Analyses

We used linear mixed effects models with the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017) in RStudio (RStudio Team, 2016) to analyze the relationships between variables. All variables with their levels are summarized in Table 3. All variables, and most importantly the dependent variable, were binary, which motivated the use of variants of the generalized linear models with binomial error structure (i.e., command: family ‘binomial’). Importantly, linear mixed effects models allow us to use raw data instead of (problematic) transformed proportions (see Jaeger, 2008 for more details).

Table 3. Analysis

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Levels</th>
<th>Predictor variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic content of gestures</td>
<td>Entity/Action</td>
<td>Information status</td>
<td>Brand-New/Inferable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Definiteness</td>
<td>Indefinite/Definite Nominal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clause structure</td>
<td>More/less specialized</td>
</tr>
</tbody>
</table>
We ran 15 analyses in total. The first, simplest model included no predictor variables, and the last, most complex model included all three predictor variables with two interactions. Between these models, we also ran all other logically possible combinations of the predictor variables. All models also contained ‘subject’ as a random factor. We then compared the AIC values (the Akaike information criterion) of all models in order to determine the model with the best predictive accuracy (Long, 2012). Specifically, AIC is an estimate which allows us to define the model that will be best if applied to out-of-sample data. Lower AIC values are better than higher values (Long, 2012; see Appendix A for a list of all models and their ranking according to AIC values).

3 Results

The results are based on 462 gestures that aligned with first mentions and either focused on ‘entity’ or on ‘action’. On average, participants produced more gestures focusing on entity (74%, $SD = 11\%$) than on action (26%, $SD = 11\%$). We tested whether the three independent variables information status, definiteness, and clause structure had an effect on the use of ‘entity’ versus ‘action’ gestures. The model with the lowest AIC value included all three predictor variables with an independent effect of clause structure and an interaction between information status and definiteness. Figure 7 presents the predicted proportions of ‘entity’ gestures occurring with brand-new versus inferable referents in more versus less specialized clause structures.

![Figure 7: Proportion of ‘entity’ gestures by definiteness encoding for brand-new versus inferable referents in more versus less specialized structures (with Confidence Intervals)](image)
The analyses revealed an independent effect of clause structure. This indicates that first mentions in more specialized clause structures are more likely to be accompanied by gestures focusing on entity information (82%). Conversely, first mentions in less specialized clause structures are more likely to be represented by gestures focusing on action information (69%; $EST = 2.48$, $SE = 0.32$, $z$-value = 7.85, $p = .000$).

Furthermore, there was an independent effect of definiteness ($EST = 0.98$, $SE = 0.34$, $z$-value = 2.87, $p = .004$), and more importantly, a significant interaction between definiteness and information status ($EST = -1.14$, $SE = 0.58$, $z$-value = -1.96, $p < .05$). This suggests that referents expressed by definite/indefinite nominals are represented differently in gesture depending on their information status. Post-hoc analyses of contrasts (with the multcomp package; Westfall, Tobias, Rom, Wolfinger & Hochberg 1999) show that brand-new referents expressed by indefinite nominals are more likely to be represented by ‘entity’ gestures than by ‘action’ gestures (91% vs. 79% in more specialized structures and 46% vs. 24% in less specialized clause structures; $EST = 0.98$, $SE = 0.34$, $z$-value = 2.87, $p = .008$). There was no difference for inferable referents (79% vs. 82% in more specialized clause structures and 24% vs. 27% in less specialized clause structures; $EST = -0.16$, $SE = 0.46$, $z$-value = -0.34, $p = .929$).

Figures 8 and 9 illustrate the main patterns. The speaker in Figure 8 produces an ‘entity’ gesture for a brand-new entity expressed as indefinite nominal in a more specialized clause structure. The speaker in Figure 9 produces an ‘action’ gesture for an inferable entity expressed as a definite nominal in a less specialized clause structure.
and then there is a basket

Figure 8. Example of a shaping gesture (entity focus) for a brand-new referent expressed with an indefinite nominal in a more specialized clause structure (an existential/locative)

das Ei
dotter ist im Begriff in die Schüssel zu fallen
‘the egg yolk is about to fall into the bowl’

Figure 9. Example of a path gesture with the right hand (representing a trajectory) for an inferable referent instantiated as definite nominal in a less specialized clause structure (an intransitive construction)
4 Discussion

The aim of the present study was to examine the variation in the semantic content of gestures (entity/ action) that co-occur with first mentioned entities in discourse. More specifically, we examined the relationship between the semantic content of gestures and information status of first mentions (brand-new/inferable), nominal definiteness (indefinite/definite), and the clause structures in which first mentions are embedded (more/less specialized for the introduction of referents).

In line with the predictions, the results show that there is a relationship between the semantic content of gestures and nominal definiteness together with information status. Specifically, gestures with a focus on entity tended to align with brand-new referents encoded as indefinite nominals. In comparison, gestures with a focus on action tended to align with inferable referents encoded as definite nominals. These findings are in accordance with previous research on this topic (Foraker, 2011; Wilkin & Holler, 2011), and critically also extend it to the context of first mentions.

Importantly, however, in line with the second prediction, the results also reveal that the semantic content of gestures is sensitive to the clause structure in which first mentions are embedded. Entity gestures occurred more often with first mentions of referents in more specialized clause structures for the introduction of referents (e.g., ‘there is/she has a broom’). Action gestures, on the other hand, occurred more often with first mentions of referents in less specialized clause structures for the introduction of referents, but focused on events (e.g., ‘she takes out a bow’ or ‘a bow comes flying by’). These results are generally in line with previous research showing that the way information is syntactically and lexically packaged within or across clauses is relevant for the semantic content of gestures (e.g., Özyürek et al., 2005). Importantly, however, the current results extend those findings by examining new clausal contexts and by specifically focusing on each gesture in its clausal context.

Finally, the results are also commensurate with studies showing that variations in sentence construction, such as transitivity, influence the way that gestures represent information (McNeill, 1992; Parrill, 2010; see also Debreslioska & Gullberg, 2017; Debreslioska et al., 2013). The important difference between studies on the relationship between gesture representation and transitivity and the current study is that the former is concerned with the question of ‘how’ gestures represent, whereas the current study deals with ‘what’ gestures represent. Furthermore, studies on transitivity and gesture typically concentrate on events (Parrill, 2010), whereas the current study compares first mentions in structures expressing events (including both transitive and intransitive constructions of events) to structures focusing on the existence of referents (such as ‘presentative’ structures). These results thus complement previous research on the introduction of referents in (in)transitive events. For instance, Debreslioska and Gullberg (2017) show that C-VPT gestures are more likely to be used for introducing
new referents in transitive constructions. Debreslionska et al. (2013) further show that O-VPT gestures are more likely to occur with reintroduced referents in intransitive events. The new insight from the current study is that those gestures aligning with (re)introductions of referents within events (C-VPT and O-VPT gestures) typically express information about ‘action’. And furthermore, that this tendency changes once we consider referents in clauses that do not express events, but rather focus on the existence of a referent. In the latter context, gestures instead tend to focus on ‘entity’ information.

The findings therefore suggest that all three domains examined in the current study (information status, nominal definiteness, and clause structure) need to be considered when we examine information expressed in gestures. In particular, the interplay between clause structure and the semantic content of gestures highlights the parallelism between the two modalities. That is, when a more specialized clause structure focuses on the existence of a referent, then gesture also focuses on entity information. When a less specialized clause structure focuses on an event in which a referent is involved, then gesture also focuses on action information involving the referent.

This contrast between more and less specialized clause structures for the introduction of referents is similar to the distinction that Du Bois (1980) makes between clauses that are in descriptive or in narrative mode, respectively. Clauses in the narrative mode advance the story whereas clauses in the descriptive mode typically do not have this function, but rather describe entities, their location or relationships to other discourse entities (see also McNeill, Cassell & Levy, 1993; McNeill & Levy, 1982; McNeill, Levy & Pedelty, 1990). Thus, it seems that representational gestures focusing on action (and co-occurring with clauses in the narrative mode) function to advance the story and gestures focusing on entity information (and co-occurring with clauses in the descriptive mode) function to describe characteristics of entities that do not necessarily advance the story to the same degree. From this point of view, it also makes sense that action gestures pattern with definite nominals/inferable entities, and entity gestures pattern with indefinite nominals/brand-new entities. When entities 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 entities are brand-new as indicated by indefinite nominals, the speaker is more likely to first indicate the properties of the entity in question before moving on to advancing the story.

Finally, the study highlights the importance of examining gestures in both their narrow and wider contexts. Gesture does not link to speech at a one-word-one-gesture level, but gestures are rather related to ‘conceptual affiliates’ (De Ruiter, 2000; Gullberg, 1998). This is to say that a gesture can be tightly related to a referential expression with which it aligns (e.g., Figure 8, where a gesture is used to shape the form of the entity ‘basket’). But the gesture can also be sensitive to the clausal context in which it is embedded (e.g., produced in C-VPT/O-VPT depending on clause transitivity;
McNeill, 1992). The current study particularly highlights this double influence in the context of discourse by showing that gestures are shaped by local and global markings of newness just like speech. Depending on its information status in discourse, a referent will be marked locally in terms of nominal definiteness and also globally by a variation in clause structure. The semantic content of gestures follows this pattern by co-varying with both local and global markings. Through this parallelism the tight connection between the modalities comes to light, and highlights in particular how, by working together on both word and clause levels, gestures and speech together achieve cohesion in discourse.

5 Conclusion

In conclusion, the present study provides evidence that the tight link between speech and gestures is observable on the word, clause, and the discourse level. This is reflected in the interplay between the semantic content of gestures and information status with definiteness on the one hand, and the variation of the semantic content of gestures with clause structures that fulfill different discourse functions on the other. This parallelism between speech and gestures provides further important evidence for the view that the two modalities are tightly linked, and form an integrated system (Kendon, 2004; McNeill, 1992).

6 References


7 Appendices

Appendix A. Model selection

<table>
<thead>
<tr>
<th>Model</th>
<th>Specification</th>
<th>Predictors</th>
<th>AIC</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gesture ~ 1 + (1</td>
<td>Subject)</td>
<td>none</td>
<td>728.7</td>
</tr>
<tr>
<td>2</td>
<td>Gesture ~information status + (1</td>
<td>Subject)</td>
<td>1</td>
<td>640.8</td>
</tr>
<tr>
<td>3</td>
<td>Gesture ~definiteness + (1</td>
<td>Subject)</td>
<td>1</td>
<td>706.2</td>
</tr>
<tr>
<td>4</td>
<td>Gesture ~ clause structure + (1</td>
<td>Subject)</td>
<td>1</td>
<td>581.0</td>
</tr>
<tr>
<td>5</td>
<td>Gesture ~ information status + clause structure + (1</td>
<td>Subject)</td>
<td>2(simple)</td>
<td>518.8</td>
</tr>
<tr>
<td>6</td>
<td>Gesture ~ information status + definiteness + (1</td>
<td>Subject)</td>
<td>2(simple)</td>
<td>631.2</td>
</tr>
<tr>
<td>7</td>
<td>Gesture ~ clause structure + definiteness + (1</td>
<td>Subject)</td>
<td>2(simple)</td>
<td>574.2</td>
</tr>
<tr>
<td>8</td>
<td>Gesture ~ information status * clause structure + (1</td>
<td>Subject)</td>
<td>2(interaction)</td>
<td>520.6</td>
</tr>
<tr>
<td>9</td>
<td>Gesture ~ information status * definiteness + (1</td>
<td>Subject)</td>
<td>2(interaction)</td>
<td>626.8</td>
</tr>
<tr>
<td>10</td>
<td>Gesture ~ clause structure * definiteness + (1</td>
<td>Subject)</td>
<td>2(interaction)</td>
<td>575.4</td>
</tr>
<tr>
<td>11</td>
<td>Gesture ~ information status + clause structure + definiteness + (1</td>
<td>Subject)</td>
<td>3(simple)</td>
<td>515.8</td>
</tr>
<tr>
<td>12</td>
<td>Gesture ~ information status * clause structure + definiteness + (1</td>
<td>Subject)</td>
<td>3(interaction)</td>
<td>517.4</td>
</tr>
<tr>
<td>13</td>
<td>Gesture ~ information status + clause structure * definiteness + (1</td>
<td>Subject)</td>
<td>3(interaction)</td>
<td>517.0</td>
</tr>
<tr>
<td>14</td>
<td>Gesture ~ information status * definiteness + clause structure + (1</td>
<td>Subject)</td>
<td>3(interaction)</td>
<td>513.9</td>
</tr>
<tr>
<td>15</td>
<td>Gesture ~ information status<em>clause structure</em>definiteness + (1</td>
<td>Subject)</td>
<td>3(interaction)</td>
<td>518.7</td>
</tr>
</tbody>
</table>

Appendix B: List of entities taken into consideration

Fairies, table, cake, candles, broom, mop, bucket, dress, mannequin, wands, stairs, belt, needle(s), thread, scissors, piece(s) of cloth, collar, bow(s), cloth triangles, bowl, milk can, sugar dots, flour bag, icing bag, sugar hearts, dough, sugar sprinkles, eggs, spoon, sugar bowl, saltshaker, dust, leaves, stars/sparks, pot, water, cleaning agent, cook book, shoes, basket, box, match

Parts/content of entities:

Parts of broom (e.g., broomstick), parts of egg (e.g., egg shell), parts of fairies (e.g., hands, buttocks), parts of dress (e.g., upper, lower part, sleeves), parts of cake (e.g., upper part), parts of cloth (e.g., a piece of it), parts of stairs (e.g., stair head), salt, sugar, milk, flour
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.