‘What is where?’ or ‘Where is what?’

Bimodal spatial descriptions in English and German

Angela Schlatter

Supervisor: Prof. Dr. Marianne Gullberg

Centre for Language and Literature, Lund University
MA in Language and Linguistics, English
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

November 2018
To Lara
Abstract

The aim of this study was to replicate Carroll et al. (2000) in order to investigate the planning processes in language production by contrasting English and German spatial descriptions, and to explore the contribution of gestures in the process of discourse conceptualization. Carroll et al. (2000) have shown that there is cross-linguistic variation in the perspective under which information is selected and interrelated in texts, specifically in English and German spatial descriptions due to differences in the structure of the linguistic system of the respective languages. Moreover, Carroll et al. (2000) have demonstrated that the domains ‘object’ and ‘space’ play a central role in interrelating information for expression and creating cohesion in texts. As gestures and speech form an integrated system and reflect the way information is encoded in speech, it was assumed that the analysis of gesture handshape and the temporal alignment of gestures with speech would provide a fuller picture of event conceptualization in English and German spatial descriptions (cf. Gullberg, 2011). To test Carroll et al.’s (2000) claims and the contribution of gestures in the planning process, an empirical study was conducted to elicit speech and speech-associated gestures in a director-matcher task. Based on previous findings, it was hypothesized that gestures would differ across languages in terms of their form and timing with speech. The speech results in Carroll et al. (2000) were successfully replicated lending support to their findings that in English an object-based perspective is central for information organization in spatial descriptions, while in German a spatially-based perspective is more frequent. In the current study, this difference in perspective was reflected in the linguistic forms chosen in reference introduction, maintenance, and word order which form a typological pattern (Carroll et al., 2000). For gestures, the difference in perspective was not manifested in the gestural handshapes, but in the temporal alignment of the gestures with speech. Most gestures in both languages were articulated with a ‘figure-incorporating’ handshape (cf. Gullberg, 2011) due to the nature of the task. In English, gestures were most frequently aligned with the object, whereas in German, gestures aligned with the object or both the object and the locative expression in speech. The same pattern for gesture form and alignment in both languages was also observed for the post hoc analysis of sequential
gestures (i.e. gestures which immediately followed a post-stroke hold). In conclusion, the gesture analysis supports the successfully replicated speech findings and provides additional evidence that speech-associated gestures contribute both to the semantic content as well as the structural properties of an utterance.

Keywords: language-specific perspectives, event conceptualization, information structure, spatial descriptions, speech-associated gestures, English, German
Acknowledgements

The completion of this work would not have been possible without the professional and personal assistance of many people.

First and foremost, I would like to express my sincere gratitude to my supervisor Marianne Gullberg for her guidance, understanding and patience throughout the process of this thesis. I am fortunate to have been able to profit from her expert knowledge on gestures, which always pointed me in the right direction and her valuable comments on earlier drafts of this thesis.

I also wish to thank my study counselor Satu Maninnen for her assistance and advice during my time at Lund University.

A big thank you to the Humanities Lab, especially to Peter Roslund for his technical support, without whom the experiment and the analysis would not have run so smoothly. A special thanks goes to the confederates of this experiment, Soraya Makki and Vera Trojansky, who were always flexible and took time in their busy schedule to assist me in the lab. I am thankful to all the participants for showing their interest in my study and taking part in my experiment.

A special mention also goes to Sandra Debrèsloska who provided me with helpful tips for my experiment. It was a pleasure to take part in her experiment and to not only discuss gesture research, but also exchange experiences of motherhood. I am thankful for Yu Bin who came to my rescue when I desperately needed a babysitter during the recording sessions.

A big thank you also to my colleagues in Malmö and Lund who made my time in Sweden an unforgettable experience

The final thank goes to my family and friends near and far who always supported and encouraged me to follow through with the completion of this work. For this and much else, I cannot thank them enough.
Table of contents

List of figures........................................................................................................................................ vii
List of tables........................................................................................................................................ vii
Abbreviations......................................................................................................................................... ix

Chapter 1 Introduction ......................................................................................................................... 1

Chapter 2 Theoretical background.................................................................................................... 5
  2.1. The replicated study ..................................................................................................................... 5
  2.2 Semantic model of information organization .............................................................................. 7
    2.2.1 Conceptualization and planning in the language production process .............................. 7
    2.2.2 Quaestio and perspective-taking in language ................................................................ 11
  2.3 Basic concepts and categories in locative relationships ............................................................ 13
  2.4 The existential construction: between existence and location ................................................ 19
    2.4.1.1 The existential there-construction in English .............................................................. 21
    2.4.1.2 Existential constructions in German: es gibt- and sein-constructions ....................... 22
  2.5 Gestures ...................................................................................................................................... 26
    2.5.1 An integrated system: gesture and speech .......................................................................... 26
    2.5.2 Classification and composition of gesture movements .................................................... 28
    2.5.3 Spatial relationships and gesture ....................................................................................... 30
  2.6 Summary .................................................................................................................................... 32

Chapter 3 The current study .............................................................................................................. 34
  3.1 Research questions and hypotheses ............................................................................................ 34
  3.2 Method ...................................................................................................................................... 37
    3.2.1 Participants ........................................................................................................................ 37
    3.2.2 Materials .......................................................................................................................... 38
    3.2.3 Procedure ........................................................................................................................ 39
    3.2.4 Coding and Analysis ......................................................................................................... 43
List of figures

Figure 1. Processes involved in language production (adapted from Warren, 2017, p. 16) ........ 8
Figure 2. The cat hunted the mouse. ......................................................................................... 9
Figure 3. Diagram of gesture movement levels proposed by Kendon (1980) (adapted from McNeill, Pedelty & Levy, 1990, p. 209). ........................................................................... 29
Figure 4. Schematic representation of the experimental set-up ......................................................... 40
Figure 5. Handshape category for figure-object ............................................................................. 49
Figure 6. Handshape categories for spatial location .......................................................................... 50
Figure 7. Example gesture sequence and rest position of the hands .................................................. 65
Figure 8. Examples of gesture sequences in English speaker 1 .......................................................... 66
Figure 9. House in rural landscape (practice picture). .................................................................... 94
Figure 10. City center (picture 1). .................................................................................................. 95
Figure 11. House in changing rural landscape (picture 2). ............................................................... 95
Figure 12. Screenshot of consent form in English .......................................................................... 102
Figure 13. Screenshot of consent form in German. ......................................................................... 103

List of tables

Table 1. Talmy’s (2000) concepts to describe Motion events (adapted from Saeed, 2016, p. 273) .................................................................................................................................................. 13
Table 2. Distribution of main frames of reference across a language sample (adapted from Levinson and Wilkins, 2006b, p. 543). .................................................................................................................. 16
Table 3. Experimental items for each picture (x=experimental). .......................................................... 44
Table 4. Coding categories for speech (following Carroll et al., 2000) ........................................... 44
Table 5. Coding categories for gestures ............................................................................................ 51
Table 6. Comparison of aggregated proportions for existential and locational constructions in reference introduction. .......................................................................................................................... 53
Table 7. Mean proportions for reference introduction for every analysis category ....................... 53
Table 8. Aggregated proportions for *es gibt* and *SEIN constructions* in the German data........ 54
Table 9. Comparison of aggregated proportions for *object-oriented* and *space-oriented*
expressions in reference maintenance................................................................. 56
Table 10. Mean proportions for reference maintenance for every analysis category............. 57
Table 11. Comparison of aggregated proportions for *object-specific* and *object-neutral*
expression in picture division.............................................................................. 59
Table 12. Mean proportions for spatial concepts and subspaces division for every analysis
category.............................................................................................................. 60
Table 13. Comparison of aggregated proportions of *relatum-theme* or *theme-relatum* word order.
.......................................................................................................................... 60
Table 14. Mean proportions for word order for every analysis category. .......................... 61
Table 15. Mean proportions for handshape in English and German. ................................. 62
Table 16. Mean proportions of gestural alignment in English and German......................... 63
Table 17. Mean proportions of gestural alignment OBJ occurring with lexical noun (N) or
without (-N). ........................................................................................................ 63
Table 18. Mean proportions for gestural alignment and *object-handshape* in English and
German.................................................................................................................. 64
Table 19. Mean proportions for gestural alignment and *space-handshape* in English and German.
.......................................................................................................................... 64
Table 20. Mean proportions for *overall gesture sequences* in English and German........... 67
Table 21. Aggregated proportions of sequential gestures for *gesture handshape*................ 67
Table 22. Aggregated proportions of sequential gestures for *speech alignment*............... 67
Table 23. Aggregated proportions of sequential gestures for *handshape and speech alignment*. 68
Table 24. Examples of gesture sequences in both the English and German data............... 69
Table 25. Participant information (English). ..................................................................... 92
Table 26. Participant information (German)...................................................................... 93
Table 27. Task instructions for participant. ...................................................................... 104
Table 28. Task instructions for confederate................................................................... 104
Table 29. Task instructions for participant. ...................................................................... 106
Table 30. Task instructions for confederate.............................................. 106
Table 31. Wordlist for each picture (in English). ........................................ 108
Table 32. Wordlist for each picture (in German)......................................... 108

**Abbreviations**

1 first person
3 third person
ACC accusative
ADV adverbial
DAT dative
DEMP demonstrative pronoun
EX existential
F feminine
GEN genitive
L1 first language
L2 second language
L3 third language
LOC locative
N noun
NP noun phrase
NOM nominative
OBJ object
PL plural
PROADV proadverbial
PRF perfect
PRS present
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTCP</td>
<td>participle</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SP</td>
<td>space</td>
</tr>
<tr>
<td>V2</td>
<td>verb second</td>
</tr>
</tbody>
</table>

*Note:* The glossing of all the examples throughout the thesis follows the “Leipzig Glossing Rules”. 
Chapter 1 Introduction

“A written discourse is to a face-to-face conversation as a stuffed grizzly bear is to a live one”. (Clark, 1994, p. 987)

Spatial concepts and the way in which language structures space in context has generated a lot of research in semantics, specifically in cognitive linguistic as well as cognitive science in general. That is, the strong interest in space has constituted the basis for an ongoing debate on the relationship between language and thought. Specifically, it has been argued that, on the one hand, the way speakers divide space is a reflection of the cognitive structure (cf. e.g. Talmy, 2000). On the other hand, arguments have been provided that languages use language-specific concepts to structure space providing a strong case for linguistic relativity (cf. e.g. Levinson, 2003). Previous studies have not only focused on the meaning of spatial expressions (e.g. Landau & Jackendoff, 1993), but also on the way in which spatial concepts are mapped onto linguistic form and are used to structure space (e.g. Talmy, 1983; 1987). Overall, spatial relationships can be divided into two broad categories. The first includes dynamic motion events (e.g. John ran into the room) and the second static locative relationships (e.g. John is in the room) (Tutton, 2013a, p. 25ff). On the one hand, there has been a considerable amount of research and cross-linguistic analyses within Talmy’s conceptual framework of spatial semantics and motion events (Talmy 2000), which includes his highly influential dichotomy of satellite- and verb-framed languages that has been further developed to a more gradient system (e.g. Kita & Özyürek, 2003; Kopecka & Pourcel, 2005; Kopecka, 2006; Lemmens, 2005; Özyürek & Kita, 1999; Slobin, 2004; 2006; Zlatev & Yangklang, 2004). On the other hand, static locative relationships are a less investigated area. Rather than focusing on the locative expressions speakers use to talk about static spatial relationships (e.g. Tutton, 2011; 2012; 2013a; 2013b; 2016), this study is more interested in the process of conceptualization and how information is organized for expression, particularly in spatial descriptions produced by native English and German speakers.
The main interest in this thesis is the cross-linguistic expression of spatial information in both speech and gestures (i.e. the spontaneous visible bodily actions, especially movements with the arms and hands, which people produce when they talk (cf. Gullberg, 2011; Kendon, 2004; McNeill, 1992)). In particular, the current study seeks to replicate a study by Carroll, Murcia-Serra, Watorek and Bendiscoli (2000) examining possible differences across languages in conceptualization in a spatial description task. When speakers plan to talk about events, they must select the relevant information for expression and how to encode this information in speech. This has been generally known as ‘conceptualization’, ‘event construal’ and ‘perspective-taking’ (Gullberg, 2011, p. 167). Previous research undertaken by von Stutterheim and colleagues has revealed that there is cross-linguistic variation in the perspective under which information is selected and interrelated in texts (i.e. a structured series of related sentences (cf. von Stutterheim, 2003)). These differences have been described in particular for English and German in spatial description tasks (e.g. Carroll 1993; Carroll & von Stutterheim, 1993; Carroll, et al., 2000). In these studies, text production is understood as a form of problem solving where the text is considered to be an answer to a question or task (i.e. ‘quaestio’) that underlies the communicative event. The communicative tasks to be solved vary in kind and provide constraints on “the type of information selected for expression, the assignment of topic-focus structure, the type of relations established between informational units and the choice of linearization principle” (von Stutterheim, 2003, p. 184).

In addition to replicating Carroll et al. (2000), the current study aims to look at what additional insights gestures can provide about such conceptualization and considers the relationship between gestures, speech and language. The analytic approach taken in this thesis is based on the arguments that gestures provide access to spatial cognition and conceptualization more generally. Specifically, it has been argued that gesture vary cross-linguistically, not only due to cultural differences, but also due to differences between the lexicon and the structure of languages, as gestures are linked to speech. That is, previous studies have shown that gestures reflect both the conceptual mental model of what is being depicted and the linguistic encoding of that model (cf. e.g. Gullberg, 2011; Kita & Özyürek, 2003). In other words, gestures can relate
either to the pre-linguistic conceptualization of what is being talked about, to the linguistic formulation of this model, or to both. These claims are based on the assumption that gestures are an integral part of the interactional event, which cannot be studied simply by looking at a written transcription. In other words, gestures and speech are understood to form a unity (cf. Gullberg, 2011; Kendon, 1972; 1980; 2004; McNeill, 1992).

The aim of this thesis is thus twofold. Taking a cognitive linguistic approach, the current study first tries to replicate the speech results in Carroll et al.’s (2000) empirical study. Specifically, Carroll et al. (2000) have shown that the spatial and object domains play a central role in interrelating information and choosing the grammatical means to describe static locative relations in English and German. In addition, Carroll et al. (2000) have argued that the contrasts in selecting and structuring information in speech can be generalized to assume contrasts at the level of abstract principles of perspective-taking which are rooted in language-specific patterns of a typological nature. Secondly, through the analysis of speakers’ concomitant gesture production, this study aims to shed new light onto the cross-linguistic differences in perspective-taking in information selection and organization in both English and German descriptions. Taking previous research into account (e.g. Carroll, 1993; Carroll & von Stutterheim, 1993; Carroll et al., 2000; Gullberg, 2011), the following ‘overarching’ research questions might be posed:

To what extent are the planning processes required for the production of coherent spatial descriptions in English and German language-specific and related to structural properties of the respective languages?

What do gestures reveal about the production process, specifically about how native English and German speakers conceptualize and lexicalize spatial descriptions?

The main focus of the thesis is information structure and discourse cohesion as multimodal phenomena. To answer the research questions an experiment was conducted to elicit speech and gestures in English and German spatial descriptions. The design of the study is a partial replication of Carroll et al.’s (2000) study, focusing only on spatial descriptions by native English and German speakers.
This thesis is divided into six chapters. The theoretical background is presented in Chapter 2, which outlines the framework of this study and introduces the previous results this thesis seeks to replicate. The aims of the study and specific hypotheses are outlined in Chapter 3, which explains the methodology and design of the experiment. Chapter 4 presents the speech and gesture results, whereas Chapter 5 provides a discussion of the results in light of previous research and the specific hypotheses formulated supplemented by some concluding remarks about the limitations of this study and implications for future investigations. Finally, general conclusions are given in Chapter 6.
Chapter 2 Theoretical background

In this chapter, the theoretical frameworks as well as empirical studies providing the basis for this investigation are introduced. The chapter is divided into four subsections. Firstly, Carroll et al.’s (2000) study which this thesis aims to replicate and the semantic model of information organization applied in Carroll et al. (2000) will be introduced. Secondly, basic concepts within spatial semantics will be briefly summarized. Thirdly, the existential construction in English and German, and finally, theories and previous research on gestures and spatial relationships will be presented.

2.1. The replicated study

The main aim of this study is to replicate Carroll et al.’s (2000) study in order to investigate how spatial information is conceptualized and organized for speech production in a spatial description task in English and German. It also seeks to shed some new light on previous findings by investigating the contribution of gestures in the planning process.

A number of studies have shown that when presenting the same task to speakers of different languages, they display variation in the perspective under which information is selected and interrelated within the text (e.g. Carroll, 1993; Carroll & von Stutterheim, 1993; Carroll et al., 2000; Carroll & Lambert, 2003; von Stutterheim, 2003; von Stutterheim & Nüse, 2003). On the one hand, possible factors were considered to be culturally determined traditions or styles in presenting information implemented by institutions, such as schools. On the other hand, the structure of the linguistic systems used for encoding information that determines the presentation of information was investigated and found to be another reason.

To take a case in point, Carroll et al. (2000) were interested in investigating the close interaction between typological features of languages and information structure and why advanced
second language learners of German, who show a high level of proficiency in formal syntactic terms, do not show native-like patterns in their organization of descriptive texts. Carroll et al. (2000) found that the selection of linguistic means used in spatial description tasks in English, Romance languages (i.e. French, Italian and Spanish) and German are guided by principles that are perspective driven and associated with patterns of grammaticization. Moreover, Carroll et al. (2000) demonstrated the central role of the spatial and object domains in interrelating information in descriptive texts. Specifically, in the comparison of spatial descriptions, language-specific preferences were detected in what they call an object-based perspective (e.g. “I'm going to describe a beautiful city. There are tall narrow buildings like you would see in a European city. There's a canal running through it and there are lots of trees”), which plays a significant role in information organization in both English and Romance languages, and a spatially-based perspective (e.g. “Starting on the left-hand side there is an old castle; further on toward the center there is a small square”), which predominantly occurred in German (Carroll et al., 2000, p. 445). Carroll et al. (2000) note that in German it is also possible to take a thematic perspective and focus more on objects in the description.

The difference in perspective between the respective languages is reflected in the selected linguistic means used in reference introduction, maintenance, the division of the picture into subspaces and word order. The object-based perspective leads to the use of existentials in reference introduction and nominal forms in reference maintenance. The spatially-based perspective shows a high rate of locationals in reference introduction, whereby through the use of adverbial forms objects are treated in their status of being located at a space in reference maintenance. Moreover, Carroll et al. (2000) point out that the spatially-based perspective in German is also supported by grammaticized means to express this function. That is, the ‘proadverbials’, such as davor ‘there in front’, constitute a paradigm which makes it possible not to mention the object in reference maintenance. Instead, the object only features implicitly in the form of the spatial adverb there (p. 460). Carroll et al. (2000) indicate that second language learners of German have acquired the spatial perspective required in almost all functions. That is, learners do not make sufficient use of proadverbials; the overall principle that requires their use is not yet fully anchored in the learner
variety. Hence, Carroll et al. (2000) conclude that structures in languages that reflect central principles of information organization are difficult to acquire for second language learners, who need to recognize different clusters of form-function relations that range over different domains.

Carroll et al. (2000) carried out their analysis within the framework of a ‘semantic model of information organization’ (cf. Klein & von Stutterheim 2002; von Stutterheim 1997; von Stutterheim & Klein 1989). Within this framework it is assumed that “information selection and packaging in context is determined by perspective-driven processes of information organization” (Carroll et al., 2000, p. 444). Moreover, a text is understood as an organized structure that answers a question or ‘quaestio’ which is produced once a perspective has been selected. In the following, important notions and concepts within this adopted framework will be further elaborated.

2.2 Semantic model of information organization

2.2.1 Conceptualization and planning in the language production process

Current models of language production group relevant tasks involved in producing spoken language into three levels: (1) ‘conceptualization’, (2) ‘formulation’ and (3) ‘articulation’ (cf. Levelt 1989), as illustrated in Figure 1 by Warren (2017, p. 16). In the process of conceptualization, presumed to be pre-linguistic, speakers use their world knowledge about the current situation to sort out their ideas. The result of this process is a ‘pre-verbal message’. To put it in another way, the language production process starts with some abstract idea of what the speaker wants to say (i.e. ‘mentalese’ or ‘language of thought’, cf. Johnson-Laird, 1983) before the process of formulation starts in which the elements of language are put together to express this idea. Finally, in order to verbally produce an utterance, the speaker goes through a process of articulation involving the speech apparatus (Warren, 2017, p. 16ff).
The current framework of a ‘semantic model of information organization’ (cf. Klein & von Stutterheim, 2002; von Stutterheim, 1997; von Stutterheim & Klein, 1989) concerns itself primarily with the process of *conceptualization* (von Stutterheim, 2003, p. 185). Warren (2017) illustrates the process of conceptualization with the following example in Figure 2, which represents the idea a speaker wants to communicate (*the cat hunted a mouse*). In other words, the pre-verbal message will reflect the intention to communicate the concept of hunting (i.e. the *action* of one animal on another), the concept of the cat as the *agent* of the action (i.e. the entity carrying out the action) and the concept of the mouse as the *theme* or *patient* (i.e. the entity which undergoes the action). Moreover, the message will also include information about what the speaker wants to highlight (e.g. focusing on the agent or the theme) (p. 17).
The planning process within conceptualization occurs on two levels: ‘macro’- and ‘microplanning’. On the one hand, macroplanning (or ‘global planning’) is the process of choosing from a series of speech acts. A speech act can be defined as the performance of some action through saying something. Part of knowing how language works includes knowing which speech act is appropriate in achieving a particular goal in a given context and for a given addressee. Furthermore, speech acts need to be sequenced in order to achieve a communicative goal. This involves ‘linearization’, which refers to choosing the order in which information should be expressed. On the other hand, microplanning (or ‘local planning’) refers to the process of determining the perspective and information structure which is most appropriate for a given speech act, especially what should be highlighted as new or topical information (Warren, 2017, p. 23).

Von Stutterheim (2003) highlights that “all these steps in the planning process are perspective-driven (cf. Carroll et al., 2000; Klein & von Stutterheim, 2002) and the choice of a consistent perspective ensures coherence at the text level” (p. 186). Moreover, languages differ in their preferences to exploit the various options available (Carroll & von Stutterheim, 2003, p. 1012). The underlying assumptions within the semantic model of information organization, together with the findings in previous studies by von Stutterheim and her colleagues (cf. von Stutterheim, 2003).

---

1 The picture in Figure 2 has been retrieved from: https://www.desktopbackground.org/wallpaper/download-animated-cute-fish-mobile-phone-wallpapers-hd-images-free-984144 [26 August 2018]
Stutterheim, 2003), are compatible with Slobin’s (1991) *thinking for speaking hypothesis* through which he demonstrates that the grammatical organization within a language leads the native speaker to adopt a specific perspective on the information to be expressed in a given communicative context. In Slobin’s words:

> The language or the languages that we learn in childhood are not neutral coding systems of an objective reality. Rather, each one is a subjective orientation to the world of human experience, and this orientation *affects the ways in which we think while we are speaking* (Slobin, 1991, p. 23, emphasis in the original).

To put in another way, speakers’ choices of information are guided through the linguistic categories afforded by their language, specifically by the habitual use of linguistic categories that select for certain types of information to be expressed when thinking for the act of speaking. ‘Thinking’, in Slobin’s (1991) discussion is tightly linked to language and refers to the ‘thinking’ which is carried out online during the process of speaking (p. 11).

Consequently, empirical studies within the framework under discussion directly tie to current debates about the relationships between language and thought. One of the central questions in the debate on linguistic relativity concerns the role of language in the language production process, specifically concerning the nature of the conceptualizer. Three positions can be distinguished according to von Stutterheim and Nüse (2003). The ‘moderate’ position argues that conceptualization is a component of language production (i.e. language in use) which is always based on language-specific principles as argued in Slobin’s (1991) ‘thinking for speaking’ hypothesis. The ‘moderate’ position contrasts with two ‘radical’ positions. The first ‘radical’ position considers the process of conceptualization as language free and universal (e.g. Jackendoff, 1990), whereas the second considers the conceptualizer as language based in nature (e.g. Levinson, 1997); a Whorfian view which claims that the way we think about the world in general in situations in which language is not necessarily being used is determined by our linguistic and cultural background (Slobin, 2003, p 157). After all, important for the current framework is that language-dependent conceptualization is relevant at the global and local level of message planning (von Stutterheim & Nüse, 2003, p. 852).
2.2.2 Quaestio and perspective-taking in language

Carroll et al. (2000) stress that language production is always done from a *perspective*. The notion of perspective is closely related to the concept of ‘viewpoint’, especially in visual art, in which it is associated with “the spatial position at which the observer is or imagines to be” (Klein & von Stutterheim, 2002, p. 62). Viewpoint also plays a role in language, for instance in the use of deictic terms or in relation to aspect. In order not to confuse the different definitions of viewpoint, Klein and von Stutterheim (2002) use the term ‘L-perspective’ (L for language) to differentiate it from ‘V-perspective’ (v for visual). L-perspective reflects a particular way of presenting what a speaker wants to say involving a series of interrelated decisions on many different levels, such as structural, lexical and contextual. In Klein and von Stutterheim’s (2002) words, perspective-taking in language “is determined by highly abstract cognitive principles which determine the choice of words and constructions against the background of shared assumptions among the interlocutors” (p. 62).

Drawing on a concept from classic rhetorics, Klein and von Stutterheim (2002) and von Stutterheim and Klein (1989) have proposed a model which assumes that texts (i.e. oral output) are defined as organized structures that answer a guiding question: the *quaestio*. The question which underlies the production of an utterance or longer text may be explicitly asked by an interlocutor in a dialogue or be implicitly given either by the speaker himself/herself or the given communicative situation. Moreover, it can also be understood as the speakers’ intention (cf. von Stutterheim, 1997). The utterances which directly contribute to answering the question constitute the ‘main structure’ of the text, while those which give additional information are called its ‘side structures’, such as the speaker’s subjective comments and evaluation or background information. Importantly, the quaestio imposes constrains on possible perspectives a speaker can take and guides the speaker in selecting and organizing information for expression and mapping it onto conceptual domains including person/object, temporal, spatial, modal and predicate domains. To put it in another way, the quaestio imposes constraints on a possible answer, but it does not determine the answer, leaving the speaker a certain amount of freedom. If the speaker ignores the
constraints then this either leads to side structures or to specific rhetorical effects (Klein & von Stutterheim, 2002).

In particular, Carroll and von Stutterheim (1993) exemplify how the quaestio constrains the mapping of information in the aforementioned domains with the following communicative task: “What happened to x at time t1, t2, etc.?" (p. 1013). An answer to this question requires a conceptualization of a situation as a series of events unfolding in time. In other words, the corresponding text is construed as a narrative, in which information is arranged on the basis of the chronological order of events. That is, the conceptual domain ‘time’ provides the basis for deciding in what order information should be selected for mention or ‘linearized’ in the text (Carroll and von Stutterheim, 1993, p. 1013).

A description task requires an information structure that provides an answer to the overall question “What is x like?” (Carroll et al., 2000, p. 445). Hence, the question calls for a representation of the entity under description that facilitates a description of its properties (Carroll & von Stutterheim, 1993, p. 1014). Moreover, Carroll and von Stutterheim (1993) point out that in descriptions, unlike in narratives, there is no prototypical sequence that can serve as a basis for defining the order in which information can be selected and linearized in the text. Rather, the ‘spatial’ and ‘object’ domain play a central role in interrelating information expressed in descriptions in English and German (p. 1014). That is, on the one hand entities can be described from a perspective centering on their inherent and functional properties. In other words, entities might be grouped with respect to their properties such as function, type, size, etc., constituting an object-based perspective. On the other hand, entities can also be selected for mention based on the regions of space they share forming a spatially-based perspective. Consequently, an object based perspective leads to the organization of information on the basis of an overall question “What is where?”, while a spatially-based perspective proceeds on the basis of the question “Where is what?” (Carroll et al., 2000, p. 446).
2.3 Basic concepts and categories in locative relationships

This subsection identifies the spatial concepts underlying locative relationships presenting important works on motion events and spatial relationships. Static locative relationships can be either conceived as a type of Motion event (cf. Talmy, 2000) or as the relationship between immobile objects (cf. Hendriks, Watorek & Giuliano, 2004).

In general terms, “locative relationships establish the location of an object (or a group of objects) in relation to one or more other objects” (Tutton, 2013a, p. 27). A common point of reference in investigating concepts underlying spatial relationships is the work of Talmy (1975; 1983; 1985; 2000) on motion events along with his conceptual framework of spatial semantics. Especially his account on how spatial information and motion are cross-linguistically encoded (i.e. his famous dichotomy of satellite- versus verb-framed languages) has provided the basis for further theoretical and empirical investigations (e.g. Kopecka & Pourcel, 2005; Kopecka, 2006; Lemmens, 2005; Slobin, 2004; 2006; Zlatev & Yangklang, 2004 to name just a few). In particular, Talmy (2000) has identified key semantic components associated with motion events, as presented in Table 1, namely, the Figure, the Ground, Motion, Path and Manner exemplified in (1) by Saeed (2016, p. 273).

Table 1. Talmy’s (2000) concepts to describe Motion events (adapted from Saeed, 2016, p. 273)

<table>
<thead>
<tr>
<th>Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure</td>
<td>the object being located in relation to another object (the Ground)</td>
</tr>
<tr>
<td>Ground</td>
<td>the reference object</td>
</tr>
<tr>
<td>Motion</td>
<td>the presence of motion or location as such in the event</td>
</tr>
<tr>
<td>Path</td>
<td>the course followed or the site occupied by the Figure object in relation to the Ground object</td>
</tr>
<tr>
<td>Manner</td>
<td>the type of motion</td>
</tr>
</tbody>
</table>

In example (1), Charlotte is the Figure, the crocodile is the Ground, away from is the Path and the verb swim encodes the Manner of motion. In static locative relationships only the first four concepts will be relevant. They are exemplified in (2), by Tutton (2013a, p. 27).
(1) Charlotte swam away from the crocodile.
   Figure Manner Path Ground

(2) The tablecloth is on the table.
   Figure Motion Path Ground

In other words, the tablecloth is the Figure, the object being located. The table is the reference object (i.e. the Ground), whereas the copular verb is (‘to be’) encoding Motion (i.e. the presence of location) and the preposition on lexicalizing Path link the Figure and the Ground syntactically. Talmy (2000) highlights: “we treat a situation containing motion and the continuation of a stationary location alike as a motion event (with a capital M)” (p. 25). Essentially, Talmy (2000) considers static locative relationships, exemplified in (2), as a type of Motion event.

Levinson and Wilkins (2006a) note that once a Figure and a Ground are identified in space, it becomes necessary to specify a direction or angle relative to the Ground, from which the Figure may be found when the Figure and the Ground are separated in space. This kind of directional specification requires some reference point or form of coordinate system, which has been termed frames of reference including three types intrinsic, relative and absolute (cf. Levinson, 1996; 2003). To take a case in point, in an intrinsic frame of reference the location of an object is defined in relation to the part of another object (e.g. its side, back, front, etc.), as exemplified in (3a). Another way to specify an angle is to use the viewer’s own bodily coordinates. In other words, the location of an object is expressed in relation to both the viewpoint of the perceiver and the position of another object, as shown in (3b). A third way to specify angles is to use fixed bearings (e.g. cardinal directions: North, South, East and West). Hence, the position of the ball is described as independent of the speaker or any part of the box, as in (3c) (Levinson & Wilkins, 2006a, p. 4).
Furthermore, Levinson and Wilkins (2006a) note that not all languages use all three frames of reference and each frame of reference may be represented by quite distinct concepts across languages (p. 5). Levinson and Wilkins (2006b) exemplify the distribution of frames of references across languages, as in Table 2.

2 The pictures of the ball and the box in example (2) have been retrieved from: 
Table 2. Distribution of main frames of reference across a language sample (adapted from Levinson and Wilkins, 2006b, p. 543).

<table>
<thead>
<tr>
<th>Frame of Reference</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute, relative and intrinsic</td>
<td>Tamil, Yukatek, Tiriyò, Ewe, Kilivila</td>
</tr>
<tr>
<td>Absolute and intrinsic</td>
<td>Warrwa, Arrernte, Jaminjung, Yélî, Dnye, Tzeltal</td>
</tr>
<tr>
<td>Relative and intrinsic</td>
<td>Japanese, Dutch, Dutch</td>
</tr>
</tbody>
</table>

Firstly, Levinson and Wilkins (2006b) suggest that the intrinsic frame of reference might be the only one that could be considered universal. Nearly all languages, including English and German, use such segmentations in spatial descriptions and it appears to be the first frame of reference systematically used by children (e.g. Piaget & Inhelder, 1956). Levinson and Wilkins (2006a) indicate that it requires some kind of division of the ground object into facets, from which the other object can be related to. The criteria by which objects are partitioned are quite similar in the English and German languages including, for instance, “functional orientation”, “normal direction of motion”, etc. (e.g. “the front of the television is the side one watches”, whereas “the front of a truck is the direction in which it moves”) (p. 21).

Secondly, Levinson and Wilkins (2006b) propose that relative frames of reference have evolved out of intrinsic systems. Therefore, if a language has a relative system, it has an intrinsic system as well, usually with shared lexemes (p. 543). The relative frame of reference involves more complex mappings from the observer’s own axes (front, back, left, right) onto the ground object. In English and German sentences such as (4a,c) might be ambiguous at times since terms like front and left are used in both intrinsic and relative frames of reference. This ambiguity is resolved in both languages, for instance, by using a possessive for the intrinsic interpretation, such as the man’s left, as shown in (4b,d).

Lastly, the absolute frame of reference usually requires fixed bearings that have to be directly available to all speakers of a community without adhering to maps or compasses (Levinson & Wilkins, 2006a). Both English and German use the word for North, but few speakers can easily point to the north. In addition, this is a rather uncommon frame of reference in small scale spatial descriptions in English and German (p. 21). Consequently, like Japanese and Dutch, English and
German belong to the language group which predominantly uses the relative and intrinsic frames of references.

(4) a) The tree is to the left of the man.

b) The tree is to the man’s left.

c) Der Baum ist links vom Mann.

‘The tree is to the left of the man.’

d) Der Baum ist zu-r Linken des Mannes

‘The tree is to the man’s left’.
Concerning different types of locative relationships, Hendriks et al. (2004) propose three sub-categories: (1) ‘static general localization’, (2) ‘dynamic general localization’ and (3) ‘change of localization’ events, exemplified in (5a-c) by Hendriks et al. (2004, p. 109).³

(5)  a)  **static general localization**

Reskio  est  au  bord  du  lac
Reskio  be.PRS.3SG  at-DAT  edge-DAT  of-the-DAT  lake-DAT

‘Reskio is at the edge of the water.’

b)  **dynamic general localization**

Reskio  fait  du  patinage  sur  le  lac
Reskio  do.PRS.3SG  of-the  skating  on  the-DAT  lake-DAT

‘Reskio skates/is skating on the lake.’

c)  **change of localization**

la  petite  fille  sort  de  l’eau
the-F.NOM  little-F.NOM  girl-F.NOM  exit.PRS.3SG  of  the-DAT  water-DAT

‘The little girl comes out/is coming out the water.’

In example (5a) the Figure (Reskio) is presented as a stationary entity, which is located in relation to the edge of the lake (the Ground). The locative relationship between the Figure and the Ground (the Motion) is stable and unchanging and encoded by the inflected form of the verb être (‘to be’). Both in example (5a) and (5b) the Figure remains within the boundaries of the Ground entity. However, in example (5b) the Figure is in a constant motion. In example (5c) a boundary crossing event takes place in which the Figure (the little girl) is changing the location (i.e. from in the water to out of the water). In the present study, mainly locative relationships which belong to the category

³ In their paper, Hendriks et al. (2004) named the categories originally in French: (1) ‘localization générale statique’, (2) ‘localization générale dynamique’ and (3) ‘changement de localization’. The translations of the category names have been adopted from Tutton (2013a, p. 28).
of static general localization are of interest. However, dynamic locative relationships can occur as well in the descriptions, since the chosen pictures (i.e. a city scene and a landscape scene) also prompt the salience of motion in the conceptualization of location, as exemplified in (6).

(6) Going along the road which runs in front of the pink house there is a little child (E.7.2).

In short, Hendriks et al. (2004) understand static location as the relationship between immobile objects in space, as shown in (5a). In turn, for Talmy (2000) the three types of localization described by Hendriks et al. (2004) fall under the umbrella of Motion events.

2.4 The existential construction: between existence and location

The following subsection is concerned with the grammatical means available in English and German to refer to the location of objects and their existence in space. Specifically, the existential construction in the respective languages will be presented as the German existential construction differs syntactically from the English one.

English *there is*, French *il y a* and German *es gibt* are typical examples of the existential construction shown in (7), which has been intensively studied in cross-linguistic, comparative, and diachronic approaches (e.g. Abbott, 1993; Bentley et al., 2015; Clark, 1978; Creissels, 2014; Hartmann, 2008; Lyons, 1967; McNally, 1997; Pfenninger, 2009; Weinert, 2013; inter alia).

(7) *There are* pandas in China.

*Il y a* des pandas en Chine.

In China *gibt es* Pandas.

Even though there is a lot of literature on existential constructions, most authors do not formulate a definition. McNally (2011) attempts to define existential constructions:

The term existential sentence is used to refer to a specialized or non-canonical construction which expresses a proposition about the existence or presence of something or someone (McNally, 2011, p.1830).
The main discourse function of existential constructions is to present new information (Biber et al., 1999). The semantics of existentials involves drawing the addressee’s attention to existence (or non-existence) of the entity or entities typically denoted by indefinite NPs, as shown in (8) and (9). In addition, existentials can assert existence or presence in a location, as shown in (9) by Weinert (2013). In fact, Biber et al. (1999) note that in conversation, 40% of existential there-constructions in English are post-modified, most commonly by place adverbials. In turn, in locative constructions the existence of the entity is just implied and definite nominals are used suggesting that the object referred to is already given through previous mention in the discourse, as shown in (10).

(8) There are blue-eyed cats.

(9) There is a cat in the garden.

(10) The cat is in the garden.

Clark (1978) has pointed out that in many languages existential constructions are systematically related to locative and possessive constructions, exemplified in (11) and (12) by Clark (1978), in terms of word order, in the verbs used and in their locative origin. However, the exact nature of this alternation, the cross-linguistic typological distribution of recurrent patterns between the three constructions and the demarcation between the cognitive domains EXISTENCE and LOCATION are still being debated (e.g. Creissels, 2014; Koch, 2012).

(11) There is a book on the table.

   The book is on the table.

(12) Tom has a book.

   The book is Tom’s.
2.4.1.1 The existential *there*-construction in English

In English, a clause with the existential *there* has the following structure, as shown in (13) by Biber et al. (1999). According to Huddleston and Pullum (2002), existentials usually have the verb *be*, while presentationalals have some verb other than *be*, as shown in (14) by Huddleston and Pullum (2002). The noun phrase following *be* is usually indefinite and referred to as the ‘notional subject’ (Biber et al., 1999, p. 943ff) or ‘displaced subject’ (Huddleston & Pullum, 2002, p. 1391).

(13) there + be + indefinite NP (+ place or time position adverbial)

There is a book on the table.

(14) There remain many problems.

Existential *there* is a function word, a dummy pronoun, which derives historically from the locative adverb *there*. Huddleston and Pullum (2002) note that locative *there* is an intransitive preposition contrasting with *here* and has deictic and anaphoric uses. In the construction concerned, *there* has been bleached of its locative meaning and reanalyzed as a pronoun (p. 1391). According to Biber et al. (1999) it differs in three respects from the locative *there*. Firstly, it is normally phonologically reduced. Hence, it can never be the focus. Secondly, it has lost its original locative meaning (see also Abbott, 1993; Bolinger, 1977; Lakoff, 1987; Lyons, 1975). That is, an existential *there* can be used in the same clause as *here* or locative *there* without any contradiction or tautology, as illustrated in (15) and (16) by Biber et al. (1999). In other words, it has the status of an empty grammatical element (Biber et al., 1999, p. 944), a dummy pronoun similar to some usages of the pronoun *it* (Huddleston & Pullum, 2002). Thirdly, Biber et al. (1999) note that syntactically the existential *there* functions as a grammatical subject rather than as an adverbial. It is placed before the verb in declarative sentences and can be used in question tags, as shown in example (16).

(15) *There*’s more gravy *here*.

(16) *There*’s still no water *there*, is there?

As mentioned earlier, the notional subject or displaced subject is typically an indefinite noun phrase. However, definite notional subjects or displaced subjects might occur as well, as shown in
(17) by Biber et al. (1999), in particular, when a series of elements are introduced, often marked explicitly by a conjunction or a linking (e.g. first) or additive adverbial (e.g. too) (Biber et al., 1999, p. 947). As mentioned above, the majority of existential clauses, especially in conversation, contain a form of the verb *be*, which may be preceded by auxiliaries or semi-modals: *has been, will be, is to be, is supposed to be, used to be* (Biber et al., 1999). Moreover, existential clauses often contain a time or place adverbial, usually at the end of the clause, but clause-initial placement is also possible, as illustrated in (18) and (19). Hence, Huddleston and Pullum (2002) distinguish between bare (20a) and extended (20b) existentials (p. 1393).

(17) First there was the scandal of Fergie romping with John Bryan, pictured exclusively in the Mirror.

(18) There are some people near the fountain (E.3.1).

(19) On the left hand side there is a group of people (E.6.2).

(20) a) There’s no milk (again).

     b) There’s one copy on the table.

2.4.1.2 Existential constructions in German: *es gibt*- and *sein*-constructions

The Duden reference grammar (Wöllstein, 2016) does not refer specifically to any existential constructions. A comprehensive diachronic analysis of written English and German existential constructions is provided by Pfenninger (2009) who examines examples from contemporary fiction and their German translations including elicited translations. According to Pfenninger (2009), High German does not incorporate a prototypical existential construction equivalent in syntactic, pragmatic function or semantics to the English existential *there*-construction. Lastly, Weinert (2013) presents a synchronic, corpus-based examination of spoken and written German with regard to the distribution and function of presentational/existential *es gibt*- and *sein*-constructions, such as *da SEIN, locative SEIN, es SEIN*, and *zero-locative SEIN*, exemplified in (21)-(25) by Weinert (2013, p. 40ff).
(21) **es gibt NP**

a) *Es gib-t* blauäugige Katze-n.

It give.PRS.3SG blue-eyed-ACC cats-ACC.PL

‘There are blue-eyed cats.’

b) *Es gibt* in Athen ein Katzenmuseum.

It give.PRS.3SG in-DAT Athen-DAT a-ACC cat-museum-ACC

‘There is a cat-museum in Athens’.

(22) **da SEIN NP**

a) *Da ist* ein-e Katze.

There be.PRS.3SG a-NOM cat-NOM

‘There is a cat.’

b) *Da ist* ein-e Katze im Garten.

There be.PRS.3SG a-NOM cat-NOM in-the-DAT garden-DAT

‘There is a cat in the garden.’

(23) **es SEIN NP**

*Es ist* ein-e Katze im Garten.

It be.PRS.3SG a-NOM cat-NOM in-the-DAT garden-DAT
‘There is a cat in the garden.’

(24) **zero-locative SEIN NP**

\[
\begin{align*}
\text{Waren} & \quad \text{noch} & \quad \text{zwei} & \quad \text{Blitzbirnchen.} \\
\text{Be.PST.3.PL} & \quad \text{still} & \quad \text{two-NOM} & \quad \text{flash-bulbs-NOM (DIMINUTIVE)}
\end{align*}
\]

‘Were still two flash-bulbs.’ *(location: in the camera)*

(25) **locative SEIN NP**

\[
\begin{align*}
\text{Im} & \quad \text{Garten} & \quad \text{ist} & \quad \text{ein-e} & \quad \text{Katze.} \\
\text{In-the-DAT} & \quad \text{garden-DAT} & \quad \text{be.PRS.3SG} & \quad \text{a-NOM} & \quad \text{cat-NOM}
\end{align*}
\]

‘In the garden is a cat.’

According to Weinert (2013), the range of presentational and existential constructions and verbs is potentially rather wide in German. In particular, Weinert (2013) points out that *es gibt* is equally frequent in spoken and written German, while *SEIN* structures are typical of spoken German only, with *da SEIN* being the most frequent one. Besides, *da* can also be used with other main verbs, as exemplified in (26) by Weinert (2013). In turn, *es SEIN* - and *zero-locative SEIN*-constructions are marginal in spoken data.

(26) Wir **hatt-en** gerade einen Vorfall. Da ist ein

\[
\begin{align*}
\text{Wir} & \quad \text{have-PST.1.PL} & \quad \text{just} & \quad \text{an-ACC incident-ACC.} & \quad \text{Da} & \quad \text{be.PRF.3SG} & \quad \text{a-NOM} \\
\text{Hund} & \quad \text{in-s} & \quad \text{Rad} & \quad \text{ge-lauften.} \\
\text{dog-NOM} & \quad \text{into-the-ACC} & \quad \text{bike-ACC} & \quad \text{PTCP-run}
\end{align*}
\]

‘We just had an incident. There was a dog that ran into the bike.’
Weinert (2013) notes that in terms of function, example (21a) asserts existence per se (i.e. ontological existence), while examples (21b) and (22)-(25) assert existence or presence in a location. According to Weinert (2013) *es gibt-* and *SEIN-*constructions should be considered as manifesting existential and locative meaning to different degrees, from (21a) asserting existence to a high degree to (25) asserting primarily the specific location of an entity. To note, *es gibt* is not historically associated with presence in a location (in contrast to *there* and *da*). The existential meaning of *es gibt* can be seen as an extension of giving, such as the manifestation or provision of an object (cf. Pfenninger, 2009).

After all, Weinert (2013) concludes that *es gibt* and *da SEIN* could be considered prototypical presentational/existential constructions in German, which occur with almost equal frequency in spoken German. In turn, *locative* *SEIN* indicates the specific location of an entity. While a locative element is often present in *da SEIN* constructions, it is frequently backgrounded. Hence, *da SEIN* appears to be positioned somewhere between *es gibt* and *locative* *SEIN* along the existential-locative/presence continuum. That is, *es gibt* can assert ontological existence, while *da SEIN* can probably not or might be only marginally possible, as shown in (27) by Weinert (2013, p. 74). In contrast, *es gibt* is questionable when used to focus on entities that are situationally present, as demonstrated in (28) by Weinert (2013, p. 74). Such cases as in (28) are usually reserved for *da SEIN*, since *da* is deictic. Moreover, the occurrence of *da* in clauses with main verbs can be related to the wider usage of *da* as an attention-drawing deictic in spoken language (Weinert, 2013, p. 73-75).

(27) a)  
\begin{align*} 
\text{Es} & \quad \text{gib-t} & \quad \text{einen} & \quad \text{Gott.} \\
\text{It} & \quad \text{give-PRS.3SG} & \quad \text{a-AVV} & \quad \text{god-ACC} 
\end{align*}

b)  
\begin{align*} 
?\text{Da} & \quad \text{ist} & \quad \text{ein} & \quad \text{Gott} \\
\text{There} & \quad \text{be.PRS.3SG} & \quad \text{a-NOM} & \quad \text{god-NOM} 
\end{align*}

‘There is a god.’
2.5 Gestures

This subsection focuses on how speakers accompany their speech with gestures. Essentially, gestures are seen as an integral part of the interactional event, which cannot be fully studied by simply looking at speech alone. A brief overview on theoretical considerations as well as empirical investigations on the relationship between gestures, speech and language will be presented.

2.5.1 An integrated system: gesture and speech

Before establishing the reasons why gestures should be considered when investigating oral spatial descriptions, the term ‘gesture’ needs to be defined. When speakers talk, they make movements with their hands, arms and head. These movements, or gestures occur spontaneously in talking about everyday events and are inseparable from speech. Hence, they are referred to as speech-associated gestures following Gullberg (2011), which are not a mere by-product of the speech production process, but are part of the utterance itself (Kendon, 2004) and help the speaker to achieve a communicative goal. In Gullberg (2006a)’s words, speech-associated gestures are

4 Within the field of gesture studies, different terms are used to refer to the same concept, such as for instance ‘gesticulations’ (Kendon, 1986) or ‘co-speech gestures’ (McNeill, 1992). This paper will use the label ‘speech-associated gestures’ (Gullberg, 2011).
“symbolic movements related to ongoing talk and to the expressive effort or intention (what you are trying to say)” (p.104). Moreover, they occur in very close temporal and semantic synchrony with speech and are used to convey both semantic and pragmatic information. In other words, gestures and speech are intricately linked to each other. This systematic link is revealed in the gestures’ synchrony with the co-expressive part of speech, such that gestures and speech express the same underlying idea unit at the same time (Gullberg, 2011; Kendon, 1986; 2004; McNeill, 1992). Hence, utterances should be considered in terms of both speech and gesture units (Kendon, 2004).

Existing research on both language perception and production have further shed light onto the close relationship between the two modalities. For instance, studies on language comprehension show that gestures affect the perception and interpretation of speech as well as memory (e.g. Kelly, Barr, Breckinridge Church, & Lynch, 1999; Kelly, McDevitt & Esch, 2009, Kelly, Özyürek & Maris, 2010). For example, Kelly et al. (1999) found that speech and gesture mutually disambiguate the meanings of one another. Further evidence for an integrated system are also provided by studies on early language development. If speech and gestures are an integrated system from the beginning or whether they are initially separate is not fully understood yet. Nevertheless, studies point to a parallel development of both speech and gestures (cf. Gullberg, de Bot & Volterra, 2008 for a review on gestures and language development). Moreover, Iverson and Goldin-Meadow (2005) indicate that infants use deictic gestures already at the earliest stages of language development whereby gestures pave the way for future language development. That is, complementary and supplementary gesture-speech combinations reliably predict the onset of two-word combinations. Moreover, in speech disfluencies and in stuttering gesture production is also disrupted (Graziano & Gullberg, 2018; Mayberry & Jaques, 2000).

The link between speech and gestures is also observed in the co-variation of speech-associated gestures and different typological and structural patterning of spoken languages (e.g. Gullberg, 2009; 2011; Kita & Özyürek, 2003; Özyürek & Kita, 1999 discussed below). This also extends to discourse, such that gestures tend to co-occur with speech representing new or focused information (e.g. McNeill, 2000a). Hence, cross-linguistic differences in discourse organization
are also reflected in language-specific gesture patterns (e.g. Gullberg, 2003; 2006b). Consequently, gestures mirror linguistic choices both in terms of linguistic, structural choices as well as information structure (Gullberg, 2011, p.169).

2.5.2 Classification and composition of gesture movements

The current study follows Kendon (2004)’s suggestions to consider gestures to be multifunctional varying along several dimensions at the same time; gestures can have referential, pragmatic and interactional functions simultaneously. Moreover, this thesis is mainly interested in what Kendon (2004) has termed referential gestures (cf. ‘representational gestures’ McNeill, 1992). Their main function (among other dimensions) is to refer to and represent discourse referents, such as their properties (e.g. size, shape, etc.) or spatial relations to other referents (e.g. next to, above, etc.) through iconicity or deixis (e.g. Kita, 2000; McNeill, 1992). Referential gestures that are indexical in nature have also been termed ‘deictic’ or ‘pointing’ gestures. They can be both concrete by indicating objects present in the environment (exophoric) or abstract by pointing to a seemingly empty space establishing a virtual object or pointing back at such a virtual object (endophoric). Referential gestures that are more iconic in nature can represent entities by depicting their properties (e.g. size, shape) or their relationship to other entities (e.g. next to) (Kita, 2000).
Gestures have an internal structure and can be considered in terms of *units, phrases* and *phases* (Kendon, 2004). Kendon’s (1972; 1980) initial levels of gesture movements have been summarized by McNeill, Pedelty and Levy (1990), as shown in Figure 3. When a speaker gestures the articulators employed undertake a movement excursion, where the articulators are moved away from the rest position toward a region of space and eventually are moved back again to a relaxed position. The gesture unit includes this entire excursion from the moment the articulators begin to move away until they finally return in a position of relaxation (Kendon, 2004, p. 111). Importantly, a gesture unit may be composed of several gesture phases. The nucleus phase, the *stroke*, occurs when the articulators reach a peak furthest away from the relaxed position at which “the expression of the gesture” is articulated with most clarity. The stroke is the part of the gesture which is synchronized with the linguistic elements co-occurring with it (McNeill, 1992, p. 83). The phase of movement preceding the stroke is called *preparation* and the phase following is called *recovery* or *retraction*.

Kendon (2004) indicates that while the gesture phrase contains only one stroke, the recovery movement, as the hand relaxes and returns to some rest position, is not really part of the gesture phrase; however, it still belongs to the gesture unit which contains the gesture phrase. Hence, gestures, that is strokes, may directly follow each other before the hand arrives again in a rest position. Sometimes a stroke is followed by a phase in which the articulators are halted in the
final position in which they arrived at the end of the stroke. This is called a *post-stroke hold*. Post-stroke holds can occur when the co-expressive speech might be delayed. Likewise, the position following the preparation phase may be briefly held until the stroke begins. These *pre-stroke holds* may occur when the stroke onset might be delayed for some reason (McNeill, 1992). This phrasal organization of gesture movements proposed by Kendon (1972; 1980; 2004) and adopted by McNeill (1992) provides a way to identify and code gesture units.

2.5.3 Spatial relationships and gesture

Numerous studies have examined how speakers gesture when they conceptualize and talk about spatial information. It has been shown that speakers gesture more frequently when they talk about spatial topics than about abstract, non-spatial ones (e.g. Alibali, 2005). Moreover, previous research on language, space and gesture has demonstrated that gestures reflect the way in which information is encoded in speech (e.g. Gullberg, 2011; Kita & Özyürek, 2003; Özyürek & Kita, 1999).

To take a case in point, Özyürek and Kita (1999) as well as Kita and Özyürek (2003) have examined the gestures of English, Japanese and Turkish speakers in the domain of voluntary motion. In the latter study, the gestures of English, Japanese and Turkish speaker were examined when describing a character’s trajectory from one building to another. The arc-like trajectory of the character’s movement is readily encoded in English (in the verb *swing*), but neither in Japanese nor in Turkish. Kita and Özyürek (2003) showed that English speakers express this arc-like trajectory more frequently in gestures than Japanese or Turkish speakers. Likewise, Özyürek and Kita (1999) previously reported that English speakers’ speech-associated gestures looked different from the corresponding Turkish speakers’ gestures. English is a satellite-framed language and expresses manner and path of motion in one single verb clause (e.g. *roll down*), and in one single gesture encoding both manner (*roll*) and direction of motion (*down*) in one movement. In turn, Turkish is a verb-framed language, which encodes the manner and the direction of motion in two
separate verbal clauses (e.g. descend rolling) and typically in two gestures encoding the direction (descend) and manner of motion (rolling) separately.

Similar results were also reported by Gullberg (2011) who examined caused motion events, specifically placement events in French and Dutch. Particularly, French and Dutch differ in their encoding of placement events in that French uses a general verb mettre ‘put’ and encodes locative information in prepositional or adverbial adjuncts, while Dutch encodes locative information across different posture verbs zetten ‘set’ and leggen ‘lay’ and prepositions. Unlike the general verb put, these posture verbs specifically consider the object’s (i.e. the Figure’s) configuration to the Ground. Gullberg (2011) demonstrates that the respective languages show language specific event conceptualization in both speech and gestures. That is, the difference in verb semantics between French and Dutch in terms of placement events is also reflected in the speakers’ gesture patterns, namely in their handshapes. Specifically, Dutch speakers are more likely to incorporate figure object information in their gestures than French speakers, who are more likely to encode only simple path information in their gestures.

As far as gesture use in locative descriptions is concerned, Striegnitz et al. (2009) have investigated how gestures express spatial perspectives in route descriptions. They point out that the perspective gestures represent is generally determined by discourse context. Moreover, Miller and Franz (2005) report that bimanual gestures generally occur more frequently in spatial descriptions than in general routine descriptions, in which unimanual gestures were detected more frequently. Tutton (2011; 2012; 2013a; 2013b; 2016) has examined locative expressions, such as prepositions, highlighting a complementary role of gestures in relation to speech in static locative descriptions. In particular, he points out that gestures can reveal additional, unlexicalized spatial information attended to by the speaker at the time of utterance production, for instance, when using the English preposition next to in speech (Tutton, 2013b). The preposition next to in English encodes the adjacency of two objects while remaining neutral as to directional information. Hence, speech-associated gestures can provide additional information about lateral (left/right) direction in the description of the spatial location of objects. Tutton (2011) also reported similar findings for the English preposition on and the French preposition sur ‘on’. Moreover, Tutton (2012) has
argued that the analysis of speech alone cannot clarify whether a speaker conceptualizes an object as a Figure or as a Ground. Tutton (2012) found that speakers recurrently establish the lexical Ground in their gesture space and concludes that these objects are therefore simultaneously ‘gestural Figures’. Finally, a complementary role of gestures has also been demonstrated by studies interested in how speakers express object size and relative position (e.g. Beattie & Shovelton, 1999a; 1999b; 2001; 2006). However, they highlight that this was only observed in the semantic domains of relative position (i.e. the position of the entity relative to something else) and size (i.e. the size of the entity).

### 2.6 Summary

This chapter sets the basis for the present experimental investigation by introducing the study this thesis aims to replicate as well as the theoretical frameworks and concepts underlying the research questions and hypotheses. Studying information organization in texts requires us to look more closely at how information is mapped onto conceptual domains and how information is linked within and across domains as the text unfolds. The present analysis will especially focus on the object and spatial domains in English and German description tasks as they seem to be important domains in determining the selection of spatial concepts, in particular intrinsic and relative frames of reference, when describing the location of objects, as shown by previous research (cf. Carroll, 1993; Carroll & von Stutterheim, 1993; Carroll et al., 2000). As the methods, especially the procedure and the analysis, in Carroll et al.’s (2000) study are sparsely described, it is necessary to test the validity of their claims. Only then can we attempt to generalize their findings for native English and German speakers to larger populations. Moreover, in order to provide a full picture of communicative events this study will also take speech-accompanying gestures into account, as they occur in very close temporal and semantic synchrony with speech and are related to the expressive intention of what a speaker tries to communicate (cf. Gullberg, 2011; Kendon, 2004; McNeill, 1992). Essentially, gestures reflect the way information is encoded in speech. Spoken
discourse can only be fully analyzed once speech-accompanying gestures are taken into account as well.
Chapter 3 The current study

In the following section the research questions and hypotheses will be introduced followed by the presentation of the design of the study including information about the participants, the materials, the procedure, as well as the coding and analysis of the data.

3.1 Research questions and hypotheses

In comparisons of spatial descriptions in English and German, previous research (Carroll, 1993; Carroll & von Stutterheim, 1993, Carroll et al., 2000) demonstrates that speakers of different languages show language-specific perspectives in information organization. Specifically, Carroll et al. (2000) suggested that an object-based perspective predominates in information organization in English, whereas a spatially-based perspective plays a more significant role in German. In particular, an object-based perspective proceeds on the basis of the question “What is where?”, whereas a spatially-based perspective is built on the question “Where is what?” (p. 446).

This study seeks to replicate Carroll et al. (2000), specifically the language-specific preferences between English and German in terms of information organization in spatial description tasks. Furthermore, it explores whether such preferences in the way information is conceptualized and organized for expression in English and German will also be visible in gesture production, and it investigates whether gestures reveal additional aspects of what perspective English and German speakers take into account when giving a spatial description of an image. On the basis of previous research (Carroll, 1993; Carroll & von Stutterheim, 1993; Carroll et. al, 2000; Gullberg, 2011) the following research questions are posed.

To what extent are the planning processes required for the production of coherent spatial descriptions in English and German language specific and related to structural properties of the respective languages?
What do gestures reveal about how native English and German speakers conceptualize and lexicalize spatial descriptions?

More specifically:

RQ1) What roles do the domains ‘object’ and ‘space’ play in interrelating information and establishing cohesion in English and German descriptive texts?

RQ2) What factors guide the selection of spatial concepts and the choice of linguistic form in English and German?

RQ3) How are language-specific perspectives in information organization reflected in gesture behavior, specifically in the gesture form and timing with speech?

The aim of the study is thus twofold. On the one hand, the study tries to replicate the speech data reported in Carroll et al. (2000) in order to verify the validity of the previous results. On the other hand, the study is interested in exploring the role of gestures in information structure, in order to get a fuller picture of what information is taken into account when speakers provide picture descriptions. Moreover, this study might provide additional evidence that speech and gesture form an integrated system. To my knowledge, none of the previous studies have investigated the role of gestures in language-specific perspectives and information organization in a cross-linguistic comparison of English and German spatial descriptions. Hence, the gesture analysis is more exploratory in nature.

The first and the second research questions relate to the speech analysis, whereas the third research question refers to the gesture analysis. In accordance with previous results (Carroll, 1993; Carroll & von Stutterheim, 1993; Carroll et al., 2000; Gullberg, 2011) the following hypotheses are suggested for each of the corresponding ‘subordinate’ research question above:

H1) If in descriptive tasks concepts are selected which relate to intrinsic features of the object under description and associated spatial concepts are selected accordingly, the main organization principle of a text is based on the object and other entities themselves. In turn, if features are selected for mention on the
basis of the regions of space they share, the main organization principle of a text is based on concepts related to the spatial location of objects.

H2) If native speakers of English and German show different preferences in information organization in descriptive tasks, the linguistic means used in reference introduction, maintenance, the division of the picture into subsections and word order should align with the language-specific frames of reference; that is, an object-based perspective in English and a spatially-based perspective in German.

H3a) If an object-based organization plays a central role in information structure in descriptions in English, English speakers’ gestures should reflect this object focus in handshapes that reflect physical properties of the figure object by tracing or molding the shape of an object (i.e. ‘figure-incorporation’, cf. Gullberg, 2011), and align with information in speech about the figure objects, thus providing evidence for language-specific perspectives in information organization.

H3b) If a spatially-based organization plays a central role in information structure in descriptions in German, German speakers’ gestures should reflect this spatial focus in non-object related handshapes (e.g. a pointing handshape or a flat and relaxed hand, cf. Gullberg, 2011) that incorporate movements towards the regions of space where objects are located, and align with information in speech about the location of objects.

The first and the second hypotheses derive directly from previous research (cf. Carroll, 1993; Carroll & von Stutterheim, 1993; Carroll et al., 2000). Von Stutterheim (2003) writes that the means used in English and German “reflect language specific principles which are perspective driven and are associated with patterns of grammaticization in core areas such as the syntactic subject, word order constraints and the morphosyntactic structure of spatial expressions” (p. 186).
The third hypothesis derives from the fact that gestures reflect linguistic choices and that speech and gestures are semantically and temporally coordinated expressing closely related meanings at the same time. (Gullberg, 2011; Kendon, 2004; McNeill, 1992). The following subsection introduces the empirical study designed to elicit spatial descriptions in English and German.

3.2 Method

3.2.1 Participants

In total, 18 native English (n=7) and German (n=11) speakers were interviewed in the Lund University Humanities Lab in Lund, Sweden. Financial compensation was not provided; however, everybody was given a small box of chocolates as a symbolic compensation. Bilingual speakers could not participate. Background information for each participant is summarized in Appendices A (English) and B (German). Participants were included in the analysis if they mentioned a minimum of three items on the given wordlist and produced at least five gestures.

3.2.1.1 Native English participants

The native English speaker group consisted of 7 participants (3 females, mean age 30;9 years, age range 27-35) all tested in Sweden. Two participants came from Canada, two from the United States, and three from the United Kingdom. It was assumed that the different English varieties spoken by the native English participants should not influence the results of the given picture description task. On average, the English-speaking participants had been living abroad for 3;3 years (range 1.5-9 years). The participants were recruited through social media (Facebook) and

---

5In Carroll and von Stutterheim (1993) the majority of participants were North American English speakers (28/32) and the rest were British English speakers. Carroll et al. (2000) do not indicate where their participants come from.
personal networks at Lund University and the University of Copenhagen. All participants worked or studied at one of the respective universities.

3.2.1.2 Native German participants

Eleven native German speakers volunteered to take part in the study (7 females, mean age 26;5 years, age range 24-31). Participants came from Germany and were tested in Sweden. They were recruited through social media (Facebook), recruitment in the university cafeteria and personal networks at Lund University, Malmö University and the University of Copenhagen. The participants all worked or studied at one of the mentioned universities. One German participant was excluded from analyses due to failure to comply with important inclusion criteria (i.e. the participant was bilingual), and another two participants because of poor data quality (i.e. one participant mentioned only one item on the wordlist in her descriptions and another participant produced less than five gestures). This led to a group of 8 native German speakers (5 females, mean age 25;9, range 24-30). On average, they had been living abroad for 3;4 years (range 0.6-10 years).

3.2.2 Materials

3.2.2.1 Stimulus pictures

Since this study is a partial replication of Carroll et al. (2000), the same task was used in which native speakers of the respective languages describe pictures representing different spatial scenes. The same stimulus material was used as in Carroll et al. (2000) in order to ensure the replication of the speech data obtained in that study. Specifically, the participants described a picture showing the old center of a town taken from the series *Hier fällt ein Haus, dort steht ein Kran oder Die Veränderung der Stadt* [Here falls a house, there stands a crane or how towns are changing] (Müller, 1976) as used in Carroll et al.’s (2000) study (see Appendix C). However, in addition, a second picture and a practice picture were chosen, which were both taken from the series *Alle Jahre wieder saust der Presslufthammer nieder oder Die Veränderung der Landschaft* [Every year the jackhammer drifts down or how the landscape is changing] (Müller, 2016). They depict the
same pink house in a changing setting (see Appendix C). The pictures were shown to the participants in the original format (85.1 cm x 31.5 cm)\textsuperscript{6} in physical form. These additional pictures were selected after piloting. They were all drawn by the same artist, and have a similar composition as the original picture, where some objects seem more prominent than others enhancing the likelihood that different viewers concentrate on similar objects.

3.2.2.2 Language History Questionnaire

An off-line questionnaire was distributed after each recording session to gather biographical information and information on general language usage from the participants. English speaking participants received a questionnaire in English, whereas German speakers filled out the same questionnaire in German. Through the questionnaire results it was possible to ensure that the participants were engaged in active use of their native language and illustrate what other languages they might be using on a regular basis. Specifically, participants filled out an adapted version of the Language History Questionnaire (Gullberg & Indefrey, 2003) with some additions from the Language History Questionnaire 2.0 (Li, Zhang, Tsai & Puls, 2013). The resulting questionnaires in the respective languages include questions on general participant background, language history and language usage, as shown in Appendices D (English) and E (German).

3.2.3 Procedure

Data were obtained from a picture description task based on previously seen stimulus pictures, as in Carroll et al. (2000). However, to make the task better suited to gesture elicitation, it was administered in this study as a director-matcher task, rather than just a speaker-listener task as in Carroll et al.’s (2000) study. Eisenbeiss (2010) points out that in a director-matcher task, speakers (i.e. directors) do not just simply pass on information to a passive listener, but direct listeners (i.e. matchers) in a way that they can actively find a particular stimulus from a set of stimuli or follow

\textsuperscript{6} Carroll et al. (2000) do not specify in what form the picture was provided and if the picture was shown in the original format or if just part of the image had been shown in their study.
the directors’ instructions (p. 23). In the present study, the director was instructed to describe a picture from memory in as much detail as possible so that the matcher was able to find the described picture from a series of similar pictures. The director (i.e. describer) was a naïve participant, while the matcher (i.e. listener) was a confederate (i.e. research assistant) who was instructed about her specific role prior to the experiment. For each language, a native speaker of the respective languages functioned as a confederate. The same two native speakers participated as confederates in all the recording sessions in order to ensure the same conditions for every participant. This way, the task was carried out under similar conditions for both language groups.

In compliance with ethics regulation at the Humanities Lab at Lund University, prior to data collection, consent to record audio and video was obtained from the participants as well as the confederates. Consent forms in the respective languages can be found in Appendices F (English) and G (German). The participants were told that they would be taking part in a study about picture descriptions and recognition. The study’s interest in gestures was not mentioned in advance. However, debriefing about the actual goal of the study took place shortly after the end of the experiment.

The describers were asked to sit across from the listener. Both received instructions. The experimental set-up is illustrated in Figure 4.

Figure 4. Schematic representation of the experimental set-up.
Two cameras captured the subjects’ heads and torso. Specifically, two cameras were placed to film the describer and the listener from a semi-frontal position to see their hands, face and part of their profile on the video recording. In this way, one camera could capture most movements in all three dimensions. No attempts were made to hide the cameras. To ensure good audio quality, audio was recorded with an additional voice recorder. Moreover, in order to make sure the subjects’ hands were free to gesture, both the describer and the listener were asked to sit on an armless chair. Before giving the instructions, the experimenter engaged in a brief warm-up with the describer and the listener that consisted of small talk in the target language in order to relax the participants and increase the likelihood of gesturing.

In order to replicate the speech results, the *instructions* were formulated as closely as possible to Carroll et al.’s (2000) instructions\(^7\). The instructions were given in written form in the respective languages to both the participants and the confederates (as shown in Appendices H [English] and I [German]) and were partly orally repeated after they had read the instructions. The describer had the task of describing the three stimulus pictures to the listener, who was seated opposite. First, the describer was asked to look at each picture and to memorize what was depicted in the pictures. A *wordlist*\(^8\) (see Appendix J) with the seven most important items on each picture was handed out to ensure that all of the participants would talk about the same objects. The describers were able to look at the pictures as long as they needed. The listener was asked to turn away in order to make sure the listener could not see the pictures. Keeping the describer in the belief that there was a genuine listener who did not know what the pictures looked like ensured that the describer would describe the pictures with as much detail as possible. When the describer was ready, the listener turned around and faced forward toward the describer who was told (1) to

\(^7\) Carroll et al. (2000, p. 446) do not give a very detailed description of the actual procedure and the task given to the participants. Hence, the instructions for the current experiment do not correspond exactly to the ones given in Carroll et al.’s (2000) study.

\(^8\) In Carroll et al.’s (2000) study participants were not given a wordlist. The items on the wordlist were chosen based on the descriptions given during the piloting of the material.
describe the scene depicted in the picture from memory⁹, and (2) to specify the location of the objects on the wordlist as well as other objects they remembered so that the listener could visualize what the picture looked like as noted by Carroll et al. (2000, p. 446). During the descriptions, pictures and wordlists were out of sight in order to avoid that the describer would point at them. The order of the picture presentation was counterbalanced. In other words, half of the describers were asked to begin describing the city scene, while the other half was asked to start with the landscape scene. As no effect was expected from randomizing the wordlist items (cf. Tutton, 2013), the items on the wordlist appeared in the same order for every describer.

During the session, neither the listener nor the describer was permitted to ask any questions. Beforehand, the listener was instructed not to cross legs or arms but to put their arms on the upper thighs in order to avoid that the describer would mirror the listener’s body position, which might be unfavorable for gesture production (cf. for example Chartrand & Bargh, 1999). Following every description, the listener was required to identify the picture that had been described to them from a selection of similar pictures (see Appendix L)¹⁰, which were in a smaller format than the original pictures by Müller (1976; 2016) and placed on the table.

Prior to the actual test session, a practice run was conducted in order for both the describer and the listener to ask clarification questions and practice their tasks. During the practice run, three selection pictures (see Appendix K; shown in A4 format to the participants), were given to the listener after the description. The listener was asked to choose the picture just described and show their choice to the describer in order for the describer to better grasp their task.

---

⁹ Carroll et al. (2000) do not specify if the descriptions were made from memory or not. However, in a similar study by Carroll and von Stutterheim (1993) descriptions were given from memory.
¹⁰ The pictures in Appendix L were shown to the listener in the same format as depicted.
3.2.4 Coding and Analysis

ELAN, a digital video tagging software was used to transcribe and code the data (ELAN, http://tla.mpi.nl/tools/tla-tools/elan/, Sloetjes & Wittenburg, 2008). ELAN enables a frame-by-frame analysis of movements as well as of sound.

Each participant produced two descriptions (one for picture 1 and one for picture 2) yielding 14 English descriptions and 16 German descriptions in total. Each description file has been numbered and labeled following the pattern: language + participant number + picture number, e.g. E.1.1 or G.1.1, in order to anonymize the participant’s personal information. To take a case in point, example (29) was uttered by the English-speaking participant with participant number 3 describing the second picture.

(29) There is a pink house in the center. In front of it, there is a tree. Underneath the tree in the foreground, there is like a white cat that seems to be rolling on the floor (E.3.2).

All descriptions were transcribed verbatim from digital video using English and German standard orthography. Pauses, hesitations and other disfluencies were not noted in the transcription. A gloss is provided for the German descriptions following the Leipzig glossing rules. In order to maintain a degree of homogeneity across the data set, the focus of the analysis was the experimental target items (labelled x) only. Two items (i.e. x-motorcycle (x-Motorrad) in picture 1, and x-family (x-Familie) in picture 2) were added to the wordlist during the analysis, since all the participants included these items in their descriptions. Hence, the target items for the analysis include the objects presented in Table 3.
Table 3. Experimental items for each picture (x=experimental).

<table>
<thead>
<tr>
<th>Picture 1</th>
<th>Picture 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-cars</td>
<td>x-house</td>
</tr>
<tr>
<td>x-tram</td>
<td>x-tree</td>
</tr>
<tr>
<td>x-bus stop</td>
<td>x-cars</td>
</tr>
<tr>
<td>x-fountain</td>
<td>x-factory</td>
</tr>
<tr>
<td>x-newsstand</td>
<td>x-excavator</td>
</tr>
<tr>
<td>x-canal</td>
<td>x-silos</td>
</tr>
<tr>
<td>x-restaurant</td>
<td>x-skyscraper</td>
</tr>
<tr>
<td>x-motorcycle</td>
<td>x-family</td>
</tr>
</tbody>
</table>

Clauses selected for the analysis include the first complete mention of the target items (in bold) as well as subsequent mentions of the target items (in italics) when introducing other objects, as shown in example (29).

3.2.4.1 Speech coding

In order to replicate Carroll et al.’s (2000) results, the speech analysis closely followed the established analysis categories in Carroll et al. (2000), namely (1) reference introduction, (2) reference maintenance, (3) the division of the picture into subspaces, and (4) information structure, as summarized in Table 4 (Carroll et al., 2000, p. 447).

Table 4. Coding categories for speech (following Carroll et al., 2000).

<table>
<thead>
<tr>
<th>(1) reference introduction</th>
<th>(2) reference maintenance</th>
<th>(3) subspaces</th>
<th>(4) word order</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX</td>
<td>N</td>
<td>f</td>
<td>t/r</td>
</tr>
<tr>
<td>existential</td>
<td>noun</td>
<td>front</td>
<td>theme-relatum</td>
</tr>
<tr>
<td>LOC</td>
<td>DEMP</td>
<td>b</td>
<td>r/t</td>
</tr>
<tr>
<td>locative</td>
<td>demonstrative pronoun</td>
<td>back</td>
<td>relatum-theme</td>
</tr>
<tr>
<td>EX-LOC</td>
<td>ADV</td>
<td>FG</td>
<td></td>
</tr>
<tr>
<td>existential-locative</td>
<td>adverbal</td>
<td>foreground</td>
<td></td>
</tr>
<tr>
<td>LOC-EX</td>
<td>PROADV</td>
<td>BG</td>
<td></td>
</tr>
<tr>
<td>locative-existential</td>
<td>proadverbial</td>
<td>background</td>
<td></td>
</tr>
<tr>
<td>OTHER</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reference introduction. In a spatial description task the main linguistic means expected are existential constructions (example 30), and presentational constructions (including attributive as in (31) or possessive forms as in (32)), and locational constructions (example 33). Existentials and presentational were grouped as existentials in this study. These options differ in that existentials (EX) direct attention toward the object and its existence, as in (30) or indicate presence via an attribute of another object, as in (31) reflecting an object-based perspective. In contrast, locationals (LOC) introduce objects according to their location directing attention toward spatial information, as exemplified in (33) by Carroll et al. (2000).

(30) There is a house.
(31) There is a house with a turret.
(32) The house has a turret.
(33) In the center is an old fountain.

In the current analysis, existentials preceded (as in 34) or followed (as in 35) by a locative phrase were distinguished from existentials as in (30-32). Hence, five coding categories were determined: EX, LOC, EX – LOC, LOC – EX or other.

(34) In front of the house there are three parked cars (E.5.2).

(35) Dann gib-t es eine Strasse vor
dem rosa Haus
the-DAT pink-DAT house-DAT

‘Then there is a street in front of the pink house.’ (G:3.2)

Reference maintenance. Possible forms found in reference maintenance in spatial description tasks include the object-based full noun phrases (NP), and demonstrative pronouns (DemP) or the spatially based adverbials (ADV) such as there (in German da) as well as proadverbials (Proadv). Specifically, the phrases analyzed are objects maintained in the function of relatum (36). Full noun phrases and demonstrative pronouns relate to the object maintained in its status as an object, as
shown in (36) and (37) by Carroll et al. (2000). Specifically, these forms relate to objects and the places they delimit.

(36) There is a dormer window up at the top of the house; beside the house [relatum] is a clock.

(37) There is a fountain; beside that [relatum] is a bus stop.

(38) There is a market at the far end of the square; they seem to sell flowers there.

In turn, the use of adverbials, such as there (38), is not linked to any specific object or area of space delimited by its outer boundaries, but rather to the area that constitutes the far end of the square, as exemplified by Carroll et al. (2000). After all, no reference is made to the object functioning as relatum.

Moreover, in German it is possible to treat maintained objects in locative descriptions in terms of the spaces they delimit, as shown in (39). This option is grammaticized in German but not in English. Expressions that relate to a space (i.e. ‘proadverbials’) form a highly productive paradigm covering spaces such as e.g. davor ‘there in front’, dahinter ‘there behind’, daneben ‘there beside’, darüber ‘there above’ and darunter ‘there below’ (Carroll et al., 2000, p. 453). In the Duden reference grammar (Wöllstein, 2016) we can find prepositional adverbs or pronominal adverbs (i.e. proadverbials) as a subgroup of adverbials for which two different terms exist. The latter one focuses on the function, while the former focuses on the form. In other words, formally, they are made up of an adverb (hier, da, wo ['here’, ‘there’, ‘where’]) and a preposition (von, auf, während, etc. ['of’, ‘on’, ‘over’, etc.]). Hence, the name prepositional adverb. From a functional perspective they stand as a ‘replacement form’ (pre-form) for a prepositional phrase, e.g. vor dem Baum-> davor shown in example (39).

(39) In der Mitte des Bild-es befinde-t sich
In the-DAT middle-DAT the-GEN picture-GEN be located-PRS.3SG itself
dieser Baum. Da-vor ist ein Acker mit
this-NOM tree-NOM. There-in front be.PRS.3SG a-NOM field-NOM with
‘In the middle of the picture is this tree. In front of it there is a field with a small white cat on it.’ (G.6.2)

In sum, the options available are ‘object-oriented’ (nominal) and ‘space-oriented’ (adverbial) (Carroll et al., 2000, p. 453). Hence, the following coding categories were chosen: N (nouns including personal pronouns and relative pronouns), DEMP (demonstrative pronouns), ADV (adverbials) and PROADV (proadverbials).

Division of picture into subspaces. An object-based perspective can be further reflected in the way the picture is being divided into subspaces, such as foreground and background (in German Vordergrund and Hintergrund). In other words, the concepts foreground and background are object based in that they are closely related to inherent features of the object as a picture. Moreover, they cannot be used nor in English neither in German to locate an entity, for instance, ‘in the front half of a car’, exemplified in (40) by Carroll et al. (2000, p. 448).

(40)  * at the foreground of the car

(41)  in front of the tree, the ball, the car, the picture, etc.

In contrast, the concepts front and back (in German vorne and hinten) can not only be applied to a picture, but more generally to a range of spaces, shown in (41) by Carroll et al. (2000, p. 448). Consequently, Carroll et al. (2000) considered the tendency to use more frequently the concepts foreground and background encoding a perspective on space that is more sensitive to specific object features in comparison to front and back, which are more object-neutral (p. 448). In short, the coding categories for subspaces were f (front), b (back), FG (foreground), BG (background).

Word order. With respect to word order two sequences are possible when comparing the orders for the thematic roles theme (i.e. Figure in Talmy (2000)’s terminology) and relatum (i.e. Ground in Talmy (2000)’s terms), as demonstrated in (42) and (43) by Carroll et al. (2000, p. 448). Carroll
et al. (2000) explain that in locative descriptions, the entity located takes the role of the theme and the entity that serves to locate it forms the relatum (p. 448).

(42) There is a newspaper stand (theme) beside the bus stop (relatum).

(43) Beside the bus stop (relatum) there is a newspaper stand (theme).

On the one hand, the speaker can decide to draw attention first to the existence of the object in the domain of discourse before locating it, as in (42), or first to draw attention to the place or space at which it is located, as in (43). Both orders are possible in both languages under investigation (Carroll et al., 2000, p. 458). The chosen coding categories for word order were: t/r (theme-relatum) and r/t (relatum-theme).

3.2.4.2 Gesture coding

The transcription and coding of gestures was conducted in several steps. Firstly, all the movements that are gestures were identified, i.e., all hand, arm, leg and hand movements regardless of size, location or complexity, excluding ‘object manipulations’ (McNeill, 1992) and ‘self-adaptators’ (Ekman & Friesen, 1969), such as for instance nose scratching or stroking the hair. Specifically, in all target item clauses gesture strokes and post-stroke holds were identified with sound turned off. The stroke is defined as the phase of the gesture movement, in which effort and shape are displayed with most clarity. In other words, it is the meaningful and expressive part of the gestural movement where the spatial excursion of the limb reaches its apex. The phase of movement leading up to the stroke is called preparation, while the recovery is the phase of movement that follows the stroke as the hand is relaxed or withdrawn. Sometimes, the stroke may also be followed by a phase in which the articulator is kept in the position at which it arrived at the end of the stroke. This has been termed the post-stroke hold, which seems to be a way by which the expression conveyed by the stroke may be prolonged to allow the rest of the co-expressive speech to be uttered before the hand goes into retraction or while the next gesture is articulated (Kendon, 2004; McNeill, 1992). As the analysis focuses on examining what the meaningful part of the gesture is aligned with in speech, both strokes and post-stroke holds were taken into account. In order to determine onset
and offset of gesture strokes, changes in the parameters of shape, placement of the hand, trajectory of motion and tension of the hands were considered (cf. Kendon, 2004).

In a second step, still with sound turned off, the form of the gesture was coded. Specifically, gestures were coded for perspective, i.e. the shape of the hand as either ‘figure–incorporating’ (Gullberg, 2011) labelled with object, or encoding simply the location of the ‘figure objects’ (Gullberg, 2011) labelled as space. Following Gullberg (2011) gestures displaying handshapes reflecting the figure object by tracing or molding the shape of the figure object were coded as object (‘figure-incorporating’, Gullberg, 2011), as shown in Figure 5.

In turn, gestures articulated with a pointing hand shape (i.e. an extended index finger) or no particular hand shape (i.e. ‘a relaxed, floppy or flat hand’, Gullberg, 2011,) indicating a particular location in the gesture space of the speaker were coded as space (i.e. location-encoding), as exemplified in Figure 6. The two coding categories are mutually exclusive for each gesture.
Excluded from the analysis were rhythmic (i.e. beats/pragmatic) and emblematic gestures (e.g. ‘listing gestures’, i.e. when a speaker is counting her fingers to list the objects in the picture) (Kendon, 2004; McNeill, 1992) because they do not represent figure objects or their spatial locations.

In a third step, with sound turned back on, speech that co-occurred exactly with the stroke and any post-stroke hold was transcribed for every identified gesture unit. Following common practice in gesture coding (e.g. Kendon, 2004; McNeil, 1992), gestures were only considered as co-occurring with the full speech element if they covered a vocalized syllable of the aligned expression. For example, if the stroke co-occurred with ‘Br’ in Brunnen ‘fountain’, which is not a full syllable, the expression was not considered to be aligned with the gesture. Thus, if the stroke co-occurred with the full syllable ‘Brun’, it was to be accounted for as aligning. Moreover, the speech alignment was further categorized into alignment with the object (OBJ), the locative (LOC), BOTH (i.e. the object and the locative expression) or OTHER. To note, gestures were coded as aligning with the object, if they co-occurred with the lexical noun, or solely with the article or the adjective referring to the object. Table 5 summarizes the gesture coding categories.

Figure 6. Handshape categories for spatial location.
Table 5. Coding categories for gestures.

<table>
<thead>
<tr>
<th>(1) gs perspective</th>
<th>(2) speech alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>object</strong></td>
<td>OBJ</td>
</tr>
<tr>
<td>object-incorporating</td>
<td>object</td>
</tr>
<tr>
<td>handshape</td>
<td></td>
</tr>
<tr>
<td><strong>space</strong></td>
<td>LOC</td>
</tr>
<tr>
<td>location-encoding</td>
<td>locative</td>
</tr>
<tr>
<td>handshape</td>
<td></td>
</tr>
<tr>
<td><strong>BOTH</strong></td>
<td></td>
</tr>
<tr>
<td><strong>OTHER</strong></td>
<td></td>
</tr>
</tbody>
</table>

3.2.4.3 Statistical analysis

In the following result section descriptive statistics, specifically mean proportions, are provided throughout. If not otherwise indicated, the mean proportions are always based on the mean over all participants and not over all the data in order to make each participant count equally instead of making every single data point count the same. Specifically, mean proportions were computed for reference maintenance, reference introduction, subspaces and word order in the speech analysis, as well as gesture handshape and temporal alignment of gestures with speech. Moreover, standard deviations (SDs) are provided to indicate the amount of dispersion of the individual data points. Calculations are based on the experimental items only for both the speech and the gesture analysis. Finally, in order to facilitate a direct comparison to the speech results in Carroll et al.’s (2000)11 study, aggregated proportions are provided as well for the speech measures.

A final note is that for practical reasons, the data have not been coded by a second coder for interrater reliability. This would have been preferable but was not possible for the project due to financial and temporal limitations.

---

11 Carroll et al. (2000) do not indicate if percentages are based on mean proportions over all the participants or over all the data.
Chapter 4 Results

4.1 Speech results

For speech, the analysis will be presented following the order in Carroll et al. (2000), including reference introduction, reference maintenance, division of the picture into subspaces, and word order (see Table 4). Aggregated proportions will first be presented in comparative tables to allow for direct comparisons to Carroll et al.’s (2000) study. Second, mean proportions for every analysis category will be given as well to allow for an inspection of dispersion over participants.

4.1.1 Reference introduction

There are three main forms whereby referents can be introduced in picture descriptions. Existentials can do this in two ways; first, by providing a statement asserting existence and second, by indicating presence via an attribute of another object or a possessive as exemplified in (30)-(32). In Table 6 and 7 existentials and presentational have been summed up as existentials. A third means of introducing a referent is with a locational, which asserts the location of the referent at a specific space only, as exemplified in (33). In the latter case, the referents existence is only implied, but not explicitly stated. Table 6 provides aggregated percentages comparing the current study with Carroll et al.’s (2000) findings, whereas Table 7 presents the mean proportions over all the participants for the present study.

The data show that English speakers have a preference for existential and presentational constructions in reference introduction (74.7%), whereas locationals are the preferred form in German (67.7%), as shown in Table 6. These findings correspond well to the findings in Carroll et al. (2000), in which English speakers used existentials most often (78.5%) compared to German speakers who preferred locationals (71.8%).
Table 6. Comparison of aggregated proportions for *existential* and *locational* constructions in reference introduction.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>existential %</th>
<th>locational %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>German</td>
<td>English</td>
</tr>
<tr>
<td><strong>this study</strong></td>
<td>138</td>
<td>130</td>
<td>74.7</td>
</tr>
<tr>
<td><strong>Carroll et al. (2000)</strong></td>
<td>371</td>
<td>252</td>
<td>78.5</td>
</tr>
</tbody>
</table>

Table 7 shows the mean proportions for the same categories broken down in more detail. As in the aggregated analysis, English speakers also make use of locationals (22.3%); however, in the majority of cases in conjunction with an existential construction following the locational (38%), as exemplified in (47). Existentials (aggregated mean 95%) are much more frequent in the English data than attributive-possessive forms (aggregated mean 5%). The category *other* occurred only 3 times in the English data (mean 2.5%, SD 0.05), whereas in the German data none of the phrases were categorized as other. Moreover, locationals are the preferred form in the German data (62.6%), while existentials are less often used in German (37.4%), as demonstrated in Table 7.

Table 7. Mean proportions for reference introduction for every analysis category.

<table>
<thead>
<tr>
<th></th>
<th>existential</th>
<th>locational</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EX</td>
<td>EX-LOC</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>Mean %</td>
</tr>
<tr>
<td><strong>English</strong></td>
<td>20</td>
<td>17.3</td>
</tr>
<tr>
<td><strong>German</strong></td>
<td>17</td>
<td>15.6</td>
</tr>
</tbody>
</table>

Specifically, in the German data, attributives, which occur with objects that were introduced via locational constructions, as in (44), occur a bit more frequently (aggregated mean 56%) than
existentials (aggregated mean 44%) as in (45). Similarly, Carroll et al. (2000) report that the majority of presentationals in their German data are attributive-possessive forms, which occur together with locationals. One German speaker in the present study selected an object-based pattern of information organization consistently throughout the description of the first picture, as illustrated in example (46). According to Carroll et al. (2000) an object-based perspective occurred with a frequency of 1 out of 20 speakers in their collected data. Finally, as shown in Table 8, in the majority of existential clauses in German the construction with da was the preferred one (19.7%).

Table 8. Aggregated proportions for es gibt and SEIN constructions in the German data.

<table>
<thead>
<tr>
<th>da SEIN</th>
<th>da – other verb</th>
<th>da – es gibt</th>
<th>es gibt</th>
<th>es SEIN</th>
<th>loc SEIN</th>
<th>loc – other verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Mean %</td>
<td>n</td>
<td>Mean %</td>
<td>n</td>
<td>Mean %</td>
<td>n</td>
</tr>
<tr>
<td>11</td>
<td>8.5</td>
<td>14</td>
<td>10.8</td>
<td>2</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>23.8</td>
<td>67</td>
<td>51.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(44) In der Mitte ist ein Marktplatz mit ein-em Brunnen und ein-em Zeitungsstand
    a-DAT fountain-DAT and a-DAT newsstand

‘In the middle is a marketplace with a fountain and a news stand.’ (G.2.1)

(45) Dann gib-t es ein-en Brunnen direkt da-hinter.
    Then give-PRS.3SG it a-ACC fountain-ACC directly there-behind

‘Then there is a fountain directly behind it.’ (G.6.1)

(46) Das nächste Bild sieh-t komplett anders aus.
    The-NOM next-NOM picture look-PRS.3SG completely different aus.
    (prefix)
Das ist die Stadt, aber ich schätze so zwischen 20er und 60er Jahr-e. zwischen 20-ies und 60ies years-DAT.PL

Da sind alte Autos auf der Strasse und alte Motorräder.

Da fließt n kleiner Kanal durch die Stadt.

Und da ist so n klein-er Springbrunnen mit ner Statue oben drauf.

‘The next picture looks completely different. It is a city, but I guess like between the 20ies and 60ies. There are old cars and old motorcycles on the street. There is a canal running through the city.’ And there is like a small fountain with a statue on top of it.’ (G.1.1)

(47) Down by the lake there is like a willow tree (E.7.2)

4.1.2 Reference maintenance

The linguistic means available for reference maintenance are the object-oriented nominals (nouns and demonstrative pronouns) and the space-oriented adverbials (adverbials and proadverbials). Table 9 provides aggregated proportions comparing the current study with Carroll et al.’s (2000) findings, whereas Table 10 presents the mean proportions for every analysis category for the
present study. In line with Carroll et al.’s (2000) results, the object-oriented expressions are the preferred means used in English in the current study (93.2%), whereas the space-oriented option predominates in the German data (73.9%).

Table 9. Comparison of aggregated proportions for *object-oriented* and *space-oriented* expressions in reference maintenance.

<table>
<thead>
<tr>
<th></th>
<th><em>n</em></th>
<th><em>object-oriented %</em></th>
<th><em>space-oriented %</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>German</td>
<td>English</td>
</tr>
<tr>
<td><strong>this study</strong></td>
<td>73</td>
<td>88</td>
<td>93.2</td>
</tr>
<tr>
<td><strong>Carroll et al. (2000)</strong></td>
<td>89</td>
<td>144</td>
<td>76.4</td>
</tr>
</tbody>
</table>

The more detailed analysis draws an even clearer picture of the language specific preferences in reference maintenance, as summarized in Table 10. Similarly to Carroll et al.’s (2000) study, in English reference maintenance via the adverb *there* occurs in only 6.4% of all utterances that specify the location of a referent (compared to 6% in Carroll et al. (2000)), as in example (48). Moreover, as in Carroll et al.’s (2000) results, there are no occurrences in the current English data that correspond to the way in which adverbs are used in reference maintenance in German, as illustrated in (50). In turn, in German, cases in which nouns and pronouns occur in reference maintenance, as shown in example (49), are not restricted to an object-based perspective as reported in Carroll et al.’s (2000) study. Nevertheless, demonstrative pronouns do not occur in reference maintenance in the current German data.
Table 10. Mean proportions for reference maintenance for every analysis category.

<table>
<thead>
<tr>
<th>object-oriented</th>
<th>space-oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>DEMP</th>
<th>ADV</th>
<th>PROADV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>English</td>
<td>51</td>
<td>74.2</td>
<td>0.34</td>
<td>17</td>
</tr>
<tr>
<td>German</td>
<td>23</td>
<td>20.3</td>
<td>0.24</td>
<td>0</td>
</tr>
</tbody>
</table>

(48) And just in front of the news stand there is like a small place for bicycles to go. And there is one bicycle there with very small wheels (E.5.1).

(49) And next to the stop on the left side.


‘There is a small news stand next to the stop on the left side. And next to this news stand is again a big old building.’ (G.3.1)

(50) The apple tree is also still there. There is a white cat under the tree.’ (G.8.2)
4.1.3 Use of spatial concepts to divide the picture into subspaces

A third factor reflecting the perspective underlying information structure considered in Carroll et al.’s (2000) study are the type of spatial concepts used to divide the object being described into sections. The alternatives available, illustrated in (51-54), are the concepts of *foreground* and *background* (in German *Vordergrund* and *Hintergrund*), which delimit spaces on the sagittal axis or the more widely applicable concepts *front* and *back* (in German *vorne* and *hinten*), which can be applied not only to the picture, but to a wide range of other objects as well (e.g. tree, house, car, etc.). The former two refer to an object-specific perspective, and the latter two refer to a more object-neutral perspective. Table 11 provides aggregated proportions to compare the current results with Carroll et al.’s (2000) findings. Table 12 illustrates the mean proportions for every analysis category (*foreground/background* and *front/back*).

(51)  Und  im  Vordergrund  an  der  Kreuzung
     And  in-DAT  foreground-DAT  on  the-DAT  crossroad-DAT

     sieh-st  du  eine  Tramhaltestelle
     see-PRS.3SG  you-3SG  a-ACC  tram stop-ACC

     ‘And you can see a tram stop at the crossroad in the *foreground.*’ (G.3.1)

(52)  In the *background* there is some houses as well as like a skyscraper (E.2.2).

(53)  Rechts  vorne  steh-t  ein  schwarzes  Motorrad.
     Right  front  stand-3SG  a-NOM  black-NOM  motorcycle-NOM

     ‘There is a black motorcycle on the right side in *front.*’ (G.8.1)

(54)  In the *back* of the house there is a swing set (E.6.2).
In Carroll et al.’s (2000) study, speakers showed clear preferences in the concepts selected, as summarized in Table 11.

Table 11. Comparison of aggregated proportions for object-specific and object-neutral expression in picture division.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>object-specific %</th>
<th>object-neutral %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>German</td>
<td>English</td>
</tr>
<tr>
<td>this study</td>
<td>38</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>Carroll et al. (2000)</td>
<td>14</td>
<td>33</td>
<td>64.3</td>
</tr>
</tbody>
</table>

In Carroll et al.’s (2000) English sample, the preferred options were *foreground* and *background* (64.3%), which reflect inherent features of the picture as an object. In contrast, in the German data, there was a preference for the terms *vorne* ‘front’ and *hinten* ‘back’ (63.3%), which are neutral to the object in question. The current findings are not in accordance with previous results by Carroll et al. (2000). Instead in the present study both languages showed an almost equal preference for object-neutral concepts over object-specific concepts. As illustrated in Table 12, English speakers used the object-neutral concepts *front* (53.4%) and *back* (22.5) more than half of the times. Likewise, German speakers predominantly used *vorne* (56.3%) and *hinten* (35.2%) to divide the picture into subsections. The object-specific concepts *foreground/Vordergrund* and *background/Hintergrund* rarely occurred in English (8.6% versus 15.5%) and even less in German (1.3% versus 7.2%).
Table 12. Mean proportions for spatial concepts and subspaces division for every analysis category.

<table>
<thead>
<tr>
<th></th>
<th>object-specific</th>
<th></th>
<th>object-neutral</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>foreground</td>
<td>background</td>
<td>front</td>
<td>back</td>
</tr>
<tr>
<td>n</td>
<td>Mean %</td>
<td>SD</td>
<td>n</td>
<td>Mean %</td>
</tr>
<tr>
<td>English</td>
<td>4</td>
<td>8.6</td>
<td>0.12</td>
<td>4</td>
</tr>
<tr>
<td>German</td>
<td>1</td>
<td>1.3</td>
<td>0.04</td>
<td>4</td>
</tr>
</tbody>
</table>

4.1.4 Word order

The final category analyzed in this section relate to the order of the thematic roles theme and relatum. Table 13 summarizes aggregated proportions to compare the current results with those reported in Carroll et al. (2000). Table 14 provides a more detailed description of mean proportions for every analysis category. Table 13 shows that the results of the word order analysis in the current study comply with Carroll et al. (2000).

Table 13. Comparison of aggregated proportions of relatum-theme or theme-relatum word order.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>relatum-theme %</th>
<th>theme-relatum %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>German</td>
<td>English</td>
</tr>
<tr>
<td>this study</td>
<td>122</td>
<td>125</td>
<td>58.2</td>
</tr>
<tr>
<td>Carroll et al. (2000)</td>
<td>341</td>
<td>224</td>
<td>51.3</td>
</tr>
</tbody>
</table>

In both languages, the orders theme-relatum and relatum-theme are possible in principle. However, the analysis shows that in German the preferred order is relatum-theme (78.8%) as shown in Table
14; meaning that spatial information that locates the introduced entity precedes its mention in the clause, as shown in (55).

(55) Vor dem Haus (relatum) ist wie so ein kleiner Acker (theme).

In front of the-DAT house-DAT be-PRS:3SG like such a-ACC field-ACC

‘In front of the house is like kind of a small field’. (G.10.2)

(56) To the left of that (relatum) there is a news stand (theme) (E.6.1).

(57) There’s like three cars (theme) lined up outside of it (relatum) (E.2.2).

In contrast, Table 14 shows that in English the orders theme-relatum (43.7%) and relatum-theme (56.3%), demonstrated in (56) and (57), occur with almost equal likelihood.

Table 14. Mean proportions for word order for every analysis category.

<table>
<thead>
<tr>
<th></th>
<th>relatum-theme</th>
<th>theme-relatum</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>English</td>
<td>71</td>
<td>56.3</td>
</tr>
<tr>
<td>German</td>
<td>100</td>
<td>78.8</td>
</tr>
</tbody>
</table>

4.2 Gesture results

The analysis of gestures provides results for gestural handshapes and the temporal alignment of gestures with speech. Moreover, a combined analysis of handshape together with temporal alignment will be provided as well. Lastly, a post hoc analysis on gesture sequences will be
presented. In the following sections mean proportions refer to the mean over all the participants if not otherwise indicated. Again, SD refers to standard deviation and \( n \) refers to raw numbers only, specifically to the sum of all gestures produced by all the participants.

4.2.1 Handshape

Handshapes were coded as either figure-incorporating (object) or encoding the location of objects (space). Figure-incorporating handshapes were identified by gestures which molded or traced the shape of the object, while location-encoding handshapes were identified as pointing gestures or gestures with a relaxed and floppy hand. Table 15 summarizes the results of the handshape analysis. In both languages, figure-incorporating and location-encoding handshapes occur with almost equal frequency; however, figure-incorporating handshapes are slightly more common, and especially so in German (English: 50.9\% and German: 59.4\%).

Table 15. Mean proportions for handshape in English and German.

<table>
<thead>
<tr>
<th></th>
<th>object</th>
<th>space</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>English</td>
<td>50.9</td>
<td>0.22</td>
</tr>
<tr>
<td>German</td>
<td>59.4</td>
<td>0.20</td>
</tr>
</tbody>
</table>

4.2.2 Temporal alignment of gestures with speech

Speech that exactly aligned with the gesture stroke was coded as expressing \( OBJ \) (object), \( LOC \) (locative), \( BOTH \) or \( OTHER \). In English half of the gestures aligned with the object (50.2\%), followed by the locative expression (24.5\%) and \( BOTH \) (21.1\%), as shown in Table 16. In German most gestures aligned with the object (42\%), followed by \( BOTH \) (35.1\%) and the locative (21.3\%). In both languages most gestures aligned with objects, but in German, gestures also frequently
aligned with both locative and object. In English 11 gestures were coded as OTHER (4.2%), while in the German data only 3 gestures aligned with something else (1.6%).

Table 16. Mean proportions of gestural alignment in English and German.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>OBJ</th>
<th>LOC</th>
<th>BOTH</th>
<th>OTHER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean %</td>
<td>SD</td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>English</td>
<td>264</td>
<td>50.2</td>
<td>0.12</td>
<td>24.5</td>
<td>0.12</td>
</tr>
<tr>
<td>German</td>
<td>233</td>
<td>42.0</td>
<td>0.16</td>
<td>21.3</td>
<td>0.10</td>
</tr>
</tbody>
</table>

A gesture was coded as aligning with the object if it co-occurred with the lexical noun or with the article or an adjective referring to the lexical noun. Nevertheless, in both languages, in the majority of cases gestures coded as aligning with the object actually co-occurred together with a lexical noun (English: 84.5% and German: 89.4%), as shown in Table 17.

Table 17. Mean proportions of gestural alignment OBJ occurring with lexical noun (N) or without (-N).

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>OBJ (N)</th>
<th>OBJ (-N)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>English</td>
<td>124</td>
<td>84.5</td>
<td>0.11</td>
</tr>
<tr>
<td>German</td>
<td>97</td>
<td>89.4</td>
<td>0.12</td>
</tr>
</tbody>
</table>

4.2.3 The relationship between handshape and temporal alignment

Taking a closer look at the relationship between handshapes and the temporal alignment with speech, the slight difference between the two languages becomes even more apparent. Table 18
and 19 summarize the relationship between handshapes and alignment separately for figure-incorporating and location-encoding handshapes, respectively. As shown in Table 18, in both languages most figure-incorporating handshapes aligned with the object in speech (English: 32.3% and German: 29.2%). The second most frequent alignment category in the German data was BOTH, which co-occurred most often with a figure-incorporating handshape (23.1%). In contrast, in English the co-occurrences of a figure-incorporating handshape with both the object and the locative was almost three times less likely as in German (8.3%).

Table 18. Mean proportions for gestural alignment and object-handshape in English and German.

<table>
<thead>
<tr>
<th></th>
<th>OBJ + obj</th>
<th>LOC + obj</th>
<th>BOTH+ obj</th>
<th>OTHER + obj</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean %</td>
<td>SD</td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>English</td>
<td>32.3</td>
<td>0.15</td>
<td>7.9</td>
<td>0.09</td>
</tr>
<tr>
<td>German</td>
<td>29.2</td>
<td>0.11</td>
<td>6.7</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>0.03</td>
<td>1.6</td>
<td>0.03</td>
</tr>
</tbody>
</table>

In contrast, the location-encoding handshape aligned most frequently with the object in English (17.9%), followed by the locative expression (16.7%) and BOTH (12.4%). In the German data most location-encoding handshapes aligned with the locative (13.5%) followed by the object (12.8%) and BOTH (12.6%), summarized in Table 19.

Table 19. Mean proportions for gestural alignment and space-handshape in English and German.

<table>
<thead>
<tr>
<th></th>
<th>OBJ + sp</th>
<th>LOC + sp</th>
<th>BOTH + sp</th>
<th>OTHER + sp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean %</td>
<td>SD</td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>English</td>
<td>17.9</td>
<td>0.10</td>
<td>16.7</td>
<td>0.12</td>
</tr>
<tr>
<td>German</td>
<td>12.8</td>
<td>0.15</td>
<td>13.5</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.03</td>
<td>0.5</td>
<td>0.02</td>
</tr>
</tbody>
</table>
4.2.4 Post hoc analysis of gesture sequences

During the course of the analysis a pattern within the English and German gesture data occurred which seemed interesting to further investigate. That is, it was observed that speakers articulated a gesture and then left the hand in place, holding the final shape of the gesture (i.e. a post-stroke hold), while the other hand continued to represent further content along with speech (i.e. a sequential gesture). It was assumed that the informational content co-occurring with the hold and the other hand was relevant to the organization of information in descriptions. Given that speakers may use gestures to create a spatial lay-out of the picture they are describing, it was assumed that an analysis of gesture sequences could provide additional indications regarding the role of language specific-perspectives and information structure in English and German spatial descriptions. Hence, a post hoc analysis of post-stroke holds and sequential gestures was undertaken. It was hypothesized that if speech and gestures form an integrated system and if gestures are sequentially related to each other within a gesture unit, they should provide a visual map of information structure in speech.

![Figure 7. Example gesture sequence and rest position of the hands.](image)

Specifically, in a post hoc analysis of gesture sequences, only the gestures, which sequentially followed and co-occurred with a post-stroke hold, were considered and analyzed quantitatively and qualitatively. That is, gestures were additionally coded if they occurred relative to another gesture or not (i.e. relation to another gesture: yes, no), exemplified in Figure 7. In Figure (7a), the speaker’s left hand remains in the position of the post-stroke hold while the sequential gesture
is articulated with the speaker’s right hand. In order to distinguish the post-stroke hold from the general rest position of the hand (cf. Figure (7b), the handshape as well as the position in the gesture space and the tension of the hand were considered.

![Image](image-url)

(x-tree, E.1.2) | (x-family, E.1.2)
---|---
(x-news stand, E.1.1) | (x-motorcycle, E.1.1)

Figure 8. Examples of gesture sequences in English speaker 1.

In the English data, 14.7% of all the gestures followed and co-occurred with a post-stroke hold, whereas in the German data only 7.1% of all the gestures were sequentially related to a post-stroke hold, as illustrated in Table 20. In English, one speaker produced 75% of all sequential gestures (see Figures 7 & 8), while in German the occurrences of sequential gestures were more evenly distributed over all the participants. In fact, English speaker 1 halted her left hand in the same position during the whole description of both pictures, as shown in Figure 8.
Table 20. Mean proportions for overall gesture sequences in English and German.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean %</td>
<td>SD</td>
<td>Mean %</td>
</tr>
<tr>
<td>English</td>
<td>264</td>
<td>14.7</td>
<td>0.29</td>
</tr>
<tr>
<td>German</td>
<td>233</td>
<td>7.1</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Further inspection of the sequential gestures in the English and German data shows that most sequential gestures were articulated with a figure-incorporating handshape (English: 69.7% and German: 71.4%), as shown in Table 21.

Table 21. Aggregated proportions of sequential gestures for gesture handshape.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>object %</th>
<th>space %</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>33</td>
<td>69.7</td>
<td>30.3</td>
</tr>
<tr>
<td>German</td>
<td>14</td>
<td>71.4</td>
<td>28.6</td>
</tr>
</tbody>
</table>

In terms of alignment with speech, provided in Table 22, on the one hand, in English, more than half of the sequential gestures (57.6%) temporally co-occurred with the object in speech. In the German data, on the other hand, the sequential gestures aligned with equal likelihood with the object (35.7%) as well as with the locative (35.7%), closely followed by both the object and the locative (28.6%).

Table 22. Aggregated proportions of sequential gestures for speech alignment.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>OBJ %</th>
<th>LOC %</th>
<th>BOTH %</th>
<th>OTHER %</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>33</td>
<td>57.6</td>
<td>12.1</td>
<td>27.3</td>
<td>3.0</td>
</tr>
<tr>
<td>German</td>
<td>14</td>
<td>35.7</td>
<td>35.7</td>
<td>28.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Similarly, in a closer examination of the relationship between gesture handshape and the temporal alignment of the sequential gestures with speech, the patterns previously reported for the whole gesture data set can be detected for sequential gestures as well, as shown in Table 23 (cf. Table 16 & 18). That is, in English most sequential gestures aligned with the object and were articulated with a figure-incorporating handshape (42.2%). In German, sequential gestures were most frequently articulated with a figure-incorporating handshape and aligned equally likely either with the object (28.6%) or both the object and the locative expression (28.6%). In addition, German speakers also produced quite frequently sequential gestures, which were articulated with a location-encoding handshape and aligned with the locative in speech (21.4%).

Table 23. Aggregated proportions of sequential gestures for handshape and speech alignment.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>OB+ obj</th>
<th>LOC+ obj</th>
<th>BOTH+ obj</th>
<th>OTHER+ obj</th>
<th>OBJ+ sp</th>
<th>LOC+ sp</th>
<th>BOTH+ sp</th>
<th>OTHER+ sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>33</td>
<td>42.2</td>
<td>6.1</td>
<td>15.2</td>
<td>3.1</td>
<td>18.2</td>
<td>3.1</td>
<td>12.1</td>
<td>0.0</td>
</tr>
<tr>
<td>German</td>
<td>14</td>
<td>28.6</td>
<td>14.3</td>
<td>28.6</td>
<td>0.0</td>
<td>7.1</td>
<td>21.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

In sum, in both English and German, most gestures following a post-stroke hold were articulated with a figure-incorporating handshape and align with the object in speech. In German, sequential gestures articulated with a figure-incorporating handshape also co-occurred frequently with both the object and the locative. In other words, in both languages, most of the sequential gestures directly referred to the figure objects being described, as shown in Table 24.

Taking a closer look at the individual occurrences of sequential gestures in Table 24, it becomes apparent that the gestures which remain in the post-stroke hold position align with the relatum in speech in both the English and German data. In turn, the sequential gestures commonly align together with the post-stroke hold with the theme in speech as exemplified in Table 24. To note, the gesture strokes of the sequential gestures and the post-stroke holds of the previous gestures together aligned with the transcribed speech in parentheses, as shown in Table 24. In most instances of a post-stroke hold, the halted hand has a flat handshape (except in example c) in Table
24) and is generally in a vertical position toward the speaker. However, the hand may also be halted horizontally to the speaker’s body, as shown in b) in Table 24. The sequential gestures were articulated with an object-incorporating handshape, except for example c) where the first sequential gesture was articulated with a location-encoding handshape and the second with a figure-incorporating handshape.

Table 24. Examples of gesture sequences in both the English and German data

<table>
<thead>
<tr>
<th>Example</th>
<th>Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>(screenshot description)</td>
<td>(speech that temporally aligned with both the sequential gesture and the post stroke hold in [ ])</td>
</tr>
<tr>
<td>a)</td>
<td>In the middle there is it’s an intersection (relatum). [There’s a tram (theme) coming towards the middle of the intersection]. <em>(x-tram, E.4.1)</em></td>
</tr>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>b)</td>
<td>There is in the foreground of the picture (relatum) [the tram line (theme) is going, the tram (theme) is going from a tram stop that says MQ Escher]. <em>(x-tram, E.5.1)</em></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>c)</td>
<td>In der Mitte ist halten dieser Marktplatz (relatum) [und links da-von] this-NOM middle-DAT be-PRS.3SG just market place-NOM and left there-of ‘In the middle is this market place and to the left of the market place’ <em>(x-canal, G.2.1)</em></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>
ist auch ein kleiner, also wie ein Kanal, ein Bach (theme)
Chapter 5 Discussion

This chapter summarizes and discusses the results of the current study in more detail. Firstly, the research questions, hypotheses and results will be briefly summarized. Secondly, the results and hypotheses will be discussed in light of theories and previous studies presented in the background section. Thirdly, attempts will be made to explain individual differences among the participants. Finally, the limitations of the present design will be shortly reviewed together with some possible outlooks for future investigations in the area of spatial descriptions and language-specific perspective taking in event conceptualization in both speech and gestures.

The aim of the study was to replicate Carroll et al. (2000) in order to investigate the planning processes in language production by contrasting English and German, and to explore the contribution of gestures in the process of discourse conceptualization. As the current study is a replication study, the research questions and hypotheses for the speech analysis were directly adopted from Carroll et al. (2000). It was assumed that, if the domains ‘object’ and ‘space’ play a central role in information organization in spatial description tasks, texts in English should be organized according to an object-based perspective, whereas in German a spatially-based perspective should guide the way information is organized. Hence, the linguistic means in reference introduction, maintenance, the division of the picture into subspaces and word order should be in line with the chosen frame of reference. Moreover, it was hypothesized that, if speech and gestures form an integrated system, language-specific perspectives in information organization should be reflected in the gesture form and alignment with speech. Specifically, an object-based perspective in English descriptions should be reflected in the use of figure-incorporating gesture handshapes which align with information about figure objects in speech. Conversely, a spatially-based perspective in German was presumed to be reflected in the use of pointing gestures or relaxed and flat gesture handshapes which align with information about the location of figure objects.
5.1. Speech results

Native speakers of English tended to adopt an object-based perspective in the descriptive task provided and to focus on the entities in the picture by making statements about the existence of entities. This object-based perspective led to the use of existentials in reference introduction and nominals in reference maintenance. In Carroll et al. (2000), the selection of concepts to divide the picture into sections also followed the object-based pattern. However, in the current study, the English speakers more frequently used the object-neutral concepts *front* and *back*, instead of the expected object-specific terms *foreground* and *background*. A possible explanation for this finding might be that most items on the wordlist were located in the ‘middle ground’ of the picture. In other words, it was not necessary to divide the picture into subsections; thus, the terms *foreground* and *background* were less relevant in locating these entities. The word order patterns in English showed no specific preference in either study. The order theme-relatum and relatum-theme occurred with almost equal frequency.

Native speakers of German predominantly followed another option in organizing information and focused on spaces in the picture. In other words, the entities were selected for mention based on the regions of space they share. This perspective included a high rate of locationals in reference introduction and adverbials and proadverbials in reference maintenance. In reference maintenance, objects were also treated in their status of being located in a specific spatial area. In German, this option is grammaticized in the form of proadverbials, which constitute a paradigm (e.g. *davor* ‘there in front’, *daneben* ‘there next to’, etc.). The object (e.g. *der Brunnen* ‘the fountain’; *in front of the fountain, next to the fountain*) was not explicitly stated in reference maintenance, it was only implicitly mentioned in the form of the spatial adverb *there*. German speakers’ ‘preoccupation’ with location was also reflected in the preferred word order (relatum-theme). In particular, the option to mention information about the location of entities before their actual mention in the clause was the most frequent way to introduce entities. Moreover, an object-based perspective is also a possible option in German, but did not constitute the norm. That is, the spatial-based perspective was the ‘habitual’ or ‘default’ option. Specifically, the object-based perspective occurred in 1 out of 16 descriptions in the current data. In Carroll et al. (2000), 1 out
of 19 speakers chose a thematic description. Slobin (1991) notes: “speakers so rarely make use of options that differ from the norm” (p. 17).

To take a case in point, one native German speaker chose an object-based pattern throughout the description of the first picture using predominantly existentials in reference introduction and nominals in reference maintenance. The consistent use of an object-based perspective throughout the description of the first picture reflects the choices made during the process of conceptualization, in which information is selected and organized for expression before information is grammatically and phonologically encoded and ensures coherence in the descriptive text. Carroll et al. (2000) point out:

The factors that determine the choice of linguistic form must be in place prior to accessing the repertoire of linguistic structures (the lexicon), if consistency and coherence is to be maintained with respect to the options selected. Thus, some factor other than the mere range of linguistic means available in the lexicon must come into play when determining the selection of one or the other set of options (Carroll et al., 2000, p. 463).

To put it differently, the preferred patterns in the respective languages ensure consistency throughout discourse organization; specifically, through the conceptualizer at the global level, within the planning process of language production. Hence, when carrying out a complex verbal task such as a description, speakers generate a conceptual structure that guides the way information is selected, organized and formulated in the production process. Carroll et al. (2000) point out that a unifying perspective guides the options selected and is linked to grammaticized forms such as “the syntactic category of subject (in coding function such as topic), the syntactic V2 feature, and the morphosyntactic structure of spatial expressions” (p. 462).

In particular, Carroll et al. (2000) indicate that the selection of an object-based perspective in English is determined by the role of the syntactic subject. Generally, English has an SVO word order, in which the order is very rigid. That is, the word order determines the syntactic roles of the constituents within a sentence. The subject always comes before the verb and the object after it (Velupillai, 2012, p. 282). Moreover, in English, the syntactic subject normally encodes topic information. In other words, word order favors information organization towards categories that
can map into the subject role, such as entities in a picture. Hence, attention is directed to the domain of objects rather than space. In contrast, in German, there is no dominant order for the constituents subject, verb and object; although the finite verb is always in the second position in a canonical main clause (i.e. verb second (V2) language). Even though the slot before the verb is reserved for topic information, both object and spatial information can be found in this slot and function as topic in information structure. Finally, as mentioned, above proadverbials are a grammaticized category forming a separate paradigm in German. That is, proadverbials refer to entities in terms of their location in reference maintenance; an option which is not available in English. Consequently, English and German differ in their structural features and generate conceptual structures that vary in the overall perspective taken on the communicative task at hand. In other words, native English and German speakers differ in the way they ‘think while they speak’ in spatial description tasks (Carroll et al., 2000, p. 462ff).

Interestingly, then, the speech results in Carroll et al. (2000) have been successfully replicated in this current study, lending support to hypotheses (H1 and H2) directly derived from Carroll et al. (2000). The results of the speech analysis overall align with Carroll et al. (2000) and support their claims that in English an object-based perspective is central for information organization in English, while in German the spatially-based perspective is more frequent. This difference in perspective is reflected in the linguistic form chosen in English and German to describe pictures. That is, the linguistic means used in reference introduction, reference maintenance, and word order in spatial description tasks in English and German form a typological pattern. In Carroll et al.’s (2000) words:

The study shows that the selection of linguistic means in the [respective] languages is guided at the global level of information organization by unifying principles that lead to clusters of form-function relations. The principles guiding the selection process are perspective driven and are associated with patterns of grammaticization (see Carroll, 1993; Carroll & von Stutterheim, 1993) (Carroll et al., 2000, p. 460).
5.2 Gesture results

For gestures, the handshape analysis showed that in English, the figure-incorporating and location-encoding handshape occurred with almost equal likelihood, whereas in German the figure-incorporating handshape was slightly more frequent. In terms of temporal alignment with speech, in English, gestures were most frequently aligned with the object and were articulated with a figure-incorporating handshape. In German, gestures aligned most often with the object or both the object and the locative expression in speech and were articulated with a figure-incorporating handshape. The same pattern was also observed for the post hoc analysis of sequential gestures (i.e. gestures which followed a post-stroke hold). All in all, most sequential gestures directly referred to the objects under description.

The results of the gesture analysis do not support hypotheses (H3a) and (H3b) completely, since native English and German speakers do not show any difference in gesture handshape in the current descriptive task. Most gestures in both languages were articulated with a figure-incorporating handshape. Thus, most speakers focused on the objects. A possible explanation for the absence of language specific behavior in handshapes might be the nature of the task itself, the specific formulation of the task (see Appendix H & I), and the provided wordlist (see Appendix J) which both put a strong focus on the objects on the pictures. Generally, a descriptive task requires an answer to the overall question: ‘What is X like?’ Therefore, an answer focuses on representing the entities on the picture and their properties (Carroll et al., 2000). Moreover, in the instructions it is specifically noted that speakers should pay particular attention to the objects on the wordlist. Hence, the nature of the task, the instructions and the given wordlist made it a pragmatic necessity for speakers to focus on the objects in their descriptions.

Nevertheless, importantly, the analysis of gesture alignment with speech did detect language specific structures and perspective-taking. As noted above, in English most gestures were articulated with a figure-incorporating handshape and aligned with the object in speech, while in German most gestures were articulated with a figure-incorporating handshape and aligned either with the object or both the object and the locative expression in speech. As pointed out above,
German speakers did not just indicate the existence of the objects in space, but also focused on locating these objects under description in a specific location in space. Formulated in terms of an overall question, English speakers thus organize information in spatial descriptions on the basis of the guiding question: ‘What is where?’; and German speakers follow the overall question: ‘Where is what?’ (Carroll et al., 2000). Hence, the difference in perspective between English and German was not manifested in the gesture handshapes, but in the temporal alignment of the gestures with speech. In Gullberg (2011), language-specific perspectives were driven by different verb semantics and therefore gestures differed cross-linguistically in their meaning, which was manifested in different handshapes. In the current study it is mainly the difference in information structure that provides evidence for cross-linguistic differences in the conceptualization of the same subject matter. In particular, the gesture alignment reflects the structural characteristics of the two languages and provides evidence that speech and gesture form an integrated system. That is, the fact that most gestures were articulated with a figure-incorporating handshape and temporally coincided with the object in speech or both the object and the locative in speech provides evidence that speech and gesture express the same underlying idea unit. Even though speech and gesture belong to different modalities, they are co-expressive. Hence, no evidence for a complementary role of gestures in spatial descriptions was found.

The post hoc analysis of gesture sequences supports the findings that the alignment of gestures with speech reflects the structural differences between English and German and that speech and gesture form an integrated system. That is, in English most sequential gestures were articulated with a figure-incorporating handshape and aligned with the object in speech, whereas in German sequential gestures were mostly articulated with a figure-incorporating handshape and aligned either with the object or both the object and the location. Thus, most sequential gestures refer directly to the objects under description. This observation is surely again related to the nature and formulation of the given task. Moreover, the qualitative analysis showed that the post-stroke hold functions as the relatum (i.e. Ground) in the visual modality and corresponds to the relatum in speech. In other words, the post-stroke hold refers to the entity whose location is of relevance. The sequential gesture, in turn, aligns with the theme (i.e. Figure) in speech and functions as the
theme in the visual modality. That is, the sequential gesture refers to the reference entity in relation to which the location of the Figure (or theme) is specified. To put it differently, the information status of the respective entities in speech is reflected in the relation of the sequential gestures and the post-stroke holds which together provide a visual map of information structure in speech. Consequently, the gestural component of an utterance contributes both to the referential content as well as to the structure of an utterance because speech and gesture form an integrated system expressing the same underlying information unit.

The participants all have a very diverse language background, and there is a lot of variation between the individual speakers and their descriptions. As previously mentioned, German speaker 1 selected an object-based perspective to describe the first picture, but for the description of the second picture German speaker 1 chose a spatial-based perspective. A possible reason for the change of perspective between the two descriptions might be that the speaker needed to warm-up in the target language before being fully engaged in native speaker mode. This speaker noted in the questionnaire that he uses English on a daily basis in most of the domains listed, whereas the usage of German has moved to marginal domains in his daily language usage. Furthermore, the standard deviation for gesture sequences in the English data was much higher than for those in the German data. This is due to the fact that English speaker 1 produced most of the gesture sequences. The left hand was always put in a halted position indicating the relatum, as shown in Figure 8 and 9. This is a rather extreme example of how speakers sometimes prolong a gesture with a post-stroke hold and probably does not reflect the norm.

5.3 Limitations of the study and outlooks for the future

It is important for possible future investigations to discuss the weaknesses that occurred in the current study in order to avoid such drawbacks in the future.

Due to time and resource constraints, as well as the explorative nature of the gesture analysis, only a small number of participants for each language under investigation were recruited. Therefore, the representativeness of the chosen subsamples of participants is one of the limitations
of this study. Participants were selected through convenience sampling and social network sampling, which are quick and easy recruitment methods. A major disadvantage of convenience sampling is that it is a non-probability method, which is rather limited in its generalizability and possible statistical inferences, as the subjects do not represent a total population of native English and German speakers. Larger samples would allow drawing more reliable inferences (Buchstaller & Khattab, 2013, p. 75-81). Further, only descriptive statistics have been applied in the analysis, which only describe trends. Consequently, generalizations of the current findings should be made carefully. Future studies should try to interview more participants per language group in order to draw more reliable conclusions.

Moreover, unlike in Carroll et al. (2000), participants were not interviewed in their first language (L1) environment, but in a second (L2) or even third language (L3) environment. That is, all the participants lived and studied either in Denmark or Sweden, where they are exposed to several other languages. Consequently, none of the participants were monolinguals. That is, all of the German speakers had a fairly advanced knowledge of English according to their self-assessment and used English on a daily basis at the university, at work, to read and write, as well as to talk to friends and colleagues. In addition, 3 out of 7 English participants also had some intermediate knowledge of German according to their self-assessment but did not use German on a daily basis. Therefore, the participants’ knowledge of a second language might have also had some influence on their first language and consequently on their speech and gesture production in the given descriptive task (cf. Brown & Gullberg, 2008). Specifically, Brown and Gullberg (2008) indicate that bidirectional interaction between the L1 and L2 in speech and gesture can occur even if speakers only have an intermediate proficiency in the second language. In other words, the current English and German data might not represent how monolingual native English and German speakers would talk and gesture about spatial description. A future study might take the participants’ language background more into account and collect data in the native countries, e.g. in Germany and England.

Unfortunately, the methods in Carroll et al. (2000) were not explained in great detail, especially the procedure, the actual task, the instructions given as well as the dimensions and form
of the stimulus material. For replicability and transparency issues, the methods in Carroll et al. (2000) should have been presented more comprehensively. Moreover, following Tutton (2013a), the scale of the space to be described might be taken into consideration as well in the future, since it affects spatial language (i.e. indoor vs. outdoor scene). After all, concerning the procedure of the experiment, improvements are necessary for a future study. That is, taking also the wordlist out of sight from the describer was a consideration error. In order to make sure that all the participants would mention all the objects on the wordlist and to simplify the analysis it would have been better to leave the wordlist in front of the participants during the descriptions. Moreover, following Tutton (2013a), participants should have been given a time limit for their descriptions to ensure that all of the descriptions are about the same length. In addition, the wording of the instructions for the descriptive task should be revised again in order to avoid that speakers are biased in their perspective-taking.

Finally, for the reliability of the coding and analysis of the data, especially of the gesture data, it would have been better to have a colleague or research assistant coding the gestures as well in order to determine the interrater agreement. However, due to temporal and financial constraints of the present study, this was not possible.
Chapter 6 Conclusions

This study was able to successfully replicate Carroll et al.’s (2000) results providing support for their conclusions. In particular, the cross-linguistic comparison of the way information is organized for expression in a spatial description task in English and German demonstrates that the domains ‘object’ and ‘space’ play a central role in interrelating information for expression and creating cohesion in texts. Specifically, the grammatical and lexical structure of locative expressions in English and German reflect language-specific preferences in conceptualizing subject matter for expression given the same communicative task. Moreover, as suggested by the ‘thinking for speaking’ hypothesis’ (Slobin, 1991), the language-specific perspectives reflect the ‘habitual’ or ‘default’ means that native speakers of English and German regularly attend to in lexicalizing spatial descriptions. The gesture analysis supports the speech results and findings in Carroll et al. (2000) and provides additional evidence that speech-associated gestures contribute both to the semantic content as well as the structural properties of an utterance. Thus, speech and gesture function together as a co-expressive whole.

In addition, Carroll et al. (2000) demonstrate how the cross-linguistic comparison of event conceptualization has important implications for second language acquisition. That is, even though very advanced second language learners of German, whose native language is English, do not show any formal errors in speech, their spatial descriptions reveal that they have not shifted completely their perspective to the target language. That is, they have not yet acquired the overall principle in information organization which necessitates the use of proadverbials in reference maintenance. In other words, they exhibit what has been termed a ‘discourse accent’ (Gullberg, 2011). Moreover, Gullberg (2009) has also shown how gestures contribute important insight into event-conceptualization in a second language. Hence, the analysis of gestures could provide supplementary clues for future studies in order to investigate to what extent proadverbials are conceptualized in L2 German speakers.
After all, gestures are an important part of natural human language, especially in conversational interaction. Another area of research that might profit from insights into language-specific perspectives in organizing information and multimodal interaction is computer science. Great efforts have been undertaken so far to create interfaces that can process natural human language. Quek et al. (2002) note:

For human-computer interaction to approach the level of transparency of interhuman discourse, we need to understand the phenomenology of conversational interaction and the kinds of extractable features that can aid in its comprehension (Quek et al., 2002, p. 172). As the role of gestures is still sparsely explored in spatial descriptions, there is a lot of untouched ground for future studies to investigate. Taking the limitations of the design of the current study into account, the findings and observations made in this analysis hopefully have provided some considerations for future investigations of gestures and language-specific perspective-taking in spatial description tasks. As the role of gestures in spatial description tasks has not been investigated extensively in the past, it seems necessary for future investigations to elaborate in this area. In the current gesture analysis, only handshape and temporal alignment were considered in the coding of the data. It would be interesting in the future to investigate how gesture-viewpoint ('character-viewpoint' vs. 'observer-viewpoint', cf. McNeill, 1992) is used by speakers in spatial descriptions and how it might differ cross-linguistically in describing and locating objects. In addition, future studies might also extend to investigate typologically different languages, which also make use of abstract frames of references in spatial encodings of Figure and Ground.
References


Kelly, S. D., Barr, D. J., Breckinridge Church, R., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40(4), 577-592. ISSN: 0749596X.


Tutton, M. (2012). When and why the lexical Ground is a gestural Figure. Gesture, 12(3), 361-386. doi:10.1075/gest.12.3.04tut


Appendices

Appendix A. General background information for English participants

Table 25. Participant information (English).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age</th>
<th>Handedness</th>
<th>Education Level</th>
<th>Occupation</th>
<th>Country of Origin</th>
<th>Country of Residence</th>
<th>Length of stay abroad</th>
<th>Ø level of Swedish or Danish (0-5)</th>
<th>Other languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>32</td>
<td>R</td>
<td>MA</td>
<td>unemployed</td>
<td>USA</td>
<td>Sweden</td>
<td>8</td>
<td>Swedish level 3</td>
<td>Spanish</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>27</td>
<td>L</td>
<td>BA</td>
<td>MA Student</td>
<td>USA</td>
<td>Sweden</td>
<td>1.5</td>
<td>Swedish level 0</td>
<td>Chinese, Spanish</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>33</td>
<td>R</td>
<td>PhD</td>
<td>Post Doc Researcher</td>
<td>UK</td>
<td>Denmark</td>
<td>1.5</td>
<td>Danish level 0</td>
<td>French, Japanese</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>28</td>
<td>R</td>
<td>MA</td>
<td>PhD Student</td>
<td>Canada</td>
<td>Denmark</td>
<td>3.5</td>
<td>Danish level 3</td>
<td>German</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>29</td>
<td>R</td>
<td>MA</td>
<td>Project manager</td>
<td>UK</td>
<td>Sweden</td>
<td>3.5</td>
<td>Swedish level 4</td>
<td>Czech, German</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>32</td>
<td>R</td>
<td>PhD</td>
<td>Post Doc Researcher</td>
<td>Canada</td>
<td>Denmark</td>
<td>2</td>
<td>Danish level 2</td>
<td>French, Spanish</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>35</td>
<td>R</td>
<td>PhD</td>
<td>Post Doc Researcher</td>
<td>UK</td>
<td>Denmark</td>
<td>6</td>
<td>Danish level 1</td>
<td>French, German</td>
</tr>
</tbody>
</table>

F: female, M: male, L: left, R: right, BA: Bachelor, MA: Master
Appendix B. General background information for German participants

Table 26. Participant information (German)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age</th>
<th>Handedness</th>
<th>Education Level</th>
<th>Occupation</th>
<th>Country of Origin</th>
<th>Country of Residence</th>
<th>Length of stay abroad</th>
<th>Ø level of Swedish or Danish (0-5)</th>
<th>Other languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>27</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>3</td>
<td>Swedish level 0</td>
<td>English, French</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>30</td>
<td>R</td>
<td>MA</td>
<td>PhD Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>10</td>
<td>Swedish level 4</td>
<td>English, French</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>27</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>2</td>
<td>Swedish level 3</td>
<td>Arabic, English, Icelandic</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>25</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>2</td>
<td>Swedish level 0</td>
<td>English, French, Japanese</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>24</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>0.6</td>
<td>Swedish level 0</td>
<td>Chinese, English</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>24</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Denmark</td>
<td>1.5</td>
<td>Danish level 0</td>
<td>English, French, Spanish</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>24</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>0.6</td>
<td>Swedish level 1</td>
<td>English, Italian</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>26</td>
<td>R</td>
<td>BA</td>
<td>MA Student</td>
<td>Germany</td>
<td>Sweden</td>
<td>7.5</td>
<td>Swedish level 4</td>
<td>English</td>
</tr>
</tbody>
</table>

F: female, M: male, L: left, R: right, BA: Bachelor, MA: Master
Appendix C. Stimulus material

Figure 9. House in rural landscape (practice picture).
Figure 10. City center (picture 1).

Figure 11. House in changing rural landscape (picture 2).
Appendix D. Language history questionnaire (in English)

Language History Questionnaire

Sbj: Date:

Below are questions about your education, profession, and language use. Please answer these questions as completely as possible.

General Background:

1. Age: Years

2. Sex (Circle one): Male/ Female

3. What is your level of education (high school, university degree (BA or MA), PhD, professional training):

4. What is your profession (e.g., student, lawyer):

5. Where were you born?
   a. Have you only lived in the county of your birth? (Circle one): YES/ NO
   b. If no, where else have you lived?

6. Where do you currently live?

7. If your country of origin and country of residence are different, then how long have you been in the country of your current residence?
Language History:

8. What is/are your native language(s)?

9. Please list any other languages that you know below. For each, rate how well you can use the languages on the following scale:

Not Good 1 2 3 4 5 Very Good

<table>
<thead>
<tr>
<th>Language</th>
<th>Speaking</th>
<th>Listening</th>
<th>Writing</th>
<th>Reading</th>
<th>Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. For each of the languages listed, which do you use on a daily basis for the following activities and for how many hours per day?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Language(s)</th>
<th>Hours/ day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading (newspaper, book, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading (textbooks, journal articles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lectures/Seminars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watching TV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening to radio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to family members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to friends</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to work colleagues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email/Whatsapp/Fb messenger/Twitter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browsing the internet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other: (please specify)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
11. For each of the languages listed, which do you use on a **regular basis** for the following activities and for how many hours per month?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Language(s)</th>
<th>Hours/ day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading (newspaper, book, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading (textbooks, journal articles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lectures/Seminars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watching TV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening to radio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to family members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to friends</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to work colleagues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email/Whatsapp/Fb messenger/Twitter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browsing the internet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other: (please specify)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12. If you have any other comments about your language experience and language usage, please leave them below.
Appendix E. Language history questionnaire (in German)

Hintergrund Fragebogen

Teilnehmernummer: Datum:

Es folgen einige generelle Fragen über Ihren Bildungshintergrund, Beruf und sprachlichen Hintergrund. Bitte beantworten Sie die Fragen so genau und vollständig, wie möglich.

Genereller Hintergrund

1. Alter: Jahren
2. Geschlecht: Männlich Weiblich
3. Bildungsgrad (z.B. Realschulabschluss, Gymnasium, Berufsausbildung, Studium (BA, MA), Doktorat, etc.):

4. Beruf (z.B. Student, Anwalt, etc.):

5. Was ist Ihr Geburtsort?
   a. Haben Sie seit Ihrer Geburt an Ihrem Geburtsort gelebt? JA NEIN
   b. Wenn nein, wo haben Sie sonst noch gelebt?

6. Wo wohnen Sie zur Zeit?

7. Sollte sich Ihr Geburtsort von Ihrem jetzigen Wohnsitz unterscheiden, geben Sie bitte an wie lange Sie schon in Ihrem momentanen Wohnort leben.
Sprachlicher Hintergrund:

8. Welches ist/sind Ihre Muttersprache(n)?

9. Bitte geben Sie unten an welche andere Sprachen Sie sonst noch sprechen. Anhand der folgenden Skala geben Sie bitte an wie gut Sie jede Sprache sprechen:

<table>
<thead>
<tr>
<th>Sprache</th>
<th>Sprechen</th>
<th>Hören</th>
<th>Schreiben</th>
<th>Lesen</th>
<th>Grammatik</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. Welche der angegebenen Sprachen benützen Sie täglich? Bitte geben Sie an für welche Aktivitäten Sie die Sprache(n) benutzen und für wie lange?

<table>
<thead>
<tr>
<th>Aktivität</th>
<th>Sprache(n)</th>
<th>Stunden/Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesen (Zeitung, Buch, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesen (Fachliteratur)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vorlesungen/Seminare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernsehen gucken</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radio hören</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unterhalten mit Familie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unterhalten mit Freunden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unterhalten mit Arbeitskollegen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email/Whatsapp/Fb messenger/Twitter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet browse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderes: (bitte spezifizieren)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
11. Welche der angegebenen Sprachen benützen Sie **regelmässig**? Bitte geben Sie an für welche Aktivitäten Sie die Sprache(n) benutzen und für wie lange?

<table>
<thead>
<tr>
<th>Aktivität</th>
<th>Sprache(n)</th>
<th>Stunden/Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesen (Zeitung, Buch, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesen (Fachliteratur)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vorleseungen/Seminare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernsehen gucken</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radio hören</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unterhalten mit Familie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unterhalten mit Freunden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unterhalten mit Arbeitskollegen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email/Whatsapp/Fb messenger/Twitter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet browsen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderes: (bitte spezifizieren)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12. Falls Sie noch Ergänzungen zu Ihrem sprachlichen Hintergrund haben, bitte fügen Sie diese hier an.
Appendix F. Consent form (in English)

![CONSENT FORM]

Study picture descriptions

Consent Form
I hereby give my permission to Angela Schlatter, Centre for Languages and Literature, Lund University, Sweden, to use today’s recordings (audio and video) for the following purposes:

(Please tick the appropriate box, "☐", if you give your permission.)

☐ 1. analyses for scientific research;

2. as illustrations of the above scientific research in professional seminars, lectures, conferences, and in scientific publications:
   ☐ as still photographs;
   ☐ as video clips.

My anonymity is guaranteed. Under no circumstances will my personal identity be revealed to anybody other than the above mentioned scientific researcher (e.g. no names will be used in presentations of the recording).

Name  Signature  Date

Participant number: ____________________

Figure 12. Screenshot of consent form in English.
Appendix G. Consent form (in German)

Figure 13. Screenshot of consent form in German.
Appendix H. Instructions (in English)

Table 27. Task instructions for participant.

<table>
<thead>
<tr>
<th>Task (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>You will see two pictures. For each picture, try to memorize what is depicted in the picture. Take as much time as you need. In particular, pay attention to the objects on the wordlist. Afterwards,</td>
</tr>
<tr>
<td>(1) describe the scene depicted in the picture (without the wordlist) and</td>
</tr>
<tr>
<td>(2) specify the location of things so that the hearer can visualize how the picture looks like.</td>
</tr>
<tr>
<td>The wordlist should not limit you in your description. Describe as much as you remember. We will practice the whole task once, so that you can ask any clarification questions. During the experiment you cannot ask any questions anymore.</td>
</tr>
</tbody>
</table>

Table 28. Task instructions for confederate.

<table>
<thead>
<tr>
<th>Task (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Smalltalk questions:</td>
</tr>
<tr>
<td>1. What do you study/work?</td>
</tr>
<tr>
<td>2. Where are you from?</td>
</tr>
<tr>
<td>3. How do you like Lund/Sweden?</td>
</tr>
<tr>
<td>b) While the participant is describing the picture:</td>
</tr>
<tr>
<td>1. Keep your hands on your thighs</td>
</tr>
<tr>
<td>2. Never cross your arms in front of your body</td>
</tr>
<tr>
<td>3. Look like as if you are interested and as if you are carefully listening</td>
</tr>
<tr>
<td>4. Do not cross your legs</td>
</tr>
<tr>
<td>5. Smile in between, nod, or say ‘aha’</td>
</tr>
<tr>
<td>c) After every picture description you will get a selection catalog. Look through the catalog and write down the number of the picture described. Try to think a little longer at times, or sometimes take less time, so that is seems realistic.</td>
</tr>
</tbody>
</table>
**Task (C):**

Choose the right picture and write the corresponding number into the box.

**Picture 1:**

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
</table>

**Picture 2:**

|   |
Appendix I. Instructions (in German)

Table 29. Task instructions for participant.

<table>
<thead>
<tr>
<th>Aufgabe (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sie werden zwei Bilder sehen. Prägen Sie sich für jedes Bild ein, was auf dem Bild dargestellt ist. Nehmen Sie sich so viel Zeit, wie Sie brauchen. Achten Sie insbesondere auf die Objekte auf der Wörterliste.</td>
</tr>
<tr>
<td>(1) Beschreiben Sie anschließend die im Bild dargestellte Szene (ohne Wörterliste) und (2) geben Sie den Ort der beschriebenen Dinge an, damit der Zuhörer sich vorstellen kann wie das Bild ausschaut.</td>
</tr>
</tbody>
</table>

Table 30. Task instructions for confederate.

<table>
<thead>
<tr>
<th>Aufgabe (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Samlltalk Fragen:</td>
</tr>
<tr>
<td>1. Was studierst du?/Was arbeitest du?</td>
</tr>
<tr>
<td>2. Woher kommst du?</td>
</tr>
<tr>
<td>3. Wie gefällt es dir in Lund?</td>
</tr>
<tr>
<td>b) Während der Teilnehmer das Bild beschreibt:</td>
</tr>
<tr>
<td>1. Hände immer auf den Oberschenkeln</td>
</tr>
<tr>
<td>2. Arme niemals verschränken vor dem Körper</td>
</tr>
<tr>
<td>3. Interessiert/ aufmerksam gucken</td>
</tr>
<tr>
<td>4. Beine nicht übereinander schlagen</td>
</tr>
<tr>
<td>5. Zwischendurch lächeln, nicken, oder „aha“ sagen ist ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aufgabe (C):</th>
</tr>
</thead>
</table>

Wählen Sie das richtige Bild aus und schreiben Sie die entsprechende Nummer ins Kästchen rein.

<table>
<thead>
<tr>
<th>Bild 1:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Bild 2:</th>
</tr>
</thead>
</table>

107
Appendix J. Wordlists

Table 31. Wordlist for each picture (in English).

<table>
<thead>
<tr>
<th>Practice Picture</th>
<th>Picture 1</th>
<th>Picture 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>house</td>
<td>cars</td>
<td>house</td>
</tr>
<tr>
<td>tree</td>
<td>tram</td>
<td>tree</td>
</tr>
<tr>
<td>train line</td>
<td>bus stop</td>
<td>cars</td>
</tr>
<tr>
<td>village</td>
<td>fountain</td>
<td>factory</td>
</tr>
<tr>
<td>bridge</td>
<td>newsstand</td>
<td>excavator</td>
</tr>
<tr>
<td>creek</td>
<td>canal</td>
<td>silos</td>
</tr>
<tr>
<td>pond</td>
<td>restaurant</td>
<td>skyscraper</td>
</tr>
</tbody>
</table>

Table 32. Wordlist for each picture (in German)

<table>
<thead>
<tr>
<th>Übungsbild</th>
<th>Bild 1</th>
<th>Bild 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haus</td>
<td>Autos</td>
<td>Haus</td>
</tr>
<tr>
<td>Baum</td>
<td>Strassenbahn</td>
<td>Baum</td>
</tr>
<tr>
<td>Bahnlinie</td>
<td>Bushaltestelle</td>
<td>Autos</td>
</tr>
<tr>
<td>Dorf</td>
<td>Brunnen</td>
<td>Fabrik</td>
</tr>
<tr>
<td>Brücke</td>
<td>Zeitungsstand</td>
<td>Bagger</td>
</tr>
<tr>
<td>Bach</td>
<td>Kanal</td>
<td>Silos</td>
</tr>
<tr>
<td>Teich</td>
<td>Restaurant</td>
<td>Hochhaus</td>
</tr>
</tbody>
</table>
Appendix K. Selection pictures for practice run
Appendix L. Selection catalogue