The status of gesture in cognitive-functional models of grammar
The Status of Gesture in Cognitive-functional Models of Grammar

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geesteswetenschappen
op woensdag 18 januari 2017 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Kasper Ivo Kok

geboren te Amsterdam
promotor: prof.dr. A.J. Cienki
copromotoren: prof.dr. M. Hannay
                      prof.dr. J.L. Mackenzie
Acknowledgements

This dissertation owes its existence to the Netherlands Scientific Organization (NWO) and the initiative taken by the 'Commissie Cohen' to invest in research in the humanities.

My utmost thanks go out to my supervisor Alan Cienki. Alan, it has been a pleasure to have you on my side as an academic mentor. You have never ceased to nurture my development as a researcher, or to provide me with food for thought. I still remember how refreshing it was – after having been exposed almost exclusively to form-oriented linguistics – to meet someone who makes claims like 'the meaning of and is embodied' or 'language is a system that expands and contracts'. I have always taken great inspiration from your enthusiasm for new and old ways of thinking about language, for building bridges and breaking molds, and for embracing new theories and methods. Your faith in my capacities has been an important motivation throughout the course of the project. Thanks for guiding me by means of the right, subtle cues at the right time, and for understanding that not being told what to do is the way for me to get things done.

Mike and Lachlan, the same is true for you. Thank you for your trust in my way of working independently, while being available at any time for consultation. Mike, I am grateful for your role in determining the scope of the project. And for opening my eyes to the many responsibilities that come with linguistic formulation. Lachlan, you are to blame that when I close my eyes and think of a clause, I see a colorful, multilayered, semi-transparent object with a bunch of satellites (actual, metal ones) floating around it. Special thanks also go to Lourens de Vries, Sherman Wilcox, Cornelia Müller, Marc Swerts and Kees Hengeveld for their willingness to serve on the reading committee.

I have been fortunate to be surrounded by colleagues and fellow PhD researchers who always provided social support. Suwei, I am deeply impressed by your courage and by the way you have developed as a researcher over the years. Your input has been of great value to my work and your sharp questions have made me reconsider many of the core issues (as you probably know, admitting that directly is not always my
Yao, your curiosity and diligence are admirable, and I can only be confident that you have all it takes to make these years a success. Break a leg, both of you. Arjan, I’m happy to have converted you into the first classicist ever to include models of gesture in an analysis of the Ancient Greek preterite. Bertie, Kashy, Gea, Petra, Monique, Nel, Margreet, Hennie, I have good memories of co-teaching, co-organizing and co-drinking with you. I gratefully acknowledge the important contribution of Marco Last to the grant application. In addition, I am thankful to the management and the administrative staff of the Vrije Universiteit, and to those who keep the buildings in good shape.

The months I spent in Bielefeld were the most productive of all. I am greatly indebted to Kirsten Bergmann and Stefan Kopp for welcoming me in their lab and granting me access to the SaGA corpus. It was great to be part of the SCS group and to enjoy the enthusiasm, intellect and humor of its members.

I also owe gratitude to the FDG community. Kees Hengeveld, Hella Olbertz, Riccardo Giomi, Núria Alturo, thank you for being open to the angle that I brought in during the colloquium sessions. The same goes for the scholarly communities of gesture studies and cognitive linguistics. I am thankful to Cornelia Müller, Irene Mittelberg, and the entire MaMuD crowd for their feedback and for inspiring workshops. I would also like to thank my former academic mentors and colleagues, in particular Ben Bergen, for demonstrating that productivity and high quality research can go alongside taking things easy. And I shouldn’t forget to mention that some of the most helpful feedback on my work, which has certainly improved its quality, came from the anonymous reviewers of my journal articles.

The process of writing this dissertation would have been much harder without the unconditional support of my family. I have always taken great inspiration from my sister Silvia and her dedication to nursing, which I consider to be among the most respectable of professions. I am, of course, also thankful to the members of Oosterpark Zaterdag 1 and other friends for keeping me in good spirit. But, at the risk of sounding formaggioso, there’s only one person who deserves my final and foremost words of gratitude. Chiara, during all these years you have put up with my selfish determination to achieve more than the
necessary, my stubbornness, and my impenetrable bubbles of thought. Thank you 'thousand' for confronting me with myself and growing with me, not against me. Or, in the words of I.C.: *Lui conobbe lei e se stesso, perché in verità non s'era mai saputo.*
Contents

CHAPTER 1. INTRODUCTION ................................................................. 13

1.1 Background and context ............................................................. 13
  1.1.1 Language viewed from a functional and cognitive perspective .... 14
  1.1.2 Language viewed as a multimodal activity ......................... 15
1.2 Scope and objectives ................................................................. 17
1.3 Overview of the dissertation .................................................... 18
1.4 Journal and conference publications ........................................ 19

CHAPTER 2. COMPARATIVE EVALUATION OF THE MODELS .......... 21

2.1 Why FDG and CG ................................................................. 21
2.2 Functional Discourse Grammar ................................................. 23
  2.2.1 Background and starting points ........................................ 23
  2.2.2 Architecture and basic principles ..................................... 27
  2.2.3 Analytical tools .............................................................. 35
2.3 Cognitive Grammar ................................................................. 40
  2.3.1 Background and starting points ........................................ 40
  2.3.2 Basic principles .............................................................. 41
  2.3.3 Analytical tools .............................................................. 47
2.4 Points of convergence ............................................................... 53
  2.4.1 Departure from formalist assumptions ................................ 53
  2.4.2 Embedding a in theory of language use ............................... 54
  2.4.3 Psychological motivations ................................................ 55
  2.4.4 Universally applicable tools .............................................. 56
2.5 Points of divergence ................................................................. 56
  2.5.1 Moderate versus extreme departure from formalist models .... 56
  2.5.2 Language as a cognitive system versus language as a vehicle for communication ......................................................... 57
  2.5.3 Modularity vs. gradience .................................................. 58
  2.5.4 Patterns vs. processes ...................................................... 60
  2.5.5 Conceptual vs. semantic structure ..................................... 60
  2.5.6 Formal versus heuristic notations ...................................... 61
CHAPTER 3. SPEAKERS’ GESTURES IN FDG AND CG

3.1 Motivations for a multimodal approach
3.1.1 Systematicity
3.1.2 Language-specificity of gesture
3.1.3 Functional and structural interfaces between speech and gesture
3.2 Current approaches
3.2.1 Systemic Functional Grammar
3.2.2 Construction Grammar
3.2.3 Phrase structure-based models
3.2.4 Shortcomings of current models
3.3 Functional Discourse Grammar and gesture research
3.3.1 Gesture at the Interpersonal Level
3.3.2 Gesture at the Representational Level
3.3.3 Gesture and Encoding
3.3.4 Initial observations
3.3.5 A first sketch of a multi-channel FDG
3.4 Cognitive Grammar and gesture research
3.4.1 Points of convergence
3.4.2 Gestures as symbolic structures
3.4.3 Degrees of schematicity in gestural form and meaning
3.4.4 Gestures and CG’s dimensions of construal
3.4.5 Gestures and CG diagrams
3.5 Remaining challenges
3.5.1 (Re)defining grammatical relevance
3.5.2 Multifunctionality
3.5.3 Defining type-token relations
3.5.4 Temporal coordination as a dimension of linguistic structure
3.6 Summary

CHAPTER 4. MAPPING OUT THE MULTIFUNCTIONALITY OF SPEAKERS’ GESTURES

4.1 Background and motivations
4.1.1 Multifunctionality in language and grammar
4.1.2 Functional models applied to gesture
4.2 A crowdsourced approach to functional characterization
4.2.1 Corpus and stimuli
CHAPTER 4. GESTURAL MULTIFUNCTIONALITY .......................................................... 122
  4.2 Functions of interest ................................................................................. 122
  4.2.2 Functions of interest .......................................................................... 122
  4.2.3 Participants and filtering ...................................................................... 124
  4.2.4 Procedure .............................................................................................. 125
  4.2.5 Reliability of the sample and performance diagnostics ...................... 127
  4.2.6 Descriptive statistics ........................................................................... 129
  4.3 Mapping out gestural multifunctionality .................................................. 130
  4.3.1 The multifunctionality of a single gesture ........................................... 130
  4.3.2 Mapping out the gesture functionality space ...................................... 133
  4.3.3 A closer look at semantic multifunctionality ...................................... 135
  4.3.4 A closer look at semantic-pragmatic multifunctionality ...................... 137
  4.3.5 Degrees of functional prominence ...................................................... 138
  4.4 Discussion and conclusion ..................................................................... 140

CHAPTER 5. PERCEPTION OF GESTURAL FUNCTIONS WITH AND WITHOUT SPEECH .............................................................. 143

  5.1 Previous research on the comprehension of gestures without audio ....... 144
  5.2 Methods and participants ...................................................................... 147
  5.3 Results ...................................................................................................... 148
  5.3.1 The object reference function .............................................................. 148
  5.3.2 The shape representation function ...................................................... 152
  5.3.3 The movement representation function ............................................ 155
  5.3.4 The meta-communicative function ..................................................... 158
  5.4 A glance at the remaining functions ....................................................... 161
  5.5 Summary and conclusion ...................................................................... 163

CHAPTER 6. MAPPING FUNCTIONS TO FORMS .................................................. 165

  6.1 Form annotation ...................................................................................... 167
  6.1.1 Handedness ......................................................................................... 167
  6.1.2 Handshape ........................................................................................... 169
  6.1.3 Orientation ........................................................................................... 171
  6.1.4 Hand Position ....................................................................................... 172
  6.1.5 Movement type ..................................................................................... 174
  6.1.6 Movement Direction ........................................................................... 175
  6.1.7 Temporal connectedness ...................................................................... 177
  6.2 Function-form relations .......................................................................... 178
  6.2.1 Formal correlates of shape depiction ................................................ 180
  6.2.2 Formal correlates of object reference ............................................... 182
  6.2.3 Formal correlates of place reference ................................................ 185
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.4 Formal correlates of process depiction</td>
<td>187</td>
</tr>
<tr>
<td>6.2.5 Formal correlates of quantity signaling</td>
<td>190</td>
</tr>
<tr>
<td>6.2.6 Formal correlates of uncertainty display</td>
<td>193</td>
</tr>
<tr>
<td>6.2.7 Formal correlates of prominence signaling</td>
<td>195</td>
</tr>
<tr>
<td>6.2.8 Beyond isolated gestures</td>
<td>197</td>
</tr>
<tr>
<td>6.3 Discussion: from (weak) statistical trends to linguistic primitives</td>
<td>199</td>
</tr>
<tr>
<td>6.4 A note on individual differences</td>
<td>204</td>
</tr>
<tr>
<td>6.5 Summary and conclusion</td>
<td>206</td>
</tr>
<tr>
<td>CHAPTER 7. GESTURE-ACCOMPANIED WORDS AND CONSTRUCTIONS</td>
<td>209</td>
</tr>
<tr>
<td>7.1 Gesture-accompanied linguistic structures</td>
<td>209</td>
</tr>
<tr>
<td>7.1.1 Methods</td>
<td>211</td>
</tr>
<tr>
<td>7.1.2 Lemma-level analysis</td>
<td>213</td>
</tr>
<tr>
<td>7.1.3 POS-level analysis</td>
<td>223</td>
</tr>
<tr>
<td>7.1.4 Bigram-level analysis</td>
<td>228</td>
</tr>
<tr>
<td>7.2 Temporal dynamics</td>
<td>231</td>
</tr>
<tr>
<td>7.2.1 The effect of the choice of time window</td>
<td>233</td>
</tr>
<tr>
<td>7.2.2 Relative timing between words and strokes</td>
<td>238</td>
</tr>
<tr>
<td>7.3 Summary and conclusion</td>
<td>244</td>
</tr>
<tr>
<td>CHAPTER 8. MULTIMODAL GRAMMARS AT WORK</td>
<td>249</td>
</tr>
<tr>
<td>8.1 Introduction</td>
<td>249</td>
</tr>
<tr>
<td>8.2 Example analyses</td>
<td>251</td>
</tr>
<tr>
<td>8.2.1 Shape tracing for attribution</td>
<td>251</td>
</tr>
<tr>
<td>8.2.2 Shape tracing for predication</td>
<td>259</td>
</tr>
<tr>
<td>8.2.3 Repeated pointing in combination with spatial description</td>
<td>263</td>
</tr>
<tr>
<td>8.2.4 Relative movement tracing</td>
<td>272</td>
</tr>
<tr>
<td>8.2.5 Quantity specification and approximation</td>
<td>279</td>
</tr>
<tr>
<td>8.2.6 Gestural focus allocation</td>
<td>285</td>
</tr>
<tr>
<td>8.2.7 Temporal scenario construction</td>
<td>290</td>
</tr>
<tr>
<td>8.2.8 Meta-communicative signaling</td>
<td>295</td>
</tr>
<tr>
<td>8.3 Discussion: merits and drawbacks of FDG and CG</td>
<td>300</td>
</tr>
<tr>
<td>8.3.1 Strengths of FDG over CG</td>
<td>301</td>
</tr>
<tr>
<td>8.3.2 Strengths of CG over FDG</td>
<td>302</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1 Background and context
Language is immensely multifaceted. Its complexity is apparent in the heterogeneity of linguistics as a field of study, comprising many different schools of thought, disciplines and methodologies. Each branch of linguistic theory has reduced one or more dimensions of language to a model-sized object of inquiry. Some of these models focus exclusively on formal structures, others on the relations between linguistic elements and aspects of communication and cognition. The units of analysis range from written to spoken expressions, and from individual sounds to words, clauses, and longer stretches of discourse. However, in practice, these different aspects of language have not always gained balanced attention.

As discussed below, two strong biases can be identified in traditional models of language: a bias toward form as opposed to function, and a bias toward written as opposed to spoken language.¹ The traditional hegemony of the theory of generative syntax has moved other facets of linguistic expression to the background. However, in the second half of the previous century, new research paradigms were developed, aimed at reviving interest in aspects of language that had traditionally remained out of focus. This dissertation is concerned with

¹ By ‘traditional models’, I refer to the theories that dominated language research in the early and mid-twentieth century in North America and Europe, in particular variants of Chomsky’s generative grammar (e.g. Chomsky 1957, 1965).
bridging the outcomes of two of the most significant paradigm shifts: a transition from form-only models to those that study linguistic form and function in an integrated fashion, and a transition from a unimodal to a multimodal conception of language.

1.1.1 Language viewed from a functional and cognitive perspective

Formal models of grammar, which have long dominated linguistic theory, have taken sentence structure as the essential characteristic of human language (e.g. Chomsky 1957, 1965). The prevailing assumption in this paradigm is that the formal aspects of language (syntax in particular) can be studied in their own right, independent of their role in language use. That is, the subject matter in this tradition is the linguistic knowledge (competence) of individual speakers, not the way this knowledge is instantiated in actual communication.

In more recent decades, form-only approaches have been criticized for disregarding the functional constraints that are imposed on human language. Broadly speaking, such constraints can be assumed to come from two directions. On the one hand, language is a system that is used to communicate between people in purposeful ways. As Tomasello (1998: xiv) puts it, “[t]he overall function of language is communication in the sense that language evolved for purposes of communication phylogenetically, and it is learned for purposes of communication ontogenetically.” On the other hand, language is acquired and processed by the human brain. A close connection can therefore be assumed to exist between linguistic structure and cognitive processes such as categorization, association, and attention.

These two sets of constraints on linguistic structure are central in two, partially overlapping fields of inquiry: cognitive and functional linguistics.² Functional linguistic frameworks aim to study linguistic

---

² Whereas the term ‘functional linguistics’ is sometimes used as encompassing all non-formalist models, including cognitive approaches, I here use it to refer to a specific set of discourse-oriented models, which will be further characterized in Chapter 2. As I discuss in later chapters, the boundary between functional and cognitive approaches is not very sharp; most current cognitive and functional theories accept both cognitive and communicative principles as determinants of linguistic structure. The difference is mostly one of focus.
form with recourse to the functions that language serves in contextualized interaction. That is, they aim to bring the structures and communicative functions of linguistic elements into correlation with each other. Cognitive linguistic approaches to grammar are generally compatible with this aspiration, but their focus is on explicating the ways in which linguistic structures reflect cognitive processes. Taking principles from cognitive psychology as the basis, all formal levels of language are analyzed as expressing the way that language users ‘construct’ meaning.

Both functional and cognitive models of grammar have moved away from the view of language as being like a mathematical system, towards a view of language as a cognitively motivated, dynamic activity that occurs in social interaction. However, neither strand of linguistic theory has fully distanced itself from another salient bias in formal linguistics: a predominant focus on written language.

1.1.2 Language viewed as a multimodal activity
An influential dogma in traditional linguistic theory is that written text provides an adequate model for any other form of language, including spoken expression. A number of critiques on this position were expressed in recent decades (Brazil 1995; Linell 1998; McCarthy & Carter 2006). McCarthy and Carter (2006), for instance, demonstrate that dialogical turns in face-to-face conversation often involve linguistic structures that are not accepted as grammatically correct in classical models (e.g. incomplete clauses). They argue that such elements of spoken language should no longer be ignored in linguistic theory: “speakers and hearers do not ignore them – they carry a sizeable share of the communicative load” (McCarthy & Carter 2006: 30 after Hockett 1986). On this account, theories of written grammar – in particular those concerned with constituent structure – cannot be seamlessly extended to spoken forms of language.3

Remarkably, the discussion on whether spoken grammar should be divorced from written grammar has largely overlooked one of the most distinguishing characteristics of spoken language: its potential to

---

3 See Biber et al. (2002) for a milder account.
be accompanied by meaningful bodily gestures. Proponents of either position seem to have followed the traditional assumption that gestures are to be set apart as matters of linguistic 'performance', residing outside the purview of linguistics (Cienki in press). Undoubtedly, this bias at least partially results from the relative convenience of gathering and dealing with written data as compared to video-recorded material. However, along with the development of interaction-oriented models of linguistic structure and increased technological possibilities, acknowledgment of the linguistic relevance of bodily expression is rapidly increasing. Accumulating evidence suggests that from the early stages of linguistic development, children communicate through multiple semiotic channels simultaneously (e.g. speaking and pointing; Andrén 2010; Tomasello et al. 1984). The same holds for face-to-face communication by adults, where language use can involve not just sound but also various types of manual and facial behaviors. Such gestures can serve important roles in the production and comprehension of linguistic utterances, contributing to the semantics as well as the pragmatics (Kendon 1994, 2004; McNeill 1992; Streeck 2009). In this light, the difference between the linguistic and the paralinguistic proper is not as straightforward as previously supposed. A truly adequate model of language use, therefore, should not see verbal language as a complete and independent semiotic system, but rather as part of a multimodal repertoire for communication (Cienki 2012).

There is an increasing interest in the question of how models of grammar can do better justice to the realities of spoken language and its multimodal nature. However, it is still a significant challenge for linguistic theories to model the semantic and pragmatic import of co-verbal behaviors and their interaction with speech. So far, most attempts to reconcile research on grammar and gesture have remained somewhat programmatic (e.g. Muntigl 2004); the incorporation of gestures in grammatical theory has rarely been elaborated in depth in the context of specific theories of grammar. In this thesis, I aim to fill this gap by assessing how cognitive-functional models of grammar can accommodate speakers’ gestures as part of their analytical framework.
1.2 **Scope and objectives**

The overall objective of this dissertation is to assess the potential status of gesture in cognitive and functional models of language structure. In its broadest form, the research question can be formulated as: What properties should a model of grammar have that is maximally informative of how speech and gesture interrelate during linguistic communication? To address this question, I evaluate how current cognitive-functional linguistics models of grammar can be developed to incorporate a view of language as multimodal, and which aspects of these models are most valuable for this purpose.

To allow for in-depth investigation, some concessions have been made with respect to the scope of the dissertation. First, the focus is on two specific models, which will be studied in depth. The theory of Cognitive Grammar (henceforth CG; Langacker 1987, 1991, 2008a) provides a focal point as the most thoroughly developed cognitive linguistic framework of language structure. Functional Discourse Grammar (henceforth FDG; Hengeveld & Mackenzie 2008) will be the focus as one of the most advanced frameworks representing the functional branch of grammatical theory. A second concession concerns the range of the co-verbal behaviors taken into account. The focus is primarily on manual gestures, which have been more extensively documented than other articulators, and are generally more versatile. Other bodily behaviors (shoulder shrugs, eye gaze, head nods) and intonation patterns are acknowledged as relevant and will be mentioned on occasion, but are not at the core of the endeavor.

Given this scope, three more specific objectives can be formulated. The first is to provide a theoretical assessment of the degree to which the principles underlying FDG and CG are compatible with the ways gestures have been studied in the literature. The second is to gain empirical insights into some of the issues and challenges that arise when attempting to define the role of gesture in these frameworks. The third is to apply FDG and CG to a range of functionally diverse spoken-gestured utterances, thus revealing the strengths and weaknesses of each of the models.
1.3 Overview of the dissertation

In accordance with the sub-goals formulated, the dissertation consists of three parts: a detailed theoretical assessment (Chapters 2 and 3), a number of quantitative studies that address some of the theoretical challenges (Chapters 4, 5, 6 and 7) and a set of micro-analyses (Chapter 8). More specifically, Chapter 2 motivates the choice of the theories of interest and introduces them to the extent relevant for current purposes. It offers a close comparison of FDG and CG in terms of their underlying assumptions and principles. This evaluation helps us to understand along which dimensions theories of grammar can be differentiated. Chapter 3 investigates, for both FDG and CG, how the categories and distinctions postulated relate to the expressive potential of manual gesture. On the basis of a review of the current gesture literature, a close comparison is given of the potential of each model for dealing with multimodality, and a number of critical challenges are raised. These challenges are investigated empirically in Chapters 4, 5, 6 and 7. Chapter 4 addresses the pervasive multifunctionality of gestural expression on the basis of an experimental perception study. It reveals to what extent gestures are perceived, by naïve attenders, as performing multiple linguistic functions simultaneously. Chapter 5 is concerned with the question of whether gestures receive the same functional interpretation when presented with or without audio. It extends the study reported on in Chapter 4 with a condition where the judges do not have access to the spoken component of the utterance. Chapter 6 investigates form-function mappings in gestural expression. It reports on a number of correlations between the functional perception ratings from Chapter 4 and the formal parameters of the gestures in the stimulus set. Chapter 7 takes a closer look at the linguistic side of the corpus from which the gesture stimuli were extracted. It asks whether the verbal structures that occur in the temporal vicinity of manual gestures are different from those found in the speech-only segments of the corpus. The results are discussed in the light of the possible linguistic functions that gestures are capable of performing. The second part of the chapter examines statistical patterns in the relative timing of spoken and gestured elements of expression. In combination, the theoretical and empirical parts of the dissertation fuel the assumptions
of Chapter 8. In this chapter, I apply FDG and CG to a diverse set of multimodal utterances. The analyses yield specific insights into the role of gesture in cognitive-functional models of grammar and put the representational tools of the models to the test. In addition, I discuss how each of the models could be modified to better accommodate gestural behaviors. Chapter 9 summarizes the main findings and discusses their implications for multimodal grammar as a field of study.

1.4 Journal and conference publications

Some of the research presented in this dissertation has been published in the articles listed below. Where applicable, this is indicated in footnotes at the beginning of the corresponding chapters.


Kok, K. I., Bergmann, K., & Kopp, S. (2015). Not so dependent after all: functional perception of speakers’ gestures with and without speech. In G. Ferré & M. Tutton (Eds.), *Proceedings of the fourth Gesture and Speech in Interaction conference* (pp. 167-172). Nantes, France. Available at https://hal.archives-ouvertes.fr/hal-01195646

In addition, the following, thematically related articles have been published or accepted for publication during the course of the PhD project.


Chapter 2. Comparative evaluation of the models

2.1 Why FDG and CG

The objective of evaluating the status of gesture in grammatical theory unavoidably raises the question: what is grammar? Lay definitions like 'the rules of language' or 'how language is structured' are not all that far removed from the way linguists think about grammar. However, it has proven rather difficult to define linguistic rules and structures in a way that is at the same time generalizable (applicable across different languages) and specific enough to capture the many subtleties of linguistic expression. As a result, there is a myriad of models of grammar on the market, with widely diverging assumptions as to what kinds of rules and structures exist, how they relate to each other and to other aspects of the human mind, and how they are best described. This dissertation will not pursue an evaluation of all of these models. Instead, it presents an in-depth inquiry into the grammar-gesture relationship, with the focus on two specific models: Functional Discourse Grammar (FDG; Hengeveld & Mackenzie 2006, 2008) and Cognitive Grammar (CG; Langacker 1987, 1991, 1999, 2008a).

The selection of FDG and CG as models in focus was guided by two main motivations. The first is the conviction that the cognitive-functional branch of linguistics provides the most fruitful starting point for understanding the relationship between language and gesture. Because speech and gesture show a great deal of overlap in terms of
their functionality, and less so in their form, purely form-oriented frameworks were discarded a priori as candidate models of interest. It may in principle be possible to apply such models to analyze gestural forms (see e.g. Fricke 2013), but form-only approaches generally fall short in detailing how speech and gesture are combined for communicative purposes. It is, moreover, questionable whether such models, in particular the theory of generative grammar, would accept gesture as part of language in the first place, or would regard it as a mere performance factor. For these reasons, the set of candidate models was limited to those that study linguistic structure and function in an integrated fashion.

The second motivation is that FDG and CG occupy somewhat opposite regions in what Gonzálvez-García and Butler (2014; 2006) call ‘functional-cognitive space’. That is, within the cognitive-functional strand of linguistics, they are quite far removed in terms of their assumptions, principles and goals. Given the aim of the dissertation to yield insights that can be generalized across cognitive-functional models, the relative distance between FDG and CG makes them appropriate as points of focus. Generally speaking, FDG can be characterized as a structural-functional model, whereas CG is a central example of a cognitive linguistic approach to grammar (for a comprehensive comparison of functional and cognitive linguistics, see Butler 2008a; Butler & Gonzálvez-García 2005; Nuyts 2004b, 2005, 2007). Because the divide between these sets of approaches is crucial to the content of the dissertation, I adopt the terms cognitive and functional as referring to two separate groups of theories, which will be further outlined below. Note that this use of these terms may deviate from other literature; it is not uncommon to find the term functional as an umbrella term for all non-formalist models.

The following sections introduce the basic motivations, theoretical assumptions and analytical tools of each of the two models in focus. The final sections of this chapter evaluate in what ways they are alike and different. To foreshadow some of the most important conclusions; there are salient differences between FDG and CG in terms of (1) the degree to which they depart from formalist assumptions, (2) how the functions of language are defined, (3) which macro-functions
Comparative evaluation of the models | 23

(symbolic vs. communicative) are prioritized in grammatical analysis, (4) the centrality of the notion of conceptualization, (5) the assumed degree of rigidity of the categories used to describe language, and (6) the nature of the notation systems. These points of divergence form the basis for the more detailed assessment of the potential of each of the models to incorporate speakers’ gestures, set out in the remainder of this dissertation.

2.2  Functional Discourse Grammar

2.2.1  Background and starting points

Functional Discourse Grammar (FDG; Hengeveld & Mackenzie 2006, 2008; Keizer 2015) is a modified version of Functional Grammar (Dik 1978, 1997). It was developed with the aim of solving some of the problems faced by Dik’s original model (Hengeveld 2004a). FDG is affiliated with Systemic Functional Grammar (Halliday 1994) and Role and Reference Grammar (Van Valin 1993), as well as more foundational work by Greenberg (1966) and Givón (1995). FDG presents itself as a ‘form-oriented function-to-form’ approach. Its ambition accords with that of predecessor Dik’s Functional Grammar:

Within the functional paradigm, one attempts to reveal the instrumentality of language with respect to what people do and achieve with it in social interaction. [...] [Language] must be studied within the framework of the rules, principles and strategies which govern its natural communicative use (Dik 1986: 11).

FDG aims to provide a motivated account of language structure in relation to language’s functions in situated communication. It considers a wide range of linguistic functions, largely motivated by foundational work in functional linguistics and typology (for taxonomies of the different functions of language in communication see Bühler 1934/1990; Halliday 1976; Jakobson 1960). To stay faithful to its ambitious scope, FDG aspires to be open to, and broadly compatible
with, other disciplines that are related to language processing and communication.

Notwithstanding its functional orientation, FDG also shares assumptions with traditional, formalist models of grammar. It adheres to the classical compartmentalization of linguistic levels of representation (pragmatics, semantics, morphosyntax and phonology), which is disputed in many cognitive linguistic models of grammar. Its concern with semantic and pragmatic factors, moreover, is strictly limited to the extent that these factors have an immediate impact on the grammatical structure of a given language. That is, following the ‘principle of formal encoding’, FDG’s analyses are restricted to those aspects of cognition and communication that are encoded in the syntax, morphology or phonology. Because some of FDG’s starting points are not all that far removed from generative linguistics, it has been described as “occupying a position halfway between functional and formal approaches to grammar” (Keizer 2015: 13). The rationale behind FDG can be further elaborated in terms of its dedication to being pragmatically, psychologically and typologically adequate.

2.2.1.1 Pragmatic adequacy

A central conviction held by FDG is that language structure is to be understood in relation to a broader pragmatic model of verbal interaction. Its dedication to pragmatic adequacy is sustained by various facets of the theory. Most saliently, FDG holds that language is structured to serve its dual functionality of being at once a system for organizing thought and performing social action. It allows for describing linguistic elements in both ‘actional’ and semantic terms. Noun phrases, for instance, are simultaneously seen as expressing (sub)acts of reference, i.e. a type of action performed by the speaker, and as denoting some referential content. These two, complementary perspectives are integral to the organization of the model, which contains a separate Interpersonal Level and Representational Level of analysis. The Interpersonal Level is seen as primary: the most basic unit of analysis is a communicative Move, i.e. a cohesive contribution to the discourse.

4 Morphology and syntax are studied as part of the same level of form.
A further characteristic of the theory that serves the pursuit of pragmatic adequacy is the postulation of a Contextual Component. Although context is not seen as part of the grammar itself, FDG acknowledges that linguistic structure is often contingent upon the communicative setting in which an utterance is produced. A final note with respect to pragmatic adequacy is that although FDG in principle covers a broad range of discursive and rhetorical relations, it is most commonly applied to analyze the structure of clause-level phenomena. Thus, despite what its name might suggest, it is not a grammar of the structure of the discourse. It merely aspires to be compatible with theories of rhetorical relations and pragmatic rules, but remains in essence a theory of linguistic structure.

2.2.1.2 Psychological adequacy
As an epistemological consequence of seeing language as a type of human activity, FDG is (broadly) concerned with the processes that govern the production of language. Its architecture, as discussed in Section 2.2.2 below, is designed to satisfy the assumption that "[a] model of grammar will be more effective the more its organization resembles language processing in the individual" (Hengeveld 2004b: 367). Accordingly, FDG's analytical tools take inspiration from psycholinguistic models of speech generation, in particular Levelt's (1989) stepwise intention-to-articulation model. However, following similar argumentation as with respect to pragmatic adequacy, FDG's relationship with psycholinguistics is merely one of compatibility. FDG itself should not be interpreted as a model of the speaker. Adopting a function-to-form architecture is merely thought to reveal "the sequence of steps that the analyst must take in understanding and laying bare the nature of a particular phenomenon" (Hengeveld & Mackenzie 2008: 2). More specifically, FDG's analytical architecture mirrors Levelt's model in that it works from function to form. A full analysis of an utterance first involves the relevant Formulation operations, i.e. the pragmatic and semantic aspects of the message that receive grammatical coding, and subsequently the Encoding operations, i.e. the morphosyntactic and phonological realization of the message. In principle, however, the model does not preclude a reversed interpretation, where it is used as a parser...
26 | Chapter 2

from form to function, mirroring language comprehension (Hengeveld & Mackenzie 2008: 2; Mackenzie 2012).

2.2.1.3 Typological adequacy
FDG’s third defining commitment is typological adequacy. Hengeveld & Mackenzie’s (2008) seminal book on the theory discusses linguistic examples from over 150 languages, which serve to motivate and evaluate its assumptions. The postulation of FDG’s hierarchical architecture, and the assumed set of linguistic primitives on each of the levels, were inspired by an extensive inquiry into the ways in which semantic and pragmatic elements of expression are coded in languages around the world. For these reasons, FDG claims to be “a theory that is capable of providing a framework for the enunciation and comparison of universals and of offering lines of explanation” (Hengeveld & Mackenzie 2008: 32). These universals are not fixed structures or rules, but typically take the form of implicational hierarchies (if a language has element A, then it also has B). As for the status of these universals, Hengeveld and Mackenzie remain rather agnostic: although it is argued that universal tendencies may point back to “aspects of the cognition that drives linguistic communication” (Hengeveld & Mackenzie 2008: 35), no strong position is taken as to whether these aspects of cognition are to be seen as innate biases.

2.2.1.4 Stratification
Being structure-oriented, FDG takes a strong interest in the hierarchical relations that exist among elements of language. One type of hierarchical organization, already mentioned above, holds between the four postulated levels of linguistic organization:

Interpersonal Level > Representational Level > Morphosyntactic Level > Phonological Level.

This hierarchy is cumulative: it is assumed that pragmatics governs semantics, that pragmatics and semantics together govern morphosyntax, and that phonology is governed by all three other levels. The separation between the first two layers, which is perhaps the most
Comparative evaluation of the models

Conspicuous (Butler 2008b), is considered relevant for distinguishing descriptive from interaction-oriented elements of communication. For instance, certain adverbs (e.g. honestly and really) can be used either descriptively, modifying the semantic content of an utterance (1), or as an expression of emphasis or commitment, pertaining directly to the linguistic action that is being carried out (2). In (1), the adverb is to be analyzed at the Representational Level, whereas in (2), it is seen as Interpersonal in nature.

(1) He answered the question honestly
(2) I've honestly never heard about it

A second type of hierarchical structure pertains to the organization of the elements within each of the levels. Based on typological evidence (scope phenomena and layer-specific morphological markers), a hierarchy of different units of analysis is assumed for all four levels. This makes it possible for the analyst to study the principles that map meaning onto form for any given language. To decide at which level or layer of the model an element of an utterance is to be analyzed, FDG makes extensive use of linguistic 'tests'. In (2), for instance, we can be sure in assigning honestly to the Interpersonal Level because it cannot felicitously be replaced by in an honest manner.

2.2.2 Architecture and basic principles

A fully detailed description of FDG's architecture remains outside the scope of the dissertation. The current section gives a brief overview of the architecture of the model, only exemplifying aspects of it that are relevant to later chapters. A visual representation of the architecture of FDG is given in Figure 2.1(adopted from Hengeveld & Mackenzie 2006: 369).

FDG's architecture embodies the assumption that grammar needs to be studied as situated in a larger model of linguistic interaction. We see that the Grammatical Component – the central part of the model – is surrounded by three 'extra-linguistic' components: the Conceptual Component, the Contextual Component and the Output Component. The Conceptual Component is thought of as containing a pre-linguistic
representation of the communicative intention and the corresponding conceptual content (cf. Levelt’s ‘conceptualizer’). This representation is not yet tailored to any kind of linguistic constraints. As paraphrased by Hengeveld & Mackenzie (2006: 386), it may for instance contain “the desire to impart some bad news to the Addressee and concurrently to show sympathy.” The Conceptual Component feeds into the Grammatical Component, which is responsible for encoding this communicative intention into a linguistically expressible form.

The Contextual Component represents those aspects of the situational or discourse context that have direct influence on the form of an utterance. This may include the sex of the interlocutors, previous discourse referents and the location of referenced objects in the immediate environment (e.g. proximal vs. distal). Detailed accounts of the contextual component, as well as discussions of its relation to the Conceptual Component, are given by Connolly (2009) and several contributions in Alturo et al. (2014). The Output component – the lower end of the pipeline – represents all ‘extra-linguistic’ features of linguistic expression. These are, in FDG terms, aspects of speech that do not map directly onto a digital aspect of pragmatics or semantics, such as the speech-rate and tone of voice.

In addition to its external structure, i.e. its embedding relative to the other components, the Grammatical Component has an internal structure. FDG distinguishes four levels of analysis, as shown in Figure 2.1. The Interpersonal and Representational Levels together can be seen as roughly covering Slobin’s (1996) ‘thinking for speaking’: on these levels, the output of the conceptual component is transformed into a representation that conforms to the semantic and pragmatic categories and structures available in a given language. The Interpersonal Level deals with the pragmatic-actional dimensions of language use, i.e. those that pertain to the role of an utterance in the interaction between the speakers. This includes the speaker’s communicative intention and assumptions about the common ground, as well as the overall structuring of the discourse (note that the term Interpersonal is used in a broader sense than in Halliday’s Systemic Functional Grammar; it also encompasses Halliday’s textual metafunction). The Representational Level is seen as purely semantic, covering the structure of the (real or
fictive) situation that an utterance refers to. The remaining two strata – the Morphosyntactic and Phonological levels – are responsible for the formal aspects of the utterance. Because the focus of this dissertation is mainly on the Formulation levels and their relation to gestural expression, the Encoding levels do not receive detailed consideration. The following subsections discuss the organization of the Interpersonal and Representational Levels in more detail.
2.2.2.1 The Interpersonal Level

The organization of the Interpersonal Level can be represented as in Figure 2.2 (adopted from Keizer 2015: 32).

![Figure 2.2 The organization of the Interpersonal Level](image)

The largest unit of analysis is the Move (M), defined as “an autonomous contribution to the interaction” (Hengeveld & Mackenzie 2008: 50). Moves can, for instance, take the form of a conversational turn or of a paragraph in written text, provided it makes a unified contribution towards a discursive goal. In English, their initiation can be marked by expressions such as anyway, or to cut a long story short. The communicative units that comprise Moves are called Discourse Acts (A). These commonly correlate with individual clauses (e.g. the utterance Monday will be a gloomy day. Tuesday we might see the sun appear. consists of two Discourse Acts). However, a Discourse Act can also be smaller than a clause (e.g. Yikes!) or larger (when two bits of information are presented as a coherent whole, e.g. I went shopping and bought four hats). Relations between Discourse Acts, such as concession (X, but Y) or motivation (because X, Y) are represented on this layer.

Discourse Acts are further characterized by an Illocution (F), Speech Participants (P₁, P₂), and a Communicated Content (C). The Illocution represents “the conventionalized interpersonal use [of a Discourse Act] in achieving a communicative intention” (Hengeveld & Mackenzie 2008: 68). In English, Illocutions that receive grammatical marking are Declarative, Interrogative and Imperative (see Hengeveld & Mackenzie 2008: 75 for a list of various other illocutions that are
grammatically marked in languages across the world). Interactives (e.g. Hey!) and verbs that characterize performatives (e.g. I hereby pronounce you husband and wife) are also analyzed at the layer of the Illocution. The speech Participant variables P1 and P2 specify elements of an utterance that directly express the speaker’s and addressee’s role in the illocution (e.g. I hereby pronounce you ...).

The Communicated Content (C) can be thought of as the actional dimension of the semantic load of an utterance. That is, it is not the semantic representation itself, but rather “the totality of what the speaker wishes to evoke in his/her communication with the addressee” (Hengeveld & Mackenzie 2006: 376). Ways of indicating the newsworthiness of the information expressed are analyzed at this level (e.g. markers of focus vs. background), as are aspects of the speaker’s subjective attitude towards the information (e.g. unfortunately, I'm sick). The Communicated Content comprises Subacts of Reference (R) and Subacts of Ascription (T). The former correspond to the act of evoking an entity, for instance by means of uttering a noun phrase or pronoun. The latter correspond to the evocation of a predicate, for instance by uttering a verb or adjective. As with the Communicated Content, it is the very action of referring and ascribing that is represented here, not the content itself. The layer of Referential Subacts is the locus of linguistic markers of the presumed identifiability of the entity evoked (e.g. the bike vs. a bike). Examples of linguistic elements relevant to the layer of Ascriptive Subacts are morphemes that express the speaker’s commitment to the accuracy of the ascription (e.g. yellowish) as well as emphatic elements (really expensive).

2.2.2.2 The Representational Level
Whereas all utterances can be analyzed at the Interpersonal Level, the Representational Level only comes into play when an expression contains a Communicative Content. This holds for all utterances that contain descriptive elements, but excludes exclamatives such as ouch!. The structure of the Representational Level, as represented by Keizer (2015: 35), is shown in Figure 2.3.
Chapter 2

The Representational Level is concerned with semantics. Its role is to account for “all the formal aspects of a linguistic unit that reflect its role in establishing a relationship with the real or imagined world it describes” (Hengeveld & Mackenzie 2006: 673). The representational layers are organized according to the different types of semantic entities that they denote. In its definition of semantic categories, FDG appeals to the semantic theory of Lyons (1977), supplemented where relevant with additional categories.

The highest semantic layer, the Proposition (p), corresponds to Lyons’ third-order entities. These are “mental constructs, such as pieces of knowledge, beliefs, and hopes” (Hengeveld & Mackenzie 2006: 377). Propositions may take the form of a single phrase (that idea is fallacious), a full clause (Jody dislikes cockroaches), or a responsive particle (yes). Linguistic structures that express propositional attitude (Jody might dislike cockroaches) or reveal a proposition’s evidential source (as I’ve noticed, Jody dislikes cockroaches) are analyzed as operating on this layer. A Propositional Content may include one or more Episodes (ep). These are sets of situations (i.e., States of Affairs, see below) that are temporally and/or spatially connected and can, as a whole, be located in absolute time. The postulation of the Episode as a layer of analysis is mainly motivated by the existence of languages that have specialized markers on all verbs in successive clauses that describe temporally or spatially connected situations (Hengeveld & Mackenzie

![Figure 2.3 The organization of the representational level, adopted from Keizer (2015: 35)]
Comparative evaluation of the models | 33

2008: 157). In English and many other languages, tense markers can be attributed to the layer of the Episode, as well as expressions that locate (series of) events in absolute time (e.g. two years ago, I carried out some fieldwork and wrote a series of papers on beavers and their dietary restrictions).

States of Affairs (Lyons’ second order entities) are the individual events or situations that comprise Episodes. They can be located in space and in relative time and they can be evaluated in terms of their reality status. States of affairs often correspond to individual clauses (e.g. the car crashed into the wall; Jody is short), but they may also be expressed by a single lexeme (e.g. accident). Linguistic elements relevant to this layer are those that specify the place of occurrence (e.g. here, the car crashed), relative time (e.g. then), frequency of occurrence (e.g. usually), duration (e.g. slowly), reality status (e.g. maybe) or cause (e.g. because of the snow). Grammatical aspect is also analyzed at this layer in Keizer’s (2015) recent textbook on FDG. This dissertation follows Keizer’s version of FDG, which is simplified relative to Hengeveld & Mackenzie’s (2008: 181-215) work, where aspect and modality are analyzed in terms of an additional layer.

States of Affairs are typically composed of lower level entities, in particular Individuals and Lexical Properties. These are the linguistic items that contribute the lexical content to the expression’s meaning. Individuals, which correspond to Lyons’ first-order entities, are concrete, tangible entities (Hengeveld & Mackenzie 2008: 236). In some cases, Individuals are evoked without lexical description (e.g. This is a nice bike. Or do you prefer that one?). More commonly, however, they are specified by one or more Lexical Properties. A simple noun phrase like the dog, for instance, is analyzed as denoting an Individual that is characterized by the Lexical Property dog. Lexical Properties may also characterize relations between two or more Individuals, as in the dog attacks the cat. Moreover, Individuals and Lexical Properties can be combined into more complex descriptions. For instance, in the big dog, the designated Individual is analyzed as modified by the nominal Property dog, which is in turn modified by the adjectival Property big. FDG’s notational formalisms for analyzing such configurations of semantic units are explained in Section 2.2.3.
In addition to Individuals and Lexical Properties, FDG recognizes the semantic variables Location (l) and Time (t), Manner (m) and Quantity (q). These are distinguished as separate units, because they receive special grammatical treatment in some languages. In English, for instance, the choice of relative pronoun can reveal whether an entity is conceived of as an Individual or a Location (e.g. compare the building that I like and the building where I live). As with other variables, these categories are only given consideration in FDG if they receive specialized grammatical expression in the language of interest.

2.2.2.3 Morphosyntactic and Phonological Levels
Because the Morphosyntactic and Phonological Levels are of little interest to the content of this dissertation, the discussion of these levels will not go beyond a brief sketch. Unlike the levels discussed above, the morphosyntactic and phonological levels belong to the Encoding component of the model. They are seen as subservient to the Interpersonal and Representational Levels; their task is to create a formal representation of the output of the preceding levels. The Morphosyntactic Level deals with all aspects of the linear structure of an utterance. As its name suggests, morphology and syntax are not treated as fundamentally distinct levels of patterning. The Phonological Level converts the input of the Morphosyntactic Level into a form that can receive expression through speech, writing or signing. It includes a repository of morphemes, combinatory templates, and intonation patterns (to the extent that these are conventionally associated with Interpersonal or Representational units, e.g. clause-final rising pitch for interrogative illocution in English). Like the other levels, the Morphosyntactic Level and Phonological Level are hierarchically organized. The details of their organization remain outside the scope of this dissertation.

Some notes are in order with regard to the interaction between the different levels. The main reason for FDG to postulate four levels as separate strata of analysis is that no straightforward one-to-one mappings exist between them that apply across languages. Each language has its own ways of projecting aspects of function onto form, and even within languages it is often difficult to formulate clear rules.
This lack of overall ‘transparency’ between the different layers can be seen as a natural outcome of the kinds of constraints that languages are shaped by. With respect to syntax, for instance, Anstey (2002: 5) notes that “syntax reflects semantic distinctions, albeit imperfectly. Imperfection is to be expected since syntax is intermediate between semantics (non-linear and layered) and phonetics (linear and non-layered).” Nonetheless, some basic relations between the different levels can be noticed. For instance, a general correspondence exists between the order of the units as they appear in the Formulation layers and the order in which they are encoded and articulated.

### 2.2.2.4 Primitives and Functions
Finally, FDG postulates a set of ‘primitives’ of linguistic knowledge for each of the four levels, from which grammatical representations are constructed. As shown in Figure 2.1 (the boxes on the left-hand side), the primitives relevant to the Formulation layers are Frames, Lexemes, and Interpersonal and Representational Operators. Frames determine the number of participants in a predication. Lexemes and operators are lexical and grammatical means, respectively, of characterizing Interpersonal and Representational variables. In addition, most variables have a slot for Functions, which define relations between units at the same layer. Within the layer of the Communicated Content, for instance, there are specific Functions for Topic-Comment (pertaining to the communicative point of departure for the utterance) and Focus-Background (marking elements of the Communicated Content as new or significant in the discourse). On the layer of the Individual, Functions serve to indicate the semantic roles of the participants of the predication, such as Actor (A), Undergoer (U) or Location (L) (see Keizer 2015: 133 for an overview of how these relate to more specific semantic roles). Functions relevant at other layers are not discussed here, as they are of little relevance for the analyses presented in the upcoming chapters.

### 2.2.3 Analytical tools
The formalisms that FDG employs are inspired by predicate logic. The version of the formalisms used in this dissertation roughly follows the
simplified notation introduced by Keizer (2015). It involves a number of minor amendments to Hengeveld & Mackenzie's original notation system. Like Keizer, I replace closing variables by superscripts and use numbers (1, 2, 3,) instead of letters (i, j, k) for variable indexing. In addition, I emphasize hierarchical relations with indents, as shown in (5) and beyond. This serves to promote the accessibility of the formalisms in general, and can be especially insightful when analyzing on what the level(s) of linguistic structure co-speech gestures play a role (see Chapter 8). In its most general form, the structure of each layer is:

\[(\pi \alpha; [\text{head}]; \sigma \upsilon_\phi)\]

where:

- $\alpha$ is the name of the layer/variable
- [head] is the core element of the variable
- $\sigma$ is the slot for modifiers.
- $\pi$ is the slot for operators
- $\varphi$ is the slot for functions.
- the subscript in position $X$ (1, 2, 3 ...) designates individual variable tokens.

To illustrate the use of FDG’s notation conventions, I apply them to a number of simple, invented examples (4)-(12). First, I focus on the Representational Level.

(4) They

(5) Vase

(6) Tall vase

\[(mx_1)\]

\[(1x_1:\ (f_1; \text{vase}) \ ^x)\]

\[(1x_1:\ (f_1; \text{vase})\ :\ (f_2; \text{tall}) \ ^x)\]
Comparative evaluation of the models | 37

(7) The vase is tall

\( e_1: [ \)
\( f_1: \text{tall} \)
\( 1x_1: \)
\( f_2: \text{vase} \)
\( x_0 \)
\( ]^c \)

(8) He drops the vase

\( e_2: [ \)
\( f_1: \text{drop} \)
\( 1Mx_j_A \)
\( 1x_2: \)
\( f_2: \text{vase} \)
\( x_0 \)
\( ]^c \)

When an Individual is evoked through a non-descriptive element (e.g. they), the head is analyzed as absent; the modifier slot \( \sigma \) remains empty (4).\(^5\) Since the pronoun is grammatically marked for number, the plurality operator \( m \) is inserted in the operator slot. Where an Individual is specified by a simple noun (5), the representation contains an Individual that is modified by a Lexical Property. The operator 1 indicates that the designated Individual is singular in number. In the analysis of an adjective-noun pair like tall vase, the Property vase is itself modified by the Property tall. This semantic embedding is represented in (6).

In (7), we see the analysis of a copular clause that contains the same lexical elements as the noun phrase in (6). The clause as a whole denotes a State of Affairs (\( e_1 \)). Because the verb to be is seen as non-descriptive, it is not represented at the Representational Level. As seen in (7), the State of Affairs contains a predication frame that combines a one-place Property (tall) and an Individual (the vase). The semantic role of the individual \( x_1 \) is indicated by the function \( U \) (Undergoer). In the analysis of he drops the vase in (8), the State of Affairs is characterized by

\(^5\) In addition to empty heads (variables with no lexical modifiers), FDG recognized absent heads, for instance in the case of indefinite pronouns.
a two-place predication frame, which contains a Lexical Property (*drop*) and two Individuals (*he, the vase*). The analysis is in essence similar to (7), but it contains a verbal instead of adjectival Property, and an additional argument. Note that the assignment of the semantic role Actor does not rest on the assumption that the dropping of the vase is volitional; the macro-role Actor also includes the semantic role Force, which corresponds to the non-voluntary instigation of an action (Keizer 2015: 133). Because the first participant (*he*) is not characterized lexically, the head of the x-variable remains empty. The grammatical characterization of this participant is represented by operators for gender (M, masculine) and number (1, singular).

The notational principles illustrated in (4)-(8) apply to the Interpersonal Level as well. However, some conventions pertaining to the Interpersonal Level require further introduction. To illustrate these, I briefly discuss the representation of the sentence in (8) (*He drops the vase*) on the Interpersonal Level.

\[(9) \quad (M_1: \[
(A_1: \[
(F_1: \text{DECL})
(P_1)
(P_2)
(C_1)
]\{A_1\})
\]

From an actional perspective, the utterance is a Move \((M_1)\) that comprises a single Discourse Act \((A_1)\) (9). As seen in Figure 2.2, Discourse Acts are characterized by an Illocution \((F_1)\), speech participants \((P_1 \text{ and } P_2)\) and a Communicated Content \((C_1)\). Because the declarative Illocution in (8) is not explicitly verbalized, it is analyzed as having an abstract head (see Hengeveld & Mackenzie 2008: 75 for a full listing of possible abstract heads of illocutions). The use of square brackets indicates that the Illocution, Participants and Communicated Content are not seen as being in a hierarchical relationship.
A representation of the lower layers of the Interpersonal Level, i.e. those that constitute the Communicated Content $C_1$, is given in (10).

\[(10) \quad C_1: \left[ \begin{array}{c}
(T_1) \\
(+id \, R_1) \\
(+id \, +s \, R_2)
\end{array} \right]
\]

The representation of the Communicated Content simply lists the Subacts in a serial manner. To avoid redundancy with the Representational Level, semantic relations between the different entities evoked are not represented here. The identifiability operator $+id$ applies to both Referential Subacts, since both the pronoun he and the noun phrase the vase indicate contextual definiteness. The second participant, in addition, contains a specificity operator $+s$, as it refers to a specific instance of a given category, not to the category itself. The reason that markers of identifiability and specificity are represented on the Interpersonal Level is that they do not contribute to the description of the situation referred to, but rather reflect the speaker’s assumptions as to whether the entities referred to are part of the common ground.

Because FDG representations can become rather intricate, shorthand notations are used on occasion. It is common practice to only represent those variables in detail that are of theoretical interest to the phenomenon under investigation. Moreover, to represent the semantic content that corresponds to Discourse Acts or Moves, paraphrases can be used. One might for instance use the shorthand representation in (12) to display the content of (11). In this example, the Conc function indicates a concessive relation between the two Discourse Acts.

\[(11) \quad \text{“although the symptoms were clearly compatible with adrenal insufficiency, the diagnosis of hyperpigmentation had been overlooked”}
\]
A more elaborate application of FDG to attested corpus data follows in Chapter 8. The following section introduces the basics of Cognitive Grammar and, for the sake of comparison, provides CG analyses of the same set of linguistic examples as analyzed above.

2.3 Cognitive Grammar

2.3.1 Background and starting points
Cognitive Grammar (Langacker 1987, 1991, 1999, 2008a) was developed in close interaction with other theories in cognitive linguistics, including frame semantics (Fillmore 1982), idealized cognitive models (Lakoff 1987), and force dynamics (Talmy 1988). Langacker’s approach to grammar is inspired by radical disagreement with generative linguistics, in particular with regard to its general neglect of conceptualization as a driving force behind linguistic structure. CG shares a number of key commitments with other cognitive linguistic approaches. These can be referred to as the generalization commitment, the cognitive commitment and the usage-based commitment (as formulated by Lakoff 1990).

The generalization commitment holds that many, if not all aspects of language can be understood in terms of a limited number of mechanisms and principles. CG takes issue with traditional theories that have assumed separate modules for different types of linguistic knowledge (e.g. phonology, syntax and semantics, etc.). Instead, it seeks to describe language in terms of structures that apply across domains, and does not subscribe to the assumption that form and meaning can sensibly be studied as distinct systems.

The cognitive commitment refers to the dedication to characterizing language, and its underlying principles and mechanisms, in a way that is consonant with what is known from other disciplines about the human mind and brain. A central motivation for taking
Comparative evaluation of the models | 41

domain-general cognition as a starting point, as Langacker (1987: 12-13) puts it, is that “we have no valid reason to anticipate a sharp
dichotomy between linguistic ability and other aspects of cognitive
processing”. CG’s analytical devices, accordingly, are motivated by
processes and structures revealed by other disciplines in the cognitive
sciences, in particular Gestalt psychology. CG rejects the idea, held by
certain formalists, that pursuing ‘representational elegance’ is sufficient
to assure the psychological plausibility of a linguistic theory.

The usage-based commitment, finally, holds that linguistic
knowledge should not be studied in isolation from language use. Instead,
linguistic knowledge is seen as consisting in abstractions over
commonalities among actual instances of linguistic communication. This
commitment has at least three important (and somewhat controversial)
theoretical ramifications. First, meaning is thought of as subjective and
embodied: linguistic categories do not directly reflect objective
structures in the ‘external world’, but are mediated by the way
individuals conceptualize their environment. Second, because the usage
events from which linguistic knowledge is derived are always situated in
context, there is from a CG point of view no obvious divide between
semantic and more general conceptual knowledge. Hence, along with
other cognitive linguistic theories (e.g. Fillmore 1977), CG subscribes to
the notion of “encyclopedic semantics” (Langacker 1999). A third,
related corollary of the usage-based paradigm is that no strict boundary
is assumed between semantics and pragmatics. Given that language use
is inherently interactive and contextualized, CG assumes that the
experiential knowledge that underlies linguistic patterns subsumes
language’s “multifaceted interactive function that includes
communication, manipulation, expressiveness, and social communion”
(Langacker 1999: 14). In other words, semantic structures in CG do not
only include encyclopedic knowledge structures, but also dialogue
management, illocutionary force, information structure, and other such
factors.

2.3.2 Basic principles
In his foundational books on Cognitive Grammar (1987, 1991) and
subsequent works, Langacker elaborates a comprehensive framework
for grammatical analysis, which draws on the three main commitments of cognitive linguistics. The current section outlines the most important principles and assumptions of the model.

2.3.2.1 Grammar as symbolic
Taking inspiration from Saussurean semiotics, CG’s most central assumption is that grammar can be conceived of as a structured repository of symbolic units. That is, CG holds that all aspects of linguistic structure can be analyzed in terms of form-meaning associations: pairings of a phonological structure (the phonological pole) and a semantic structure (the semantic pole). As already mentioned above, what constitutes these poles may go well beyond what is traditionally considered the realm of phonology and semantics. In line with the usage-based paradigm, CG maintains that symbolic structures are abstracted from communicative experiences, which “consist [...] of a comprehensive conceptualization, comprising an expression’s full contextual understanding, paired with an elaborate vocalization, in all its phonetic detail” (Langacker 2001: 144).

The symbolic structures that constitute grammatical knowledge can be characterized in terms of three, closely related dimensions. The first is their phonological complexity: some structures have a simple, atomic phonology (e.g. lexemes like *dog* or suffixes like *-er*), whereas others have a complex, composite structure (e.g. multi-morphemic words or idioms such as *all of a sudden*). Second, some symbolic units have a fully specific phonological pole (e.g. *bird, kick the bucket*), while others have open slots (e.g. *the Xer the Yer*) or are almost entirely schematic (e.g. the double object construction Subj-Verb-Obj1-Obj2). A third parameter along which symbolic structures differ is the degree to which they have become conventionalized. Linguistic structures that occur with high frequency are likely to be further entrenched in the memory of language users, and further conventionalized in a community of speakers, than less frequent structures are. Thus, in line with other usage-based models (Barlow & Kemmer 2000; Bybee & Hopper 2001), CG holds that whether or not elements of expression are established as units of grammar is ultimately a matter of degree.
An important implication of the symbolic view on grammar is the ‘content requirement’, which posits that all elements of grammar, including morphology and syntax, have some form of semantic import. In sharp contrast to formal theories, CG maintains that grammar is not self-contained, but “reduces to the structuring and symbolization of conceptual content and thus has no autonomous existence at all” (Langacker 1999: 1). The semantic load of basic grammatical categories is characterized in terms of maximally abstract patterns of human experience. The grammatical class of nouns, for instance, is thought of as rooted in the experience of perceiving (groups of) entities as bounded or holistic (see 2.3.3 for more detail). Altogether, we see that CG advances a monostral view on grammar; it holds that all aspects of language structure essentially boil down to symbolic structures, which are defined by both phonological and semantic characteristics.

2.3.2.2 Meaning is conceptualization

A second major tenet of CG is that meaning amounts to conceptualization. CG rejects the objectivist perspective on semantics, i.e. the view that meaning is to be studied in terms of reference. Instead, it adopts a view of linguistic meaning in terms of mental processes. In its treatment of semantics, CG appeals to Fillmore's (1975) theory of conceptual networks and semantic frames. It assumes that the experiential knowledge that makes for the substance of linguistic meaning comprises a number of closely connected cognitive domains. These may subsume anything that constitutes mental sensations, including “[our] experience of time and our capacity for dealing with two-and-three-dimensional spatial configurations [as well as] domains associated with the various senses: color space, pitch scale, temperature sensations and so on” (Langacker 1991: 4). The representation of the word *foot*, accordingly, includes abstractions from perceptual and motoric experiences related to feet, as well as more abstract properties corresponding to its use in communication (e.g. combinatory restrictions of the grammatical class of nouns).

At least equally relevant for grammatical organization is the notion of *construal*: the way the conceptual content evoked by an expression is structured. CG maintains that the form of an utterance
mirrors the construal that is imposed on it by the conceptualizer (i.e., the speaker or hearer). Dimensions along which semantic construals can differ include the following (Langacker 1987, 2002):

1) **Profile-base relationship.** Essential to the meaning of an expression, according to CG, is the relation between what is explicitly referred to (the profile) and the conceptual knowledge that lies at the basis of its interpretation (the base). The meaning of *nine o'clock*, for instance, is meaningful in relation to the conceptual frames related to time, numbers, daily routines, etc., but does not refer to these notions explicitly.

2) **Level of specificity.** A conceptualized entity or situation can be described in different degrees of detail. The sentence *something happened* may refer to the same situation as *Vurnon was given a red boat for his birthday*, but is construed with a different degree of specificity. In CG terms, the former is more schematic, whereas the latter is more granular.

3) **Trajector-landmark alignment.** Whenever the profile contains two or more participants, the linear nature of language requires one of them to be selected as primary. The primary participant, typically expressed in subject position in English, is called the ‘trajector’ in CG. The secondary participant is referred to as the ‘landmark.’ The selection of a trajector-landmark relation is seen as critical to an expression’s meaning. Therefore, pairs of sentences like *Jody’s car is next to the tree* and *the tree is next to Jody’s car* are not seen as semantically equivalent, despite the fact that they hold true under the same circumstances.

4) **Subjectivity-objectivity.** Elements of linguistic expressions can involve different degrees of recourse to the ground, i.e. the contextual circumstances in which the linguistic interaction takes place. Elements of an expression that make no explicit or implicit reference to the ground are described as objective (e.g. *Amsterdam, 2016*). If the interpretation of a linguistic element is contingent upon the properties of the ground, it is said to be subjective (e.g. *here, this year*).
5) **Mental scanning.** Certain distinctions between linguistic elements (e.g. verbs, nouns) are described in CG in terms of the way the semantic content is ‘scanned’ by the conceptualizer. The same event may for instance be construed as a process (*Jody runs*) or as a reified entity (*the running of Jody*). In the former case, the evoked conceptualization is assumed to involve ‘sequential scanning’ – its content is accessed in a serial, state-by-state fashion. The nominalized expression, in contrast, is theorized to involve ‘summary scanning’: the conceptual content is presumed to be accessed in a simultaneous, holistic manner (Langacker 1987: 145).

### 2.3.2.3 Compositional structure

As a theory of grammar, CG is concerned with the way linguistic units combine into larger structures. Compositional structure is analyzed in terms of the integration of symbolic units, through a process of elaboration (Langacker 1987: 304). Elaboration entails that a schematic element of one of the component structures, which serves as an 'elaboration site' (*e-site* for short), is in correspondence with another, more specific element. The schematic element is said to be elaborated by the more specific one: the latter adds granularity to the semantic characterization of the former. As a basic example, consider the semantic pole of the verb *to jump*. Because its meaning cannot be conceived of without the presumed existence of an actor (someone who performs the jumping), it includes a salient e-site as part of its semantics. When combined with a noun phrase in subject position (*the athlete jumps*), the subject noun elaborates this e-site: it characterizes the actor in finer detail. A fundamental assumption in CG is that the process of elaboration takes place across all domains of linguistic analysis. In phonology, for instance, consonants are analyzed as evoking an e-site, because they require the presence of a vowel to yield an audible speech segment.

Structures that have an e-site (e.g. verbs, adjectives, consonants) are referred to as dependent structures, whereas structures that can exist in their own terms (e.g. nouns, vowels) are said to be autonomous. Some structures are almost entirely dependent; they serve as ‘templates’
for the assembly of larger constructs. The combination of a noun phrase and a verb in simple clauses, for instance, is analyzed as governed by a ‘constructional schema’ that has no specific phonological features by itself, but serves to structure the semantic content in a specific way. Some such schemas bear resemblance to syntactic rules or phrase-level templates (in FDG terms), except that in CG, they are not seen as ontologically distinct from other linguistic elements – they are theorized to derive from abstractions from language use, in the same way as lexical items or other structures. A final note regarding compositionality is that the meaning of a linguistic construct is not only defined by the end product of elaboration, but also the incremental process of composition. The compositional path, i.e. the totality of steps that lead up to a composite structure, is part of the full meaning of an utterance.

2.3.2.4 Context and discourse in CG
Discourse-related processes have been acknowledged as central to grammar ever since the early days of CG (Langacker 1987, 1991), but have received most attention in relatively recent years (Langacker 2001, 2008a, 2013). The way a speech event connects to its situational context is analyzed in terms of grounding: the establishment of a relation between the conceptualized situation and the communicative circumstances (time, place, etc.). Applied to nominals, grounding entails a shift from a type conception of an entity (reference to a category) to an instance conception (singling out a specific token). On the clause level, grounding implies that a connection is established between the current situation and the situation described. This relation can for instance be a temporal one (e.g. in the case of tense marking), or one of likelihood (e.g. in the case of modal verbs).

In recent works, Langacker further elaborates on language’s unavoidable embedding in discourse, with special attention to the dynamic dimension of language use (Langacker 2001, 2008a). He emphasizes the incremental nature of linguistic meaning: “as the discourse unfolds [...] the interlocutors negotiate a series of discourse frames, each produced by updating the previous one” (Langacker 2008a: 281). This dynamic view of language entails acknowledgement that part of the meaning of linguistic structures resides in their relation with
preceding and upcoming material. Discourse, accordingly, is modeled as a series of connected discourse frames (or attentional frames), which are approximately clause-sized and accessed in a sequential manner. With each successive discourse frame, the newly evoked content is anchored to the representation that was set up by previous events, and possibly projects one or more imminent frames. The fact that the elements responsible for discursive anchoring (elsewhere called thematic or topical elements) are often relatively simple in structure is seen as a natural solution to the problem of mapping complex conceptual content onto a linear series of attentional frames. Attentional processes, for this reason, are believed to be paramount to linguistic structure: “I would strongly deny that the meaning contributed by attentional framing lies outside the proper scope of linguistic semantics” (Langacker 2001: 158).

2.3.3 Analytical tools
CG uses diagrammatic notation conventions. In view of space, this section does not introduce these in full detail, but considers some of the facets that are most relevant for later chapters. In Figure 2.4, we see how CG represents the semantic poles of nouns, adjectives and verbs in their most schematic form (adopted from Langacker 2008a: 99).

The square in Figure 2.4a represents the maximally general semantic category ‘entity’. Entities are generalizations over any type of
experience, including the other conceptual categories in Figure 2.4. Because all linguistic elements evoke an entity of some sort, this semantic category does not have a grammatical correlate in and of itself. However, the notion of entity is relevant to the definition of other semantic notions, as discussed later on in this section. The circle in Figure 2.4b represents the conceptual category THING, which is theorized to underlie the class of nouns. Its most prototypical instantiation is a concrete physical object, but it encompasses a much broader area of conceptual space: anything that is conceptualized as a product of reification or grouping (Langacker 2008a: 105). In Figure 2.4c, we see a schematic depiction of a NON-PROCESSUAL RELATIONSHIP – the conceptual correlate of adjectives. It includes two entities and an atemporal relationship between them. A closely related category, seen in Figure 2.4d, is that of PROCESSUAL RELATIONSHIP. These relationships have an additional, temporal dimension, and are thought to correspond to the grammatical class of verbs. Even static verbs (to sit, to be) are analyzed as evoking a PROCESSUAL RELATIONSHIP, as they are theorized to activate conceptual content in a dynamic, rather than holistic way (they evoke sequential scanning; see Section 2.3.2.2). Note that the number of stages depicted in Figure 2.4d is somewhat arbitrary; the temporal development of the conceptualized relation is continuous, not stepwise. Figure 2.4e shows a shorthand notation for the semantics of a full proposition. The schema combines a single THING and a PROCESSUAL RELATIONSHIP, although this representation is not limited to intransitive clauses.

To illustrate these and further notation conventions, I here go through CG-based analyses of the examples for which we have already seen FDG representations in (4)-(8). In Figure 2.5, we see how simple adjectives, nouns and adjective-noun pairs may be represented.

---

Small capitals are used in CG for the semantic characterization of grammatical classes.
For the noun *vase*, a basic representation is given in the lower-right diagram. The heavy circle indicates that the noun profiles a *THING*. The characterization of this *THING* by the word *vase* is abbreviated with the letter *v* here. The lower left diagram shows a CG notation of the adjective *tall* (inspired by Langacker 1987: 133). Being an adjective, *tall* is analyzed as profiling a *NON-PROCESSUAL RELATIONSHIP*. Because the nature of this relationship can be more concisely represented through depiction that through description, it is shown pictorially: the diagram shows that the meaning of *tall* includes a spatial correspondence between the lower edge of some object (or, technically, a *THING*) and the lower edge of a conceived vertical axis. In addition, it assumes a correspondence between some point in the higher region of the vertical axis and the upper edge of the *THING*. Thus, the diagram emphasizes that the conceptual substrate of ‘tallness’ is defined relative to a conceptual base, namely the human experience of visual perception along a vertical axis. The use of pictorial representations, importantly, should not be misunderstood as an attempt to depict what mental imagery 'looks like'. Rather, CG’s diagrams provides a heuristic device that can be helpful for analyzing the semantic substrates of lexemes and other linguistic structures.

Figure 2.5 An example of basic compositionality in CG
In the top part of the figure, we see how the composite meaning of *tall vase* is accomplished. Because the *THING* that is evoked by *tall* is schematic, it is represented as an *e-site* (the notation of *e-sites* as a hatched area, introduced by Langacker in 1987 is adopted here for clarity). This *e-site* is in correspondence with the noun *vase*, as represented by the dotted, horizontal line. In other words, the noun *vase*, which has a rather specific semantic pole, elaborates the *e-site* evoked by the adjective *tall*. To emphasize the direction of the elaboration while keeping the diagrams manageable, the process of elaboration is marked by a single dotted arrow (cf. the notation with two separate lines for correspondence and elaboration used by Langacker 1987 and elsewhere). The integrated semantic representation, displayed in the top part of the figure, shows that the noun phrase, as a whole, profiles a *THING* that belongs to the semantic category 'vase' and has a *NON-PROCESSUAL RELATIONSHIP* with a specific region of the vertical axis of spatial experience. Overall, the diagram incorporates three of CG's key concepts: a profile-base relationship, schematicity and elaboration.

In Figure 2.6, we see that a full representation of the proposition in (7) (*the vase is tall*) involves CG's analytical tools for grounding and attentional framing. The bottom left diagram represents the meaning of the definite article *the* (Langacker 2008a: 285). The rounded rectangles represent two attentional frames that are accessed in (not necessarily immediate) succession. As indicated by the dotted line, the semantics of the definite article evokes a correspondence between an instance of a *THING* in the current attentional frame (where the noun phrase is uttered) and a *THING* that has figured in a previous discourse frame. The instances are depicted as circles with points in them, to make them visually distinguishable from entire classes of entities (following Langacker 1987). The ellipse below the attentional frames represents the ground, i.e. the current interactional circumstances, including the speaker (S) and hearer (H). The dashed arrows represent the allocation of the attention of the interlocutors to the *THING* that is schematically referred to. In the construction of the noun phrase *the vase*, this element

---

7 This representation does not explicitly acknowledge that the entity referred to can be contextually salient for another reason than having been referred to earlier in the discourse.
becomes elaborated by the more specific conceptual content of the noun *vase* (middle-left diagram).

The right side of the diagram displays a schematic representation of the verb *to be*. It is analyzed as denoting a schematic PROCESSUAL RELATIONSHIP between two entities. Followed by *tall*, one of these entities is elaborated by the more specific content of the adjective. As a result, the phrase *is tall* profiles a PROCESSUAL RELATIONSHIP between a THING (to which the tallness applies) and a conceived segment of the vertical axis. The THING that is implicitly referred to by the adjective *tall*, here shown as an e-site, subsequently becomes elaborated by the referential content of *the vase*. The emerging semantics of *the vase is tall*, shown at the top of the diagram, differs in at least two ways from that of *tall vase*. One difference concerns the profile-base relationship: whereas

---

8 The verb *to be* does not necessarily predicate over THINGS, but may also pertain to relationships, as in *green is nice.*
the noun phrase *tall vase* only profiles the *THING* that it designates, the proposition in Figure 2.6 profiles two entities and their respective relationship. Second, the profiled relationship is construed as processual, whereas the relationship evoked by *tall vase* does not develop over time.

Finally, I turn to the representation of *he drops the vase*, the final sentence of which we have seen an FDG representation (8). A simplified CG analysis of the sentence is given in Figure 2.7. To avoid redundancy with previous figures, only the final state of the compositional path is represented.

Like the definite article, the personal pronoun *he* schematically refers to a contextually salient entity. The diagram on the left, which assumes that the pronoun is used anaphorically, shows that the semantics include a previous moment of shared attention to some instance of a *THING* (Langacker 2001: 168). Because gender in CG is seen as a class that is primarily defined by distributional characteristics, and only very partially by semantic features, I simply use the orthographic representation *he* for this category instead of a visual heuristic. The representation of *the vase* is equivalent to the analysis of the previous example (Figure 2.6). The placement of the two participants in the same attentional frame is not meant to suggest that they were introduced at the same point in the discourse – the alternative option of using of two separate rounded rectangles would entail an arbitrary decision as to which was introduced first. The lexical meaning of the verb *drops* is represented iconically: the diagram emphasizes one dimension of the meaning of the verb *to drop*, namely an increasing vertical distance between the trajector (tr) and the landmark (lm) over time. Although this representation may appear somewhat convoluted at this point, CG’s potential for graphically depicting spatial aspects of semantic relations can be useful for analyzing gestural aspects of language, as we will see in Chapters 3 and 8.
Before turning to the inclusion of gesture in the models, the remainder of this chapter provides a comparative evaluation of FDG and CG on a number of relevant features. I will first go over some ways in which the models are alike, and subsequently discuss a range of differences with respect to their assumptions, goals and analytical methods.

2.4 Points of convergence

Before turning to the inclusion of gesture in the models, the remainder of this chapter provides a comparative evaluation of FDG and CG on a number of relevant features. I will first go over some ways in which the models are alike, and subsequently discuss a range of differences with respect to their assumptions, goals and analytical methods.

2.4.1 Departure from formalist assumptions

Perhaps the most significant commonality between FDG and CG is their outright rejection of some of the core assumptions of generative linguistics. Both theories take issue with the idea that syntax is to be studied as independent of other aspects of language. More generally, both models rest on the assumption that language is shaped by, and to be understood in relation to, cognitive and communicative constraints. In CG terms, language is seen as inextricably linked to its ‘semiological function’ (symbolizing thoughts through ostensive behaviors) as well as its ‘interactive function’ (engaging in social interaction) (e.g. Langacker 1998). FDG assumes a similar, dual motivation for language structure, as apparent in the divide between the Interpersonal Level and the Representational Level.

More specific formalist assumptions are also rejected by both frameworks. Neither FDG nor CG postulates a ‘deep structure’, nor is
there a place for transformational processes or movement rules such as those stipulated in generative grammar. With regard to the issue of innateness or universal grammar, both CG and FDG take a rather cautious stance. The extensive typological review that underlies FDG’s architecture reveals some apparently universal tendencies (for example, if a language accepts an Undergoer as subject, then it also accepts an Actor), but these are not purported to reflect innate cognitive biases. Rather, the hypothesis is entertained that universal patterns reflect commonalities in the communicative salience of certain types of structures across linguistic communities (Hengeveld & Mackenzie 2008: 35). CG also takes a rather agnostic position in this regard. Langacker (1987: 12-13) denies that the inextricable link between language and cognition implies the existence of an innate faculty for language. His argument is that even if genetically determined predispositions for language exist, linguistic experience is so crucial for its development that it is wrongheaded to study language structure without assuming a fundamental role for other cognitive domains.

2.4.2 **Embedding a in theory of language use**

Both FDG and CG maintain that grammar is best understood in relation to language use, including situational context, discourse relations, patterns of intonation, and other factors that have traditionally been left out of the realm of grammatical theory. FDG explicitly aspires to achieve pragmatic adequacy, taking account of the various ways in which discursive and situational context can influence the structure of an utterance. CG presents itself as a usage-based theory, because it draws upon the assumption that linguistic knowledge emerges from abstractions over ‘usage events’. Thus, the symbolic structures that make up grammatical knowledge are assumed to comprise abstractions over the social contexts in which they are instantiated.

In practice, it is noticeable that the grammatical analyses found in both the FDG and CG literature concern relatively decontextualized data. Most analyses follow the longstanding linguistic tradition of using invented, isolated sentences as data. Whereas some practitioners of FDG have moved towards the use of corpora in recent years (e.g. Hengeveld & Keizer 2011; Keizer 2015), detailed descriptions of the situational embedding of the data are not common. Moreover, the notion of
discourse, when taken into account, is often restricted to the verbal utterance that directly precedes the utterance examined, leaving such matters as genre and register out of consideration. Thus, although FDG and CG are presented as discourse-oriented theories, neither theory fully lives up to this commitment in terms of the choice and description of linguistic data.

2.4.3 Psychological motivations
Some degree of commitment to psychological reality is shared between CG and FDG. As discussed in Section 2.3.1, principles from cognitive psychology are at CG’s basis. Langacker draws heavily on Gestalt psychology and employs notions such as imagery, categorization and attention to account for aspects of linguistic structure. Nonetheless, explicit reference to specific psychological research is somewhat sparse.

On the FDG side, we also see an interest in cognition, despite the fact that most of its practitioners do not subscribe to the cognitive linguistic paradigm. As discussed above, FDG maintains that a model of grammar is most effective if the analytical steps prescribed mirror the sequence of cognitive processes of the language user. Taking Levelt’s speech production model as an inspiration and using process-like vocabulary to describe relations between levels, FDG creates the impression of being oriented toward speech production. Hengeveld & Mackenzie (2008: Ch.1) emphasize, however, that FDG is not to be understood as a theory of psychological processes.

Thus, both CG and FDG commit to psychological adequacy, but present themselves as theories of grammar, not cognition. Whereas this quasi-psychological approach has its advantages, it has also been a source of criticism for both accounts. FDG’s architecture, especially its strictly modular organization, has been argued to be at odds with current-day psychological literature (Anstey 2008; Butler 2008a; Inchaurralde 2004, 2005). Likewise, some of CG’s key notions (e.g. summary vs. sequential scanning) have been criticized for lacking empirical support (Broccias & Hollmann 2007). The prevailing response is that theories of grammar, even cognition-oriented ones, should not be interpreted as making claims about psychological processes, but this
point may not be equally satisfying to cognitive scientists as it is to (some) linguists.

2.4.4 Universally applicable tools
As a corollary of studying language in relation to cognitive and functional constraints, both CG and FDG strive to be applicable to any language in the world. From a CG perspective, the main reason to assume universal applicability is that the cognitive operations it draws on are shared by people in all cultures. For instance, it can safely be assumed that humans across cultures perform cognitive operations such as abstraction and categorization (even though we might not all categorize in exactly the same way). Whereas the bulk of CG research concerns English, the framework has also been applied to Slavic languages (Cienki 1995), Chinese (Zhang 2014) and Uto-Aztecan (e.g. Langacker 2008a: 182), among others.

On the side of FDG, universal applicability is fostered by postulating categories that reflect the commonalities in language use across communities. FDG goes much further than CG in its cross-linguistic orientation: it uses over 150 different languages as the empirical ground for the theory. The levels and categories postulated in FDG are motivated by a systematic assessment of the functional and structural commonalities and differences between these languages.

2.5 Points of divergence

2.5.1 Moderate versus extreme departure from formalist models
A general discrepancy between FDG and CG concerns the extent to which they deviate from formalist assumptions (see e.g. Butler 2008a for a discussion of moderate vs. extreme functionalism). As is evident from the discussion of the basic assumptions of the models, FDG represents a relatively moderate functional position, whereas CG is on the more extreme side of the spectrum. In fact, FDG’s modular view on language remains in some ways compatible with Chomsky’s assumptions on self-containedness: the different levels of linguistic structure are represented as modules with their own internal organization. Syntactic structures are seen as ‘primitive’ aspects of
grammatical knowledge that are ontologically independent (although connected with) semantic structures. Thus, FDG holds on to the traditional lexical-compositional view of grammar, which has been disputed in cognitive linguistics (see e.g. Taylor’s 2003 critique of the ‘building block metaphor’ of linguistic structure). Butler (2003: 30) for this reason characterizes FDG as belonging to the category of ‘generative’ functional models: “they are generative in the sense that they provide an interlocking set of rules and principles whose import and relationships are made absolutely explicit and which is capable of assigning structures and interpretations to linguistic expressions.”

CG’s more extreme functional position is evident from its rejection of the existence of (morpho)syntactic rules independent of their instantiations. CG holds that syntax essentially reduces to a set of cognitive routines, each of which has inherent semantic and pragmatic value. Its approach to grammar is ‘monostratal’: the basic units of analysis are form-meaning pairings, which cross-cut phonology, syntax, semantics and pragmatics. Further indications of its more radical position, discussed in the following sections, include its denial of a sharp boundary between semantics and pragmatics and between linguistic and non-linguistic knowledge.

2.5.2 **Language as a cognitive system versus language as a vehicle for communication**

In principle, both CG and FDG acknowledge that grammar can be conceived of as a system on the individual level (as a cognitive mechanism) as well as on the community level (as a socially motivated system for interaction). In the FDG literature, for instance, we see a strong orientation towards typology – suggesting a community level view – but yet also reference to models of individual speakers. In CG, too, the notion of grammar seems to be defined on two different levels at once: Langacker (1987: ch. 2) explicitly maintains that both entrenchment in an individual’s mind and conventionalization in a community are necessary preconditions for expressive elements to acquire a grammatical status.

Nonetheless, FDG and CG clearly differ in their primary orientation. As we have seen, CG’s main selling point is its cognitive
orientation. Although the notion of conventionalization is theoretically relevant, CG does not devote much attention to processes that take place on the community level. For FDG, in contrast, the social-communicative dimension of language is seen as the primary motivational force behind linguistic structure.

A consequence of this difference in orientation is that the models have different assumptions about the relation between semantics and pragmatics. In FDG, semantics and pragmatics are studied as separate levels of analysis. The latter is dealt with on the Interpersonal Level, which covers a wide range of phenomena, including rhetorical relations, information structure, illocutionary force, and the expression of commitment and emphasis (Hannay & Hengeveld 2009). Semantics is seen as subservient to a communicative intention, but is nevertheless studied in its own terms: the meaning of linguistic elements is analyzed “in isolation from the ways these are used in communication” (Hengeveld & Mackenzie 2008: 129). CG, in contrast, posits a continuum between pragmatic and semantic aspects of linguistic meaning. Semantics is not approached from an 'external' vantage point, as in FDG, but with significant attention to the role that individual minds play in mediating the relationship between words and their referents. The semantic pole of grammatical structures is assumed to simultaneously comprise referential, discursive and interpersonal dimensions of meaning. In practice, however, CG’s most established analytical tools relate to traditional semantic phenomena (e.g. THINGS, PROCESSES); its application to phenomena like speech acts and rhetorical relations has not received nearly as much treatment in the literature.

2.5.3 Modularity vs. gradience
Like all models of language, FDG and CG strongly rely on the categorization of linguistic elements. The discreteness of these categories and their borders, however, constitutes a point of dispute. FDG adheres to a traditional, compartmentalized view of language, with separate levels of representation for pragmatics, semantics, morphosyntax and phonology. The grammar proper is defined as having a sharp boundary: it is limited to ‘digital’ (opposition-based) aspects of language, while ‘analog’ aspects are considered extra-grammatical
(Hengeveld & Mackenzie 2006: 388). Furthermore, FDG takes a
dichotomous view on the grammar-lexicon nexus, as reflected in the
distinction between modifiers and operators. In stark contrast, CG holds
that many, if not all of the distinctions that are made in traditional
linguistic theory (e.g. grammar vs. lexicon, semantics vs. pragmatics) are
ultimately a matter of degree and should be studied in an integrated
manner. All categories relevant to grammatical analysis, moreover, are
explicitly claimed to have a prototype structure, with central and less
central members, and no sharp boundaries are assumed to exist
between them.

One way of understanding this divergence is in terms of the
trade-off between theoretical adequacy and practical applicability.
Whereas some of CG’s claims might be better justified by the
psychological literature (strictly modular models of the mind do not
receive much support in current cognitive science), FDG’s rigid character
creates a greater potential for generating replicable analyses and explicit
hypotheses. Depending on one’s research goals, this advantage of a
discrete-modular view of language might outweigh its cognitive reality:

If a distinction is relevant to the description and analysis of
linguistic utterances, we ought to find a way of employing, and
therefore defining, it in the theory. Thus, if we find that the
grammatical/lexical distinction can help us to account for
certain differences in linguistic behaviour, not reflecting this
distinction in the FDG model would seriously weaken the
model’s explanatory and psychological adequacy. Therefore,
although we know the difference between lexical and
grammatical to be non-discrete, we nevertheless need to draw a
line; this needs to be done in a principled and consistent way, on
the basis of well-defined criteria, and for each language
individually. (Keizer 2009: 52)

A consequence of this stance is that FDG’s assumptions are explicit
enough to be tested and falsified on the basis of linguistic observations.
It can be argued that this aspect of the model renders FDG, in some
ways, a more scientific enterprise than CG. On the other hand, by
adhering to a lexical-compositional view of language, the range of phenomena that FDG can accurately account for might be comparatively limited, in particular when it comes to constructions that are on the border of lexicon and grammar (Anstey 2008; Keizer 2009).

2.5.4 Patterns vs. processes
A further point of divergence pertains to the conception of the relation between grammar and the processes of language production and comprehension. The most general characterization of this difference is that FDG (together with most other functional models) is process-oriented, whereas CG (and other cognitive models) is pattern-oriented (Nuyts 2005, 2007). In other words, whereas both models dispense with the generativist assumption that grammar is merely a matter of competence, there is no general consensus on whether grammar is to be seen as a declarative or procedural system.

As seen in Figure 2.1, FDG decisively adopts a procedural architecture. In the FDG literature, grammar is often described as a process where the output of the one level ‘feeds into’ another level (see especially the dynamic interpretations of FDG, e.g. Bakker & Siewierska 2004; Hengeveld & Smit 2009; Inchurralde 2004; Mackenzie 2004, 2010, 2012; Smit 2009). An alternative view on the relation between grammar and language production, proposed by Mackenzie (2004), is that the Interpersonal Level is procedural in nature, but the Representational Level represents a declarative memory storage. In CG, grammar is seen as a repository of symbolic structures; although conceptualization is seen as a dynamic process, no explicit connections are made with models of speech production or comprehension. Thus, CG takes a view on grammar that is essentially declarative.

2.5.5 Conceptual vs. semantic structure
FDG and CG also take opposite positions on the relation between semantic and conceptual structure. FDG posits a sharp dichotomy; it assigns conceptual and semantic information to distinct components of the model. The Conceptual Component is regarded as the ‘motivational force’ behind the Formulation operations, but it is seen as non-linguistic by itself. That is, the Grammatical Component only includes the
Comparative evaluation of the models | 61

conceptual material that receives explicit formal expression. CG, in contrast, refutes the idea that semantics can be divorced from the conceptual system. It considers meaning to be ‘encyclopedic’ by nature, and makes no distinction between coded and evoked aspects of meaning. Experiential knowledge of any kind may be relevant to grammatical organization, including perceptual, motoric and emotive sensations. Nonetheless, no full overlap is assumed between linguistic and non-linguistic knowledge: semantic representations are seen as a subset of the conceptual system, tailored to community-specific conventions (Langacker 1987: 99).

Interestingly, both accounts have been in dispute for their positions on this issue. The rigid divide proposed by FDG was criticized on various occasions. Anstey (2008), for instance, argues that it is often impossible to determine what information is conceptually relevant versus semantically relevant, and suggests that the formulation layers should be conceived of as conceptual. Butler (2008b) and Nuyts (2004b), furthermore, accuse the Conceptual Component of being an unstructured ‘waste-box’ that encompasses too many different types of information (conceptual, affective, interactional, discursive etc.). On the CG side, the denial of a dichotomy between a conceptual and semantics was vigorously contested by Levinson (1997), who argues that it is fundamentally flawed to equate semantic knowledge to conceptual knowledge. According to Langacker (2008a: 40), however, Levinson’s criticism rests on a misinterpretation of the claim. CG does not deny that thoughts need to be organized in a special way when using language, but rather that there is no fundamental difference in the nature of the knowledge involved in language and other cognitive activities.

2.5.6 Formal versus heuristic notations
A final notable difference pertains to the notation systems associated with the two models. As is clear from sections 2.2.3 and 2.3.3, the representational tools used in FDG and CG are in many ways dissimilar. FDG employs formalisms that are derived from predicate logic, whereas CG uses what can be called ‘heuristic illustrations’. The notational differences mirror some of the points of theoretical and methodological divergence between the models. FDG’s notations force the analyst to
simultaneously take the interactional and representational dimensions of language into account, and promote close attention to hierarchical relations between the elements of an utterance. CG, on the other hand, enables a more in-depth analysis of semantic structure in terms of construal operations, in a way that is compatible with conceptual semantics. Notational ramifications of the difference in orientation are revealed by comparing the examples seen in 2.2.3 and 2.3.3. FDG, which does not see itself as called upon to deal with lexical semantics, simply represents predicates by their form (e.g. ‘f: tall’). CG’s notation system, by contrast, offers elaborate tools for displaying aspects of the underlying conceptual substrate (e.g. a vertical axis for the semantic pole of tall, as in Figure 2.5).

Depending on one’s research goals, both systems may have their benefits and drawbacks. Whereas FDG’s treatment of semantics could be considered comparatively shallow, FDG practitioners might hold that some of the dimensions of construal that CG is concerned with (e.g. the specificity of lexical items) are grammatically irrelevant. On the practical side of the coin, we see that FDG’s notations are maximally explicit, while CG diagrams generally involve a higher degree of flexibility and, arguably, some degree of arbitrariness. It is, for instance, not always obvious which conceptual structures are sufficiently relevant to be included in a diagram, and what amount of pictorial detail is appropriate.

2.6 Summary and a note on reconciliation

FDG and CG have developed from comparable motivations, but in fundamentally different ways. The most general characterization of their deviance is that CG has taken the functional position to a more extreme level. Being part of the cognitive linguistic enterprise, CG has a more individualistic view on grammar, and is explicitly geared towards compliance with the gradient nature of linguistic categories. FDG, being structure-oriented and taking typology as a primary source of inspiration, has developed a more interaction-oriented view on grammar, with relatively rigid analytical tools. As argued above, the differences discussed in the current chapter cannot be evaluated as better or worse independent of one’s research interests. Nonetheless,
the survey presented in this chapter leads to the general conclusion that FDG and CG have developed somewhat complementary strengths: FDG’s analytical architecture promotes balanced attention to the different levels of linguistic organization, and offers a wide range of built-in tools available for social-interactive factors. CG, on the other hand, provides a more thorough account of the relation between grammar and the cognitive processes that underlie conceptualization, and does justice to the various non-discrete facets of linguistic knowledge.

Finally, I will briefly consider the question of whether an integration of the two models is possible and, if so, desirable. The literature has seen several pleas for a reconciliation of cognitive and functional perspectives on language structure. Nuyts (2004a: 135), for instance, argues that combining the two perspectives “[...] is not a matter of finding a balance: If one wants to understand language, one has to be concerned as much as possible with both, for one needs to understand the full import of each.” Although Nuyts asserts that a unification of cognitive and functional perspectives need not be impossible, there are some hurdles to overcome:

In terms of ‘ontology’, both notions relate to different but equally inherent aspects of language: they are two sides of the same coin. The functional dimension refers to the ‘operational’ side: what language as a behavioral system does for a human being. The cognitive dimension refers to the ‘infrastructural’ side: what kind of system language is in terms of the human ‘biological’ setup (broadly speaking) and how it is ‘built up’ [...]. In terms of scientific epistemology, the two dimensions are not on equal footing. One cannot ‘see’ the infrastructure. Being concerned with the functional dimension of language is epistemologically antecedent to being concerned with the cognitive dimension, and getting a grip on the latter is only possible via an understanding of the former. (Nuyts 2004a: 135)

The discrepancy pointed out by Nuyts has not impeded the pursuit of greater convergence between cognitive and functional models of grammar. FDG’s psychological adequacy has been critically evaluated
(e.g. Inchaurralde 2004) and various attempts have been made to include insights from pattern-oriented models in FDG. Anstey (2008: 852), for instance, advocates a reconciliation of FDG with a Construction Grammar perspective. He proposes a hybrid model, which he argues can help to counter the problem faced by FDG of having to formulate a vast set of ‘mapping rules’ between function and form:

A plausible solution [...] is to shift the burden to the speaker, who has knowledge of multiple, partially overlapping pathways from conceptualization to expression. This view of the organization of grammatical knowledge, which privileges vertical conglomerations of all the formal and functional information necessary for successful communication over horizontal modules of all the behavioral and coding structures necessary for efficient algorithmisation, is known as construction grammar. [...] Just as the lexicon is a tripartite architecture of conceptual, syntactic and expression, so are higher levels of organization, from phrases, idioms, clauses, acts, moves, and so forth. Such an approach therefore for FDG could be called Constructional FDG (CFDG). (Anstey 2008: 852; emphasis in original)

Similar proposals, with variable degrees of compromise to FDG’s current principles, were made by Inchaurralde (2005), Keizer (2009, 2014) and Butler (2012). To date, however, the actual application of the proposed modifications by practitioners of FDG remains somewhat limited.

There have also been developments in the opposite direction. Cognitive linguistic models of grammar, including CG, are being put under increasing pressure to ‘recontextualize’ meaning in a more social setting (Croft 2009; Geeraerts 2010). Croft (2009), for instance, advocates that the basic tenets of cognitive linguistics need to be reformulated in social-interactional terms. He argues that the cognitive linguistic tenet ‘meaning is construal’ needs to be transformed into ‘meaning is construal for communication’, and applied accordingly. In line with this suggestion, cognitive linguistics has seen the emergence of interaction-oriented branches (e.g. cognitive sociolinguistics; Geeraerts
et al. 2010), as well as dialogical interpretations of CG (Langacker 2013; Zima 2013).

Given that FDG and CG rest on partially incommensurable principles and assumptions, the current dissertation does not pursue the development of an integrated model. Instead, I further evaluate both models with respect to a second significant paradigm shift that has taken place in linguistics over recent years: taking not only verbal, but also co-verbal modes of expression into account. Because speakers’ gestures are cognitively and functionally related to speech in numerous ways, as I will discuss in the next chapter, we can expect that both FDG and CG provide ways of analyzing their relation to language structure. On the basis of the comparative discussion presented above, it can be assumed that both frameworks are in principle open to the incorporation of co-verbal behaviors, but shed different light on the potential status of gesture in grammar. Exploring how multimodality can be incorporated in either model therefore provides novel, complementary ways of understanding gestural expression. In addition, such an assessment presents a test of whether the models of interest truly live up to their commitment to cognitive and communicative embedding. The dimensions along which cognitive-functional models of grammar diverge, outlined above, form the basis for the following chapters, which present a detailed assessment of how FDG and CG can be developed into models of ‘multimodal grammar’.
Chapter 3. Speakers’ gestures in FDG and CG

During spoken interaction, speakers express themselves through various communicative channels, most of which have traditionally remained outside the purview of grammatical theory. Instead, manual and facial aspects of spoken language use have been put forward as ‘paralinguistic’ or as mere matters of ‘performance’ (Chomsky 1983). Over the past decades, this bias toward written language has increasingly given way to a more multimodal view on grammar. Along with empirical advances in gesture studies, consensus has been growing that a comprehensive theory of language structure needs to incorporate more than the verbal channel alone.

The aim of this chapter is to evaluate the interface between grammar and co-speech gesture from the perspectives of FDG and CG. Although this dissertation is primarily concerned with manual gestures, most of the arguments presented apply to certain forms of facial behaviors and body posture as well. After reviewing recent literature on the grammatical potential of co-speech gesture, I discuss how the various ways in which co-speech gesture has been studied align with the main categories and distinctions postulated in FDG and CG. Finally, I

---

9 Some of the material presented in this chapter also appears in adjusted format in Kok (2016) and Kok and Cienki (2016).
10 See Allwood (2002) for a comparative discussion of different types of bodily behaviors from a semiotic perspective.
discuss some challenges that arise when taking the multimodal grammar enterprise seriously.

3.1 Motivations for a multimodal approach

McNeill (1992: 19) characterizes co-speech gesture as a mode of expression that is "opposite" to speech in certain ways. Whereas speech is segmental, combinatorial and systematic, he argues, gestures are "global-synthetic" and "idiosyncratic" in character, conveying meaning in a "holistic" fashion. This view, which has had a strong influence on gesture studies, appears to be at odds with a grammatical approach to gesture, corroborating Chomsky's (and others') skepticism about this matter:

Presumably there is a system, a set of principles, that determines the nature of the gestural system and the way in which it interacts with the language system, but whether those principles should be called a grammar is another question. I would think that it is a dubious metaphor, because it leads one to expect commonality of structure, and that is very much an open question (Chomsky 1983: 41).

In recent decades, nonetheless, evidence has accumulated that gestural modes of expression have more in common with speech than is traditionally assumed. In this section I discuss three ways in which gestures are relevant to theories of grammar.

3.1.1 Systematicity

Most cognitive-functional linguists would agree that a necessary precondition for an element of expression to be a candidate subject of grammatical theory is for it to have a certain degree of systematicity in the relation between its form and function. Under this criterion, different types of gesture can be ascribed some degree of grammatical potential. Kendon (1988) classifies various types of manual behavior in terms of how conventional or sign-like (in semiotic terms) they are. Of those manual behaviors that occur in concurrence with speech, so-called emblems or quotable gestures are at the most systematic end of the
Speakers’ gestures in FDG and CG | 69

spectrum. These are gestures with a stable form, that can be attributed a meaning within a given culture in the absence of verbal context (Efron 1972; Ekman & Friesen 1969; Kendon 1988). Examples are gestures such as the thumbs-up, which generally conveys the meaning of positive evaluation, at least in many cultures in the Western world, or the gesture whereby the index finger and middle finger are extended into a V-shape, with the palm of the hand facing outward, to say ‘peace’. The class of speech-linked gestures (e.g. those that complement speech in deictic utterances; McNeill 1992, 2000b) is somewhat more variable in its forms and meanings. Pointing gestures, for instance, can be performed in various ways and can serve a range of different semantic and interactive functions, depending on the verbal and cultural context in which they are deployed (Bavelas et al. 1992; Kita 2003).

Spontaneous gesticulation, i.e. those gestures that are produced spontaneously along with speech, is not fully unconstrained either. Despite the fact that gesticulation is profoundly creative and tailored to the context of the conversation, some patterns and clusters of such gestures have been identified. These were coined ‘gesture families’ (Kendon 2004) or ‘recurrent gestures’ (Bressem & Müller 2014; Müller 2004; Müller et al. 2013). The cyclic gesture, for instance, whereby the hand makes a repetitive outward rotating movement around the wrist, is linked to a limited set of pragmatic contexts, namely word search, description and request (Ladewig 2011). Based on thorough examination of German multimodal discourse, over fifteen such recurrent patterns in spontaneous gesticulation have been identified and documented (Bressem & Müller 2014). Recurrent gestures fundamentally differ from emblems in having a flexible form and a meaning that is partly contingent upon the accompanying speech. Cienki (2015) incorporates this category into Kendon’s continuum, represented as follows:

![Figure 3.1](image.png)

Figure 3.1. Different types of gesture ordered according to their degree of conventionality, according to Cienki (after Kendon 1988).
Experimental research has provided additional evidence that gestures encode semantic and pragmatic meanings in a relatively systematic fashion. Numerous studies have found that the form of co-speech gestures systematically reflects the discourse structure (Levy & McNeill 1992), linguistic viewpoint (Parrill 2010; Stec 2012), the type of semantic information conveyed (Beattie & Shovelton 1999), and many more such factors (see McNeill 1992, 2000c). Of course, one should be cautious not to conclude from this body of research that gestural expression is even remotely as crystallized as the verbal component of language. McNeill (1992) is not wrong in stating that speech and gesture are quite different in terms of their degree of discreteness and autonomy as a semiotic channel. In spite of these differences, however, it stands to reason that gestures demonstrate some degree of systematicity on various levels of analysis.

3.1.2 Language-specificity of gesture
A question of particular interest to typology-oriented grammarians is to what extent gestural expression is language-specific. A substantial amount of cross-cultural work has been conducted to address this question (for a review, see Kita 2009). One rather robust finding is that speakers of different languages have distinct repertoires of emblematic gestures (Efron 1972; Ekman & Friesen 1969; Kendon 1992; Payrató 1993; Sparhawk 1978). In Dutch culture, for example, indicating that the food that is one is eating is tasty is done by rapidly waving a hand next to the cheek, but in Italian, this is done by twisting the tip of the index finger back and forth into the cheek. Comparative inquiry suggests that the way people use their body for making deictic reference is not stable across cultures either. Many different formal variants of index finger points are known (Kendon & Versante 2003; Wilkins 2003), as well as cases of pointing by means of the lips (Enfield 2001; Sherzer 1973) and the nose (Cooperrider & Núñez 2012). In certain indigenous languages, the use of gestures appears strongly entrenched as a means of managing discourse perspective (Sandoval 2014) or as part of the numeral system (De Vries 1994).

Intercultural differences in gesturing have furthermore been observed on a more abstract level, for instance in the way people use
their hands while presenting spatial descriptions. Whereas in most cultures people use their own body as a reference point when using gestures to represent spatial scenes, records exist of communities that deploy an absolute frame of reference when doing so (Haviland 1993; Levinson 2003). These findings, notably, are only relevant to the current discussion to the extent that they reflect communicative conventions rather than conventions in imagery. It is important to be aware that regularities in gestural expression do not always reflect linguistic patterns, but could instead emerge from commonalities in the way people conceptualize their environment (Okrent 2002; Parrill 2010).

Nonetheless, the existence of culture-specific, communicative patterns in gestural expression is evident, especially when it comes to emblems and indexical gestures. At least to some extent then, co-verbal expression involves learned and partly arbitrary knowledge, manifested in codified symbols as well as more schematic expressive patterns.

3.1.3 Functional and structural interfaces between speech and gesture

Both in terms of the cognitive processes underlying their expression (McNeill & Duncan 2000) as well as in terms of their role in communication (Enfield 2004, 2009; Kendon 2004; Streeck 2009), speech and gesture are best thought of as strongly intertwined rather than as autonomous mechanisms. As Kendon (2004: 127) puts it, “[i]n creating an utterance that uses both modes of expression, the speaker creates an ensemble in which gesture and speech are employed together as partners in a single rhetorical enterprise” (emphasis original). Accordingly, one might argue that gestural expression should not be treated as a mere addendum to the grammar of the verbal channel, but as an integral part of linguistic structure. Grammar, in other words, should be thought of as multimodal.

For this claim to be a substantitive one, it is important to refine what is meant by the notions of ‘grammar’ and ‘multimodal’. Fricke (2008, 2012, 2013) lays out a number of motivations and theoretical assumptions required for treating grammar as a multimodal system. She notes that multimodality can be interpreted in two ways: as pertaining to the medium (e.g. visual or auditory) or to the ‘mode of expression’
(e.g. text, image). When dealing with the speech-gesture connection, both types of multimodality play a role: speech is acoustic (medium) and textual (mode), whereas gesture is visual and kinesic.\(^{11}\) A second question addressed by Fricke is how to determine whether the simultaneous expression of speech and gesture is multimodal, as opposed to merely ‘multimedial’. A defining criterion of a multimodal system, according to Fricke, that it shows signs of code manifestation and/or code integration. In simplified terms, code manifestation is the expression of a functional or grammatical category in two different modes; code integration is the phenomenon whereby two modes are combined as parts of a single message.

With respect to speech-gesture multimodality, the literature has shown many examples of both of these phenomena. In an experimental study by Hadar and Krauss (1999), participants were asked to annotate spontaneous gestures in terms of their ‘lexical affiliate’ – the word or phrase they relate to most. The authors report that gestures were more often considered co-expressive with verbs, nouns and prepositions than with other grammatical categories (see also Chapter 7). Not only single words and phrases, but also syntactic relations can be co-expressed by gestures, however. So et al. (2009) report that the location in which iconic co-speech gestures are produced often mirrors basic semantic relations (who did what to whom). In addition, recent work has shown that gestures produced in concurrence with speech reflect higher-level grammatical functions. Verbal expressions of grammatical aspect (Duncan 2002), negation (Harrison 2009), modality (Schoonjans 2014a) and plurality (Bressem 2012) tend to go along with specific gestural patterns in certain languages.

A clear example of what Fricke (2008) calls code integration is the case where gestures intersect with the syntax of speech. Gestures that occur at the moment speech is being interrupted may function as a substitute for verbal elements and take over their syntactic function (Enfield 2009; Ladewig 2012; Slama-Cazacu 1976; Streeck 1993). In a comprehensive analysis of this phenomenon, Ladewig (2012) finds that

\(^{11}\) Compare, for instance, the simultaneous processing of speech and text, which involves two media but only a single mode of expression.
gestures most often occupy noun and verb slots in utterance-final position. Exactly which type of grammatical element a gesture corresponds to cannot always be conclusively determined, however. The gesture in (13), discussed by Ladewig (after McNeill 2005: 5), can for instance be interpreted as replacing either a non-finite ing-clause (e.g. flying out the window) or a prepositional phrase (e.g. into the garden or like this).

(13) Sylvester went... [gesture of an object flying out laterally]

A special case of code integration discussed by Streeck (2002) and Fricke (2012) is cataphoric integration, where a specific verbal item serves as a point of connection between the two modes. The most obvious examples are demonstratives and qualitative deictics (e.g. so and son in German and like this in English), which create the expectation of a deictic reference or qualitative description. In certain cases, the performance of a gesture even seems obligatory, in the sense that omitting it would render the utterance infelicitous. Utterances like look over there, and other forms of spatial deixis, are only felicitous when accompanied by a pointing gesture of some form.

Another type of code integration is the case where gestures provide semantic content while being produced in concomitance with the verbal channel. Fricke (2013) reports that gestures, when they coincide with speech, often take the role of an attribute to a nominal phrase. They may either refine verbally presented information (14) or present additional information (15).

(14) A small wooden sculpture [opposing the index finger and thumb to give a more precise qualification of the size of the sculpture, thus elaborating the meaning of the adjective small]12

(15) A small wooden sculpture [tracing a line in the air to show the contours of the shape of the sculpture in the air (a property which is not expressed verbally)]

12 These are fictive examples which serve to illustrate the arguments presented. In the following chapters, similar phenomena will be discussed in more detail on the basis of attested examples from a video corpus.
It follows that manual gestures, including those that are not themselves conventionalized, can be closely intertwined with the grammar of speech. Fricke’s criteria of code manifestation and code integration are clearly satisfied by various types of spoken-gestured phenomena. These findings call for an approach to grammar that accommodates, or at least acknowledges, the linguistic functions of co-speech gestures.

3.2 Current approaches
In spite of the predominant orientation towards written language, the linguistic potential of gesture has not gone unnoticed in the grammar literature. Protagonists of Systemic Functional Grammar were among the first to examine the multimodal dimension of grammar, first concentrating on text-image combinations (e.g. Kress 2000; Kress & van Leeuwen 1996), and later taking into consideration the relation between spoken language and gesture (Martinec 2004; Muntigl 2004). Other accounts adopted the perspective of Construction Grammar (Andrén 2010; Schoonjans 2014a; Turner & Steen 2012; Zima 2014), Eisenberg’s grammar (Bressem 2012; Fricke 2012; Ladewig 2012; Ladewig & Bressem 2013), Head-Driven Phrase Structure Grammar (Alahverdzhieva 2013), FDG (Connolly 2010; Kok 2016) and CG (Cienki 2012; Kok & Cienki 2016; Langacker 2001, 2008a; Wilcox 2004; Wilcox & Xavier 2013). Before providing an in-depth discussion of the latter two, I will briefly discuss previous work on gesture from the perspective of Systemic Functional Grammar, Construction Grammar and phrase structure-based models.

3.2.1 Systemic Functional Grammar
Halliday’s (1985) systemic functional approach, although primarily designed as a tool for text analysis, was among the first that took semiotic systems other than written language into serious consideration. Kress & Van Leeuwen’s (1996) ’grammar of image’ advances a social semiotic analysis, inspired by Halliday’s tripartite model of language (among other sources), of the structure of images and visual design. The authors entertain the view that images can be decomposed into different functional strata, and provide numerous examples of systemic analyses of text-image assemblies. Muntigl (2004)
and Martinec (2000, 2004, 2011) devote attention to the speech-gesture nexus from a systemic point of view. Muntigl (2004) approaches the issue primarily from a theoretical perspective, asking how SFG could be adapted to model the interaction between verbal and gestural elements of expression. He takes issue with McNeill’s arguments for calling gestures ‘non-linguistic’, and speculates on several types of content-expression pairs that govern gestural expression. On the basis of a brief literature review, he moreover contends that gestures have the potential to manifest the three main ‘metafunctions’ of language recognized by Halliday (1985): they can present ideational material (e.g. refer to entities and the relations between them), textual meanings (theme-rheme contrasts, e.g. projecting the upcoming utterance), and interpersonal meanings (e.g. mitigating the importance of an utterance). Brookes’ (2005) analysis of quotable gestures further attests that Hallidayan terminology can be helpful in analyzing functional analogies between the verbal and gestural components of language.\footnote{Section 3.3.1 and 3.3.2 of the current chapter evaluate how FDG’s functional levels and layers may receive gestural expression.}

Martinec (2000, 2004, 2011), in a similar vein, develops the hypothesis that gestural expression is guided by a fixed network of formal and functional choices. In his (2011) article, he argues that the inherent meaning of certain gestures is best comparable to that of ‘low-delicacy’ items of language, i.e. lexemes (or other structures) with a relatively general or abstract meaning. The fact that gestures typically require context to be interpretable does not preclude them from the language proper, since the same is true for many verbal items:

The decontextualized meaning of lexical items in language appears to be something of a myth in any case. Most lexical items in a dictionary have more than one meaning that are often defined in relation to different contexts, and this principle is reflected in, for example, Fillmore’s (e.g. 1977, 1982) lexical semantics. Without knowing such context, the meaning of a gesture for someone whose awareness might be drawn to it by
an analyst is not 'The player kicked the ball' but 'someone did/does something'. (Martinec 2011: 15)

Drawing on this view, Martinec (2000, 2004, 2011) proposes a taxonomy of the types of semantic processes that can be expressed by bodily actions, as well as the relevant formal features of the hands and arms. Contrastive features of indexical gestures, in this model, include the selection of the articulator (hand, forearm, fingers), the type of movement (directed vs. undirected) and the direction of the movement. On the ideational side, Martinec stipulates a basic distinction between kinetographic (action-related) and iconographic (image-related) meanings (based on Efron 1972) as well as contrasts between actions, states and locations and between directed and non-directed processes. Because empirical backup for the generalizability of the model is rather meager (only a handful of qualitative examples are reported as support for the claims), it cannot be seen as more than a preliminary proposal. Together with Muntigl's theoretical assessment and Brookes' (2005) functional analysis of emblems, however, this work shows how a systemic functional linguistic perspective brings some of the 'linguistic' properties of gestures – stability of form-meaning mappings and some form of compositionality – to the forefront.

3.2.2 Construction Grammar
Possible connections between gesture research and Construction Grammar (mostly with reference to Goldberg 1995) have caught the attention of both gesture researchers and grammarians. McNeill (2002) suggests that the notion of grammatical construction is useful for modeling how 'growth points' (primitive idea units) develop over time into spoken-gestured utterances (see Breslow et al. 2010 for a computational implementation of this idea). In language development research, the idea that linguistic units may combine verbal and gestural elements has a long history (e.g. Tomasello et al. 1984). The utility of the Construction Grammar framework for describing patterns of (multimodal) language development was already alluded to by Tomasello (1992). It receives more significant attention in the work of Andrén (2010), who applies notions from Construction Grammar to
account for the finding that the speech-gesture ensembles expressed by young children display various degrees of combinatorial structure.

In addition, recent years have seen an upsurge of interest in ‘multimodal constructions’ in adult speech, where the focus is on systematic co-expression of specific gestural and verbal patterns. Schoonjans (2014a, 2014b), for instance, proposes that certain German modal particles are best conceived of as multimodal in nature. One of his findings is that the particle einfach ‘just, simply’ co-occurs remarkably often in concurrence with ‘pragmatic headshakes’ (McClave 2000) and carries out similar functions. Both the particle and the pragmatic headshake are known to express subjective obviousness; they can be used to indicate that something is simply the only possibility. This finding is interpreted as evidence for a multimodal construction, characterized by verbal as well as gestural features. Schoonjans (2014b: 338-344) considers the alternative of regarding the particle and the gesture as separate constructions problematic, because this view ignores the fact that the verbal and gestural elements serve as optional components of each other’s formal and functional characterization. That is, to better capture the systematic interrelation between the two structures, it is compelling to postulate a single construction.

In recent years, various other researchers have adopted Construction Grammar as a way of framing observed correlations between verbal and gestural patterns. The availability of large collections of video data promises continuation of this line of research (Turner & Steen 2012; Zima 2014). It is worth noting, however, that similar findings have been described without explicit recourse to Construction Grammar. For instance Müller et al. (2013) describe a set of gestures with recurrent formal and discursive features, but make no explicit appeal to Construction Grammar. Thus, at least for some correlational phenomena, the added value of Construction Grammar as an explanatory framework can be called into question.

### 3.2.3 Phrase structure-based models

A third, in some ways more classical, approach is to consider how gestures may be integrated in models of phrase structure. Given Fricke’s (2008) observation that speech and gesture can combine into a single
matrix code (see Section 3.3.3), exploring the potential of syntax-oriented models to incorporate co-verbal behaviors could be worthwhile. In her account of multimodal noun phrases, Fricke (2012) appeals to the grammatical framework of the German linguist Eisenberg (1999) (later followed by Bressem 2012; Ladewig 2012). The formalisms used in this framework resemble the syntax trees found in generative linguistics, supplemented with notations for functional relations (e.g. subject, object, attribute). Ladewig (2012) shows how Eisenberg’s formalism can help to illustrate how gestures can take over the syntactic position and function of a verbal constituent in the case of interrupted speech (Figure 3.2).

![Phrase structure tree](image)

Figure 3.2 A gesture integrated in a phrase structure tree, adopted from Ladewig (2012: 91)

In the analyzed example, the speaker interrupts her speech after uttering *Ich wollte dieses* ‘I wanted this/that’ and at that point performs a gesture that represents a vertically oriented object. Ladewig entertains the possibility that the gesture is conditioned by the demonstrative pronoun *dieses*, and serves as a substitute for the noun.

Phrase structure-oriented models have also been used in computational approaches to multimodal grammar. In a recent dissertation, Alahverdzhiyeva (2013) employs Head-Driven Phrase Structure Grammar (HPSG) to model the conditions under which spoken-gestured utterances are grammatically ‘well-formed’. She
postulates a number of construction rules that guide the integration of speech and gesture into a unified utterance. These rules, largely based on previous literature and findings from a corpus study, pertain to the functional relation of the gesture in relation to the semantics of the verbal channel, as well as to the timing of the gesture relative to prosodic peaks in the speech. If these rules are satisfied, the model produces “a single syntactic tree for the multimodal action” (Alahverdzhieva 2013: 85). Overall, phrase structure-based models live up to the aspiration of multimodality better than, for instance, Martinec’s (2004) model of gesture-internal structure, which has in essence remained unimodal (gesture-only).

3.2.4 Shortcomings of current models
We have seen that current approaches to the grammar-gesture connection have somewhat complementary points of focus. Systemic functional approaches pay close attention to gesture-internal structure and emphasize the functional analogies between different semiotic channels, Construction Grammar approaches focus on the properties of specific spoken-gestured patterns, and phrase structure approaches primarily concentrate on the syntactic integration of gestures in a spoken clause. All provide support for the claim that gestures have some degree of linguistic potential, and contribute to the increasing acknowledgement of multimodal grammar as a field of study. There are, however, some ways in which current approaches fall short. First off, many writings on multimodal grammar seem to take for granted that the assumptions that underlie the model employed are compatible with what is known about gesture. Although crucial for justifying claims on the multimodality of grammar, many current accounts do not offer an in-depth discussion of the points of convergence between the gesture literature and the theoretical underpinnings of the model of interest. This sometimes results in rather bold assumptions about the analogies that exist between verbal and gestural elements of expression. When using phrase-structure models, for instance, one cannot blatantly assume a one-to-one mapping between a gesture and a word or grammatical category, as gestures might map onto multiple units simultaneously (see Chapter 4).
Second, many current approaches have remained somewhat programmatic: numerous scholars have made the claim that gestures can or should be considered part of grammar, but there is a general paucity of detailed application of grammatical frameworks to actual multimodal data. Furthermore, current proposals of multimodal grammar are often limited to the referential or attributive functions of gestures. The functional side of Eisenberg’s grammar, for instance, is largely restricted to traditional syntactic functions (e.g. subject, object, complement) and can therefore only account for a limited set of gestural functions (and only a few ways in which gestures can intersect with speech, leaving out factors such as illocution and discourse management). Especially in the cognitive-functional paradigm, one would expect a much broader coverage of gesture types and their points of interaction with the functional-stratal organization of speech. A final drawback of many current approaches is that they inherently prioritize speech over gesture. Phrase-structure based grammars, in particular, typically take verbal syntax as a starting point. This can be partly justified by the fact that speech is often the dominant channel of communication, but problems may arise when accounting for cases where speech and gesture are mutually dependent (see Chapter 8 for examples).

The remainder of this dissertation aims to overcome these drawbacks. In the current chapter, I present an in-depth theoretical analysis of the relation between gestural expression and grammar as conceived by FDG and CG. This review, in combination with the empirical findings reported in subsequent chapters, motivates the application of FDG and CG to a diverse range of spoken-gestured utterances (Chapter 8).

### 3.3 Functional Discourse Grammar and gesture research

FDG combines a distinct number of features which make it an interesting candidate for further evaluating the grammar-gesture nexus. One interesting aspect of FDG, as seen in Chapter 2, is that the model is designed to simultaneously include semantic and pragmatic factors in grammatical analysis. Since gestures often carry out representational and interactive functions in tandem (Müller 2013), this is an important
prerequisite for an adequate functional model of multimodal language structure. Second, FDG allows for grammatical analyses to be carried out in a comprehensive manner. In contrast to constructionist approaches to grammar, FDG’s basic unit of analysis is a full interactional Move (i.e., a unit of interaction, e.g. a conversational turn) rather than an individual grammatical element or construction. This enables close scrutiny of the function of a gesture in relation to a detailed functional-structural analysis of the verbal channel. Finally, FDG provides a very rigorous analytical formalism that enforces a high degree of explicitness in its analyses.

Nonetheless, the application of FDG to multimodal data is still in its infancy. Connolly (2010), in the first paper to discuss FDG’s multimodal potential, concludes that FDG is in principle compatible with a multimodal conception of discourse, provided that it is understood as a branch of semiotic theory. Like many other assessments of the gesture-grammar interface, however, Connolly’s analysis has remained rather programmatic and discusses only a few ways in which gesture (and other non-verbal modes of communication) can be relevant to the study of grammar. With the aim of achieving a more detailed view, the remainder of the present section evaluates a number of connections between gesture research and language structure as conceived by FDG. The following subsections show that the various ways in which co-speech gesture has been studied align well with the main layers of pragmatic and semantic representation postulated in FDG. These are summarized in Table 3.1 as a reminder.

Table 3.1. A summary of the most important semantic and pragmatic layers recognized in FDG

<table>
<thead>
<tr>
<th>Interpersonal level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of representation</strong></td>
</tr>
<tr>
<td>Move</td>
</tr>
<tr>
<td>Discourse Act</td>
</tr>
</tbody>
</table>
| Illocution | “the conventionalized interpersonal use [of a Discourse
Act] in achieving a communicative intention” (Hengeveld & Mackenzie 2008: 68)

Communicated Content "the totality of what the speaker wishes to evoke in his/her communication with the addressee” (i.e., the actional dimension of expressing a Propositional Content; Hengeveld & Mackenzie 2006: 376)

Ascriptive Subact “an attempt by the Speaker to evoke a property” (i.e., the actional dimension of expressing a Lexical Property; Hengeveld & Mackenzie 2006: 376)

Referential Subact “an attempt by the Speaker to evoke a referent” (i.e., the actional dimension of referring to some entity; Hengeveld & Mackenzie 2006: 376)

<table>
<thead>
<tr>
<th>Level of representation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propositional Content</td>
<td>“mental constructs, such as pieces of knowledge, beliefs, and hopes” (Hengeveld &amp; Mackenzie 2006: 377)</td>
</tr>
<tr>
<td>Episode</td>
<td>“sets of States-of-Affairs that are thematically coherent” (Hengeveld &amp; Mackenzie 2006: 378)</td>
</tr>
<tr>
<td>State of Affairs</td>
<td>“entities that can be located in relative time and can be evaluated in terms of their reality status” (Hengeveld &amp; Mackenzie 2008: 166). Also includes what is called the Configurational Property in Hengeveld and Mackenzie (2006: 379-380), i.e. the internal organization of a State of Affairs.</td>
</tr>
<tr>
<td>Lexical Property</td>
<td>A (lexically expressed) characteristic of a semantic unit (Hengeveld &amp; Mackenzie 2008: 215-236)</td>
</tr>
<tr>
<td>Individual</td>
<td>“a concrete, tangible entity” (Hengeveld &amp; Mackenzie 2008: 236)</td>
</tr>
</tbody>
</table>

In the following sections, I argue that all of the layers of pragmatic and semantic organization summarized above can receive gestural expression. Motivated by this discussion, I conclude with a proposal for a multimodal extension of FDG.
3.3.1 Gesture at the Interpersonal Level

Gestures contribute in various ways to the interpersonal dimension of language use, as follows from experimental (e.g. Bavelas et al. 1992), semiotic-communicative (e.g. Kendon 2004) as well as praxeological research (Streeck 2009). The comprehensive architecture of FDG makes it possible to bring together these resources into a single coherent framework. Here I outline how some of the interpersonal-pragmatic functions of gestures can be interpreted with respect to FDG’s layered hierarchy of pragmatic organization (see also Kok 2016).

An interactional Move can be realized by a gesture alone, for instance when waving one’s hands to greet someone or nodding to approve something. When a Move is expressed by speech, gestures are oftentimes deployed to designate its beginning or end (along with verbal markers). Streeck (2009), for instance, describes how speakers mark the end of a conversational turn – which may encompass one or more Moves – by spreading out their arms toward the addressee, as if literally handing the turn over to the interlocutor.

Gestures with a so-called parsing function are relevant with respect to Discourse Acts. These are gestures that “contribute to the marking [...] of the structure of spoken discourse” (Kendon 2004: 225). When summing up a list of points or arguments, for instance, speakers often keep track of the overall structure of the discourse (the respective organization of Discourse Acts, or clusters of them) by holding up one hand and outstretching one finger at a time in rough temporal correspondence with the points or arguments presented.

Several examples are known of conventionalized gestures that contribute to the Illocution layer (Kendon 1995; Payrató 1993). A well-documented gestural complement of interrogatives, for example, is the ‘purse hand’ in Southern Italian discourse, whereby “all the digits of the hand are held fully extended, but they are drawn together so that they are in contact with one another at their tips” (Kendon 1995: 249). Kendon also elaborates on the ‘joined hands’ gesture, which he says marks the discourse “as an appeal to others to take a certain line of action” (Kendon 1995: 261). This gesture can be analyzed as operating on imperatives, transforming what would otherwise come across as an order into a request or plea. Thus, it has a mitigating role.
With respect to the Communicated Content layer, the close connection between gesture and information structure is of interest. Some have claimed that gestures are a direct expression of high communicative dynamism, appealing to Firbas’ (1964) definition as the extent to which an utterance “pushes the communication forward” at some point (Levy & McNeill 1992; McNeill 1992: 207). Accordingly, some gestures may best be analyzable as manifestations of Focus or Comment functions operating on Communicated Contents. Drawing once again on analyses of Southern Italian discourse, Kendon (1995) furthermore characterizes the ‘finger bunch’ gesture as a marker of topic-comment oppositions in the accompanied speech. A characteristic of this gesture is that the fingers are first held together and subsequently spread out, closely coinciding with the discourse segments that serve topic and comment functions, respectively. The Contrast function, furthermore, may receive rather specific gestural coding in the form of a weighing gesture, where two Communicated Contents are opposed by holding the two hands in front of the body with the palms faced up, shaped like cups (McNeill & Levy 1982).  

Gestures (or aspects of them) pertinent to the layer of the Referential Subact include those intended to make a referent identifiable to the addressee. In the case where someone points while asking Could I have this one? (e.g. in a store), the pointing gesture can be analyzed as expressing a Referential Subact (with an identifiability operator), which is co-indexical with the Referential Subact expressed in speech. Gestural correlates of existential constructions can be relevant at this layer as well. Mittelberg & Mortelmans (2013) report that constructions such as there is in English and es gibt in German are often accompanied by a ‘palm-up-open-hand’ gesture. This recurrent gesture is known to typically “present an abstract discursive object as inspectable […] and it invites participants to take on a shared perspective on this object” (Müller 2004: 233). The performance of a palm-up-open-hand gesture in combination with an existential construction, according to this

---

14 Note that the expression on the one hand… on the other hand… plausibly has its roots here.
characterization, can be hypothesized to co-express (or further refine) the focal status of the Referential Subact.

Gestural components of Ascriptive Subacts, finally, subsume those that express the speaker’s subjective commitment to an ascribed property. Certain types of shoulder shrugs and hand gestures are known to be attuned to this purpose (Debras & Cienki 2012). The ‘swaying gesture’ (Bressem & Müller 2014), whereby the lax flat hand is repeatedly turned relative to the wrist, can be seen as a manual equivalent of approximation markers. When indicating the size of an object by holding the hands at a certain distance from each other, for example, holding the hands firm and tensed hints at a fair amount of self-reliance, whereas holding the hands lax and twisting them around the wrist conveys a lesser degree of confidence that the ascribed property is accurate. In these cases, the position of the hands can be compared to a (Lexical) Property at the Representational Level, but the degree of tension and stability of the hands relates to the Interpersonal Level at the layer of the Ascriptive Subact.

3.3.2 Gesture at the Representational Level

A substantial body of gesture research is concerned with the semantic functions of gesture, i.e. those that relate to the Representational Level in FDG. Despite the fact that studies on gestural semantics typically draw on ad-hoc categories (e.g. Beattie & Shovelton 1999) or semantic constructs from cognitive linguistics (Cienki 2005), it is possible to classify gestural patterns in terms of their relation to FDG’s representational layers.

To begin with, head nods in response to yes-no questions may be analyzed as Propositional Contents. Akin to the words yes and no, head nods can constitute entire Propositional Contents, assigning a truth value to some proposition. Furthermore, some manual gestures relate to the speaker’s degree of commitment to the proposition that is expressed verbally. One of the functions of the swaying gesture, mentioned above, is to “mark [...] ideas as uncertain or indeterminate” (Bressem & Müller 2014: 1581).

In addition, recent research has hinted at a gestural correlate of Episodes. According to Müller et al. (2013), the performance of multiple
gestures in close spatial or temporal proximity engenders the mental construction of a coherent semantic scenario (see also Chapter 6). When speakers gesture while describing a series of temporally or spatially connected events, they typically do not return their hands to a rest position in between successive gestures. In line with the definition of Episodes in FDG, the physical continuation of kinesic expression in such cases indicates that different states of affairs form a temporally or spatially coherent whole, as in (16).

(16) I first drove east and then traveled back up north [tracing a line in the air, first horizontal, then vertical, without returning the hands to rest position in between].

Gestures that relate to States of Affairs include those that convey information with respect to "the properties of their occurrence" (Hengeveld & Mackenzie 2008: 171). The relative temporal occurrence of two States of Affairs, for instance, can be reflected by the temporal ordering of the two gesture strokes, as in (16). Gestures that co-express negation, typically characterized by the horizontal outward movement of one hand with one or multiple fingers stretched (Harrison 2009; Kendon 2004), may pertain to this layer as well. Other aspects of gestures relevant here are those that correlate with grammatical aspect and thereby expose the internal structure of the State of Affairs. Various studies report that speakers of English make longer and more complex gestures when producing utterances with progressive-marked aspect than with perfective aspect (Duncan 2002; McNeill 2003; Parrill et al. 2013).

Lexical Properties and Individuals, finally, can be represented gesturally in many ways. Müller (1998a) distinguishes a number of strategies people use for accomplishing iconic reference to an object or person, for instance by molding its shape, tracing its contours or embodying the object so that the hand becomes the object itself. By virtue of this iconic potential, gestures are particularly adept at representing information related to shape and motion (17), and size (18).
The rock bounced down the cliffs [moving a fisted hand through the air to trace the rock's trajectory] (17)

I just saw a huge rat! [holding the hands apart at a certain distance, palms facing each other] (18)

Interestingly, Individuals and their corresponding Properties are often encoded in one and the same gesture. In (17), for instance, the hand represents an Individual (the rock) by embodying it, but at the same time (co-)ascripts two Properties to it (shape and spatial trajectory). This aspect of gestural expression is further investigated in Chapter 4.

3.3.3 Gesture and Encoding

As yet, the structural mechanisms underlying gestural expression are not understood in sufficient detail to make an explicit and detailed proposal on what a gestural Encoding component of FDG entails (i.e. the part of the grammar model that transforms semantic and pragmatic representations into a readily expressible form). Given the intrinsic differences in the communicative channels that speech and gesture exploit, however, it is clear that an adequate multimodal grammar requires multiple, distinct levels of form.

Despite the absence of a single, unitary formalism for describing gesture form, structure-oriented approaches have alluded to the existence of gestural 'primitives'. Some have advocated the view that gestural expression involves a number of organic kinesic patterns and handshapes (e.g. curves, loops, lines and planes), which are consistently linked to specific semantic contexts (Calbris 1990 ch. 3; 2008; Webb 1996). Provided that certain formal facets of gestures can be recombined to create novel meaningful patterns, Calbris (1990, 2011) and others suggest that gestural expression is at least partly compositional by nature (although the kind of compositionality involved here is not of the same linear-segmental nature as verbal syntax).

To get a better grip on gestural compositionality, form-oriented approaches to gesture have adopted the four phonological parameters from sign language studies: handshape, orientation, location and movement (Stokoe 1960/2005). Gesture notation systems based on these elementary form features have proved helpful in identifying
recurrent patterns in spontaneous gesticulation (Müller et al. 2013). The observation that spontaneous gestures overlap in their formation parameters has also given rise to the suggestion that gesture form is governed by a “rudimentary morphology” (Müller 2004: 3) or an “emerging morphosemantics” (Kendon 2004: 224).

A different line of research has focused on the articulator-referent mappings that underlie iconic representations in spontaneous gesticulation. Based on systematic analyses of the forms of referential gestures in relation to the spatial characteristics of their referents, it has been suggested that creative iconic gestures are not as unconstrained as previously supposed (Bergmann & Kopp 2010; Hassemer 2015; Poggi 2008). The formal make-up of iconic gestures can be modeled in terms of generative (Poggi 2008) as well as probabilistic grammatical formalisms (Sadeghipour & Kopp 2014).

In terms of their temporal-sequential organization, it has furthermore been argued that gestures follow a kind of hierarchical constituency structure (Fricke 2012, 2013). According to Fricke (2013: 743-744), the way in which gesture phases (e.g. the preparation of the gesture, the stroke phase, and the retraction of the hand into a rest position; Kendon 1972) combine into bigger units can be captured by a limited number of “syntactic rules”, which are hierarchical and recursive. Whether the notion of syntax can be extended so seamlessly to the domain of co-verbal expression, however, remains a topic of controversy.

A final notable finding is that encoding processes for speech and gesture are in close interaction with each other. A series of studies by Özyürek and colleagues (Kita 2003; Özyürek et al. 2005) shows that simultaneous speech-gesture combinations are sensitive to the typological characteristics of the language in which they occur. Speakers of verb-framed languages (Talmy 1985), which express spatial information on manner and path in separate clauses, tend to break up manner and path into two consecutive gestures. Conversely, speakers of satellite-framed languages, which express manner and path in a single clause, were found to conflate manner and path in one and the same gesture.

Altogether, at least four types of primitives can be hypothetically assigned to a gestural Encoding module: (1) lexeme-like gestural
patterns such as emblems, (2) basic phonological parameters (handshape, orientation, location and movement) that can be combined productively, (3) morpheme-like patterns such as planes, curves and lines, and (4) mimetic modes underlying iconic mappings. Nevertheless, various questions and points of dispute remain with respect to the findings presented in this section. Despite the existence of some degree of compositionality, it is irrefutable that gestural meaning is strongly constrained in a top-down fashion, i.e. by verbal and situational context. This to some extent sets gestural patterns apart from the verbal primitives postulated in FDG, such as lexemes and syntactic templates, which have stable grammatical and/or semantic properties that are relatively independent of the context in which they are attested.

3.3.4 Initial observations

The explorations outlined thus far yield some preliminary observations. First of all, there appears to be a difference in the types of gesture that relate to interpersonal layers and those pertaining to representational layers. Gestures that carry out interpersonal functions are almost without exception conventionalized to some degree – they are either emblematic or classifiable as a recurrent gesture. Lower layers at the Representational Level (e.g. Individuals or Properties), by contrast, more often involve creative forms of gesticulation. A plausible explanation for this divergence is that the hands have a natural potential for creating iconic mappings with respect to semantic aspects of conceptualization, but much less so with respect to interpersonal aspects. That is, the hands can perform referential and ascriptive functions without drawing on a conventionalized code, but by exploiting iconic relations between their physical form and the spatial features of the referent. This applies much less to interpersonal functions, which generally do not have inherent spatial or motoric characteristics that easily allow for iconic representations. The pressure for gesture types with interpersonal

---

15 Some pragmatic gestures, such as the ‘weighing gesture’ described above, are iconic in establishing a metaphoric mapping to a concrete spatial domain. However, such gestures are still conventionalized to some degree; they do not rely on iconicity in the same way as gestures with a referential or ascriptive function.
functions to become conventionalized may therefore be stronger than for those with a semantic function.

Second, the (grammatical) functions of gestures are often dependent on what is conveyed verbally (Feyereisen et al. 1988). Gestures serving as the head of a formulation layer are rare, with the exception of those embodying full Moves or Propositional Contents. In the examples discussed, gestures predominantly provided further qualifications of (or operations on) verbally encoded material.

It should be noted that the close alignment between FDG’s layers and the ways gestures have been studied in the literature does not necessarily imply that gestural expression is organized according to the exact same structure as the verbal channel. On the one hand, not all grammatical distinctions relevant to the verbal component of the grammar receive distinctive encoding in gesture. For instance, the finding that particular kinds of grammatical aspect co-occur with longer and more complex gestures (Duncan 2002) does not entail that the structure of gestural expression respects the aspectual distinctions as recognized by FDG. Conversely, gestures may encode semantic or pragmatic variables that are not expressed verbally. Whereas German does not have grammatical markers for Episodes, for instance, Episode-like semantic units can be marked by German speakers’ gestures (Müller et al. 2013; Chapter 6 of this dissertation).

### 3.3.5 A first sketch of a multi-channel FDG

From the literature reviewed, it follows that an appropriate multimodal extension of FDG entails separate, but mutually interactive formulation and encoding operations for gesture and speech. The architecture of such a model can be inspired by psycholinguistic models of speech-gesture co-production (e.g. De Ruiter 2000; Kita & Özyürek 2003), which have extended Levelt’s (1989) model of language production with a parallel component for gesture. Along the same lines, Figure 3.3 presents a possible architecture of a multimodal extension of the FDG model, following Connolly’s (2010) suggestion to treat the different modalities as constituting separate, but mutually interactive semiotic systems. Additions to the current version of the FDG model are marked with dashed lines.
The Conceptual Component remains in essence unchanged relative to current FDG. It contains the totality of the interactional and ideational material that is to be translated into ostensive signs (manifested through the auditory or the visual-manual modality). The Grammatical Component here can be understood in a broad, inclusive sense, namely as comprising the totality of socially shared knowledge for enabling contextualized meaning-making through whatever modality.

Note that the postulation of a single, multimodal grammar is not meant to imply that the operations underlying speech and gesture generation are of the same kind. Instead, as discussed above, each modality has its own operations and draws on a separate set of primitives, with distinct structural features and levels of schematicity.

The initial, rough division of communicative labor over verbal and manual modalities is determined during Macro-Planning. The
reason for including this operation in the Grammar is the ostensible existence of multimodal constructions in certain languages, i.e., fixed speech-gesture combinations that form a single expressive unit (see 3.2.2). Further distribution of labor over speech and gesture can be assumed to be established on the fly during Formulation, as a result of mutual interactions between the Speech Formulation and Gesture Formulation operations (cf. McNeill & Duncan 2000). This distribution is in part determined by the format of the conceptual content activated: conceptual material of a spatial-motoric nature is a more natural candidate for being expressed manually, whereas more abstract types of content are most readily encoded by verbal means (Kita 2000).

The Speech Encoding and Gesture Encoding operations further elaborate the output of the Formulation operations into a form that is encodable in speech and gesture, respectively. Possible primitives that this operation draws on have been discussed in Section 3.3.3. In keeping with the finding that gestural encoding of information flexibly adapts to the syntax of the verbal channel, a tight interaction can be assumed between the two Encoding components. Additional configurational templates may be required here to explicate under which conditions the sequential combination of speech and gesture is acceptable, e.g. why an utterance like (19) is generally more felicitous than (20). Further elaboration on such templates remains outside the scope of this chapter.

(19) Then the car went [tracing the trajectory of the car in the air]
(20) ??He [gesture depicting the action of walking] to the shop

The Contextual Component, finally, subsumes all elements of context that have a systematic impact on verbal and/or gestural form. This includes such factors as the (real or fictive) presence of a referent in the immediate environment, the physical position of the interlocutor, the assumed common ground and the gesture space previously used. Some of these factors are specifically relevant to gesture, whereas others may influence speech and gesture expression in similar ways.

It should be kept in mind that Figure 3.3 is not to be conceived of as a processing model in and of itself. Although inspired by psycholinguistic models of speech-gesture production, it is meant as a framework for multimodal grammatical analysis. The strong
resemblance to a production model is in accord with FDG’s philosophy that a grammatical framework is most effective when modeled (roughly) after a processing architecture (Hengeveld & Mackenzie 2008: 2-3). As follows from this tenet and the empirical evidence synthesized above, a multimodal FDG needs to reflect at least the following three empirical facts: (1) speech and gesture are components of one and the same communicative process; (2) speech and gesture are mutually interactive during formulation and encoding; (3) yet, speech and gesture involve partially independent primitives and operations. The architecture proposed in Figure 3.3 explicitly incorporates all three of these empirical facts.

An alternative architecture of a multimodal FDG was recently proposed by Alturo et al. (2016). As seen in Figure 3.4, this version does not contain a parallel structure. It embodies the assumption that spoken and gestured modes of expression draw on a shared set of frames and templates, which restrict the ways in which verbal and co-verbal elements can be combined. Like in the model proposed above, the assumption is that speech and gesture have separate sets of primitive forms and operators available for Formulation and Encoding.

The assumptions behind the two models are largely compatible; both agree that speech and gesture are governed by a single grammar. The main point of divergence seems to be whether the emphasis is on the semiotic dissimilarity between speech and gesture, or on their unity in communication. The single-tier structure proposed by Alturo et al. (2016) emphasizes the connectedness of speech and gesture as elements of the process of utterance. It helps to avoid the interpretation that speech and gesture operate as autonomous systems (this is not implied by my proposal, but I am aware that the postulation of separate modules for the two tiers can give this impression). The parallel architecture I propose in Figure 3.3, on the other hand, reminds us that speech and gesture are semiotic systems that are considerably different in nature (linear-segmental vs. visual-kinesic, etc.). More importantly, the model provides a motivation for the operational choices in Chapter 8, where I analyze spoken-gestured utterances in a three-fold manner, with two separate analyses for the contributions of the two tiers, and one for the utterance as a whole.
According to Cognitive Grammar’s usage-based perspective on language structure, the realm of the linguistic need not be limited to verbal expression. Several scholars have discussed the potential of speakers’ gestures to attain grammatical status. Langacker, himself, already alluded to the potential of certain gestures to be part of a linguistic system, maintaining that any type of expressive behavior can in principle become entrenched as a symbolic structure (Langacker 1987). This potential receives further attention from Langacker (2001, 2008a, 2013), and is evaluated in more detail by Cienki (2012, 2014, 2015). With respect to signed languages, Wilcox (Wilcox 2004; Wilcox & Xavier 2013) and Liddell (2003), among others, show that CG provides useful analytical constructs for detailing the relation between (ASL) grammar and iconicity. Others adopt CG terminology in their discussions of sequential speech-gesture compositions (Enfield 2004; Ladewig 2012) and the phenomenon of gestural iteration (Bressem 2012).

Despite the many connections between CG and gesture research made by these scholars, CG has thus far not developed so as to adopt multimodality in its basic design, nor has it become clear what it would entail to incorporate speakers’ gestures in actual cognitive grammatical

Figure 3.4 An alternative architecture of FDG, adopted from Alturo et al. (2016).

3.4 Cognitive Grammar and gesture research

According to Cognitive Grammar’s usage-based perspective on language structure, the realm of the linguistic need not be limited to verbal expression. Several scholars have discussed the potential of speakers’ gestures to attain grammatical status. Langacker, himself, already alluded to the potential of certain gestures to be part of a linguistic system, maintaining that any type of expressive behavior can in principle become entrenched as a symbolic structure (Langacker 1987). This potential receives further attention from Langacker (2001, 2008a, 2013), and is evaluated in more detail by Cienki (2012, 2014, 2015). With respect to signed languages, Wilcox (Wilcox 2004; Wilcox & Xavier 2013) and Liddell (2003), among others, show that CG provides useful analytical constructs for detailing the relation between (ASL) grammar and iconicity. Others adopt CG terminology in their discussions of sequential speech-gesture compositions (Enfield 2004; Ladewig 2012) and the phenomenon of gestural iteration (Bressem 2012).

Despite the many connections between CG and gesture research made by these scholars, CG has thus far not developed so as to adopt multimodality in its basic design, nor has it become clear what it would entail to incorporate speakers’ gestures in actual cognitive grammatical
analysis. The aim of this section is to provide a more comprehensive overview of the points of convergence between CG and gesture than is currently available, and to advance the incorporation of gesture into CG as appropriate.

### 3.4.1 Points of convergence

In exploring the potential of CG as a framework for understanding the grammar-gesture relationship, at least three aspects of the theory are of particular interest. First of all, CG views grammar as emergent from actual communication, without posing restrictions on the kind of behaviors that may constitute a linguistic system. Second, CG holds that grammatical meaning resides in conceptualization – a cognitive process that has often been hypothesized to receive expression in manual gestures as well. Third, CG draws on the assumption that many grammatical notions reflect cognitive representations that are spatial in character, which aligns with the inherently spatial expressive domain of manual gestures. In the following sections, I synthesize current literature on these and related points of convergence.

### 3.4.2 Gestures as symbolic structures

As discussed in Chapter 2, CG holds that virtually any aspect of language use may become entrenched as part of a grammatical structure, as long as it is apprehended as a common denominator in multiple communicative experiences. The semantic side of a usage event does not only correspond to the imagery (in the technical sense of CG, not limited to visual imagery) underlying the situation described, but also involves the discourse context and aspects of the interaction itself (e.g. patterns in turn taking). Likewise, on the expressive side, grammatical structures may include “both the full phonetic detail of an utterance, as well as any other kinds of signals, such as gestures and body language” (Langacker 2008a: 458). This broad conception of the (potential) grammar proper is expressed by Langacker’s (2001) representation of usage events as comprising multiple ‘channels’ at each pole (Figure 3.5).
What follows from this view is that whether or not elements of expression qualify as linguistic does not depend on the modality through which they are expressed. Rather, the grammatical potential of co-verbal behaviors is to be assessed according to their degree of entrenchment as symbolic structures in an individual’s mind and the degree of conventionalization of those symbolic structures within a given community (Langacker 2008a; Zima 2014). Taking this continuous view as a starting point, the picture emerges that some gestural behaviors are clearly candidates to be part of the linguistic system (as a general category, or of a given language; Cienki 2012), whereas others are more on the periphery of that system.

3.4.3 Degrees of schematicity in gestural form and meaning

Following CG’s usage-based perspective, Kendon’s (1980) classification of different gesture types can be reappraised in terms of their status as grammatical units. When applying CG’s view of linguistic elements – as ranging on a continuum from specific to more schematic structures – to gestures, the status of emblems is rather clear-cut. The use of emblems is to some extent language-specific, with different repertoires being employed by different cultural communities (Efron 1972; Payrató 1993). Hence, this category of gestures can be assumed to consist of symbolic units that are both entrenched and conventionalized. For most other categories of gestures, the grammatical status is only evident on a higher level of abstraction. Even for gestural behaviors that may seem rather systematic, such as manual pointing gestures, substantial variation in
Speakers’ gestures in FDG and CG

form and meaning exists. Whereas pointing gestures often involve an outstretched index finger (in most Western cultures), there appear to be no strict constraints on the shape of the hand and the degree of tension in the fingers. Likewise, in terms of their meaning, pointing gestures may serve a variety of functions: they may be used to draw the interlocutor’s attention to some contextual or abstract referent; to resolve referential ambiguity, or to give or request the turn in a conversation (Bavelas et al. 1992; Kita 2003). From a CG perspective, this variability in form and meaning across different contexts “correlates with what cognitive grammarians believe is the schematicity of grammatical meaning” (Lapaire 2011: 95). Pointing gestures thus can be said to manifest grammatical structures that are, as a category, slightly more schematic (i.e., more contextually variable) in form and meaning than emblems.

The grammatical status of the category of ‘recurrent gestures’ or ‘gesture families’ (Kendon 2004; Müller 2004) can be analyzed in similar terms. The cyclic gesture, as discussed above, does not have a fully fixed form, but rather “constitute[s] what can be called a family resemblance category of phonological structure” (Cienki 2015: 506). That is, different instances of cyclic gestures can be seen as more or less prototypical exemplars of this gesture type. In terms of their meaning, cyclic gestures are dependent on the context in which they are performed, but are nonetheless bound to a limited set of discourse situations. Ladewig (2011) characterizes this semantic commonality among cyclic gestures in terms of an Idealized Cognitive Model (ICM; Lakoff 1987) that encompasses the image schema CYCLE and a number of metaphorical extensions. Because some degree of commonality in form and meaning exists between instances of these gestures, they can be qualified as symbolic structures, with phonological and semantic poles that are largely schematic (Cienki 2012).

A final category of gestures, for which the potential grammatical status is even less obvious, is that of creative gesticulation: those gestures that are creatively generated during speech production and do not instantiate a type of emblem or recurrent gesture (note that the term ‘creative gesticulation’ as used here is narrower than Kendon’s (1988) category of gesticulation, which includes certain recurrent
patterns as well). Because these gestures do not appear to conform to clear standards of form, McNeill (1992) characterizes them as ‘idiosyncratic’. However, some common denominators exist between instances of creative gesticulation. For instance, speakers make consistent use of a small number of ‘representation techniques’ when using the hands to refer to objects and processes (e.g. molding the shape of the object, or tracing its outline in the air; Müller 1998a). Each of the representation techniques could be considered as a central characteristic of a schematic class of gestures. As Enfield (2013) argues with respect to tracing gestures:

> It may be that there are conventions which allow interpreters to recognize that a person is doing an illustrative tracing gesture, based presumably on formal distinctions in types of hand movement in combination with attention-directing eye gaze toward the gesture space. While the exact form of a tracing gesture cannot be pre-specified, its general manner of execution may be sufficient to signal that it is a tracing gesture. (Enfield 2013: 701).

In terms of their semantics, creative gestures are strongly context-dependent, but certain commonalities nevertheless exist. One recurrent feature, shared by most instances of creative gesticulation (as well as some other gesture types), is the function of modifying or highlighting some aspect of the spoken discourse. Further, there is overlap with the category of ‘speech-linked gestures’ that form part of a performed quotation or impersonation (and I was like + [gesture]). In such cases, as Cienki (2015: 505) argues, “there is the schematic form-meaning structure in place whereby such words call for some kind of depiction or illustration in order for the speaker’s point to be adequately expressed (as if the word had a slot that needed to be filled by a gesture).” Accordingly, it seems inadequate to fully dismiss creative gesticulation from the realm of (cognitive) grammar. Rather, this category can be

---

16 Also note that this term does not imply that the speaker has the intention of being creative.
considered to manifest a type of very schematic grammatical structure, sharing only very few phonological and semantic features.

In conclusion, as summarized in Table 3.2, different types of gestures can be thought of as ranging on a continuum from relatively fixed to much more flexible symbolic structures.

Table 3.2. Summary of the degrees of specificity/schematicity in the form and meaning of different types of gestures

<table>
<thead>
<tr>
<th>Gesture type</th>
<th>Emblems</th>
<th>Pointing gestures</th>
<th>Recurrent gestures</th>
<th>Creative gesticulation (as defined above)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Form</strong></td>
<td>Largely fixed within a given culture</td>
<td>More or less fixed within a given culture, with some degrees of freedom</td>
<td>Common 'formational core', but strong variation within categories</td>
<td>Flexible, but not unrestricted - bound to high-level norms and constrained by the grammar of the verbal channel</td>
</tr>
<tr>
<td><strong>Meaning</strong></td>
<td>Largely fixed within a given culture</td>
<td>Variable, but with clear commonalities - typically related to attention allocation</td>
<td>Only characterizable in abstract terms and/or with constructs such as image schemas or ICMs</td>
<td>Mostly context-dependent; generally associated with some modification of the verbal channel or emphasis of some aspect of it</td>
</tr>
</tbody>
</table>

This continuum aligns well with the CG perspective that verbal structures such as words, morphemes and syntactic constructions can be placed along a continuum of schematicity. As follows from the discussion above, categories of co-speech gestures are not in principle different from one another or from verbal structures, but rather analyzable as a set of structures that range from fixed to more variable and schematic in their form and meaning. A view of this kind may help
to answer the question whether gestures are to be seen as subject matter for linguistic theory. To echo Langacker’s (2008b) view on this matter, some gestures are clearly part of a grammatical system, others less so.

3.4.4 Gestures and CG’s dimensions of construal
A second important tenet of Cognitive Grammar is that language structure reflects conceptualization. This provides another line of convergence with the gesture literature. It is consistent with the hypotheses that gestures emerge from imagistic processes during ‘thinking for speaking’ (McNeill 1992; McNeill & Duncan 2000), ‘conceptualizing’ (De Ruiter 2007), or ‘visual-motor simulation’ (Hostetter & Alibali 2008). More specific parallels between CG and gesture research can be noticed in relation to what Langacker (1987) puts forward as the dimensions of construal that receive grammatical expression (see Chapter 2). These include specificity, perspective, focus-background and the postulation of cognitively basic semantic structures. This section outlines how each of these notions relates to speakers’ gestures.

3.4.4.1 Specificity
When communicating about the world, any referential situation needs to be construed with a certain level of specificity. The same object may, for instance, be described as a thing, or a small object, or a corkscrew, depending on the specificity of the predication (Langacker 1987: 118). The same holds for gestural expression: because our hands are not clay that can be reshaped however we please, referential gestures are inherently underspecified. Gestures that represent objects are therefore always only able to iconically show some part of the whole, and pointing gestures indicate a salient point to refer to the whole referent entity or space (Cienki 2013a).

An example of a gesture on the most specific end of the spectrum is that for ‘telephone’ – whereby the pinky and thumb are extended and the other fingers folded – which has a rather restricted domain of possible referents (namely, a telephone, perhaps even just the handset of
Speakers' gestures in FDG and CG | 101

a landline phone). Gestures with a grip handshape, whereby all five fingers are bent as if holding something in the fist, are much less specific: they most likely (but not exclusively) refer to small-sized, round or cylindrical objects. A type of gesture on the most schematic end is the palm-up-open-hand gesture (Kendon 2004; Müller 2004). Palm-up-open-hands may, via metonymy, refer to any type of object that might be held on the hand, either concrete (e.g. the physical referent of the co-expressed noun phrase) or abstract (e.g. a position in a debate).

3.4.4.2 Subjective and objective construal

Elements of linguistic expression can involve different degrees of recourse to the ground, i.e. the contextual circumstances of the linguistic interaction. Those that are construed without reference to the ground are described in CG as objective (e.g. the conceived spatial relation evoked by the lion is next to the rock); those that are contingent upon the physical location of the conceptualizer (or another aspect of the ground) have a higher degree of subjectivity (e.g. the conceived spatial relation evoked by the lion is in front of the rock).

Similar construal options have been discussed with respect to the use of referential gestures (McNeill 1992; Parrill 2010). Whereas pointing gestures typically assume a reference frame relative to the physical location in which the interaction takes place (e.g. ‘go this way’ + [pointing gesture]), other gestures are more objectively construed in the sense that the current communicative situation is irrelevant to their interpretation (e.g. the statue is shaped like this + [tracing gesture]). This difference is further captured by the proposed distinction between gestures with a character perspective and those with an observer perspective.

17 The intended meaning of a phone handshape may of course go further than mere reference to a phone, for instance when used as a full pragmatic move, as to say ‘I’ll call you’. The current argument is concerned only with a simple, referential use of this gesture (e.g. as performed while telling a story that involves the action of picking up a phone).

18 Note that neither of these gestures has an absolute reference frame: both draw upon the spatial configuration of the hand of the speaker relative to the rest of the body. Where they differ, however, is in whether the real-space physical position of the body matters to the interpretation of the utterance.
perspective (McNeill 1992). In the latter case, for instance when tracing the path of the car by moving the hand through space along a certain trajectory, the speaker construes a situation in a relatively objective manner. A more subjective construal of the same scene can be achieved when a character perspective is adopted, for instance when one mimics the action of driving in a car by impersonating hand movements of the driver holding the steering wheel.

3.4.4.3 Focus and the stage model
In CG, foreground-background distinctions are seen as vital to grammatical organization. As discussed in Chapter 2, CG conceives of meaning as residing in the relation between a profile, i.e. the center of attention, and a conceptual base, i.e. the immediately relevant context and background knowledge. This does not only apply to word meaning (the noun rim is only meaningful against the background of the concept of something having such an edge, such as a WHEEL) and syntactic relations (e.g. a verb phrase profiles a process, but requires as its conceptual base a scene that involves an agent), but can also help elucidate the way speech and gestures relate to each other semantically. As Bressem (2012) notes, the profile of a verbal expression often serves as part of the base with respect to the gesture. That is, speech often provides the relevant background frames against which the meaning of the gesture is elaborated (Figure 3.6). A gesture whereby a circle is drawn in the air, for instance, is inherently underspecified, but attains a more specific referential meaning when co-expressed with the noun rim.

Gestures may furthermore specify the verbally presented content. The different representation techniques people use for making iconic reference – molding, tracing, embodying or enacting (Müller 1998a, 2014) – profile different aspects of the referent. While referring to a rim, a gesture whereby the hands draw its contours in the air profiles its shape and outline; when using the hands as if interacting with the rim, one profiles its physical affordances.

The trajector-landmark distinction that holds between primary and secondary participants within the focus domain can also have a gestural counterpart, as demonstrated in Enfield’s (2004) description of the ‘symmetry-dominance construction’ in Lao. Enfield describes a
3.4.4.4 Gesture and basic conceptual categories

In its description of grammatical classes, CG holds that language structure reflects patterns in basic human experience, or ‘conceptual archetypes’, through which we view the world. This way of semantically characterizing grammatical categories can provide further insights into the relation between linguistic elements expressed through speech and those conveyed with the hands. In sign language research, it was adopted in order to characterize the semantics of manual expression. With respect to ASL classifier predicates, Wilcox (2004: 127) argues that “the things of Cognitive Grammar are mapped onto handshape, and process is mapped onto phonological movement” (emphasis original). In line with this view, it is possible to characterize those gestures that refer
to objects and represent their physical and/or kinesic properties in terms of their semantic relation with basic grammatical notions (Ladewig 2012). All gestures that perform the reification of some conceptual content (by construing some referent as a **THING**) evoke a conceptual structure that overlaps with that of nouns. Gestures whereby the hands of the speaker (in addition) depict some static property of the referent are in correspondence with the semantic domain of adjectives: they profile a **NON-PROCESSUAL RELATIONSHIP** between the referenced entity and some other conceptual structure. In cases where the hand moves through space to represent the motion of some entity, the gesture can be said to have a verb-like character, as it designates a **PROCESSUAL RELATIONSHIP** (Figure 3.7).

<table>
<thead>
<tr>
<th>Notation</th>
<th>Schematic characterization</th>
<th>Grammatical class</th>
<th>Related semantic potential of gestures</th>
</tr>
</thead>
<tbody>
<tr>
<td>semantic pole</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>THING</th>
<th>NON-PROCESSUAL RELATIONSHIP</th>
<th>PROCESSUAL RELATIONSHIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schematic characterization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical class</td>
<td>Noun</td>
<td>Adjective</td>
<td>Verb</td>
</tr>
<tr>
<td></td>
<td>The capacity of gestures to reify, e.g. by pointing or representing an (abstract) object</td>
<td>The capacity of gestures to attribute a non-temporal feature to some object, for instance by depicting its size, shape or location</td>
<td>The capacity of gestures to depict a temporal process, for instance when moving the hand through space as to mimic actual movement</td>
</tr>
</tbody>
</table>

Figure 3.7 A CG representation of the semantic poles (in their most schematic form; adopted from Langacker 1987) of three major grammatical classes and the features of gestures that relate to them.

Two possible caveats with respect to the proposed parallels can be noted at this moment. First, since gestural expression is not linearized in the same way as speech, it is surely possible that multiple conceptual
categories are simultaneously evoked by a single gesture. In fact, it is hard to imagine a gesture that depicts the shape of an object without simultaneously performing reification (see Chapter 5; Kok et al. 2016).

Second, it should be noted that not any movement of the hand(s) evokes a representation of a process. As Ladewig (2012) notes, the movement of the hands may instead be part of the act of reference, for instance when referring to a physical object by tracing its contours in the air. The gesture in this case can still be seen as an object-process synthesis, where the hand is some drawing utensil and the movement represents the act of drawing, but it serves a meta-referential function: not the act of drawing but the drawn object is part of the situation communicated.

3.4.4.5 Autonomy - dependence
CG distinguishes autonomous and dependent structures – the former can be described in their own terms, whereas the latter presuppose the support of another structure. A dependent structure, in other words, “refers schematically to an autonomous, supporting structure as an intrinsic aspect of its own characterization” (Langacker 2008a: 199). For an adequate characterization of the relationship between speech and gesture, the concept of autonomy-dependence alignment is of great importance. Dependence between the two tiers is pervasive in multimodal language use, and exists in two directions. Verbal constructions, in particular those involving deictic reference (e.g. look over there) or iconic specification (e.g. the German so einen X; ‘such a X/a X like this’; Fricke 2012, 2013), require the presence of some gesture (presumably a pointing gesture in the first case and a representational gesture in the second). Conversely, many manual gestures articulated during speech are functionally contingent upon the semantic and pragmatic frames evoked verbally (Wilcox & Xavier 2013). Depending on the verbal context, a cyclic gesture may, for instance, represent a cyclic movement of some object, but it may also indicate that the speaker is searching for the right words (Ladewig 2011).

It is worth emphasizing that autonomy-dependence is not always a strictly asymmetrical relationship. In the case of multimodal demonstrative utterances like the man over there [+ pointing gesture], speech and gesture are mutually dependent: the verbal component does
not only presuppose the performance of a deictic gesture, but the meaning of the gesture is at the same time elaborated by the speech (the gesture could perhaps have pointed to the woman standing next to the man). This symbiotic character of speech-gesture compositions is in accordance with Langacker’s characterization of the autonomy-dependence relationship as variably asymmetrical:

Canonically the structures in a valence relation manifest substantial asymmetry, with one of them (on balance) clearly dependent, and the other autonomous. As always, though, recognition of the prototype must not be allowed to obscure the existence of other possibilities. Nothing in the definition precludes a relation of mutual dependence between two structures, or guarantees that there will always be a significant relation of dependence in one direction or the other (Langacker 1987: 300).

Thus, CG does not predict a simple dichotomy between gesture-compatible and gesture-incompatible structures. There may instead be a continuous range of degrees to which gestures and verbal constructions presuppose the presence of one another.

3.4.5  Gestures and CG diagrams
In addition to the many theoretical points of convergence with gesture research, CG offers a notation system that is potentially interesting for representing aspects of gestural expression. Cognitive linguistic models like CG explicitly address language’s relation to spatial cognition in their representational tools. The frequent employment of diagrammatic tools creates a natural point of connection with the inherently spatial expressive domain of manual gestures (cf. Tversky et al. 2009). This potential has already been demonstrated by the natural integration of gesture research with the notion of image schemas (Cienki 2005) and Talmyan semantics (Hassemer 2015), but has not been elaborated in great detail in CG.

Nevertheless, as Cienki (in press) notices, “there is great potential for incorporating schematic images of relevant gesture forms
as part of the phonological pole [...], images which, through their form and orientation, would also inherently show the perspective of the speaker's construal." Liddell (2003) demonstrates that CG's notation conventions can be useful for analyzing the phonological and semantic structure of ASL signs. The benefit of these diagrams over text-only notations is most obvious when it comes to displaying iconic and indexical relations between phonological space and semantic space, for example when the position or movement of the hands is isomorphic to the conceptualized position or movement of some entity. This potential is demonstrated in more detail in Chapter 8.

3.5 Remaining challenges

It is clear from the previous sections that both FDG and CG have a great potential to incorporate co-verbal behaviors as part of their subject matter. However, a number of challenges need to be addressed in order to further develop the application of these models to gestural behaviors. This section first discusses a broad, overarching problem: the (re)definition of the notion of grammatical relevance. In the sections that follow, this problem is broken up into three, more granular issues: understanding the ways in which gestures are multifunctional, defining type-token relations and accounting for the dynamic dimension of speech-gesture compositions.

3.5.1 (Re)defining grammatical relevance

A commitment shared by FDG and CG, although made much more explicit in the former, is that aspects of linguistic meaning and function are only of concern to grammatical theory to the extent that they have an impact on formal expression. When applied to the verbal channel, this notion can already be a subject of debate: it is not always possible to determine exactly what conceptual and contextual factors determine the form of a given utterance. When examining co-verbal behaviors from a grammatical perspective, the notion of grammatical relevance becomes

---

19 Liddell's analyses, which extend CG notations with constructs from conceptual blending theory, differ from the ones presented in this dissertation, which draw on CG notions only.
even more difficult to operationalize. The formal characteristics of gestures are sensitive to a wealth of different factors, including such matters as the distance between the interlocutors, the social status of the addressee, the physical properties of the referent and the presence of physical constraints or utilities (e.g. holding a beverage or a map). Because gesture forms emerge from a complex interplay between communicative, biological and other types of constraints, it is not easy to single out those aspects of gestural expression that reflect some semantic or pragmatic structure.

A related difficulty is that not all co-verbal behaviors are necessarily addressee-oriented. That is, considerable asymmetry can exist between gesture production and comprehension. Certain forms of gestural expression may primarily serve to ‘externalize’ imagery and other cognitive processes onto one’s body (Okrent 2002; Pouw et al. 2014). People may for instance use their hands to facilitate word retrieval (Krauss et al. 2000), or to scaffold internal cognitive processes such as counting and numerical reasoning. Although these are, strictly speaking, functions of gestures, such behaviors do not seem good candidates for inclusion in grammatical analysis (pursuing incorporation of self-oriented functions of bodily behaviors in a linguistic theory entails a massively broadened, possibly even boundless, conception of grammar). However, even if a hand movement is not intended as part of a communicated message by the speaker, it might be interpreted as such by the addressee. To give a somewhat contrived example: at the moment a bartender sees a customer counting on his fingers before making any eye contact, he might already have understood the number of drinks that the customer wants to order. Whether the customer’s finger counting is to be regarded as a linguistic action is open to interpretation. Conversely, communicatively intended gestures are not always picked up on by the addressee. People’s disposition to integrate gesturally and verbally

---

20 As discussed in more detail in Chapter 8, some gestures might combine self-oriented and other-oriented functions. They might for instance help the speaker retrieve a lexical item and simultaneously signal to the addressee that he or she intends to ‘hold the floor’.
presented information has been shown to differ from person to person (Wu & Coulson 2014).

A possible way of dealing with this issue is to follow clear diagnostics for distinguishing speaker-oriented actions from communicative ones. A number of salient physical characteristics of “utterance-dedicated” actions have already been formulated by Kendon (2004: 11) and Enfield (2009), among others. These include certain kinesic features of the gesture as well as the direction of eye-gaze and body posture. Although these heuristics can be useful, the challenge of deciding which forms of manual expression are to be considered subject matter for grammatical analysis does not appear to have a trivial solution. Especially for FDG, the relative lack of strongly systematic, ‘black-and-white’ features of gestural expression is difficult to reconcile with the rigid distinction between coded and non-coded aspects of meaning.

In the following, the problem of determining grammatical relevance is split up into three separate issues, which will be addressed independently. These challenges are (1) how to deal with the ubiquitous multifunctionality of gestures, (2) how to define type-token relationships in gestural expression, and (3) how to model the grammatical relevance of the temporal coordination between the two channels. These issues are briefly discussed in the next section, and will be addressed empirically in the chapters that follow.

3.5.2 Multifunctionality

Some grammar-oriented work on gesture gives the impression that the meaning or function of a given gesture is always clear-cut and easily interpretable. In reality, however, most naturalistic gestures do not bear a clear correspondence to a distinct meaning or function. Instead, gestures are semiotically rich units of expression, capable of making complex, multi-layered contributions to an utterance (Enfield 2009, 2013). Whether this presents an insurmountable problem for FDG or CG remains to be seen. Both models (FDG in particular) are likely to subscribe to Dik’s assertion that
[A] functional explanation of grammatical phenomena will typically not be based on an assumption of simple form-function correlations, but will instead involve a network of interacting requirements and constraints, each of which may be understood in functional terms itself, but which interact in complex ways and in a certain sense ‘compete’ for recognition and expression in the final design of linguistic expressions. (Dik 1986: 17-18)

Provided that the same rationale applies to a functional account of co-verbal behaviors, one need not assume that each gesture is linked to a single function only. Therefore, a closer understanding is needed of the potential multifunctionality of gestures in situated communication. Chapter 4 of this dissertation reports on an empirical study aimed at gaining further insights into this issue.

3.5.3 Defining type-token relations
As far as gestures can be broken down into isolable form-function mappings, it remains an open question how these are to be defined. In FDG terms, this translates into the question of determining what the gestural ‘primitives’ are that speakers draw upon. In CG terms, the question is what kind of symbolic structures gestures instantiate, and how their phonological and semantic poles are to be represented. A difficulty here, as pointed out by Enfield (2013: 697), is that studies of gesture often focus on the meaning of a gesture as situated in a particular context, without necessarily classifying the gesture as an instance of a particular type (as analogous to a lexical entry or grammatical category). In order to avoid inconsistency in the levels of description for the verbal and gestural tier, it is important to define a token-type relation for every gesture analyzed, so that the verbal and gestural components of an utterance can be described in terms of their ‘type meaning’ (i.e. categorical) as well as their ‘token meaning’ (enriched by, and specific to the current context).

Finding the optimal set of categories for classifying gesture types for the purpose of grammatical analysis remains an unsolved problem. There is no widely established classification system for identifying gestures as tokens of types. Instead, it is common practice among
gesture researchers to tailor one’s classification scheme to the research question asked. Because some types of gestural expression (e.g. emblems) have clearly definable characteristics, whereas others have features in common only on a rather abstract level of analysis, it is furthermore important to postulate gestural categories on different levels of abstraction. For instance, some form-meaning mappings might be most adequately captured in terms of individual form parameters (e.g. hand orientation, location, etc.), others in terms of more holistic patterns (e.g. tracing gestures) and yet others in terms of theoretical constructs such as image or action schemas (Cienki 2005, 2013b; Wilcox 2004). In Chapters 5 and 6, I report on two experimental studies aimed at understanding what types of type-token relations exist for gestures performed during route description.

3.5.4 Temporal coordination as a dimension of linguistic structure

The accommodation of gesture in a grammatical model may revive the tension between pattern-oriented (static) and process-oriented (dynamic) views on grammar (Butler 2008a; Nuyts 2008). Everett (in preparation), for instance, provocingly asserts that “dynamic human gestures interact in complex ways with static human grammars to produce human language” (emphasis in original). A multimodal perspective on language, however, does not necessarily imply such a dualism. Like speech, gesture is produced dynamically and modulated on-line, but can be analyzed and represented in a static, pattern-based fashion.

A more substantial challenge follows from the fact that the real-time coordination of speech and gesture can influence the meaning of the multimodal utterance as a whole. It has often been observed that gestures are typically produced right before the element of speech to which they are most closely related (Habets et al. 2011; Kendon 1970; McNeill 1992), but this characterization is rather rudimentary. Various open questions remain with respect to the rules and constraints that govern the temporal dimension of speech-gesture composition. In fact, when modeling the structure of language as a dynamic multimodal activity, the notion of (morpho)syntax drastically increases in
complexity. Verbal and gestural elements of expression are typically not organized in a purely sequential fashion or constrained by well-defined rules or templates, but emerge from a dynamic arrangement of acoustic and visual-kinesic forms of expression.

As argued by Cienki (2015) from a CG perspective, patterns of temporal coordination ought to be considered as part of the phonological pole of the symbolic structures underlying speech-gesture compositions. It is questionable, however, to what extent CG’s images lend themselves to accurately represent the dynamic dimension of multimodal communication. McNeill (2005: 74) expresses pessimism with regard to CG’s utility in this respect, given its tight connection to the, inherently synchronic, Saussurean tradition. Cienki (2015: 511) notes with respect to CG’s notation system that “[a] trick that remains to be solved is how to display these analyses dynamically (with moving graphics), so as to better reflect the actual dynamic processes of expression and online thinking (or understanding) for speaking.” Indeed, to the extent that the temporal relation between the verbal and gestural tiers forms part of the grammatical structure of the utterance, the notation systems of both FDG and CG might need to be endowed with a more dynamic form of representation. The issue of dynamism will be addressed empirically in Chapter 7, which examines patterns in the relative timing between the elements of the two channels.

3.6 Summary
In this chapter, I have argued that a multimodal perspective on language structure can be mutually beneficial for cognitive-functional linguists and gesture researchers. The application of models of grammar to the structure of multimodal utterances, on the one hand, can provide gesture researchers with analytical tools for studying the speech-gesture relationship in a principled and rigorous manner. On the other hand, considering gesture as integral to language provides important insights for, and presents challenges to, existing models of grammar.

Both FDG and CG have a number of interesting features in view of understanding the grammar-gesture nexus in greater detail. On FDG’s side, we have seen that the explicit acknowledgment of language’s “omnipresent dual functionality” (Hengeveld & Mackenzie 2008: 30)
provides a useful analytical framework for studying points of connection between verbal and co-verbal expression. In fact, sections 3.3.1 and 3.3.2 suggest that gestures have the potential to modulate (or substitute) most, if not all of the main layers of pragmatic and semantic organization that FDG recognizes. Gesture’s role in multimodal utterance formation is often comparable to that of a lexical modifier or operator, but in the absence of speech gestures can also embody entire Moves. Although this analysis does not conclusively show that speech and gesture are sensitive to the same semantic and pragmatic factors, it corroborates the fruitfulness of an integrative, multimodal model of language structure. Motivated by these findings and by further connections between grammar and gesture research, this chapter has presented a sketch of a multimodal extension of FDG.

On the side of CG, a complementary set of theoretical and operational strengths can be observed. Due to its non-restrictive, usage-based nature, CG does not require fundamental amendments or a supplementary ‘gesture component’. Because CG conceives of grammar from the perspective of prototype theory, as having more central and more peripheral structures, it avoids having to make a rigid, arbitrary distinction between the linguistic and the non-linguistic proper. For spoken language, this aspect of the theory can be extended beyond orally produced sounds to include other behaviors. Moreover, there are numerous points of interface between gesture research and the dimensions of imagery postulated in CG (e.g. specificity, perspective), as well as a promising potential to incorporate gesture in CG’s diagrammatic notation system.

Various challenges need to be resolved in order to further advance the accommodation of gesture in either of the two models. Most importantly, further development is needed of FDG’s and CG’s analytical tools to cope with gesture’s variably systematic, multifunctional and dynamic nature. In order to make the proposed synergy between grammar research and gesture studies fully appreciable, these issues need to receive more thorough empirical attention. The next four chapters report on experimental and corpus-based studies that address the challenges highlighted here. Based on data from a large-scale video corpus, I present empirical insights into the multifunctionality of gesture
(Chapter 4), the existence of speech-independent mappings between gestural form and function (Chapter 5 and 6), and the timing of gestures relative to the grammatical items they relate to (Chapter 7). The outcomes of these studies motivate a number of key assumptions that serve to further develop the capacity of FDG and CG to deal with multimodal data.
Chapter 4. Mapping out the multifunctionality of speakers’ gestures

This chapter is devoted to one of the major challenges raised by the notion of multimodal grammar: understanding the ways in which gestures perform multiple linguistic functions at the same time. To address this issue, I first present a theoretical view on the inherent multifunctionality of speakers’ gestures, broadly inspired by frameworks in structural-functional linguistics. Then I report on a large-scale gesture perception study, designed in a way which is open to the potential for functional complexity of gestural expression. Because this study is based on group intuitions rather than interpretations of individual, trained analysts, it provides a relatively objective characterization of gestural functionality. After presenting a general exploration of the statistical trends in the data, I identify a number of clusters of functions that tend to be combined, as well as (positive and negative) correlations between pairs of functions.

---

21 Parts of this chapter have been published in Kok et al. (2016). For reasons of consistency with earlier chapters, the pronouns have been changed to first person singular.
Chapter 4

4.1 Background and motivations

As is evident from the literature reviewed in the previous chapters, gestures are functional elements of human communication. Over recent decades, various classification systems have been devised to capture their diverse functionality (e.g. Efron 1972; Lausberg & Sloetjes 2009; McNeill 1992; Wundt 1973). Having enabled the currently flourishing research tradition of quantitative inquiry, these classification systems have undoubtedly been of great value to the field of gesture research. However, the categorization of gestures into discrete functional types implicitly draws on the rather dubious assumption that gestures carry out only one function at any given time. Ecologically oriented gesture researchers have pointed out that such a ‘pigeonhole’ view on gestural functionality does not do justice to the multifarious nature of gesture use in natural discourse (e.g. Kendon 2004; Streeck 2009).

Although many gesture scholars acknowledge this problem, at least in theory, and are aware that the exhaustive classification of gestures into a small number of all-or-nothing categories is no more than a “convenient fiction” (Loehr 2004: 128), operational alternatives are sparse. Less discrete views, such as McNeill’s (2005: 40-41) proposal to regard iconicity, metaphoricity, deixis and temporal highlighting not as categories but as ‘dimensions’ of meaning that may be mixed in a single gesture, have rarely been adopted explicitly in quantitative research designs. Instead, solutions to the issue of gestural multifunctionality often go no further than the inclusion of mixed categories like ‘iconic-deictic’ in a coding scheme.

Paradigms based on semantic features (Beattie & Shovelton 1999, 2001; Bergmann et al. 2011; Bergmann & Köpp 2006; Holler & Beattie 2002) have expressed some acknowledgment of the communicative complexity of co-speech gestures, but generally leave out non-semantic functions in their design. Other studies have demonstrated that the ways the hands are used to refer to objects and actions are closely linked to pragmatic factors, such as information structure (Levy & McNeill 1992; Parrill 2010), the negotiation of common ground (Holler & Wilkin 2009) or the disambiguation of lexical elements (Holler & Beattie 2003). However, it is not clear to date how the (quantitative) extents to which gestures manifest different
Mapping out the multifunctionality of speakers’ gestures

functional categories simultaneously can be analyzed in a systematic fashion. In the following, I aim to bridge this gap by providing an empirically based characterization of gestural (multi)functionality, in a way that is inspired by, and compatible with, cognitive-functional models of grammar. Thus, I attempt to achieve some degree of convergence between ecological and experimental views on gesture functionality.

4.1.1 Multifunctionality in language and grammar

Functional approaches to language are characterized by the assumption that language structure reflects the goals of language users. Although no full consensus exists as to how these goals are best described, most functional accounts agree that language simultaneously serves a cognitive function (it organizes and represents thought) and a social function (it allows for coordination of one’s behavior with that of other people). This ambivalence was already recognized by Wilhelm von Humboldt (Humboldt 1903: 24), who concludes that “there lies in the primordial nature of language an unalterable dualism.”

Karl Bühler’s (1934/1990) Organon model, which has exerted great influence on modern functional linguistics, advances the view that language manifests a relation between three main components of communicative events: a Sender, a Receiver and a Referent. Linguistic functions can be characterized according to this model: aspects of an utterance that serve to describe objects, situations or mental states have a ‘representational’ function; those that reveal the speaker’s emotional state or attitude towards the situation described have an ‘expressive’ function; those that are aimed at directing or affecting others’ behavior or mental state perform an ‘appeal’ function. Crucially, these functions are not all-or-nothing, nor are they mutually exclusive. An important tenet in Bühler’s account is that all three functions are present to some degree in every utterance, albeit not always in equal salience. That is, any expression in some way reflects the Sender’s construal of the Referent, his or her affective evaluation of it, and the intended effect of
the message on the Receiver. From this point of view, all linguistic utterances are inherently multifunctional.

The work by Bühler and his successors has inspired various models of language structure. As discussed in Chapter 2, functional models of grammar (Dik 1978; Halliday 1985; Hengeveld & Mackenzie 2008; Van Valin 1993) hold that linguistic structures are best understood in relation to their functional contributions to the ongoing discourse. Systemic Functional Grammar (Halliday 1985), one of the most elaborate models in this field, pursues the view that language structure reflects the interplay between different ‘metafunctions’. This theory describes grammar on three levels: in terms of its role in organizing experience (the ideational metafunction); in terms of the various social relationships that are acted out through language (the interpersonal metafunction) and in terms of the way these functions are combined into text or speech (the textual metafunction). These metafunctions are further realized by a large number of ‘systems’, i.e. sets of choices available to speakers during language production. FDG also acknowledges that language structure simultaneously reflects pragmatic (interpersonal) and semantic (representational) factors, as we have seen in Chapter 2. Based on extensive typological inquiry, FDG furthermore recognizes a number of hierarchically organized layers on each of these levels of analysis. According to both models (and related ones), the surface form of an utterance emerges from the interplay of cognitive, social and contextual constraints on expression.

4.1.2 Functional models applied to gesture

Various connections and overlapping assumptions exist between functional linguistics and work by Kendon (e.g. 1980, 2004), Streeck (e.g. 2009), Enfield (e.g. 2009) and various other gesture researchers.

---

22 Utterances that do not have a referential meaning (e.g. expressives such as *Ouch!* are an exception to this.

23 Note that these three metafunctions do not map one-to-one onto Bühler’s functions. The ideational function corresponds to Bühler’s representational function. Both the appeal and expressive functions would fall under interpersonal in Bühler’s model.

24 Recall that FDG’s Interpersonal Level subsumes SFG’s interpersonal and textual metafunctions.
The work of these scholars concentrates predominantly on the meaning and function of gestures as situated in natural discourse. Nonetheless, explicit reference to, or adaptation of, the functional models discussed above is somewhat sparse. Müller (1998b Ch. 2; 2013: 214) adopts Bühler’s model to support the view that gestures can be understood in terms of their representational, expressive and appeal-related qualities. She maintains that gestures can fulfill each of Bühler’s functions as their primary purpose: “gestures are in principle functionally comparable to verbal signs: [...] they may re-present something other than themselves while at the same time expressing some inner state, being addressed towards somebody, and [...] executing speech acts and other communicative activities.”

As discussed in Chapter 3, some concrete attempts have been made at incorporating gestures in functional models of grammar. Martinec (2004) and Muntigl (2004), adopting a Systemic Functional Linguistic view, argued that some gestures can be analyzed in terms of a network of competing functional choices. The models they propose are rather programmatic, however, and based on too sparse data to validate their generalizability across gesture types. Taking Functional Discourse Grammar as a point of departure, recent work has provided further corroboration that a stratified functional model of language structure aligns well with the ways gestures have been studied in the literature (Connolly 2010; Kok 2016).

In line with McNeill’s (1992) characterization of gestures as ‘holistic’ complexes of meaning, these works suggest that no one-to-one mappings exist between gestures and their functions. Motivated by this consideration, the remainder of this chapter presents an empirical examination of the ways in which, and the degrees to which, speakers’ gestures are multifunctional. The design of the study is loosely inspired by functional linguistic models such as SFG and FDG. Its focus lies on those representational and interpersonal functions that occur in route directions, a type of discourse that is known to involve descriptive as well as directive communication (cf. Denis 1997; Lücking et al. 2013). In particular, the focus is on the functions listed in Table 4.1, discussed in 4.2.2 below, and their (co-)occurrence in a large corpus of natural and spontaneous direction-giving dialogues.
4.2 A crowdsourced approach to functional characterization

To study the functional side of gestural expression in a way that is compatible with the theoretical considerations presented in Chapter 3, the design of the experiment combines three innovative features. First, it incorporates multiple functional tiers. This makes it possible to capture the multifunctional complexity of gestures, instead of forcing the annotator to classify any gesture as performing only one function or another (cf. the LASG coding system for a comparable approach; Bressem et al. 2013). A second feature is that gestural functions are not treated as all-or-nothing. The use of Likert-scales instead of a binary coding scheme allows coders to express different degrees of certainty about their judgments (which, arguably, mirror the relative prominence of the functions expressed by a gesture). A third feature is that the functional characterization of the gesture stimuli is ‘crowdsourced’, i.e. performed by a large group of laypeople recruited over the internet. This helps to avoid reliance on the analysts’ subjective interpretation or their knowledge as trained scholars of gesture or language. Thus, the functional coding of the data can be interpreted as representative of the view of a broad community of language users. Before providing more details on the procedures for gathering the data, I introduce the corpus from which the stimuli were derived.

4.2.1 Corpus and stimuli

The stimuli were video snippets taken from the Bielefeld Speech and Gesture Alignment Corpus (Lücking et al. 2010, 2013). The construction of this corpus was initiated in 2007 as part of the “Speech-Gesture Alignment” project of the Collaborative Research Center “Alignment in Communication” (CRC 673). It consists of 25 German-spoken dialogues, spanning a total of 280 minutes of video, recorded from three camera positions. The task conducted by the recorded participants consists of two parts. First, one of the participants, the ‘route giver’, sits in front of a large video screen and watches an animation that makes it appear as if she is taking a tour through a town in a virtual reality landscape (SaGA town; Figure 4.1). The tour passes five landmarks: a sculpture, a church, a town hall, a chapel and a fountain. In the subsequent phase of the task,
the route giver is told to instruct the second participant, the ‘route follower’, to follow the same path through the town (unbeknown to the route giver, the route follower never actually has to do this). The total corpus contains nearly forty thousand words and nearly six thousand gesture units. The corpus was heavily annotated as part of the SaGA project, both on the level of speech (lemmatization, parts of speech, information structure) and gesture (gestures type, form parameters, modes of representation) (see Lücking et al. 2013). Although some of these annotation layers lend themselves to addressing specific linguistic questions, previous use of the SaGA corpus has had different aims, mostly oriented toward the design of virtual avatars and automated dialogue systems (Bergmann & Kopp 2009; Bergmann et al. 2010; Kopp et al. 2008; Lücking et al. 2010).

For the current study, a subset of this corpus was divided into small video clips. These were extracted from five dialogues, namely those for which the participants had given consent for the data to be used and presented. The start and end points of the snippets were determined by the nearest moment where the hands were in rest position or the speaker paused his or her speech. Thus, the stimulus videos contained relatively isolated units of discourse. Because the speakers often made successive gestures without returning the hands to the rest position, some video snippets contained more than one gesture. To allow each of the gestures of interest to be referred to individually, numbers were edited into the video during the stroke phases (i.e. the most energetic and/or marked phase of the gesture). Crucially, no a priori filtering was applied to the stimulus set: virtually all of the gestures performed by the route giver were treated as potential stimuli. The only exceptions were
complex gesture sequences with more than six strokes performed in quick temporal succession. These were discarded to prevent the stimulus videos from being too long for the online survey. The application of this procedure resulted in 174 videos, containing a total of 462 individually marked gestures. The average length of the videos was 10.9 seconds (SD = 3.50). For the current study, all videos were played with the sound on.

4.2.2 Functions of interest

We have seen numerous examples in Chapter 3 of linguistic functions that can be expressed gesturally. The analyses in this chapter are limited to a set of nine, listed in Table 4.1. This set of functions was selected on the basis of theoretical considerations (they relate to various levels of linguistic organization) and operational criteria (they have a high frequency in the corpus). That is, the list was prepared by assessing which of the functions that have been discussed in the literature on linguistic approaches to gesture recur with sufficient frequency in the SaGA corpus. As a consequence, it is not exhaustive: the existence of additional functions of gestures needs to be acknowledged (e.g. their role in discourse segmentation and expression of illocutionary force; Kendon 1995), but these do not occur frequently enough to be worthwhile including in the research design. Table 4.1 provides a minimal description of each of the functions of interest and gives examples of how these might be realized verbally and in gesture. It follows FDG’s dual view on language by organizing these in terms of representational and interpersonal (non-representational) functions.

Table 4.1 Examples of (sub-)functions of speech that can be realized or modified gesturally. These comprised the set of functions taken into consideration in the present study.

<table>
<thead>
<tr>
<th>Function</th>
<th>Examples in speech</th>
<th>Examples in gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refer to a concrete, tangible entity</td>
<td><em>The book</em>; <em>she</em>; <em>Mary</em></td>
<td>Certain pointing gestures; gestures that represent an object; catchments (reiterated reference to an object associated with a certain region)</td>
</tr>
</tbody>
</table>
Mapping out the multifunctionality of speakers' gestures

<table>
<thead>
<tr>
<th>Function</th>
<th>Examples in speech</th>
<th>In gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refer to a location</td>
<td><em>At your left-hand side; in London; here</em></td>
<td>Certain pointing gestures and catchments</td>
</tr>
<tr>
<td>Describe a physical property (e.g. size or shape) of an object or person</td>
<td><em>Big; round; shaped like the leaf of an oak</em></td>
<td>Gestures that depict the size or shape of a referent, e.g. by drawing its contours in the air</td>
</tr>
<tr>
<td>Describe the movement of an object or person</td>
<td><em>is spinning; turns left; rolls down</em></td>
<td>Gestures that trace an object's movement trajectory in space</td>
</tr>
<tr>
<td>Designate the amount of the referent</td>
<td><em>A; five; a couple of</em></td>
<td>Conventional number gestures (extending a specific number of fingers vertically); the use of one or two hands when referring to one or two objects</td>
</tr>
<tr>
<td>Locating an event in (real or fictive) space</td>
<td><em>[the accident happened] over here; [we watched TV] in the living room</em></td>
<td>Gestures that depict an event in a certain region of space; some pointing and placing gestures (e.g. a pointing gesture co-occurring with <em>the accident happened over here</em>)</td>
</tr>
</tbody>
</table>

**Interpersonal and meta-communicative functions**

<table>
<thead>
<tr>
<th>Function</th>
<th>Examples in speech</th>
<th>In gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal (lack of) commitment to the accuracy of some predication</td>
<td>Ten-<em>ish</em> people <em>Sort of big</em></td>
<td>Oscillating gestures (e.g. wiggling the lax hands at the wrist); shoulder shrugs</td>
</tr>
<tr>
<td>Signal importance of an element of the utterance</td>
<td>Word order; prosodic prominence; it-clefts: <em>It was John</em> who went to the university</td>
<td>Gestural expression in general (Levy &amp; McNeill, 1992); in particular certain beat gestures (e.g. those deliberately used in political speeches)</td>
</tr>
<tr>
<td>Indicate that one is having difficulty finding the right words</td>
<td>Filled pauses such as <em>uh</em> and <em>uhm</em></td>
<td>Finger snapping; cyclic gestures</td>
</tr>
</tbody>
</table>

Note that, in accordance with the functional linguistic tradition, the functions in Table 4.1 are defined relative to the utterance as a whole.
Hence, this listing is not meant as a full taxonomy of gesture functions independent of speech, but as a (non-exhaustive) set of examples of functions of linguistic elements that can be co-performed or modified gesturally. In addition, and more crucially, these functions are non-exclusive, i.e. they can be expected to be combined in a single gesture.

4.2.3 Participants and filtering

Participants were recruited through the online platform Crowdflower and received a monetary reward for their participation. Because the internet-based method that was employed lacks full experimental control, I undertook several steps to validate the reliability of the data. Four performance thresholds were established to filter out those participants who gave a strong impression of having completed the task without paying attention to the instructions or taking the assignment seriously. Trials were excluded if the participant (1) had failed to respond correctly to 20% or more of the test questions, which simply asked them to tick one of the seven boxes, (2) had taken less than seven minutes to complete the entire survey (average completion time was 30.3 minutes, SD = 13.2), (3) had a variance in their ratings of less than 0.60 (M = 1.98, SD = 0.43), or (4) had given only one or two unique ratings to all questions for five stimuli or more (non-attentive internet participants are known to complete surveys by clicking consistently on just one or two boxes). After having filtered out the participants who had failed any of these criteria, new participants were recruited in order to balance the number of trustworthy participants per stimulus. This procedure was repeated until exactly 18 participants had judged each video.

After applying the filtering procedures described below, 260 participants remained. Their reported ages ranged from 18 to 69 (M = 38.2, SD = 11.7). All were present in Germany at the moment of participation (according to their IP-address) and reported full proficiency in the German language. The participants (reportedly) had

25 http://www.crowdflower.com
26 This was confirmed on the basis of IP-addresses. It is not very likely that speakers with low German proficiency participated, because they would have been unlikely to pass the test questions, which were asked in German.
diverse daily occupations; only 42 of them (16.2%) had an academic affiliation.

4.2.4 Procedure
Before the survey was presented, the availability of well-functioning video and audio equipment was verified by means of a digital ‘captcha’, where visually and auditorily presented words had to be typed in, in order to proceed. Next, participants – sometimes referred to as “judgers” below – were presented with a set of video snippets and a number of statements about one of the gestures seen in the video (e.g. 'The hands of the speaker refer to a specific object or person'; see Table 4.2 for the full list of statements). The set of statements corresponds to the gestural functions of interest listed in Table 4.1.

Table 4.2 Statements presented to the participants (translated from German)

<table>
<thead>
<tr>
<th>Question label</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depict-Shape</td>
<td>The hands of the speaker depict the <em>size or shape</em> of an object or person</td>
</tr>
<tr>
<td>Signal-Prominence</td>
<td>The hands of the speaker show that the provided information <em>is important or deserves special attention</em></td>
</tr>
<tr>
<td>Depict-Movement</td>
<td>The hands of the speaker depict the <em>movement</em> of an object or person</td>
</tr>
<tr>
<td>Refer-to-Place</td>
<td>The hands of the speaker refer to a specific <em>place</em></td>
</tr>
<tr>
<td>Refer-to-Object</td>
<td>The hands of the speaker refer to a specific <em>object or person</em></td>
</tr>
<tr>
<td>Signal-Uncertainty</td>
<td>The hands of the speaker show that she is <em>uncertain</em>, whether she correctly depicts the properties of an object (e.g. its form, size, movement or number)</td>
</tr>
<tr>
<td>Number</td>
<td>The hands of the speaker tell something about the <em>number</em> of objects she talks about</td>
</tr>
<tr>
<td>Localize-Event</td>
<td>The hands of the speaker indicate <em>where or when</em> the described event takes place</td>
</tr>
<tr>
<td>Word-Search</td>
<td>The hands of the speaker show that she has <em>difficulty finding the right words</em></td>
</tr>
</tbody>
</table>
The task of the participants was simply to indicate, on a seven-point Likert scale, whether or not they agreed with each of these statements. The following instructions were given as guidance (translated from German):

In this survey you will be asked to answer questions on short video segments. The questions concern the relationship between the speaker’s gestures and what he/she says. Every list of questions concerns one single gesture. When multiple gestures occur in a video segment, they will be marked with numbers. Please read the questions carefully and answer on the basis of your own perception. Be aware that for some videos none of the options is applicable, while for other videos, multiple options may apply. In the video segments, someone explains a route through a city to his/her interlocutor. You can play the video as often as you want.

After having been familiarized with the setup, participants saw a webpage with an embedded video on the top of the screen and the nine statements listed in Table 4.2 (Figure 4.3). The following question was asked: ‘Do the following statements apply to gesture number X?’ where X was the number that appeared in the video in concurrence with the gesture of interest.\textsuperscript{27,28} On the right side of each question a 7-point Likert scale appeared with labels on both extremes: trifft sicher nicht zu ‘certainly does not apply’ and trifft sicher zu ‘certainly applies’.

To become accustomed to the task, participants were first given a practice item, which was the same for all participants. Subsequently, participants were presented with a block of twenty gesture videos, presented one at a time, which were randomly sampled from the total collection of stimuli. Because each block contained a different subset of

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{27} The pronouns were adjusted to the speaker’s gender.
\item \textsuperscript{28} Two additional statements were presented, but these were discarded from the analysis because of very low overall ratings. The statements were: ‘The hands of the speaker show that the depicted object is unknown to the addressee’ and ‘The hands of the speaker show for how long the event described takes place.’
\end{itemize}
\end{footnotesize}
the total collection (20 out of 462) and all analyses of interest are item-based, participants were allowed to take part in the study multiple times, with a limit of five. All redundant data points (cases where the same participant had by chance been assigned the same video twice) were excluded from the analysis post-hoc. In order to be eligible for the monetary reward, participants were required to complete the entire block of twenty stimuli. The order of the questions was randomized for each stimulus. For one out of every four videos on average, a test question was added to the survey to ensure that participants were actually reading and paying attention to the questions. These test questions simply asked the participants to tick one of the seven boxes (e.g. the second one from the right).

4.2.5 Reliability of the sample and performance diagnostics
To gain an impression of the reliability of the sample of raters with respect to larger populations, I analyzed the mean scores on each of the nine questions assigned to the practice video, which was seen and judged by all 260 participants (Figure 4.2).

![Figure 4.2 Diagnostics on reliability of sample: mean ratings stabilize as more participants are recruited](image-url)
<table>
<thead>
<tr>
<th>Stimmen Sie den folgenden Aussagen zu, in Bezug auf die Geste Nummer 2?</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Wenn das Video nicht sichtbar ist, drücken Sie die F5-Taste.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statement</th>
<th>Trifft sicher zu</th>
<th>Trifft sicher nicht zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Die Hände des Sprechers zeigen, dass das dargestellte Objekt bei ihrem Gesprächspartner unbekannt ist.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers stellen die Bewegung eines Objekts oder einer Person dar.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers geben an, für wie lange das, worüber er spricht, passiert.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers zeigen an, dass er Probleme hat, die richtige Worte zu finden.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers stellen die Größe oder Form eines Objekts oder einer Person dar.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers geben an, wo oder wann das, worüber er spricht, passiert.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers zeigen an, dass er unsicher ist, ob er die Eigenschaften eines Objekts richtig darstellt (z. B. die Form, Größe, Bewegung oder Anzahl).</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers verweisen auf einen spezifischen Ort.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers zeigen an, dass die Informationen, die er vermittelt, wichtig sind oder besondere Aufmerksamkeit verdienen.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers verweisen auf ein Objekt oder eine Person.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
<tr>
<td>Die Hände des Sprechers zeigen etwas über die Anzahl dessen an, worüber er spricht.</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
</tbody>
</table>

Figure 4.3 Screenshot of survey as presented to participants
A Pearson correlation test revealed that the average ratings on each of the questions after 18 participants provide a good estimate of the average scores as they stabilize over time. The mean ratings of the first 18 participants on the practice item correlate strongly with the mean ratings of all remaining 242 participants on the same item ($r = 0.96, p < 0.001$) Thus, these ratings can sensibly be interpreted as representing collective intuitions.

### 4.2.6 Descriptive statistics

A set of exploratory statistics is given in Table 4.3. Here and in the remainder of this dissertation, the gestural functions of interest are referred to with the labels introduced in Table 4.2 above. For instance, by ‘the scores on Depict-Shape’ I refer to the mean ratings to the first-listed statement in that table. For computing the mean scores and standard deviations, all ratings were first averaged per gesture, so that a continuous score between 1 and 7 was assigned to each gesture-function pair. These item-based mean ratings were aggregated across all gestures, yielding a grand mean, standard deviation, minimum and maximum value. In addition to conventional descriptive statistics, Table 4.3 lists the intraclass correlation coefficient (ICC), which is a measure of inter-rater agreement.

<table>
<thead>
<tr>
<th>Function</th>
<th>Descriptive statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Refer to Object</td>
<td>4.08</td>
<td>1.00</td>
<td>1.06</td>
<td>6.22</td>
</tr>
<tr>
<td>Refer to Place</td>
<td>4.18</td>
<td>0.95</td>
<td>1.06</td>
<td>6.17</td>
</tr>
<tr>
<td>Depict Shape</td>
<td>3.50</td>
<td>1.44</td>
<td>1.06</td>
<td>6.50</td>
</tr>
<tr>
<td>Depict Movement</td>
<td>2.88</td>
<td>0.83</td>
<td>1.06</td>
<td>5.28</td>
</tr>
<tr>
<td>Localize Event</td>
<td>2.77</td>
<td>0.65</td>
<td>1.06</td>
<td>6.17</td>
</tr>
<tr>
<td>Number</td>
<td>2.46</td>
<td>0.81</td>
<td>1.06</td>
<td>6.17</td>
</tr>
<tr>
<td>Signal Prominence</td>
<td>3.79</td>
<td>0.70</td>
<td>1.56</td>
<td>5.33</td>
</tr>
<tr>
<td>Word Search</td>
<td>3.38</td>
<td>0.86</td>
<td>1.61</td>
<td>6.33</td>
</tr>
<tr>
<td>Signal Uncertainty</td>
<td>3.45</td>
<td>0.81</td>
<td>1.56</td>
<td>6.06</td>
</tr>
</tbody>
</table>
We see considerable variation across stimuli for all functions. The minima and maxima cover the largest segment of the scale, although the highest end (> 6.5) is not reached for any of the questions. Means range between 2.46 (for the Number question) and 4.18 (for the Refer-to-Place question). These statistics suggest that there is sufficient variance for these data to be subjected to further statistical analysis. As for the ICC metric, we generally see low values. This might partially be due to failure to filter out noise: despite the efforts to discard the data contributed by non-attentive participants, it is possible that not all judgers were fully committed to the task, causing somewhat dispersed data patterns. In addition, it is possible that some questions were interpreted slightly differently by the different raters. This is not a fundamental problem, because the means can still be interpreted as representing typical responses by naïve viewers. Given the measures for reliability and stability of the sample (Figure 4.2), one can be confident that the mean scores provide a good estimate of a typical response in a greater population. Nevertheless, it is important to keep in mind that individual variation in the interpretation of the gestures is substantial, and that the results presented in this chapter will not always reflect the perception of every individual rater.

4.3 Mapping out gestural multifunctionality

This section empirically investigates the tendencies of the functions listed in Table 4.1 to be performed in combination with each other. I first demonstrate how the functionality of a single gesture can be characterized on the basis of the current data. Then I further examine the statistical patterns across the data set.

4.3.1 The multifunctionality of a single gesture

Before examining the more general patterns that occur in these data, I show how the functionality of a single gesture can be characterized according to the current approach. In example (21) and Figure 4.4, the speaker talks about two towers above the side aisles of a church. Meanwhile he moves both hands up, with the palms of the hands facing each other, all fingers curved. The timing of the gesture relative to the speech is represented below the speech transcription, following the
conventions described by Kendon (2004; Ch. 7). The capital letters above the transcript correspond to the timing of the video stills. The label prep stands for the preparation phase of the gesture, stroke stands for the most effortful part, and hold is the phase following the stroke, where the hands are typically held in place for some time before they are retracted to rest position.

Figure 4.4 The gesture performed in concurrence with the utterance in (21)

(21) *die Kirche hat halt ein ein Spitzdach und zwei Türme an diesen*

<table>
<thead>
<tr>
<th>prep</th>
<th>stroke</th>
<th>hold</th>
</tr>
</thead>
</table>

 prep stroke hold

**Seitenschiffen**

recovery

‘the church has well a a pitched roof and two towers at these side aisles’

Figure 4.5 shows the means and standard errors of the ratings assigned by the participants to each of the nine function statements with respect to this stimulus. 29 With regard to four out of nine potential functions,

29 The treatment of Likert-like items as continuous data (instead of ordinal) requires a note of caution. Because the underlying construct might in fact not be continuous, no absolute comparisons can be made between the distances between pairs of survey scores. The distance between a mean score of 2 and 3
there is a high degree of certainty among the 18 participants (mean rating 5 or higher). That is, strong consensus exists that the gesture refers to an object, refers to a place, describes the size or shape of an object, and provides information with respect to its number. In addition, there is a rather strong agreement that the gesture signals that the information provided is noteworthy (mean rating > 4.5). The remaining four functions consistently receive low scores: participants are generally certain that the gesture does not show that the speaker is having trouble finding words, is uncertain about what he says, depicts movement or localizes some event in space.

These data suggest that the gesture in question is rich in meaning. If the same functionality were to be expressed verbally (e.g. in German), at least five lexical or grammatical elements may have been needed: one for referring to an entity (a noun phrase or pronoun, presumably), one for indicating its number (e.g. a numeral or inflection), one for referring to its location (e.g. an adverb of place or a prepositional phrase), one for describing its shape (e.g. an adjective) and one for marking that the information given is noteworthy (e.g. a word order or intonation contour associated with discursive prominence).\footnote{Note that the gesture was in fact performed together with speech, and from the transcription one can see that speech and gesture are largely co-expressive in terms of all functions described. Only the shape-depiction aspect of the gesture is not explicitly mentioned verbally (but one may argue that it is implicit in the meaning of the word \textit{tower}).} In the following, I take this layered-gradient view on gesture functionality as a starting point to inspect the commonalities that exist between the functional profiles of the 462 gestures for which I have gathered comparable data. In particular, based on correlational patterns in the data, I look further into the general tendencies of specific functions to be co-expressed.
4.3.2 Mapping out the gesture functionality space

In order to obtain a global overview of the relations between gestural (sub-)functions, I applied Principal Component Analysis – a statistical technique for reducing the complexity of high-dimensional data by mapping them onto the axes of biggest variance (Wold et al. 1987). The first three principal components, as Table 4.4 shows, explain about 78% of the total variance within the data. The difference in informativeness between the third and subsequent components is relatively marginal. The plot in Figure 4.6 displays the Eigenvector-rotated values of all gesture stimuli on the first two principal components as points, and the coefficients of the survey questions on these components as vectors. Generally speaking, the question-vectors that point in the same direction have a similar response profile across stimuli, and points projected in the direction of any of the vectors represent gestures with high scores on the corresponding questions.
Table 4.4 Loadings of all variables (the survey questions) on the first three principal components

<table>
<thead>
<tr>
<th>Question</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal-Prominence</td>
<td>-0.42</td>
<td>0.10</td>
<td>-0.25</td>
</tr>
<tr>
<td>Localize-Event</td>
<td>-0.28</td>
<td>-0.40</td>
<td>-0.38</td>
</tr>
<tr>
<td>Refer-to-Object</td>
<td>-0.33</td>
<td>0.42</td>
<td>-0.14</td>
</tr>
<tr>
<td>Refer-to-Place</td>
<td>-0.41</td>
<td>-0.21</td>
<td>-0.24</td>
</tr>
<tr>
<td>Depict-Shape</td>
<td>-0.23</td>
<td>0.51</td>
<td>-0.03</td>
</tr>
<tr>
<td>Depict-Movement</td>
<td>-0.21</td>
<td>-0.45</td>
<td>-0.19</td>
</tr>
<tr>
<td>Number</td>
<td>-0.18</td>
<td>0.36</td>
<td>-0.26</td>
</tr>
<tr>
<td>Signal-Uncertainty</td>
<td>0.40</td>
<td>0.13</td>
<td>-0.56</td>
</tr>
<tr>
<td>Word-Search</td>
<td>0.41</td>
<td>0.05</td>
<td>-0.55</td>
</tr>
<tr>
<td><strong>Variance explained</strong></td>
<td><strong>41.3%</strong></td>
<td><strong>25.3%</strong></td>
<td><strong>11.1%</strong></td>
</tr>
</tbody>
</table>

Some noticeable patterns emerge from this analysis. For one, the spatial organization of the vectors that represent the functions suggests that the gestural ‘functionality space’ comprises three, somewhat orthogonally organized clusters. The first can be characterized as a *representational* dimension, pertaining to the capacity of gestures to refer to objects and their intrinsic properties such as size, shape and number. This dimension subsumes the survey questions Depict-Shape, Refer-to-Object and Number. The second cluster roughly represents a *spatial* dimension of gesture meaning, corresponding to a gesture’s capacity to localize objects and events in space. The third can be described as *meta-communicative*, subsuming the questions Signal-Uncertainty and Word-Search. The only survey question that does not clearly fall within any of these three clusters is Signal-Prominence. The capacity of gestures to indicate that some information is noteworthy correlates with both the representational and spatial features (see below for a more detailed analysis of these correlations), but is typically not combined with signaling uncertainty or word search.

---

31 Note that in most functional linguistic models, the signaling of word search and the display of uncertainty belong to different functional categories. To the extent that the signaling of word-search is aimed at warning the addressee of an impending delay of the discourse, as in the case of interjections like *uh* and *um* (Clark & Fox Tree 2002), the broad category label ‘meta-communicative’ used here covers Halliday’s interpersonal and textual functions.
Another noteworthy observation is that the gesture stimuli are widely dispersed throughout the entire plot. This suggests that although some functional clusters exist, most gestures fall right in between these. This is in line with McNeill’s (2005) argument that gestures often simultaneously combine iconic, deictic and pragmatic features. In the next section, I examine some of the relations between the scores on pairs of the individual questions in more detail.

4.3.3 **A closer look at semantic multifunctionality**

A first type of multifunctionality concerns patterns of co-occurrence between different types of semantic information. Figure 4.7a displays
the mean scores of all gesture stimuli on the question Depict-Shape as a function of the mean scores on Refer-to-Object. Figure 4.7b displays the mean scores on the question Depict-Movement plotted against the scores on Depict-Shape.

Figure 4.7 Scatter plots of mean scores on two pairs of semantics-related questions

A Pearson correlation test reveals a strong positive trend between the mean answers on Depict-Shape and Refer-to-Object ($r_{160} = .78$, $p < 0.001$). From the scatter plot, however, it appears that this relation is not fully symmetrical. Whereas none of the gestures in the data set were judged to depict the shape of an object without also referring to a concrete entity, there are some cases of gestures that score high on Refer-to-Object but not on Depict-Shape. Qualitative inspection of this subset of the stimuli, indicated visually in the figure by the dashed ellipse, reveals that this category includes many instances of abstract deixis: gestures that refer to verbally described objects by pointing to a location in interactional space associated with a discourse referent. The reverse pattern – high scores on Depict-Shape with low scores Refer-to-Object – does not occur: according to the participants, those gestures that evoke a physical or spatial attribute necessarily also make reference to some object or person; gestures were never perceived as isolated attributes.

In Figure 4.7b, we see a rather different picture: there is a weak negative correlation between the questions Depict-Size and Depict-
Movement ($r_{460} = -0.14, p = 0.002$). In line with this trend, the region of the plot corresponding to high scores on both questions is empty. This indicates that the stimulus set did not contain any instances of gestures that were judged to simultaneously depict the shape and the movement of an object. Although there are no reasons why such gestures could not exist in principle – one may imagine a gesture whereby the handshape refers to a pen which is moved through space to represent writing – none of the gestures in the natural spatial dialogues under investigation were judged to have these characteristics.

4.3.4 **A closer look at semantic-pragmatic multifunctionality**

The semantic multifunctionality analyzed in the previous section pertains to only one of the levels of analysis distinguished in functional linguistics: the representational function of language (or ideational metafunction, in SFG terms). However, there are reasons to believe that gestures may also conflate representational (semantic) and interpersonal-textual (pragmatic) functions. In this section, I investigate two such relations as they occur in the data.

I first address the question of whether referential gestures are typically perceived as indicating discursive prominence. As Figure 4.8a shows, there is an overall positive correlation between the scores on the corresponding questions ($r_{460} = 0.59, p < 0.001$). The majority of gestures that were judged to refer to an object or person were also judged to indicate that this act of reference has focal status. This finding is consistent with Levy and McNeill’s (1992) hypothesis that gestures are an expression of high communicative dynamism – they contribute significantly to ‘pushing the communication forward’. Given the unspecific way in which the Signal-Prominence question was asked, the current data do not reveal whether gesture performance correlates with topic-comment contrasts, or rather with a focus-background distinction (recall that FDG sees these as independent dimensions of information structure). Another way of framing these results is in terms of a relevance-theoretic view on iconic gesturing (cf. Sperber & Wilson 1986): the use of the hands to refer to an object appears to create the expectation of its own relevance as a contribution to the interaction.
Finally, I turn to the relation between spatial reference and meta-communicative signaling. In Figure 4.8b, we see that a negative correlation exists between Word-Search and Refer-to-Place ($r_{460} = -0.50$, $p < 0.001$). Thus, the potential of gestures to refer to a location is rarely combined with signaling that the speaker is having trouble finding the right words. As is already apparent from the results of the Principal Component Analysis and Figure 4.6, the abilities of the hands to refer to a place and to express ongoing management of one’s own speech appear to be mutually exclusively in the current data (as a general trend, at least). Note however, that a few exceptions to this pattern exist: we can see from Figure 4.8b that some gestures received a modestly high mean score (±5) on both questions.

4.3.5 Degrees of functional prominence
As is evident from the wide dispersion of the data in the analyses presented above, a substantial variability exists in the degree of certainty with which the different functions were ascribed to the gestures. In fact, the majority of the mean scores on the questions taken into account fall right in between the ‘certainly’ and ‘certainly not’ poles. Whereas this may to some extent be explained by interpersonal
differences in the interpretation of the gestures across participants, it also suggests that the functions carried out by any given gesture have different degrees of prominence to the observers. Any given gesture may for instance foreground a certain type of information (e.g. the shape of some object) but simultaneously provide information that is of secondary importance (e.g. the location of the object). A view of gestures as having a primary function next to various secondary functions has been advocated before, at least as a theoretical issue (Müller 1998b), but most research designs still operationalize gestural functions as ‘present’ or ‘absent’.

In Table 4.5, I quantify the multifunctionality of the stimuli, assuming a relation between the degree of certainty expressed by the raters and the salience of a gesture’s function: questions with mean scores higher than 5 on the 7-point Likert scale are assumed to correspond to a gesture’s primary functions; those with mean scores between 4 and 5 and those with mean scores between 3 and 4 are classified as secondary and tertiary, respectively. Table 4.5 shows how many such functions were attributed to the gestures in the stimulus set.

<table>
<thead>
<tr>
<th>Function type by rater certainty</th>
<th>Number of functions per gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Primary' functions (mean score &gt; 5)</td>
<td>M=0.87, SD = 0.97</td>
</tr>
<tr>
<td>'Secondary' functions (mean score between 4 and 5)</td>
<td>M=1.74, SD = 1.17</td>
</tr>
<tr>
<td>'Tertiary' functions (mean score between 3 and 4)</td>
<td>M=2.81, SD = 1.63</td>
</tr>
<tr>
<td>Accumulated</td>
<td>M=5.42, SD = 1.53</td>
</tr>
</tbody>
</table>

We see that the gestures in the corpus were typically assigned no more than one primary function. Hence, the gesture in Figure 4.4, which has four primary functions according to the definitions employed here, appears more of an exception than a rule. With regard to secondary and tertiary functions, however, multifunctionality is a much more frequent, if not ubiquitous phenomenon. When accumulating over the three categories, gestures have over five functions on average; only one out of
462 gestures was 'monofunctional' according to these criteria. Of course, the exact numbers in Table 4.4 are not very meaningful, as they depend on the number of questions included in the survey and the way these were asked (as well as on the operational choices made – recall for instance footnote 29 on the caveats of treating Likert-items as continuous). Yet, the general pattern in these data underscores another important characteristic of gestures that has often been neglected in experimental work: gestural functions are not all-or-nothing, but come in different degrees of explicitness and salience.

4.4 Discussion and conclusion

Inspired by objections to the rigidity in current functional classification systems, this chapter has advanced, operationalized and empirically substantiated the view that gestures are potentially multifunctional. The results of a large-scale gesture perception study suggest that the functional potential of the gestures in the direction-giving discourse segments used in this study involves at least three, somewhat opposed components – one pertaining to reference and representation of objects, one pertaining to space and movement, and one pertaining to meta-communicative signaling. Some of these functions can be present simultaneously in a single gesture, but with different degrees of salience.

Note that the three clusters of functions that emerged from the analysis do not strictly reflect Bühler's three categories. Hence, these data do not directly corroborate Müller's (1998b, 2013) claims about how gesture functions can be characterized in the same way that Bühler analyzes the functions of verbal expressions. On the other hand, given the operational details of the study, the current results do not present direct counterevidence to such a view either. As mentioned, the questions in the survey were tailored according to the availability of relevant data in direction-giving dialogues, and consequently more geared towards the representational function than to expression and appeal. Nevertheless, the results of the current study present some further insights into the potential of gestures to combine multiple Bühlerian functions. Provided that there is some correspondence between expression and appeal on the one hand, and the functional cluster here described as 'meta-communicative signaling' on the other,
the current data corroborate that representational and other functions of gestures are typically not perceived as equally prominent.

Some important notes of caution are in order when interpreting the data presented in this chapter. First, the results are certainly not fully independent of the research design and the way the survey questions were formulated. The set of questions included in the survey may not reflect the full functional potential of gestural expression. In addition, all stimuli came from a route description corpus – a discourse type that involves a relatively large number of concrete referential and spatial gestures and is not fully representative of everyday dialogue.

With respect to the interpretation of the results, it is furthermore important to emphasize that this study involves the de-contextualized perception of gestures by an idealized comprehender. Raters were unconstrained in time and were allowed to watch the video more than once, while being uninformed about the exact local discourse situation and preceding utterances. Therefore, and because scores were averaged over a group of raters, one should be cautious about inferring that the functional profiles of the gestures as investigated here correspond to the real-time processing of the addressee at the moment of the conversation. The characterization of the gestures here can better be compared to canonical (functional) linguistic analysis; it involves a level of description that is abstracted from actual usage and generalized over subjective experiences.

Notwithstanding the caveats mentioned, the study presented in this chapter is among the first to provide quantitatively supported insights into the pervasive multifunctionality of gestures. Moreover, by exemplifying how a more gradient and layered view of gesture function can be operationalized, it potentially has methodological implications. A setup akin to the one described here lends itself to implementation in an experimental design. Coding schemes could be more effective if their annotation tiers reflect the layers of a theoretically motivated, stratified model of language. In addition, they can be endowed with gradient scales that allow coders to express different degrees of certainty. Primary functions of gestures could be determined using a system such as the one described by Bressem et al. (2013), which goes from describing the forms of gestures to looking at their functions with
respect to the accompanying speech. Overall, it is the hope that in addition to their empirical value, the contents of the present chapter can inspire quantitatively oriented researchers to adopt a more refined view on the function(s) of gestures in situated discourse. The following chapter investigates to what extent the functional interpretation of the gestures in the corpus is dependent on the accompanying speech.
Chapter 5. Perception of gestural functions with and without speech

A further difficulty in incorporating gestures in models of grammar is that their forms are strongly adaptive to the verbal context. As noted by McNeill (1992:41), the same object or situation can be referred to by gestures with different formal characteristics. This context-dependence, according to McNeill, “contrasts with the stability of lexical forms, [which] present modifications of the core meaning in all contexts but do not give up their basic form in this process.” If no basic, context-independent forms exist at all in gestural expression, it can be expected that the meaning of a gesture cannot be understood when the access to the accompanied speech is denied. The current chapter puts this hypothesis to the test. I first summarize current evidence that certain aspects of gestural meaning can be interpreted in the absence of speech. Then, I argue that insights into this question can be advanced by adopting a (cognitive-functional) linguistic perspective on gestural meaning. Based on an extension of the study presented in the previous
chapter, I examine whether access to speech is necessary to understand what type of linguistic function a gesture performs.\footnote{Some of the contents of this chapter were published in the proceedings of the 4th Workshop on Gesture and Speech in Interaction (Kok et al. 2015).}

5.1 Previous research on the comprehension of gestures without audio

Various experimental studies have compared utterance comprehension in audio-visual and audio-only conditions (Feyereisen et al. 1988; Hadar & Pinchas-Zamir 2004; Kibrik & Molchanova 2013; Krauss et al. 1991; Swerts & Krahmer 2005). Feyereisen et al. (1988) tested whether participants were able to classify the manual gestures of speakers in naturalistic discourse as either iconic (mimicking the spatial-motoric features of some entity) or batonic (non-referential, expressing emphasis or rhythm only) when video recordings were presented with or without sound. The authors report that raters were substantially more accurate when the auditory channel was available, but performance was still above chance level when access to the speech was denied. In a second experiment, it was found that when video clips were presented without access to the audio channel, subjects were generally unable to guess the original utterance from a fixed list of options. Despite the low performance, however, a high degree of consistency was found in participants’ responses to this task. According to the authors, this suggests that the gestures in their stimulus set have intrinsically meaningful qualities.

A similar paradigm was used by Krauss et al. (1991), who presented participants with audio-muted video snippets and asked them to guess which of two given words was the gesture’s ‘lexical affiliate’ (the word or words to which the gesture relates most, semantically). Raters’ performance on this task was far from perfect, yet significantly above chance level. In a follow-up experiment, participants were instructed to assign semantic categories to a set of gestures in conditions with and without sound. It was found that when the speech was available, the classifications very closely reflected the semantic content of the accompanied speech. However, in the absence of the verbal channel, the judgments were not random either: gestures were
often assigned the same semantic category as their lexical affiliate, to which participants had no access. Given that the presence or absence of speech was consistently found to be an important factor, the authors conclude that "although gestures can convey some information, they are not richly informative, and the information they convey is largely redundant with speech" (Krauss et al. 1991:743).

In accordance with this conclusion, Hadar and Pinchas-Zamir (2004: 210) argue that the meaning of gestures is best understood in terms of different levels of ‘semantic specificity’: some gestures have a specific, lexeme-like meaning, whereas others convey meaning in a "vague and tentative" fashion. Taking the notion of semantic specificity as their point of departure, two experiments were carried out where participants had to select a word from a list that they thought related most closely to a given gesture. Participants less often chose the word that had been coded as the lexical affiliate of the gesture when the speech was muted than when the speech or a transcription of it was present. Among the ‘erroneous’ responses, moreover, visually or semantically related distractor words were chosen more often than unrelated distractors. Based on this graded effect, the authors claim that gestural behaviors can have a range of degrees of semantic specificity.

A recent study by Kibrik and Molchanova (2013) used a more contextualized task to investigate the interdependence between speech, gesture and prosody. Participants watched segments of movies or videotaped conversations in conditions with audio only, video only, prosody only, or combinations of these channels. They then answered a set of multiple choice questions about the content of the movie clips (e.g. 'What does Tamara offer Masha before the beginning of the conversation?'). In line with previous findings, it was found that although participants were more accurate when the audio was available, a substantial subset of the questions was answered correctly in the absence of speech as well. On the basis of this result, the authors argue that whereas speech might be the leading channel of information conveyance, gestures (and prosody) do carry some degree of independent semantic load.

All four of these studies report a rather ambivalent relation between speech and gesture: on the one hand, gestures alone are not as
informative as speech-gesture combinations. On the other hand, gestures by themselves provide sufficient cues for participants to score well above chance level on various types of comprehension tasks. The authors of the papers discussed here roughly agree that this ambivalence reflects the fact that gestures are semantically unspecific: they carry meaning, but only on a rather schematic level. It remains a relatively open question, however, how this level of schematicity is best characterized. What types of meaning are associated with gestural forms irrespective of speech, and on what level of abstraction?

Getting a grip on this question requires a comprehension task of a less specific character than those used in the research discussed above. Previous studies commonly assessed utterance comprehension on the basis of relatively concrete and detailed questions (e.g. ‘what is the lexical affiliate of this gesture?’ or ‘what did X say to Y?’). Thus, the notion of meaning is often simply conceived as reference to some specific entity or situation – a view that has been under dispute in contemporary cognitive-functional linguistics. An alternative approach is to characterize meaning in terms of the functional contribution of an expressive unit to the ongoing discourse. Following the considerations put forward in Chapter 3, a functional view can shed new light on the interface between verbal and gestural expression. In terms of experimental design, this raises the question whether access to speech is needed to understand what kind of semantic or pragmatic function a given gesture performs.

This chapter describes an experiment that implements a functional view on speech-gesture interdependence. It focuses on four prominent functions of gestures, which have been introduced earlier: object reference, attribution of a static property (e.g. depicting the shape or size of an object), attribution of a dynamic property (e.g. depicting a process or movement), and meta-communicative signaling (e.g. indicating difficulty in speech production). The perception of these functions by naïve observers is investigated by extending the online perception study reported on in Chapter 4 with a condition where the videos are presented without sound.
5.2 Methods and participants

The online perception experiment described in Chapter 4 was supplemented with an audio-only condition. That is, the methodological procedures were repeated in a condition where the videos were muted. Participants were again recruited via the online research platform Crowdflower and filtered according to the same performance thresholds as before (test questions, timing, variance in the data, see Chapter 4 for details). This yielded a participant pool of 347 judgers, with 16 unique judgers per gesture. Because of unbalanced data samples, 13 gestures were excluded from the analysis, leaving a total number 449 gestures to be analyzed, distributed over 173 videos. The ages of the participant pool ranged from 16 to 69 (M = 37.3, SD = 12.1), and 129 were female. All were present in Germany at the time of participation (according to their IP-address) and reported having full proficiency of the German language. The experiment has a between-subject design: participants were assigned either to the Sound-On or the Sound-Off version.

The analyses focus on a subset of four of the nine functions examined, which together cover a diverse range of the functions that gestures are capable of carrying out. Three of these correspond to prevalent semantic functions: the ability of gestures to refer to concrete entities (the Refer-to-Object function), to ascribe specific spatial properties (the Depict-Shape function) and to evoke dynamic properties (the Depict-Movement function). The fourth question concerns one of the meta-communicative functions that gestures can perform: the display of difficulty in word retrieval or formulation. Table 5.1 repeats the survey statements corresponding to these functions as presented to the participants, as well as the labels used for referring to them in the analysis section.
Table 5.1. The functions of interest in the current chapter (a subset of the full list introduced in Chapter 4)

<table>
<thead>
<tr>
<th>Question label</th>
<th>English translation of corresponding statement in the survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refer-to-Object</td>
<td>The hands of the speaker refer to a specific object or person</td>
</tr>
<tr>
<td>Depict-Shape</td>
<td>The hands of the speaker depict the size or shape of an object or person</td>
</tr>
<tr>
<td>Depict-Movement</td>
<td>The hands of the speaker depict the movement of an object or person</td>
</tr>
<tr>
<td>Word-Search</td>
<td>The hands of the speaker show that he/she has difficulty finding the right words</td>
</tr>
</tbody>
</table>

In the final section of this chapter, the remaining functions also receive attention, but their analysis is limited to descriptive statistics. Note that although the functions of interest are examined as individual variables here, they may be intercorrelated, as seen in the previous chapter.

5.3 Results

I conducted two types of analysis to gain insight into how access to speech affects the functional interpretation of the gestures. To find out whether the ratings given with and without sound are different across the board, I compared the frequency distributions of the mean ratings on all stimuli in the two conditions. Second, I conducted correlational analyses to investigate whether the gestures were assigned similar ratings when presented with and without sound. Results of these analyses are discussed in the light of qualitative inspection of the data.

5.3.1 The object reference function

As apparent from Figure 5.1a, scores on the Refer-to-Object question are dispersed throughout the full spectrum of possible ratings. The mean score over all stimuli is 4.06 (SD = 1.02) for the Sound-On condition and 3.78 (SD = 0.75) for the Sound-Off condition. A paired samples t-test reveals that this difference is statistically significant ($t_{(448)} = 18.0$, $p < 0.001$, $d = 0.31$). However, note that the effect size is rather modest: the difference in means is less than a third of the standard deviation. Thus, although gestures are more often judged to refer to an object when
perceived in the context of speech, their referential function is almost equally easily recognized without access to the verbal channel. This is further corroborated by the correlational analysis. A Pearson test reveals that the ratings in the Sound-On and Sound-Off conditions are strongly correlated ($r_{(447)} = 0.65, p < 0.001$).

![Figure 5.1 Ratings on Refer-to-Object with and without sound](image)

There is a noticeable difference between the conditions in terms of the variance in the data. The distribution of scores in the Sound-Off condition is considerably steeper than the one corresponding to the Sound-On condition (SD with sound: 1.02; SD without sound: 0.75). In combination with the correlational results, this suggests that the presence of speech does not directly contribute to the qualitative interpretation of the gesture, but rather decreases uncertainty among observers.

The difference in variance between the conditions is likely related to the large number of ‘cohesive gestures’ in the stimulus set (McNeill 2000a). Various gestures in the corpus refer to an object by placing the hand in a region of space that is associated with a certain landmark in the virtual town, without giving further specification of the shape or size of the referent. In the absence of the verbal channel, such
gestures may have the appearance of non-representational movements of the hands (e.g. beat gestures; Feyereisen et al. 1988). In addition, some representational gestures involve rather subtle, low-effort movements. In example (22), for instance, the speaker makes a referential gesture that is not clearly interpretable as such in the absence of speech (Figure 5.2).

![Figure 5.2 Example of a gesture with different scores on Refer-to-Object when presented with or without sound](image)

(22)  _hat zwei Bäume davor zwei Türen äh eine in der Mitte eine links äh zwei_  

| prep | stroke recovery |

\[(B) \quad (C)\]

_Feuertreppen_

\[\text{has two trees in front of it two doors uhm one in the middle one left uhm two fire escapes'}\]

The movement made by the speaker while referring to the two fire escapes is rather marginal. He briefly raises both hands, but then quickly withdraws to rest position. The index fingers of the hands are stretched, but do not reach full extension. In Figure 5.3, we see how this gesture is rated in the conditions with and without sound.
For most of the functions, the ratings in the two conditions are not very far apart. Those in the Sound-Off condition are generally lower, which indicates that the gesture is not easily interpretable when presented in the absence of speech. In the Sound-On condition, two functions receive substantially higher ratings than in the Sound-Off condition: Refer-to-Object and Number. From the transcript, we see that the gesture coincides directly with the articulation of the noun phrase *zwei Feuertreppen* ‘two fire escapes/stairs’. It appears that the gesture is interpreted as co-expressive with the co-articulated noun phrase in the condition with sound, in such a way that the two entities evoked (fire escapes) map onto the two raised hands, and the respective positioning of the hands shows their relative location.

Further inspection of the stimulus set reveals that for some gestures, the referential status is not even clear-cut in the presence of speech. With regard to gestures that co-occur with sentences like *over there you should go to the left* and trace the trajectory described, there is generally low agreement as to whether or not the hand refers to a person. Even though such gestures can be thought of as having a referential component, where the hand embodies the interlocutor, they could also be interpreted as purely predicative, that is, as tracing the...
movement of some entity without referring to it. The difficulty of this issue is evident from the responses to the Refer-to-Object question for utterances of this kind, which are quite heterogeneous across judges, regardless of the presence or absence of speech.

5.3.2 The shape representation function
Mean ratings on Depict-Shape do not differ to a statistically significant extent across conditions ($t_{(448)} = -1.8, p = 0.075, d = 0.052$). In fact, a correlational analysis suggests that gestures were assigned very similar ratings when presented with and without sound ($r_{(447)} = 0.84, p < 0.001$). Figure 5.4 presents the outcomes graphically.

![Ratings on Depict-Shape with and without sound](image)

Figure 5.4 Ratings on Depict-Shape with and without sound

We see that in the Sound-On condition, the distribution of mean scores has two peaks: one close to the ‘certainly not’ pole, and one corresponding to more certain ratings. Thus, for a large portion of the gestures in the stimulus set, participants were fairly certain as to whether the gesture served to depict the shape or size of an object. In the Sound-Off condition, the histogram’s shape more closely approaches a normal distribution. When the sound was muted, the overall
uncertainty rate was substantially higher, with a local peak between 4 and 5 on the Likert scale.

A possible explanation for the high degree of uncertainty in the Sound-Off condition is that the SaGA corpus contains many tracing gestures, i.e. gestures where the hand draws a line in the air as if holding a drawing utensil. Such gestures can be semantically ambiguous: tracing handshapes can either be used to draw the shape or outline of some object, or to depict the movement of some entity through space (Bressem 2012; Müller 1998a). Without access to the verbal channel, tracing gestures are therefore not always easily interpretable. The utterance in (23) contains a gesture that is ambiguous between shape depiction and movement representation when presented without audio (Figure 5.5).

![Figure 5.5 Example of a gesture with different scores on Depict-Shape when presented with or without sound](image)

\[(23) \quad \text{Also da siehst du schon quasi wenn du diese [..] das ist ne lange Allee halt} \]

\[
\begin{align*}
\text{prep} & \quad \text{stroke} & \quad \text{retraction} \\
\text{..................................................................} & \quad \text{...............................................} & \quad \text{- - - - -} \\
\end{align*}
\]

'so there already you see so to speak when you this [...] it is a long avenue well ...'

The speaker moves his hand away from his body, with the back of the hand facing the interlocutor and the fingers hanging down loosely. When the arm has reached its maximal extension (screenshot B), the fingers
are slightly extended forward. This sagittal movement is repeated twice, but the second iteration is considerably less effortful than the first. Subsequently, the hand is retracted to rest position. In Figure 5.6, we see how this gesture is interpreted in the conditions with and without sound.

![Figure 5.6 Ratings of the gesture in example (23) with and without sound](image)

The ratings in the Sound-Off condition are generally higher than those we have seen for the previous example, with particularly high scores for Depict-Movement. This suggests that without access to speech, attenders of this gesture perceive it as referring to a process. In addition, many raters in the Sound-Off condition entertain the possibility that the gesture refers to an object, signals prominence and refers to a place.

With access to speech, the gesture is given a rather different interpretation. Most conspicuously, there is a strong increase in the average rating for Depict-Shape. In addition, raters are generally convinced that the gesture refers to an object and refers to a place. This difference in interpretation is likely to be related to the fact that the gesture roughly aligns in time with the articulation of the noun phrase *[ei]ne lange Allee* ‘a long avenue’. The most common interpretation of the gesture seems to be that it refers to the described avenue and depicts its physical contours. The movement of the hands thus does not map onto a dynamic process, but serves to represent the shape of the object
referred to. In line with this interpretation, the mean ratings on the question Depict-Movement are noticeably lower in the Sound-On condition than in the Sound-Off condition.

An interesting aspect of these data is that even when the speech is present, the ratings for Depict-Movement do not drop lower than 3.5 on the Likert scale. This suggests that even in the context of the verbal part of the utterance, some raters are open to the possibility that the gesture provides movement-related information. Possibly, some judges perceive the speech and gesture as giving slightly different pieces of information; whereas the speech merely expresses that the interlocutor will see an avenue, the gesture is seen as providing the additional information of how to go about traversing it. From a cognitive grammatical point of view, it can moreover be assumed that the semantics of the verb *sehen* ‘to see’ has a dynamic component: it is conceived through sequential scanning according to the theory of CG. On this account, it could be hypothesized that the movement of the hands reflects the dynamics of the speaker’s mental access to the conceptualized material.

### 5.3.3 The movement representation function

Scores on the Depict-Movement question also diverge to some extent between conditions, but in a different direction than we have seen thus far. The mean rating across stimuli is higher in the Sound-Off condition than in the Sound-On condition ($t_{(448)} = -15.9, p < 0.001, d = 0.28$; Figure 5.7).

These data suggest that not many gestures in the corpus unambiguously depict the movement of an object or person. Strikingly, direction-tracing gestures such as those discussed earlier (e.g., those co-occurring with *go to the left*) are not consistently judged as depicting movement. Moreover, likely due to the ambiguity of tracing gestures discussed in the previous section, the Sound-Off condition has a high concentration of uncertain responses. Thus, many tracing gestures were judged to be potentially depicting movement when presented in the absence of speech, but were clearly perceived as non-dynamic in their semantics (i.e., the physical movement of the hands was not perceived as mapping onto the described movement) when presented with the audio
on. Nonetheless, the corpus contains various examples of gestures that were only judged to be movement-related in the Sound-On condition. These include many gestures that barely involve any effortful movement, such as the one in example (24) (Figure 5.8).

![Ratings on Depict-Movement with and without sound](image)

**Figure 5.7** Ratings on Depict-Movement with and without sound

![Example of a gesture with different scores on Depict-Movement](image)

**Figure 5.8** Example of a gesture with different scores on Depict-Movement when presented with or without sound
As is shown by the timestamps, the gesture stroke does not last longer than half a second. The speaker raises his right hand, but it remains in the lower segment of gesture space. The stroke consists of no more than a subtle clockwise rotation of the right hand, in coincidence with the adverb *geradeaus* 'straight ahead'. Figure 5.9 shows how this gesture is interpreted in the two conditions.

Given the subtlety of the movement, it is not very surprising that raters in the Sound-Off condition do not assign particularly high scores to any of the functions. Interestingly, the highest average rating is for Word-Search, which suggests that a pragmatic interpretation of the gesture is relatively prevalent. When perceived with the sound on, two functions stand out as receiving high scores: Refer-to-Place and Depict-Movement. The finding that the difference is greatest for Depict-Movement is understandable, given the tight temporal co-occurrence with the adverb...
geradeaus 'straight ahead'. Although the gesture contains hardly any forward-directed motion, the rotation of the wrist seems to be interpreted as co-expressing the speaker’s instruction to go straight ahead. In cases like these, access to speech appears necessary to decide whether the gesture has a dynamic interpretation (i.e. refers to motion) or expresses a pragmatic function. Nonetheless, like in the previous cases, we see that the scores in the two conditions are strongly correlated.

5.3.4 The meta-communicative function
The three functions discussed so far pertain primarily to the domain of semantics. Roughly speaking, they relate to the ability of gestures to refer to concrete entities and to ascribe spatial and dynamic predicates. The fourth and final function taken into consideration here is of a rather different nature. It concerns the capacity of gestures to signal that the speaker experiences difficulty in formulating an utterance – an activity sometimes referred to as ‘own communication management’ (Allwood et al. 2006). This section inquires whether access to speech is necessary to understand that a gesture is meta-communicative in character (Figure 5.10).

Overall, the pattern observed is not very different from what we have seen for the other functions inspected. Although mean scores are not equal across the two conditions ($t_{(448)} = 4.2, p < 0.001$), the effect size is marginal ($d = 0.20$). The scores in the Sound-On and Sound-Off conditions are again strongly correlated ($r_{(447)} = 0.53, p < 0.001$). These data suggest that a substantial overlap exists in the response profiles across conditions. It appears, as we have seen before, that the judgments in the presence of speech are more assured, but qualitatively similar to those in the video-only condition. In other words, the influence of the verbal channel on the perception of the gestures is largely limited to the reduction of uncertainty. To help understand what is behind these patterns, I here discuss a typical example of a gesture with high ratings on Word-Search in both conditions, shown in (25) and Figure 5.11.
Perception of gestural functions with and without speech

Figure 5.10 Ratings on Word-Search with and without sound

![Graph showing distribution of ratings with and without sound](image)

(a) Normalized frequency

(b) Mean ratings with sound on

Figure 5.11 Example of a gesture with different scores on Word-Search when presented with or without sound

![Sequence of gestures with timestamps](image)
sone ja ... eine Überführung ist es nicht ... aber da sind so zwei Gebäude

(A) (B) (C)

äh baulich äh verbunden also so eine son son son ... Übergang da

|~~~|************|······|

pre. stroke recovery

‘such a well ... it is not an Überführung ... but there are two of those buildings uhm structurally uhm connected so one of those those those ... overpasses there’

As seen in Figure 5.11, the preparation of the gesture starts from the end position of another gesture, where the hands were opposed to one another as if holding a long object (screenshot A). The stroke phase consists of a rapid, energetic back-and-forth movement of the hands. The left hand, as seen in screenshot C, is extended higher and stays raised for a slightly longer time than the right hand. The spoken tier gives the impression that the speaker is having trouble finding the most appropriate word for the object he describes: so eine son son ‘one of those those those’. Figure 5.12 displays the functional ratings assigned to this gesture in the two conditions.

Figure 5.12 Ratings of the gesture in example (28) with and without sound
We see a somewhat polarized distribution of scores here, with two ratings that receive high scores, and all other ratings below 3. None of the functions receives noticeably different ratings in the two conditions. This suggests that the physical properties of the gesture alone are sufficient to understand its communicative value. There is, however, a small quantitative difference between the two highest rated functions: whereas raters who did not have access to the verbal channel gave the highest scores to Signal-Uncertainty, the interpretation of the gesture as signaling Word-Search is slightly more dominant in the Sound-On condition. These data are largely in line with the hypothesis that own communication management is associated with specific formal patterns (finger snapping or rapid cyclic movement of the hands, cf. Allwood et al. 2006). On the whole, the presence or absence of speech does not have a crucial influence on the type of function that gestures are perceived to perform.

5.4 A glance at the remaining functions
In the interest of space, I will not provide detailed analyses of the five remaining functions. However, to gain an impression of whether the other functions can be recognized without speech, Table 5.2 summarizes a number of relevant statistics: the means of the ratings in the Sound-On and Sound-Off conditions, the results of a pairwise t-test and the results of a Pearson correlation test.

33 As will be discussed in more detail in Chapter 6, it is possible that raters were not always able to correctly understand the difference between these questions.
Table 5.2 Descriptive statistics and results of t-tests and Pearson correlation tests for all functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Mean (SD) with sound</th>
<th>Mean (SD) without sound</th>
<th>P-value (t-test)</th>
<th>Cohen's d</th>
<th>R-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal-Prominence</td>
<td>3.76 (0.72)</td>
<td>3.72 (0.59)</td>
<td>0.33</td>
<td>0.047</td>
<td>0.49</td>
</tr>
<tr>
<td>Localize-Event</td>
<td>2.76 (0.68)</td>
<td>2.93 (0.55)</td>
<td>&lt; 0.001</td>
<td>0.28</td>
<td>0.35</td>
</tr>
<tr>
<td>Refer-to-Object</td>
<td>4.06 (1.02)</td>
<td>3.78 (0.75)</td>
<td>&lt; 0.001</td>
<td>0.31</td>
<td>0.65</td>
</tr>
<tr>
<td>Refer-to-place</td>
<td>4.15 (0.98)</td>
<td>3.71 (0.74)</td>
<td>&lt; 0.001</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>Depict-Shape</td>
<td>3.49 (1.46)</td>
<td>3.56 (1.14)</td>
<td>0.075</td>
<td>0.052</td>
<td>0.84</td>
</tr>
<tr>
<td>Depict-Movement</td>
<td>2.86 (0.85)</td>
<td>3.46 (0.80)</td>
<td>&lt; 0.001</td>
<td>0.73</td>
<td>0.30</td>
</tr>
<tr>
<td>Number</td>
<td>2.43 (0.82)</td>
<td>2.56 (0.51)</td>
<td>&lt; 0.001</td>
<td>0.18</td>
<td>0.48</td>
</tr>
<tr>
<td>Signal-Uncertainty</td>
<td>3.44 (0.82)</td>
<td>3.23 (0.67)</td>
<td>&lt; 0.001</td>
<td>0.28</td>
<td>0.52</td>
</tr>
<tr>
<td>Word-Search</td>
<td>3.37 (0.88)</td>
<td>3.21 (0.70)</td>
<td>&lt; 0.001</td>
<td>0.20</td>
<td>0.53</td>
</tr>
</tbody>
</table>

For most of the functions inspected, the mean ratings in the two conditions differ to a statistically significant extent. Signal-Prominence and Depict-Shape are the only exceptions to this trend. However, because of the large data sample, p-values can be somewhat misleading (there is a high chance of false positives). The effect sizes (Cohen's d and r) can be more informative as to how different the ratings are in the presence or absence of speech. Many of the d-values are around 0.3 or lower (small, according to the standards formulated by Cohen 1992). Only for Depict-Movement and Refer-to-Place does the presence of speech have a substantial impact on the interpretation (with effect sizes higher than 0.5; moderate to large according to Cohen’s standards).

In the light of the discussion of the examples above, the fact that space-related functions (Depict-Movement and Refer-to-Place) are most strongly dependent on speech is understandable. Because manual gestures are unavoidably performed in some area of space, their location cannot be expected to be semantically relevant in all cases. Whether the location of the gesture actually maps onto some spatial characteristic of the scene described is often only clear when the verbal channel is accessible. In example (22), for instance, the gesture was not interpreted as spatial in the Sound-Off condition, but when perceived in the
Perception of gestural functions with and without speech

presence of the noun phrase *zwei Feuertreppen*, it was interpreted as depicting a spatial relationship. Despite the (overall modest) differences in the mean ratings between conditions, there are strong correlations between the ratings for all of the functions taken into account – most of the correlation coefficients are around 0.5 or higher. This suggests that overall, the ratings across conditions were much alike: the interpretations of the gestures were qualitatively in the same direction, but generally expressed with more confidence when the speech was present.

5.5 **Summary and conclusion**

Although the question of speech-gesture dependence has a long history, the inherent meanings of speakers’ gestures have to date not been characterized in a way that is compatible with (cognitive-functional) linguistic theory. In this chapter, I have argued that a functional view on gestural meaning could provide a fruitful level of abstraction to determine which aspects of gestural meaning can be understood without speech. On the basis of an internet-based study, I have presented a detailed comparison of the ways the gestures in the corpus are perceived when presented with and without access to speech.

The results suggest that access to the verbal channel has only a moderate influence on the perceived functions of the gestures in the corpus. Pairwise t-tests yielded rather small effect sizes (between one third and one fifth of a standard deviation) and response profiles show a great deal of overlap across conditions for all functions investigated. However, the variance in the responses between conditions was generally higher for the Sound-Off condition – participants were more hesitant and inconsistent in their responses when they judged the gestures on the basis of the visual channel alone. In all, the presence of audio seems to have little impact on the qualitative interpretation of the accompanied gestures, but it causes the raters to be more assured of their judgments.

These findings raise theoretical as well as methodological considerations. First, they suggest that the question of whether gestures constitute an independent (sub)system of communication hinges strongly on the level of abstraction one adopts when characterizing their
meaning. Using somewhat schematic, functionally oriented questions, the dependence of gestures on speech was found to be smaller than what has been concluded from experiments that have assessed gestural meaning through very specific questions (e.g. ‘what is the lexical affiliate of this gesture?’). This is an important insight, because it informs the level of description appropriate for understanding how the meanings of speech and gesture relate to each other. Instead of seeking to understand gesture as a system that functions akin to speech, one needs to ask the question: on what level of abstraction do the two channels intersect? In accordance with the results of the previous chapter, the results of this study suggest that the current way of operationalizing gestural functionality provides a fruitful way of capturing the qualities of gestures that are inherently meaningful.

A possible drawback of the current study is that the full body of the speakers was visible in both conditions. To better isolate the specific contributions of manual gestures to the functional ratings assigned, judges would need to be denied visual access to other articulators, such as body posture, facial movements and eye gaze. These behaviors can have substantial impact on the perceived function of a gesture, but their influence was not properly controlled for in the current design. Another question that remains unanswered is to what extent the perceived function of a gesture can be predicted on the basis of its formal features. The results from the current chapter suggest that the physical properties of a gesture give some cues as to what function it performs, but it remains unclear what aspects of form and function are associated with each other. The following chapter examines this question in greater detail.
Chapter 6. Mapping functions to forms

Given that the functions of gestures are to some extent interpretable in the absence of speech, the question arises whether these functions are systematically linked to patterns of form. In grammatical terms, this question translates into whether there is a set of primitive phonological or morphological structures that underlie gestural expression. Although numerous regularities and patterns have been found to exist in gestural expression, as discussed in Chapter 3, there is little consensus on how to define the elementary structures and contrastive features. McNeill (1992) even denies that gestures are subject to a ‘duality of patterning’ in the first place; he maintains that gestural forms and meanings cannot be separated in the same way as those of verbal expressions (see Stokoe 1991 for a comparable argument with respect to iconic aspects of sign languages). As a methodological consequence of this influential view, many studies on gesture employ classification systems with categories that essentially conflate function and form.

Another factor that has impeded consensus on the existence of primitive structures is that no fully established system exists for describing and documenting gestural forms. Researchers have employed a variety of different schemes for annotating gestural behaviors in terms of formal and functional parameters (Bressem 2013; Stec 2014). Classification schemes are typically tailored to the specific research
goals of the study they are employed for. Only recently have a number of attempts been made to devise more comprehensive annotation systems that can be applied across research contexts (e.g. LASG – Bressem et al. 2013; and NEUROGES – Lausberg & Sloetjes 2009). Both the LASG and the NEUROGES systems annotate form and function on independent tiers, thus deviating from McNeill’s integrative view. They prescribe a ‘phonological’ dissection of the forms of manual gestures, inspired by Stokoe’s (1960/2005) system for documenting forms in sign languages. In most coding schemes for form annotation, the most central parameters are handshape, palm orientation, hand location, and movement direction, sometimes supplemented by additional tiers that describe the dynamics and the degree of tension in the hands. The underlying assumption is that these parameters capture most of the degrees of freedom of manual expression.

The current chapter contributes to this line of research, using a methodology that is innovative in a number of ways. A first novel aspect is that the functions of interest are motivated explicitly by cognitive-functional theories of grammar (see Chapter 4). By adhering to specific models of language, the current study goes beyond approaches that have studied gestural patterns in terms of broad, generic categories (e.g. ‘iconic’ as a single class of gestures). Second, the current study does not rely on individual annotations to characterize the functionality of the gestures in the corpus, but instead uses the crowdsourced functional ratings described in Chapter 4. As argued in Chapter 4, this method has the benefit of revealing collective, rather than individual intuitions, and it accounts for different degrees of functional prominence. Third, the current approach works top-down, that is, from function to form. Thus, it has the potential to complement research with a form-to-function orientation (e.g. Kendon 2004; Müller et al. 2013). Moreover, this approach resonates with the top-down architecture of functional linguistic models such as FDG. Finally, the research design relies solely on statistical models for identifying the form parameters that are relevant for each of the gestural functions inspected. Before explaining the motivations behind, and details of the statistical methods, I first turn to the selection and annotation of the form parameters.
6.1 Form annotation

This section presents the coding scheme that was used for annotating the formal features of the gestures in the corpus. The choice of parameters and categories was roughly in concordance with earlier linguistic-semiotic approaches (Bressem et al. 2013; Calbris 1990; Webb 1996). It took as a basis the annotation manual of the SaGA corpus (Freigang & Bergmann 2013). Because this annotation system is geared toward computational modeling (Kopp et al. 2008; Lücking et al. 2013) and therefore is not entirely suitable for the current research questions, I made use of a modified version. The coding process involved four steps. First, for all gestures that had already been coded as a part of the SaGA project, the existing annotation labels were merged into a smaller number of categories, so that label frequencies would be sufficient for reliable statistical analysis (see below for details). Second, those gestures that had not yet been annotated in the original SaGA corpus were coded directly in terms of the new, simplified categories (the SaGA only contains form annotations for the gestures that were labeled as ‘iconic’; it lacks annotations for discourse-related gestures). During the third step, all annotations were reviewed to assure consistency between the recoded data and the newly coded data. Finally, to guarantee reliability of the modified coding scheme, all form parameters were cross-coded by one out of four secondary coders, based on a portion of the data set. In cases where the cross-coding procedures revealed inconsistency between the primary coder (myself) and the secondary coder, cases of disagreement were discussed in order to establish an understanding of the differences in judgment. After this discussion, another round of coding was performed by the secondary coder on a different subset of the videos. This procedure was repeated until the coders converged on their interpretation of the categories. Agreement ratings (percentages and kappa scores) will be listed below for all form parameters individually. A brief description of the coding schemes for all form features is given next.

6.1.1 Handedness

In a first round of annotation, I divided the data set into left-handed, right-handed and two-handed gestures (Table 6.1). The hands of the
speaker were regarded as ‘taking part in’ the gesture whenever they were markedly removed from rest position. The only exceptions were cases of self-adaptors, where the hands were clearly oriented toward the speaker’s own body, for instance when scratching one’s head. No cross-coding was applied, since very little subjective interpretation was involved in this round of annotation.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-handed</td>
<td>Only the left hand is moved away from rest position</td>
<td>39 (8.4%)</td>
</tr>
<tr>
<td>Right-handed</td>
<td>Only the right hand is moved away from rest position</td>
<td>200 (43.3%)</td>
</tr>
<tr>
<td>Two-handed</td>
<td>Both hands are moved away from rest position</td>
<td>223 (48.3%)</td>
</tr>
</tbody>
</table>

We see that nearly half of the gestures in the data set were performed with one hand. Of these, the vast majority were right-handed. Four of the five participants made barely any left-handed gestures at all, whereas one (presumably left-handed) participant made about fifty percent more gestures with the left than with the right hand.

The original SaGA corpus contains separate tiers for the two hands of the speakers. Because this often leads to redundant annotations – the hands typically do not act independently but constitute a single unit of expression – I assumed only one annotation value per gesture. To this end, an additional round of coding was performed where, for all two-handed gestures, coders indicated which hand they thought was dominant (if any). Of the 223 two-handed gestures, 43 were classified as left-dominant and 67 as right-dominant (88% agreement, $k = 0.55$). For the remaining 113 gestures, neither hand was clearly dominant. The latter category includes gestures whereby the hands were positioned symmetrically and shaped alike. All remaining layers of annotation were applied only to the dominant hand or, in case of equal dominance, to both hands.
6.1.2 Handshape

The original annotation manual of the SaGA corpus involves a very detailed inventory of possible handshapes. As many as seventeen possible finger configurations are distinguished, in addition to sequential combinations of them and an optional tag that specifies whether the fingers are tense or loose. Applying this annotation scheme would result in an abundance of unique labels, which is inappropriate for the intended analyses. In order to make the categorization of handshapes more manageable, I employed a substantially simplified scheme. For gestures that combined two handshapes, coders were instructed to focus only on the handshape that was in their judgment most salient. When a gesture involved more than two consecutive handshapes, it was labeled as Complex. The degree of tension in the hands was not included in the annotation scheme, because regarding tension as an independent parameter can be problematic (some handshapes are inherently more tense than others). The remaining labels were merged into five global categories, defined in Table 6.2. Inter-coder reliability, based on fifty videos, was reasonable (70%, $\kappa = 0.58$). The most common disagreements between the coders concerned the boundaries between the categories Spread-palm-straight, Spread-palm-curved and Cupped-hand, as the data contain numerous handshapes that are on the borders between these categories.

<table>
<thead>
<tr>
<th>Handshape category</th>
<th>Description</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spread palm straight</td>
<td>All fingers of the hand are extended and stretched out</td>
<td>97 (21.0%)</td>
</tr>
<tr>
<td>Spread palm curved</td>
<td>All fingers of the hand are extended, but slightly curved</td>
<td>176 (38.1%)</td>
</tr>
</tbody>
</table>
Gemures with a loose, spread palm are the most common in the corpus. These are arguably also the least effortful and the least marked. The Index-only and Spread-palm-straight categories are also rather frequent. Note that these are physiologically not far removed from the Spread-palm-curved category, but generally involve more tension in the hands.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cupped hand</td>
<td>All fingers of the hand are curved, shaped as a cup</td>
<td>46</td>
<td>10.0%</td>
</tr>
<tr>
<td>Index only</td>
<td>Only the index finger is stretched, all other fingers are curved</td>
<td>124</td>
<td>26.8%</td>
</tr>
<tr>
<td>Other</td>
<td>All other handshapes, including clenched fists and gestures whereby two, three or four fingers are stretched</td>
<td>19</td>
<td>4.1%</td>
</tr>
</tbody>
</table>
Gestures in the category Other, which covers all non-canonical handshapes (see Table 6.2 for some examples), only have 19 occurrences.

6.1.3 Orientation
The SaGA annotation manual prescribes separate annotations for the orientation of the front of the hand, the orientation of the back of the hand, and the orientation of the wrist. It distinguishes six basic directions of orientation, corresponding to the horizontal (left-right), vertical (up-down) and sagittal (behind – in front) axes. In addition, the original classification scheme includes detailed labels to deal with in-between cases and successions of orientations (e.g. Palm-towards-left/Palm-towards-body>Palm-towards-right as a single tag).

The categories used in the current classification scheme emerged from substantial simplification of the original labels. The orientation of the hand palm was taken as central (i.e. wrist orientation was not taken into account separately). The remaining categories were simplified using a strategy similar to that used in the case of handshape. In cases where the gesture involved two subsequent orientations, only the second orientation was taken into account. Sequences of more than two different hand orientations were coded as Complex. All in-between categories (e.g. Palm-left/Palm-up) were reduced to a single orientation, where the first priority was given to the sagittal axis, the second priority to the vertical axis and the third to the horizontal axis (e.g. Palm-away-from-body/Palm-down was coded as Palm-away-from-body; Palm-up/Palm-left was coded as Palm-up). This prioritization mirrors the different degrees of effort that are involved in orienting the hands along these three dimensions. The application of the revised coding protocol resulted in a total of six categories for handshape: Palm-up, Palm-down, Palm-towards-body, Palm-away-from-body, Palm-sideways and Complex (Table 6.3). Agreement across coders was substantial (76%, $\kappa = 0.69$). The most common disagreements between the coders concerned cases where the hands made subtle movements or rotations during the course of the stroke. Note that the label Palm-sideways includes both leftward and rightward orientations, so that the handedness of the speakers could not be a confounding factor.
Table 6.3 Orientation annotation

<table>
<thead>
<tr>
<th>Orientation category</th>
<th>Description</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palm sideways</td>
<td>The palm is oriented to the left or to the right during the stroke phase</td>
<td>171 (37.0%)</td>
</tr>
<tr>
<td>Palm down</td>
<td>The palm is oriented downward during the stroke phase</td>
<td>128 (27.7%)</td>
</tr>
<tr>
<td>Palm up</td>
<td>The palm is oriented upward during the stroke phase</td>
<td>33 (7.1%)</td>
</tr>
<tr>
<td>Palm away from body</td>
<td>The palm is oriented away from the body during the stroke phase</td>
<td>74 (16.0%)</td>
</tr>
<tr>
<td>Palm towards body</td>
<td>The palm is oriented toward the body during the stroke phase</td>
<td>43 (9.3%)</td>
</tr>
<tr>
<td>Complex</td>
<td>The palm has more than two orientations during the stroke phase</td>
<td>13 (2.8%)</td>
</tr>
</tbody>
</table>

It is noticeable that the three main axes of orientation – horizontal, vertical and sagittal – are approximately equally represented; all have more than one hundred occurrences. On closer inspection, the distribution of orientations along the vertical axis appears most skewed. Palm-down orientations are almost four times more frequent than palm-up orientations. Gestures in the Complex category, where the palm is oriented in three or more directions during a single stroke, are uncommon in the data set.

6.1.4 Hand Position

The coding of the position of the speaker’s dominant hand was loosely based on McNeill’s (1992: 89) grid, shown in Figure 6.1. As seen in the figure, the original grid was strongly simplified, so that each gesture could be assigned to one of five locations. To promote conformity in the interpretation of the categories, additional instructions were given to the coders concerning the borders between the different segments of the grid. The region of space right below the speaker’s shoulder was considered as the outermost inclusive border of the Central categories (note that the label Central is used for the horizontal axis, whereas Center is used for the vertical axis). The Lower regions included, by
definition, all gestures whereby the upper arm was directed downward and made an angle of ninety degrees or more with the lower arm. After explicitly establishing these criteria among the coders, reliability based on fifty videos reached 72% (κ = 0.60). Despite the rather specific instructions, most of the disagreements between the coders concerned the borders between the Center and Lower areas and those between the Central and Lateral areas – many of the gestures in the corpus are performed between or across these regions. Gesture counts per category are given in Table 6.4.

Figure 6.1 Space segmentation, loosely based on McNeill’s (1992: 89) grid
Table 6.4 Annotation of hand location

<table>
<thead>
<tr>
<th>Hand position category</th>
<th>Description</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center-Central</td>
<td>During the stroke phase, the hand is held in front of the chest</td>
<td>202 (43.7%)</td>
</tr>
<tr>
<td>Center-Lateral</td>
<td>During the stroke phase, the hand is held on chest-height, left or right of the body</td>
<td>64 (13.8%)</td>
</tr>
<tr>
<td>Lower-Center</td>
<td>During the stroke phase, the hand is held in front of the stomach or waist</td>
<td>109 (23.6%)</td>
</tr>
<tr>
<td>Lower-Lateral</td>
<td>During the stroke phase, the hand is held on waist-height, left or right of the body</td>
<td>23 (4.9%)</td>
</tr>
<tr>
<td>Upper</td>
<td>During the stroke phase, the hand is held on or above shoulder-height</td>
<td>64 (13.9%)</td>
</tr>
</tbody>
</table>

The label frequencies indicate that on both the horizontal and vertical axes, central regions are best represented. In addition, we see that lower regions are used substantially more often than upper regions. These findings are in line with those of McNeill (1992; Ch. 3), who used more fine-grained spatial categories.

6.1.5 Movement type

The original SaGA annotation scheme contains separate annotation tiers for movement of the wrist through space and for movement of the hand relative to the wrist. In the current analyses, these annotations were merged into a single tier, which subsumes the most prominent categories of both wrist and hand movement. The label Simplex-line-or-arc was assigned to gestures whereby the hand moved along a single straight or curved line. Linear combinations of two or more of such movements were classified as Complex-line-arc-sequence. If the hand was moved repeatedly along the same trajectory, it was coded as Repeated-line-or-arc. Gestures whereby the hands were rotated back and forth around the wrist were given the tag Wiggle-sway. The label Energetic-hold was assigned to gestures whereby the hands were positioned in a fixed location for several seconds, with extraordinarily high tension in the fingers and (optionally) a subtle repetitive back and forth movement. The final category is that of Punctual movement: gestures where the energy of the stroke was concentrated in an
instantaneous moment. The overall agreement for this parameter was 72% ($\kappa = 0.59$). The most common disagreements concerned movements that were so subtle that both Punctual and Simplex-line-or-arc seemed to be appropriate labels. Table 6.5 shows the frequencies of the labels in the corpus.

<table>
<thead>
<tr>
<th>Movement type</th>
<th>Description</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctual</td>
<td>The stroke is instantaneous, i.e. the effort is concentrated in a moment of negligible duration</td>
<td>160 (34.6%)</td>
</tr>
<tr>
<td>Simplex line or arc</td>
<td>The hand makes an effortful movement through space along a straight line or arc</td>
<td>177 (38.3%)</td>
</tr>
<tr>
<td>Complex line-arc sequence</td>
<td>The hand makes an effortful movement through space along a complex path, encompassing more than one line or arc</td>
<td>74 (16.0%)</td>
</tr>
<tr>
<td>Repeated line or arc</td>
<td>The hand makes a repeated (back and forth) movement through space</td>
<td>15 (3.2%)</td>
</tr>
<tr>
<td>Wiggle-sway</td>
<td>The hand is rotated back and forth around the wrist</td>
<td>17 (3.7%)</td>
</tr>
<tr>
<td>Energetic hold</td>
<td>The hand is held stationary, but with extraordinarily high tension and/or a subtle back and forth movement</td>
<td>19 (4.1%)</td>
</tr>
</tbody>
</table>

The distribution of the gestures among the Movement Type categories is more skewed than for the other form parameters seen thus far. Simplex-line-or-arc and punctual gestures are by far the most frequent, followed by the Complex-line-arc-sequence category. Of the remaining categories, none cover more than 10% of the data set.

6.1.6 Movement Direction

The coding of the direction of the movement was much akin to that of hand orientation. Again, the original annotations tend to be rather complex, including combinations of spatial axes (e.g. downward/leftward) and linear successions (e.g. downward $>$ upward $>$ downward). In merging these annotations into a smaller number of labels, all annotations with more than two consecutive movement
direction labels were coded as Complex. For gestures with one or two successive movement directions, coders selected the axis they perceived as most dominant. For movements directed along two or more axes simultaneously, the same prioritization order was used as for the orientation parameter (sagittal > vertical > horizontal). The resulting label counts are shown in Table 6.6. Inter-coder reliability was 72% (κ = 0.62). Disagreements among coders were not concentrated in specific pairs of categories, but concerned various types of diagonally oriented gestures.

### Table 6.6 Annotation of movement direction

<table>
<thead>
<tr>
<th>Movement-direction</th>
<th>Description</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctual-static</td>
<td>The hand does not undergo directed movement during the stroke phase (i.e., its position is constant or it is only moved around its own axis). This includes the movement types Punctual and Energetic-hold, as well as those Wiggle-sway gestures where the wrist only moves around its own axis.</td>
<td></td>
</tr>
<tr>
<td>Horizontal</td>
<td>The hand is moved predominantly sideways (leftward or rightward) during the stroke phase</td>
<td></td>
</tr>
<tr>
<td>Vertical</td>
<td>The hand is moved predominantly along the vertical axis (upward or downward) during the stroke phase</td>
<td></td>
</tr>
<tr>
<td>Sagittal</td>
<td>The hand is moved predominantly along the sagittal axis (toward or away from the body) during the stroke phase</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>The hand makes consecutive movements in more than two different directions during the stroke phase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>185 (40.0%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72 (15.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52 (11.3%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>87 (18.8%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>66 (14.2%)</td>
<td></td>
</tr>
</tbody>
</table>

In line with what we have seen for the Movement Type parameter, a large portion of the gestures are labeled as Punctual; many gestures do not involve any effortful movement of the hand during the stroke phase. Movements in the remaining directions (and complex movements) are fairly evenly distributed.
6.1.7 **Temporal connectedness**

A final layer of coding was implemented to capture the temporal structure of the gestures in the corpus. That is, each of the gestures was classified as being performed in isolation or as part of a sequence. Moments where the speaker returned the hands to rest position in between gesture strokes were considered to mark the beginning and ending of a gesture sequence. Connected strokes were labeled in terms of the length of the sequence they were part of, resulting in the distribution summarized in Table 6.7. The analysis of this form parameter in Section 6.2.8 only compares isolated and sequenced gestures as two general categories, without considering the length of the sequence.

<table>
<thead>
<tr>
<th>Temporal relation</th>
<th>N in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-stroke gesture</td>
<td>126 (27.3%)</td>
</tr>
<tr>
<td>Part of two-stroke gesture sequence</td>
<td>101 (21.9%)</td>
</tr>
<tr>
<td>Part of three-stroke gesture sequence</td>
<td>114 (24.7%)</td>
</tr>
<tr>
<td>Part of four-stroke gesture sequence</td>
<td>36 (7.8%)</td>
</tr>
<tr>
<td>Part of five-stroke gesture sequence</td>
<td>55 (11.9%)</td>
</tr>
<tr>
<td>Part of six-stroke gesture sequence</td>
<td>30 (6.5%)</td>
</tr>
</tbody>
</table>

As seen in the table, roughly one out of four gestures in the corpus was performed in isolation. Two-stroke and three-stroke successions are the most frequent types of sequences, together amounting to almost half of the stimulus set. Recall from the description of the data set (Chapter 4) that no gesture sequences were included that combined more than six consecutive strokes.

In all, the frequency counts presented in this section suggest that the gestures in the corpus show extensive variation in various dimensions of form. In most cases, the distribution of the gestures over the form categories is somewhat skewed: certain handshapes, regions of space, orientations and movement trajectories are more frequent than others. This alludes to some degree of systematicity on the formal side of gestural expression. An important question, yet unanswered by these statistics, is whether the observed regularities are solely indicative of physical constraints (some movements require less effort than others),
or are also motivated by linguistic-communicative principles. The following section looks deeper into the question of whether patterns in form are consistently linked with the perceived functions of the gestures.

### 6.2 Function-form relations

In this section, I bring together the functional characterization of the gestures described in Chapter 4 and the form annotations outlined in Section 6.1. Thus, I investigate whether systematic relationships exist between patterns in form and function. The research strategy adopted can be characterized as 'onomasiological' (cf. Fricke et al. 2014), as it goes from function to form. The question asked is whether the functions discussed in Chapter 4 are linked to specific patterns of form. A theoretical motivation for this approach is that it complements the 'semasiological' (form-to-function) approach that is dominant in the literature. With some notable exceptions (Bergmann & Kopp 2010), most previous studies have departed from the observation that certain formal patterns exist, and explored the functional contexts in which these occur (Kendon 2004; Müller et al. 2013). An additional, linguistic motivation for the current approach is that it resonates with the rationale behind functional grammatical models – FDG in particular – which explicitly adopt a function-to-form orientation.

The methodology can furthermore be characterized as inductive. That is, it is not hypothesis-driven, but exploratory in nature. In principle, all possible relations between functional and formal aspects of the gestures in the corpus are taken into consideration. This approach minimizes the risk of being biased by a priori intuitions or expectations. In order for the inductive strategy to be viable, it is important to be aware of the risk of combinatorial explosion. The possible number of variables on the functional side (nine different functions) and the formal side (six different levels of form annotations, and combinations of them) is too massive for all combinatorial possibilities to be considered in equal detail. As a way of avoiding this pitfall, I adopt a two-step method. First, I make an estimate, for each of the gestural functions inspected, of which form parameters are most important (i.e., statistically discriminative). Then, I take only the most important formal
parameter(s) into closer consideration. The assessment of the importance of the form parameters is accomplished through linear regression analysis. For each function, a linear model is constructed with the crowdsourced ratings (in the Sound-On condition) as dependent variable and the six formal parameters outlined in Section 6.1 as predictors. That is, for each functional dimension, the regression model is specified as:

\[
\text{Functional rating} \sim \\
\quad \text{handshape} + \\
\quad \text{hand orientation} + \\
\quad \text{hand location} + \\
\quad \text{hand configuration} + \\
\quad \text{movement type} + \\
\quad \text{movement direction}
\]

The relative importance of each of the formal parameters is subsequently determined using the 'lmg' method (Chevan & Sutherland 1991; Lindeman et al. 1980; 'relaimp' package in R), which estimates the contribution of the predictors to the explained variance by averaging over all possible orderings of the regressors. For the formal parameter(s) that make the greatest contribution to the model, function-form associations will be explored further by comparing the mean ratings per annotation level using t-tests (with non-pooled standard deviations, to account for unequal sample sizes). For instance, when the regression model reveals that the parameter Handshape is the most important predictor of a given function, I subsequently explore which of the specific handshape categories corresponds to the highest ratings. In cases where two form features stand out as most discriminative, I inspect both parameters independently. The labels used to refer to the gestural functions follow the conventions introduced in Section 4.2.4. That is, I will use phrases like the scores on Depict-Shape as a shorthand reference to the mean Likert ratings assigned to the question with this label in the online experiment (i.e., 'does the gesture depict the size or shape of an object?').
6.2.1 **Formal correlates of shape depiction**

When treating the scores on Depict-Shape as dependent variable, the full model accounts for 43.3% of the variance in the data ($F_{(24,437)} = 16.1$, $p < 0.001$). The relative contributions of each of the form features are plotted in Figure 6.2 (normalized to sum 100%).

![Relative importance of the form parameters for predicting the ratings on the Depict-Shape question](image)

We see that several form features contribute to the predictive ability of the model. The Movement Type parameter stands out as most important (28.2% of the explained variance). The other form dimensions, with the exception of Orientation, also explain at least 15%. This suggests that shape attribution is associated with a range of possibly interdependent formal characteristics.

Figure 6.3 displays the mean scores and 95% confidence intervals for each of the Movement Type categories. The charts here and below also display the Bonferroni-corrected outcomes of a set of t-tests that were conducted to compare the mean scores between all pairs of categories. Following conventions in statistical notation, a single asterisk * represents a p-value between 0.05 and 0.01, a double asterisk ** represents a p-value of between 0.01 and 0.001, and a triple asterisk *** represents a p-value lower than 0.001.
From visual inspection of Figure 6.3, it appears that gestures with different movement types are not equally likely to be interpreted as depicting the shape of an object. The inequality of the mean ratings per category is confirmed by an ANOVA test (type 3; $F_{(5,456)} = 27.8$, $p < 0.001$, $\omega^2 = 0.22$).

Two movement types are noticeable as being most strongly associated with shape depiction: Energetic-hold and Complex-line-arc-sequence. The high scores for Energetic-hold handshapes are plausibly related to the fact that this movement type is typical for the depiction strategy whereby the hands ‘embody’ the referent (Müller 1998a). By holding the hand(s) in a specific shape with high tension in the fingers, a speaker can direct attention to the shape of the hand (e.g. a cupped hand representing a bowl). The gestures in the Complex-line-arc-sequence category receive the second highest ratings. This category of movements is characteristic of another common way of shape depiction: tracing the outline of an object. As discussed in Chapter 3, tracing gestures involve the creation of a mapping between the movement of some edge of the

![Figure 6.3 Mean ratings on Depict-Shape by movement type category](image-url)
hand (e.g., a fingertip) and the physical contours of the object referred to. Complex sequences of movements are especially apt for this type of depiction, as they allow for the successive profiling of multiple dimensions or edges of an object. Simpler movements may also serve this purpose, but they can provide less spatial detail. By contrast, gestures whereby energy of the stroke is instantaneous do not allow at all for creating movement-to-shape mappings. This is a plausible explanation for the finding that the lowest ratings were assigned to gestures in the Punctual category.

6.2.2 **Formal correlates of object reference**

With the ratings on Refer-to-Object as dependent variable, the full regression model explains 38.7% of the variance ($F_{(24,437)} = 11.49, p < 0.001$). Figure 6.4 displays the estimated contributions of each of the form features to the regression model.

![Figure 6.4 Relative importance of the form parameters for predicting the ratings on Refer-to-Object question](image)

Despite the fact that the scores on the Refer-to-Object function correlate strongly with those on the Depict-Shape function, as seen in Chapter 4, a different set of form parameters shows up as most important. Handshape and Hand Position make the greatest contribution to the regression model. In Figure 6.5 and Figure 6.6, I examine the relation between these form features and object reference in more detail.
With regard to the relation between object reference and hand position, an ANOVA test suggests that the average scores per category are not equal ($F_{(4,457)} = 19.80, p < 0.001, \omega^2 = .16$). A linear relationship seems to exist between the vertical position of the hands and the likelihood that a gesture is interpreted as referring to an object: gestures in the lower regions receive the lowest scores, those in the central regions receive higher scores, and those performed in the upper regions of space were assigned the highest ratings on average. This trend is likely to be related to the fact that lower regions of gesture space are relatively unmarked and therefore less likely to be used for referential gesturing. Along the horizontal axis, there are remarkably small differences: the mean scores for lateral and central segments of gesture space are not statistically different. Thus, with respect to the relation between space and the referential function of gestures, the vertical axis appears most informative.

Figure 6.5 Mean ratings on Refer-to-Object by hand position category
Figure 6.6 expresses the relation between object reference and the handshape parameter, which makes the second strongest contribution to the model.

![Graph showing mean ratings on Refer-to-Object by handshape category](image)

The handshape categories were assigned significantly different ratings ($F(4,457) = 15.3, p < 0.001, \omega^2 = .11$). As seen in the plot, the Cupped-hand and Index-only handshapes are more strongly associated with object reference than the three other categories. A likely explanation is that both handshapes are associated with specific ways of making reference. Cupped handshapes are characteristic of the 'holding' mode of representation; gestures with this handshape can be interpreted as holding or carrying some entity on top of the palm. Handshapes with a stretched index finger (Index-only) are typical of deictic reference through finger pointing, and/or reference by tracing (recall that the coding scheme is not granular enough to distinguish between the handshapes used for pointing and tracing).

The finding that gestures with stretched index fingers are not given as high ratings as those with cupped handshapes is plausibly
indicative of a higher degree of ‘polysemy’ for the former category. Index-only gestures are not only associated with (deictic) reference to concrete objects, but also with a range of pragmatic functions, such as signaling agreement and interaction management (Bavelas et al. 1992). Cupped handshapes seem to be more unequivocally associated with the function of reference. In fact, the scores on Refer-to-Object become gradually higher as a function of the degree of finger bending; cupped hands are significantly more likely to be interpreted as referential than gestures in the Spread-hand-bent category, and yet lower scores were assigned to gestures in the Spread-hand-straight category. These results suggest a general association between handshapes with bent fingers and the reification of conceptual content.

6.2.3 **Formal correlates of place reference**

A portion of 28.7% of the variance in the ratings on the Refer-to-Place function is explained by the regression model ($F_{(24,437)} = 7.34, p < 0.001$). The relative importance of each of the regressors is displayed in Figure 6.7.

![Relative importance for Refer-to-Place](image)

Figure 6.7 Relative importance of the form parameters for predicting the ratings on the Refer-to-Place question

We see that Orientation is the strongest predictor, followed by Movement-Type. The finding that Hand Position only accounts for 13.8% of the explained variance is remarkable. Given the potential for iconic use of gestures space – speakers can position their hands in a location relative to their body to refer to an analogous region of space in the
described situation – one might expect this parameter to be the most predictive one. Figure 6.8 shows the mean ratings for the Orientation parameter, which contributes most strongly to the explained variance.

As we can see in Figure 6.15, there is not a single orientation-related correlate of spatial reference (ANOVA: F(5,456) = 12.8, p < 0.001, $\omega^2 = 0.11$). The predictive power of the hand orientation parameter is mainly driven by the low scores for gestures in the Palm Up category. The ratings for the other categories are statistically indistinguishable. The observation that spatial reference is unlikely to be accomplished by gestures with the palm oriented upward could have various underlying reasons. For one, the two sides of the hands do not have the same potential for making indexical reference: pointing or placing is much more likely to be performed with the palm of the hand than with the back of the hand. Therefore, making reference to a specific spatial region is unlikely to be achieved with the palm up (unless the location referred to occupies a higher area of the visual field, but this is uncommon in most discourse settings). An additional explanation could be that a
spatial interpretation of palm-up gestures is 'pre-empted' by pragmatic functions. Palm-up-open hand gestures are related to a number of discursive phenomena, such as signaling emphatic stress, display of reduced commitment and discourse segmentation (Ferré 2011; Kendon 2004; Müller 2004). It is possible that an interpretation of palm-up gestures as carrying out one of these functions is generally favored over an iconic or indexical interpretation.

6.2.4 **Formal correlates of process depiction**

The regression model with the scores on Depict-Process as the dependent variable has a non-zero slope ($F_{(24,437)} = 8.86, p < 0.001; R^2 = 32.7\%$). Figure 6.9 shows how important each of the form parameters is for predicting the ratings assigned for this function.

![Relative importance of the form parameters for predicting the ratings on the Depict-Movement question](image)

As can be expected, the movement-related form parameters (Movement Direction and Movement Type) are among the strongest contributors to the model's predictive capacity. In addition, we see that handshape plays a major role. Below, I examine the relation between movement depiction and the two most discriminative form parameters: Movement Direction and Handshape.
As seen in Figure 6.10, the mean scores per movement direction category diverge substantially ($F_{(4,457)} = 22.6$, $p < 0.001$, $\omega^2 = 0.16$). A somewhat obvious finding is that Punctual-static gestures are unlikely to be interpreted as representing motion. Since gestures of this type do not involve effortful movement through space, they have less potential for depicting spatial processes than other gestures. However, as we can see in Figure 6.10, the gestures with non-punctual strokes are not equally associated with movement depiction. Gestures whereby the movement is predominantly performed along the vertical axis are given lower ratings than those performed along the horizontal and sagittal axes. A possible explanation is that route direction discourse involves many descriptions and instructions that are horizontal in nature (e.g. *turn left/right*) or refer to the sagittal axis (e.g. *keep going straight*). A second reason is that many of the objects referred to in the corpus – landmarks such as trees, buildings and statues – have a salient vertical dimension. The relatively frequent occurrence of gestures that trace the vertical contours of these landmarks may have resulted in comparatively low average ratings of vertically directed gestures for the Depict-Movement...
function. As both of the explanations mentioned have to do with specific characteristics of the route direction discourse, it remains unclear to what extent the current findings can be generalized to other discursive situations.

In Figure 6.11, we see how the scores on Depict-Movement vary with the categories of the Handshape parameter \(F_{(4,457)} = 16.3, p < 0.001, \omega^2 = 0.12\). The highest scores were assigned to gestures with the palm spread out and the fingers stretched, followed by those with only the index finger extended. This pattern is in line with previous analyses of handshapes in route direction (Fricke 2007, 2014b). The two handshapes with the highest ratings roughly correspond to what Fricke (2014b) calls 'G-shape' gestures and 'palm-lateral-open-hand' (PLOH) gestures. In line with Fricke’s qualitative observations, spread hands with stretched fingers are strongly associated with movement depiction (compared to, for instance, gestures where the fingers are slightly bent). Thus, despite the fact that the physiological differences between these handshapes are quite minimal, fully and partially stretched hand palms

![Figure 6.11 Mean ratings on Depict-Movement by handshape](image)

Figure 6.11 Mean ratings on Depict-Movement by handshape
seem to have considerably distinct functional interpretations. Another conspicuous finding is that cupped handshapes are assigned by far the lowest scores. A possible explanation is that, as discussed earlier, cupped handshapes are often associated with holding a (concrete or abstract) object. Given that shape depiction and movement enactment are negatively correlated (at least in the current corpus, see Section 4.3.4), cupped handshapes are unlikely to be combined with movement-related functions. Another potential explanation of this finding is that cupped handshapes do not inherently have a clear edge that can trace the path of the movement, like the tip of the index finger in gestures with an Index-only handshape.

6.2.5 **Formal correlates of quantity signaling**

The regression model with Signal-Quantity as dependent variable explains 25.8% of the variance in the data ($F_{(24,437)} = 6.33, p < 0.001$). As seen in Figure 6.12, a small number of form parameters are responsible for the explained variance.

![Relative importance of the form parameters for predicting the ratings on the Signal-Quantity question](image)

The skewed distribution seen in Figure 6.12 suggests that the gestural conveyance of quantity information is linked to relatively stable patterns of form. The fact that Handshape and Hand Configuration are among the most predictive features is in line with the intuition that there are two

---

34 The experienced difference between these handshapes, in terms of effort required, might be larger than the ‘measurable’ difference between them.
prominent ways of giving quantitative information using the hands. The most obvious one is the use of a handshape whereby one or multiple fingers are raised vertically. A second possible way of signaling quantity (or more specifically: duality) is by using two hands to refer to two separate entities. To see whether the data support these intuitions, I inspect the two strongest predictors in more detail (Figure 6.13 and Figure 6.14).

![Ratings on Number by handshape](image)

**Figure 6.13 Mean ratings on Signal-Quantity by handshape**

A comparison of the mean ratings per category suggests that the scores on Signal-Quantity vary as a function of the handshape parameter (ANOVA: \( F_{(4,457)} = 13.7, p < 0.001, \omega^2 = 0.10 \)). Gestures in the category Other receive the highest scores on average, although due to the high variance, not all pairwise comparisons with the other categories reach statistical significance. The high mean rating is likely driven by the presence of typical number-signaling handshapes in this category (those with one or multiple fingers outstretched). However, the Other category also contains gestures that have nothing to do with quantity (see the examples in Table 6.2). It is probable that if a more granular division of
handshapes were used, more specific correlations could have been detected.

As seen in Figure 6.12, the interpretation of gestures as providing quantity information can also be predicted on the basis of the Hand Configuration parameter. Instead of examining this form parameter in isolation, it is worthwhile exploring its interaction with handshape, so that deeper insights can be obtained into the conditions under which two-handed gestures are interpreted as referring to multiple objects. Figure 6.14 displays the mean scores on Signal-Quantity by handshape and hand configuration. The category Other is excluded from this analysis, because the sample size is too small to allow for reliable statistical analysis.

A two-way ANOVA reveals a statistical interaction between handshape and hand configuration ($F_{(6,431)} = 3.42, p = 0.003, \omega^2 = 0.032$). For gestures in the Index-only and Cupped-hand categories, two-handed gestures receive significantly higher scores than single handed gestures. For the two categories that involve spread hand palms, there is no such difference: two-handed gestures are not more likely to be interpreted as conveying quantity information than one-handed gestures when the hand palms are spread out. A possible explanation for this finding is that
two-handed gestures with spread palms are interpretable as holding or carrying a single object (provided the palms face each other). Because not all handshapes are equally suitable for carrying objects – one would not hold a sizable object between two fists – it is understandable that handshape mediates the interpretation of two-handed gestures as conveying quantitative information.

On the basis of qualitative inspection of the data, a final side note on gestural number signaling can be made. In one of the conversations, a third way of conveying number information was observed: movement reiteration. The speaker in this dialogue made three consecutive circular tracing gestures while saying *drei Schleifen* 'three loops'. This way of number signaling, whereby a quantity is mapped onto the temporal-iterative structure of the gesture, is too rare to be included in the current analysis.

6.2.6 **Formal correlates of uncertainty display**

Figure 6.15 displays the contributions of each of the form parameters to the regression model with the ratings on Signal-Uncertainty as dependent variable (\(R^2 = .17, F_{(24,437)} = 3.83, p < 0.001\)).

![Figure 6.15. Relative importance of the form parameters for predicting the ratings on the Signal-Uncertainty question](image)

As seen in Figure 6.15, the Movement Type parameter accounts for more than a third of the explained variance. As in the case of quantity signaling, this alludes to the existence of a relatively fixed pattern. The
mean scores for each of the different movement type categories are shown in Figure 6.16 (ANOVA: $F_{(5,456)} = 8.6$, $p < 0.001$, $\omega^2 = 0.08$).

Gestures in the Wiggle/sway category are given higher scores than gestures with other movement types. This is in line with earlier analyses of oscillating movements as expressing vagueness or uncertainty display (Bressem & Müller 2014; Calbris 1990). In fact, back and forth rotation of the hand(s) around the wrist is the defining feature of the ‘swaying gesture’, which was found to be recurrent among German speakers in contexts that involve approximation or hedging (Bressem & Müller 2014). The second highest scores on this function are assigned to gestures that involve iterated movement along a single line or arc. This could hint at a more general pattern: some kinematic similarity exists between the Wiggle/sway and Repeated-line-or-arc categories, as both involve some form of repetition in movement.

Interestingly, the response profiles on the Signal-Uncertainty question are strikingly similar to the ones for Word-Search; the ratings
Mapping functions to forms

on these two functions are strongly correlated.\(^{35}\) A possible reason is that the formal features with which these functions are associated are too similar to be discernable using the current methodology. The lay audience that rated the gestures may not always have been able to perceive the difference between the display of uncertainty and the display of difficulty in word retrieval or formulation. It is also possible that the difference in the formal patterns related to each of these functions was too subtle to capture in the current coding scheme. Nonetheless, some formal distinctions are noticeable on the basis of qualitative inspection. Word-search gestures appear to involve larger movements with greater tension in the hands, whereas uncertainty-related gestures involve a lower level of energy and a relatively subtle rotation of the hands around the wrist.

6.2.7 **Formal correlates of prominence signaling**

The final function to be considered is prominence signaling. In Figure 6.17, we see which form parameters predict whether a gesture is interpreted as indicating that the information given is noteworthy. The regression model explains 31.9% of the variance and its predictive ability is statistically significant \((F_{(24,437)} = 8.51, p < 0.001)\).

![Relative importance for Signal-Prominence](image)

Two of the form features stand out as the strongest contributors: Movement Type \((F_{(5,456)} = 20.48, p < 0.001, \omega^2 = 0.17)\) and hand position

---

\(^{35}\) For this reason, no separate analysis is given of the Word-Search function.
(F(4,457) = 19.86, p < 0.001, \( \omega^2 = 0.14 \)). Figure 6.18 shows how the scores on Signal-Prominence vary with the movement type parameter.

Figure 6.18 Mean ratings on Signal-Prominence by movement type

The results bear a close resemblance to those seen for the Refer-to-Object function. This is not very surprising, as we have seen in Chapter 4 that prominence signaling is strongly correlated with object reference. A rudimentary trend that can be derived from the current data is that prominence is associated with the exertion of energy. Complex movements and energetic holds (high energy) are most predictive of high ratings, whereas punctual strokes and wrist-only movements (low energy) receive the lowest scores.

These findings can also be interpreted as shedding new light on the phenomenon of beat gestures. Beat gestures are often defined by their short, punctual stroke, as well as by their function of highlighting some part of the discourse (McNeill 1992). The current findings do not suggest that punctual gestures cannot have this function, but it appears that most other movement types are at least equally strongly associated with emphasis. A question that arises in this light is whether beat
gestures and referential gestures convey the same kind of prominence. The current way of assessing discursive prominence – simply asking lay participants whether they think the gesture marks the provided information as noteworthy – is possibly too coarse to capture the different types of discursive highlighting that gestures are capable of.

6.2.8 **Beyond isolated gestures**

All analyses presented thus far were aimed at form-function relationships in individual gestures. However, advancing the incorporation of gestures in grammar models such as FDG and CG, both of which reach beyond individual clauses, also entails serious consideration of structures that span across multiple gesture strokes. The question of how meaningful structures can be expressed by sequences of gestures performed over time has already received considerable attention in the literature. Most previous work concerns the repeated use of specific regions of space as a way of ‘anaphorically’ referring to previous discourse concepts (Bressem 2012; McNeill 2000a; McNeill et al. 2001). In addition, some researchers have focused on temporal structures that play out across subsequent clauses. Müller et al. (2013), for instance, posit that the performance of successive gestures can be aimed at setting up spatially or temporally coherent ‘scenarios’ that extend across multiple utterances. If this is a systematic phenomenon, one would expect that isolated gestures would be functionally distinct from those that occurred in sequences. The current data set offers a way of exploring this hypothesis by comparing the average functional ratings for isolated and sequenced gestures. Figure 6.19 gives an impression of how the scores on the Refer-to-Place function vary for gestures performed in isolation and those performed as part of a sequence. The plot shows the ratings on the Refer-to-Place question for the gestures performed by one participant as a function of their chronological order. Sequences of gestures are shown as black squares and connected with lines, while isolated gestures are displayed as white squares.
The trend seen in this visual representation is that isolated gestures are generally less often interpreted as conveying spatial information than gestures performed as part of a sequence. In addition, it appears that within the sequences, the gestures often receive similar ratings. This is in line with the hypothesis that sequenced gesture strokes can serve to incrementally build up coherent spatial scenes. To gain more comprehensive insights into the functional differences between isolated and sequenced gestures, Table 6.8 compares the average ratings for each of the inspected functions.

Table 6.8 Mean scores for isolated and sequenced gestures on all function-questions

<table>
<thead>
<tr>
<th>Function</th>
<th>M (SD) isolated</th>
<th>M (SD) sequenced</th>
<th>t</th>
<th>p</th>
<th>Sig.</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal-Prominence</td>
<td>3.40 (0.78)</td>
<td>3.94 (0.61)</td>
<td>-6.88</td>
<td>&lt;.001</td>
<td>***</td>
<td>.80</td>
</tr>
<tr>
<td>Refer-to-Object</td>
<td>3.76 (1.01)</td>
<td>4.20 (0.93)</td>
<td>3.95</td>
<td>.001</td>
<td>**</td>
<td>.44</td>
</tr>
<tr>
<td>Refer-to-Place</td>
<td>3.76 (1.03)</td>
<td>4.34 (0.87)</td>
<td>-5.66</td>
<td>&lt;.001</td>
<td>***</td>
<td>.64</td>
</tr>
<tr>
<td>Depict-Shape</td>
<td>3.12 (1.44)</td>
<td>3.64 (1.41)</td>
<td>-3.54</td>
<td>.0004</td>
<td>***</td>
<td>.37</td>
</tr>
<tr>
<td>Depict-Movement</td>
<td>2.71 (0.88)</td>
<td>2.94 (0.80)</td>
<td>-2.45</td>
<td>.015</td>
<td>*</td>
<td>.27</td>
</tr>
<tr>
<td>Number</td>
<td>2.38 (0.95)</td>
<td>2.49 (0.75)</td>
<td>-1.27</td>
<td>.21</td>
<td></td>
<td>.15</td>
</tr>
<tr>
<td>Signal-Uncertainty</td>
<td>3.55 (0.94)</td>
<td>3.41 (0.75)</td>
<td>1.45</td>
<td>.15</td>
<td></td>
<td>-.17</td>
</tr>
<tr>
<td>Word-Search</td>
<td>3.49 (0.96)</td>
<td>3.34 (0.82)</td>
<td>1.59</td>
<td>.11</td>
<td></td>
<td>-.18</td>
</tr>
</tbody>
</table>

Figure 6.19 Chronological display of gesture stroke sequences performed by one participant and their ratings on the Refer-to-Place question
Sequenced gestures receive significantly higher ratings than isolated gestures for the functions Signal-Prominence, Refer-to-Object, Refer-to-Place, Depict-Shape and Depict-Movement. These findings are in accordance with the observation that gesture sequences can function to set up a scene with multiple objects, where they display their spatial configuration and the energetic transactions that take place between these objects. The fact that a particularly large effect size is obtained for Signal-Prominence also suggests that the creation of these scenarios is typically interpreted as deserving special attention.

This analysis, although rudimentary and limited to the specifics of the current methodology, may have important theoretical implications. It yields quantitative support for the claim that gestural functionality resides not only in the intrinsic features of individual gesture strokes, but also in their temporal-linear arrangement. This is important in the face of the definition of appropriate units of analysis in multimodal extensions of models of grammar. Whereas it is obviously true that gestures do not conform to the same type of combinatorial or syntactic rules as verbally expressed units of talk, the current results suggest that their linear structure deserves serious attention.

6.3 Discussion: from (weak) statistical trends to linguistic primitives

This section has revealed various correlations between functional and formal aspects of gestural expression. On the basis of these findings, it is possible to stipulate a tentative list of 'primitive' function-form associations (Table 6.9).
Table 6.9 Gestural form-function correlations found in this chapter

<table>
<thead>
<tr>
<th>Function</th>
<th>Form (most salient characteristics in current data set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape attribution</td>
<td>Energetic hold or relatively complex movement</td>
</tr>
<tr>
<td>Object reference</td>
<td>Cupped fingers or index-only handshape; higher regions of gesture space</td>
</tr>
<tr>
<td>Location reference</td>
<td>Palm oriented in any direction other than upward</td>
</tr>
<tr>
<td>Movement depiction</td>
<td>Horizontally and sagittally directed movements; flat hands or Index-only handshape</td>
</tr>
<tr>
<td>Quantity signaling</td>
<td>Certain specific handshapes; the use of two hands in combination with a non-flat handshape</td>
</tr>
<tr>
<td>Display of uncertainty</td>
<td>Oscillation or repeated back and forth movement of the hands</td>
</tr>
<tr>
<td>Prominence signaling</td>
<td>Effortful movement, tension in the hands</td>
</tr>
<tr>
<td>Marking coherence of events</td>
<td>Kinesically connected gesture sequences</td>
</tr>
</tbody>
</table>

To some extent, the current results corroborate previous results. For instance, the association of oscillating gestures with the display of uncertainty is not a newly discovered phenomenon (Bressem & Müller 2014; Calbris 1990). Other aspects of the data are complementary to earlier findings. By dint of the function-to-form approach, this study was able to show how certain gestural functions are linked to multiple, distinct patterns of form. These patterns do not necessarily correspond to holistic units (e.g. emblems or recurrent gestures), but also lay bare more subtle form-meaning associations on the level of individual form parameters. Overall, the observations are incompatible with a view of gestural expression as idiosyncratic or fully determined by context.

Obviously, the interpretation of the patterns in Table 6.9 as linguistic primitives should be taken with a pinch of salt. The claim is not that the current study has solved the problem of defining the ‘building blocks’ of gestural expression, nor that Table 6.9 gives the most accurate description possible. Nevertheless, the postulation of a tentative list of form-meaning associations can provide a helpful starting point for
applying models of grammar to actual spoken-gestured data, as will be endeavored in Chapter 8.

Tying these findings back to the discussion on multimodal grammar is not a trivial matter. A difficulty is that the statistical trends observed are rather weak; none of the analyses conclusively points to a one-to-one mapping between function and form. In this section, I discuss three, somewhat complementary explanations for this aspect of the results and their consequences for models of grammar.

1) More clear-cut mappings between function and form exist, but these require a different level of formal and/or functional description to be disclosed.

A first possible explanation for the weakness of the statistical trends is that the operationalization of the study has its shortcomings. Stronger, rule-like patterns could exist, but the level of granularity on which gesture form and function were annotated in the current study might not be suitable for detecting these. The coding scheme and the categories used in the survey may have been too coarse, or based on inappropriate assumptions. Although the method of dissecting gestures into phonological parameters (handshape, orientation, etc.) is well established in the gesture literature, it is ultimately adopted from the domain of sign language studies (where it has also been subject to dispute, cf. Stokoe 1991). Whether this type of coding scheme decomposes gesture forms in a way that is maximally appropriate for detecting the relevant contrastive features is by no means obvious.

It is possible that other, less widely used systems are more fruitful for detecting certain grammatically relevant patterns. In Chapter 3, I briefly touched upon the work of Calbris (1990, 2011), who posits a ‘morphological’ level of form, including semi-holistic patterns such as loops and planes. This approach holds a middle ground between feature-based and holistic views of gesture form. Alternatively, gesture forms could be annotated in terms of their resemblance to Image Schemas (Cienki 2005), Talmyan semantics (Hassemer 2015), or mimetic schemas (Zlatev 2005). Unlike most current approaches, the latter explicitly takes the physical affordances of the body as a starting point.
Boutet’s (2010) physiological annotation system pursues this even further, taking the anatomy of the limbs as central to its description of gestural forms and movements. To date, it remains unclear which of the many available systems for form and function annotation is most appropriate for answering linguistically inspired questions.

In addition, it is important to realize that the employment of manual coding (as opposed to, for instance, motion capture data) always comes with some degree of human error. Even if high inter-coder agreement is achieved, the results remain dependent upon the way the categories and their boundaries are defined. Moreover, the imposition of any type of categorical structure on gestural forms and functions comes with borderline cases, which unavoidably contribute noise to the data.

A related possibility is that the coding scheme is in principle appropriate, but the current study has not looked deep enough into combinations of formal features. Only one of the analyses above has examined multiple parameters in tandem (handedness and handshape for the Signal-Quantity function; Section 6.2.5). It is probable, however, that meaningful patterns in gestural expression are characterized by more complex formal structures than individual formal parameters or pairs of them. In the regression analyses, it was often found that three or more form features contributed simultaneously to the explained variance. Unfortunately, the current study lacked the statistical power required to carry out more detailed analyses of clusters of features. It should therefore be conceded that although some associations between individual form parameters and gestural functions were found to exist, the patterns observed cannot necessarily be thought of as self-contained units of gestural expression. Instead, most of the patterns revealed in this chapter should perhaps be thought of as derivatives of larger, more complex structures.

2) Many gestures attain their meaning through their iconic potential, not through fixed form-meaning pairings. Therefore, no strong mappings between gestural function and form are to be expected.
A second explanatory factor is that gestural forms are strongly context-dependent. As discussed in the introduction of this chapter, gestures differ from words in that their form does not have a clear ‘stable core’. Their forms are adaptive to the physical features of the objects they refer to, as well as to contextual factors such as the distance to the interlocutor. Although the current findings relativized this adaptivity to some extent, the flexible nature of gestures could still offer a partial explanation for the fact that the statistical trends observed are rather modest. As gestural forms are far more variable than verbal ones, it would be naïve to expect dictionary-like relations between form and function. However, this does not preclude the possibility that gestural expression is constrained on a higher level of abstraction. The fact that various function-form connections were found in the current study suggests that gestural idiosyncrasy is to some (moderate) extent compromised by normative conventions. These appear to play out on a relatively schematic level, for instance in the form of a limited repertoire of modes of representation that speakers have available for creating iconic mappings.

3) Form-function mappings in gestural expression are primarily defined by combinatorial constraints, rather than fixed formal characteristics.

The third explanation can be thought of as a cognitive grammatical extension of the second. It holds that when looking to define gestural primitives, one should not remain fixated on the intrinsic physical features of the gestures. As Langacker (1987) and other cognitive-functional grammarians have proposed, not all units of grammar have a stable form. The grammatical class of nouns, for instance, cannot be adequately characterized only in terms of the formal features of its members (internal syntax), but is also defined by certain constraints for linear combination with other structures (external syntax). It can be argued that primitive gestural structures should also be defined along these lines; instances of gestural categories may have only a few formal characteristics in common, but are confined by the linguistic contexts in which they are produced. On this account, gestural categories are better
comparable to grammatical classes than to lexemes. If this holds true, the current research has severe explanatory limits, as it has fully neglected the verbal contexts in which the reported form-function mappings are valid. Chapter 7 of this dissertation is devoted to filling this gap.

6.4 A note on individual differences

Before concluding this chapter, I briefly turn back to a question that has already received attention in Chapter 2: should grammatical primitives be defined on the level of the community or the individual? Since the study presented here collapses across all five speakers in the corpus, it is unclear whether the reported correlations apply equally to each individual speaker. Qualitative inspection of the data suggests the contrary: at least some aspects of gestural expression appear to be strongly speaker-bound. The use of space, in particular, is subject to strong individual variation. In Figure 6.20, an example is given of how two participants use considerably different portions of gesture space, despite speaking and gesturing about the very same action and object (a circular movement around a pond in the SaGA landscape).

The speaker in (a) says *läufst du um den See rum* ‘you walk around the lake’ and traces the described trajectory with a spread hand, while holding the other hand in front of the lower torso. The speaker in (b) gives a very similar route direction: *gehst quasi drei Viertel um den Teich herum* ‘you go let’s say three quarters around the pond’. He also traces a circular line with one hand around the other, but he stretches only one finger on each hand and barely moves his hands away from rest position. Thus, although the core semantics of the two gestures are similar – both speakers trace a circular path around a specific landmark – certain dimensions of the execution of the two gestures are markedly different. A rudimentary analysis with respect to this particular phenomenon suggests that the differential use of space by the two speakers seen in Figure 6.20 is rather consistent. The speaker in (b) keeps his hands in the lower-central region during 87.5% of the gestures he performs, whereas only 34.6% of the gestures of the speaker in (a) occupy this region of gesture space.
A comprehensive, quantitative examination of the individual differences between the participants is beyond the scope of the current thesis (see Bergmann et al. 2010 for further examination of individual differences in the SaGA corpus). Nonetheless, the possible implications of individual variation for the notion of multimodal grammar are worth elaborating on here. As discussed in Chapter 3, it is not always clear whether grammar is to be conceived of as an individual or community-level system. Both models under investigation in this dissertation seem to be open to both views simultaneously (FDG is explicitly modeled after a psycholinguistic model of a speaker, but its typological orientation suggests a community-oriented view; CG assumes both conventionalization and entrenchment as defining criteria for
grammatical items). To account for speaker-specific gestural behaviors, it could however be sensible to make a more explicit distinction between systematicity on the individual level and systematicity on the aggregate level. Applied to gesture, it should be acknowledged that certain gestural behaviors are solely subject to ‘typification’ (i.e., the formation of regularities or categories by an individual), whereas others have developed into fully-fledged communicative conventions (cf. Andrén’s 2010 taxonomy of levels of gestural conventionalization). An alternative view is that individual differences simply reflect styles of gesturing. Drawing an analogy to the notion of idiolect – aspects of individual styles of speaking – some have referred to speaker-bound gestural patterns as ‘idiogests’ (Brannigan 2011; Perrez & Lemmens 2015).

The question remains whether speaker-specific patterns are a concern for grammatical theory. From an FDG perspective, differential use of gesture space can be seen as a matter for the extra-linguistic ‘output’ component, under the assumption that it does not reflect a semantic or pragmatic distinction. Protagonists of CG, however, are likely to call this assumption into question, considering that any difference in gesture execution – including those seen in Figure 6.20 – might correspond to a difference in conceptualization (e.g. the specificity or scope with which the conceptualized scene is construed). Because the issue of individual variation and grammar is far too complex to be resolved here (cf. Dąbrowska 2013; Hurford 2000), the remainder of this dissertation will simply continue to assume a community-level view on grammar, abstracting over the behaviors of all individuals in the corpus. The applicability of the results of this chapter to each and every individual speaker should be interpreted with appropriate caution.

**6.5 Summary and conclusion**

This chapter has investigated the degree to which the functions of co-speech gestures, as perceived by a lay audience, can be predicted on the basis of their formal characteristics. The approach taken to tackle this question was based on an annotation system that includes six form parameters: handedness, hand location, handshape, palm orientation, movement type and movement direction. To assess whether and how these form parameters are associated with specific functions, the ratings
obtained in Chapter 4 were treated as dependent variables in a series of regression models, where the form annotations served as predictors. The models yielded statistically significant results for all functions inspected. The form features that contributed most strongly to the models were inspected in more detail, motivating the postulation of a list of tentative form-function pairings (summarized in Table 6.9).

The data presented in this chapter altogether suggest that gesticulation is at least to some extent guided by context-independent form-meaning relationships. However, various points of discussion remain with respect to the interpretation of these data in the light of models of grammar. The statistical trends observed were all rather weak and the level of formal description might not have been optimal for capturing subtle patterns. Moreover, it cannot be taken for granted that the observed associations apply in the same way to all individual speakers. These caveats taken into account, the current chapter can be seen as a step toward the definition of primitive elements of the gestural component of language. In Chapter 8, I continue to discuss how the patterns examined here can be interpreted in terms of CG and FDG. First, I turn to a closer inspection of the spoken tier of the corpus.
Chapter 7. Gesture-accompanied words and constructions

Chapters 5 and 6 have shown that the perceived linguistic functions of speakers’ gestures are to some extent predictable on the basis of their intrinsic form features. Although this suggests that certain form-function associations are detectable when abstracting over a large number of usage instances, the fact remains that gestures are understood against a backdrop of a rich discursive context. This chapter looks in more detail at the relation between gestural and verbal expression. It compares the linguistic characteristics of spoken-only and spoken-gestured utterances in terms of the relative frequencies of different types of linguistic structures. For those linguistic units that appear most strongly affiliated with gestural expression, it moreover examines patterns in the relative timing of the two channels.

7.1 Gesture-accompanied linguistic structures

This section investigates what linguistic structures can be seen as points of connection between speech and gesture. Specifically, it asks whether gestures have a (dis)preference for occurring together with certain words, constructions or grammatical categories. Previous research has

---

36 Some of the material presented in this chapter also appears in the form of a journal article that has been accepted for publication in the International Journal of Corpus Linguistics (Kok in press)
already pointed out some verbal patterns that have a special role when it comes to multimodal expression. Recall for instance McNeill’s (1992) discussion of ‘speech-linked gestures’, which are performed in syntactic slots marked by phrases such as like in English or so in German. Others have found that certain words or constructions are remarkably often accompanied by specific manual gestures. This holds for constructions like all the way from X [prep] Y (Zima 2014), certain German modal particles (Schoonjans 2014a) and numerous other examples that have been discussed in previous chapters.

Whereas these studies have started out from specific linguistic patterns, others have pursued a ‘bottom-up’ approach to gain insights into the linguistic contexts in which gestures tend to occur. Hadar and Krauss (1999) and Morrel-Samuels and Krauss (1992) examined the distribution of the words that were tagged as ‘lexical affiliates’ of the gestures in their corpus, as well as their grammatical classes. They report a general preference for gestures to be co-expressive with nouns, verbs and prepositions. As argued before, however, the notion of lexical affiliate comes with a predisposition toward semantically loaded words on the part of the coder. That is, with this method, one does not detect words or constructions that correlate with gesture performance for reasons other than co-expressivity (e.g. words that correlate because they can be used to direct attention to a gesture). To obtain more objective and comprehensive insights into the linguistic characteristics of gesture-accompanied speech, an approach is needed that studies the speech-gesture connection without an intermediate level of human interpretation.

One way of pursuing this ambition is by using the acoustic features of the speech channel as a basis for identifying the words to which gestures related (e.g. pitch accent; Alahverdzhieva 2013). This strategy can be motivated by the fact that movements of the hands and the vocal tract are often coordinated (Treffner & Peter 2002), and provides a relatively objective heuristic for identifying the verbal affiliates of gestures. However, an acoustically based approach comes with a bias towards certain word groups (e.g. content words more often receive prosodic stress than articles) and it assumes gestures to be directly aligned in time with the words they relate to. The current
chapter pursues an alternative approach. It aggregates all words that occur in the temporal proximity of the gestures in the corpus, and compares these to the set of words that occur in unimodal contexts. The rationale behind this method can be motivated by the view that spoken-only expression constitutes a different ‘linguistic mode’ than spoken-gestured expression (Cienki 2012). The question asked, accordingly, is whether the verbal structures used in these two linguistic modes are qualitatively and/or quantitatively different. The following sections implement this research strategy on the basis of the full SaGA corpus, with respect to three different units of analysis: lemmas, grammatical categories and bigrams.

7.1.1 Methods
Three different corpora were abstracted from the SaGA data, corresponding to the three types of linguistic items that are used as units of analysis. The first corpus simply contains all lemma annotations from 23 videos in the SaGA corpus, ordered chronologically. This corpus lends itself to addressing how gestures relate to the meanings and functions of individual words. Second, a corpus was extracted that includes all part-of-speech tags corresponding to the lemmas. This level of analysis can provide an important addition because, as we have seen in Chapter 3, connections between speech and gesture may exist on more abstract levels than that of individual words. Finally, as a step toward the idea of a ‘constructional’ basis of language (assuming primitive units beyond individual words), I reorganized the data set as a collection of bigrams.

For each unit of analysis, I divided the corpus into two sub-corpora. The gesture-accompanied sub-corpus contains all items that were uttered in the temporal proximity of the gestures in the corpus, whereas the speech-only sub-corpus contains all remaining items. As a definition of ‘temporal proximity’, the current analyses assume a time window of one second before and one second after the stroke phases of the gestures (as annotated during the SaGA project). That is, a word is considered to be gesture-accompanied if there is any temporal overlap

---

37 Two videos were excluded, because the relevant data were not available for these.
between its articulation and the time frame than runs from one second prior to the onset of the gesture to one second after its offset (also if some part of the articulation falls outside this window). Previous literature has suggested that a time window of this size can be appropriate for capturing meaningful speech-gesture relationships (Leonard & Cummins 2011; Loehr 2004; McNeill 1992). However, because this assumption cannot be taken for granted and has not been validated across different types of linguistic units, the second part of this chapter explores whether varying the operational definition of speech-gesture coincidence influences the results (e.g. assuming a time window of 0, 2, 3 or 4 seconds).

To assess the (dis)preference of linguistic structures for co-occurrence with gestures, I compare the relative frequencies of all items in the speech-only and gesture-accompanied segments of the corpus. The metric used for comparing these frequencies is the Relative Frequency Ratio (henceforth RFR; Damerau 1993). This is the ratio of the normalized frequencies of a linguistic item (a word, POS-tag or bigram) in the gesture-accompanied versus the gesture-unaccompanied parts of the corpus:

$$ RFR(i) = \frac{\text{frequency of } i \text{ in gesture-accompanied sub-corpus}}{\text{number of items in gesture-accompanied sub-corpus}} \div \frac{\text{frequency of } i \text{ in speech-only sub-corpus}}{\text{number of items in speech-only sub-corpus}} $$

High values of the RFR indicate that the item occurs more often in the company of than in the absence of a gesture, taking into account its total frequency in each of the sub-corpora. Because the RFR by itself yields a skewed distribution – it is a ratio of positive numbers – the plots below are mapped onto a natural logarithmic scale. Thus, positive numbers correspond to ‘gesture-attracting’ items, i.e. items with a higher relative frequency in the gesture-accompanied sub-corpus, whereas negative numbers correspond to ‘gesture-repelling’ ones.

In order for the results to be meaningfully interpretable, it is important to take the role of chance into account. To assess which values of the (logged) RFR metric are different from what one might expect when comparing a random pair of sub-corpora, a confidence interval
was estimated using a resampling method. That is, the same analysis was applied two thousand times to pairs of randomly sampled sub-corpora of the same size as the two sub-corpora examined. This yields a distribution of the most likely values of the RFR on the basis of chance, which can be compared to the observed values. From the observed effect size and the confidence interval, a p-value was extracted (following Altman & Bland 2011), which can be interpreted as the likelihood that the observed RFR is a result of random variation. In the following, I examine which lemmas, parts of speech and bigrams have RFR values that exceed the 95% confidence interval. The findings are discussed in the light of multimodal grammatical theory.

7.1.2 Lemma-level analysis
Assuming a one second tolerance, the total number of lemmas is 17,384 in the gesture-accompanied corpus and 13,986 in the speech-only corpus. To investigate the discrepancies between the relative frequencies in each of the sub-corpora, Figure 7.1 plots the RFR values of all the words in the corpus that have at least 80 occurrences. The dashed lines represent the outer borders of the 95% confidence intervals. Note that as a consequence of the procedure adopted (division of normalized frequencies), words with higher overall frequencies in the corpus generally have lower chance baselines. Positive values, corresponding to gesture-attracting words, are shown in Figure 7.1a, while gesture-repelling words are displayed in Figure 7.1b.
We see that 22 lemmas exceed the chance baseline on the gesture-attracting side, whereas 20 words have a RFR that is significantly lower than chance. All words with a RFR that exceeds chance level are listed in Table 7.1. The words are sorted according to their degree of gesture

Figure 7.1 Relative frequency ratios of most common lemmas (on a log scale)
attraction. P-values are reported as a proxy for the statistical reliability of these results.

Table 7.1 Gesture-attracting lemmas

<table>
<thead>
<tr>
<th>Lemma</th>
<th>English translation (most common senses)</th>
<th>N in gesture-accompanied sub-corpus</th>
<th>N in speech-only sub-corpus</th>
<th>Relative Frequency Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total N</td>
<td>17,384</td>
<td>13,986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hier</td>
<td>here</td>
<td>94</td>
<td>5</td>
<td>15.13</td>
<td>4.09E-38</td>
</tr>
<tr>
<td>Seite</td>
<td>side</td>
<td>87</td>
<td>24</td>
<td>2.92</td>
<td>1.07E-7</td>
</tr>
<tr>
<td>quasi</td>
<td>kinda/so to speak (a)round</td>
<td>76</td>
<td>24</td>
<td>2.55</td>
<td>3.45E-6</td>
</tr>
<tr>
<td>rund</td>
<td>right</td>
<td>71</td>
<td>26</td>
<td>2.2</td>
<td>1.37E-4</td>
</tr>
<tr>
<td>recht</td>
<td>this/that</td>
<td>243</td>
<td>94</td>
<td>2.08</td>
<td>3.28E-11</td>
</tr>
<tr>
<td>rechts</td>
<td>(to the) right</td>
<td>288</td>
<td>122</td>
<td>1.9</td>
<td>2.1E-10</td>
</tr>
<tr>
<td>son</td>
<td>such a, a ... like this</td>
<td>202</td>
<td>86</td>
<td>1.89</td>
<td>1.42E-7</td>
</tr>
<tr>
<td>Straße</td>
<td>street</td>
<td>114</td>
<td>49</td>
<td>1.87</td>
<td>4.95E-5</td>
</tr>
<tr>
<td>drauf</td>
<td>on (top of) it/that</td>
<td>68</td>
<td>30</td>
<td>1.82</td>
<td>0.0032</td>
</tr>
<tr>
<td>groß</td>
<td>large</td>
<td>67</td>
<td>30</td>
<td>1.8</td>
<td>0.0043</td>
</tr>
<tr>
<td>links</td>
<td>(to the) left</td>
<td>239</td>
<td>108</td>
<td>1.78</td>
<td>1.34E-7</td>
</tr>
<tr>
<td>von</td>
<td>from, of</td>
<td>103</td>
<td>47</td>
<td>1.76</td>
<td>7.06E-4</td>
</tr>
<tr>
<td>gehen</td>
<td>go</td>
<td>304</td>
<td>141</td>
<td>1.73</td>
<td>3.28E-8</td>
</tr>
<tr>
<td>mit</td>
<td>with</td>
<td>111</td>
<td>52</td>
<td>1.72</td>
<td>7.52E-4</td>
</tr>
<tr>
<td>wieder</td>
<td>again</td>
<td>114</td>
<td>54</td>
<td>1.7</td>
<td>3.91E-4</td>
</tr>
<tr>
<td>nach</td>
<td>after</td>
<td>92</td>
<td>48</td>
<td>1.54</td>
<td>0.011</td>
</tr>
<tr>
<td>stehen</td>
<td>to stand</td>
<td>106</td>
<td>56</td>
<td>1.52</td>
<td>0.0088</td>
</tr>
<tr>
<td>Weg</td>
<td>street, road</td>
<td>86</td>
<td>46</td>
<td>1.5</td>
<td>0.019</td>
</tr>
<tr>
<td>ein</td>
<td>a(n)</td>
<td>594</td>
<td>361</td>
<td>1.32</td>
<td>1.41E-5</td>
</tr>
<tr>
<td>halt</td>
<td>well (discourse particle expressing plausibility or givenness)</td>
<td>163</td>
<td>101</td>
<td>1.3</td>
<td>0.038</td>
</tr>
<tr>
<td>so</td>
<td>like this, in such a way</td>
<td>405</td>
<td>265</td>
<td>1.23</td>
<td>0.0074</td>
</tr>
</tbody>
</table>
We see a variety of different types of words on the gesture-attracting side. The proximal locative adverb *hier* 'here' has by far the highest score. A prevalent view of the combination of *hier* and a pointing gesture is that pointing gestures restrict the borders of the reference domain of *hier*, which is otherwise rather vague (see Fricke 2007 for a comprehensive discussion). In this sense, it is surprising that its distal counterpart *da/dort* 'there' does not show up in this list of gesture-attracting words. Distal locatives are more likely to be infelicitous when performed without some form of hand, head or eye movement (e.g. in utterances like *look over there*). A possible explanation for the remarkably high RFR value for *hier* is that the speakers in the current discourse context (route direction) often refer to entities in fictive locations. Since the participants speak about a landscape that they cannot perceive from the room in which they are located, they often set up local, fictive scenes as to display spatial relations between the objects referred to. The word *hier* in such cases helps to establish a deictic center so that further entities can be set up in the same space (26). Thus, as participants in the current discourse situation often establish common ground via a fictive map, immediacy-related words like *hier* can be expected to be more frequent than in other discourse types.

The high rating of the nouns *Seite* 'side', *Straße* 'street' and *Weg* 'street' are also no doubt related to the specifics of the discourse situation. Route directions often involve reference to particular sides of the road or of the referenced objects ('on the left side you see ...'). In addition, phrases like *der straße folgen* 'follow the street' are particularly common.

---

38 The distal deictic *da* is generally less marked than the English locative *there*, however. The more emphatic version *dort* has too low a frequency in the corpus to be included here.
in this discourse type. The data suggest that such phrases are comparatively often accompanied by manual gestures. A plausible reason is that, by virtue of their iconic potential, gestures can be more parsimonious than words when specifying the spatial relations between objects.

The high RFR for the discourse particle *quasi* ‘kinda/so to speak’ plausibly has different underlying reasons. Since the meaning of *quasi* is interpersonal in nature, typically expressing approximation or indeterminacy, the correlation with gesture performance cannot be caused by a shared referent between the channels. Instead, the data suggest that the use of gestures is generally linked to situations where speakers are not fully able to express themselves verbally, or not fully committed to the accuracy of their formulation. In some of these cases, the lack of verbal specificity might be compensated for by manual expression. In (27), for instance, the speaker expresses a lack of commitment to the accuracy of the word *Torbogen* for describing the object she refers to, and uses her hands to display its physical contours.

(27)  *wenn du rechts von dir halt* *sonen* *Torbogen* *siehst quasi ähm dann*

    | ~~~~~~~~~~ | ***************** | ****** | ~~~~~~~ |
    prep stroke hold prep

    *musst du da rein*
    ****** ******
    stroke hold

    ‘when you see on your right well one of those arches so to speak uhm then you have to get in there’

The discourse particle *halt*, which also shows up as gesture-attracting, is not semantically loaded either. *Halt* often marks the content of an utterance as plausible or indicates that something is pragmatically given or predefined in the communicative context (Schoonjans 2014b; Thurmair 1989). However, *halt* can also be used as a placeholder, i.e. as a way to delay the discourse in order to plan an upcoming utterance. The latter use could be related to the tendency for it to co-occur with gestural expression. When *halt* is used to delay the upcoming speech in case of difficulty in lexical retrieval, gesture might be used to aid this
retrieval process or to compensate for unspecific lexical content (as might be the case in (27)). An additional possibility follows from the observation by Thurmair (1989) that *halt*, and other particles related to obviousness, can be used by speakers as a way of ‘masking’ their uncertainty. If this phenomenon is consistent in the current corpus, the gesture-attracting nature of *halt* is closely related to the strong correlation of gestural expression with *quasi*. In line with this interpretation, these two particles are often combined in a single utterance (27).

We see two adjectives listed in Table 7.1 as gesture-attracting: *rund* ‘round’ and *groß* ‘large’. These have at least two aspects in common: they have a spatial meaning and they are relatively unspecific. When performed in concurrence with these adjectives, gestures may function to further qualify their meaning, for instance indicating how big or how round a given object is (see Chapter 8 for examples). The relation between gesture and space-related words is also evident from the high number of prepositions in the list in Table 7.1: *von*, *mit* and *nach* (‘of/from’, ‘with’ and ‘after’). Although the spatial meaning of these and other prepositions is generally somewhat bleached, the corpus contains many cases of these words where they describe spatial arrangements that are simultaneously depicted using the hands.

The list furthermore contains three determiners: *dies*, *son* and *ein*. As discussed by Hole and Klumpp (2000), *son* is a fully grammaticalized article (derived from *so ein* ‘such a’) that is used to refer to an indefinite token of a contextually definite type. Following this analysis, Fricke (2012) characterizes the combination of *son* with a gesture as a ‘turning point’ between the characterization of a semantic category and the determination of a specific token. When *son* is combined with a pointing gesture directed at an extralinguistic object, the gesture renders the type contextually definite, while the token remains indefinite (that is, pointing gestures combined with *son* do not designate a specific object, but a type or class of entities for which the referenced object is typical). *Son* combined with a depictive gesture (e.g. tracing the outline of an object), also narrows down the conceptual category referred to, but typically achieves a lower degree of type-definiteness. In either case, the gesture-attraction of *son* plausibly
derives from its close relation to the potential of gestures to contribute to semantic type specification. The finding that the indefinite article *ein* is substantially more gesture-attracting than the definite article *der* (in lemma-form, including other genders) further corroborates that the gestures in the corpus more often support indefinite than definite reference. The high RFR for the demonstrative *dies* 'this/that', however, suggests that demonstrative reference is an exception to this trend.

Finally, we see a high RFR for the 'qualitative deictic' adverb *so*. *So* is a good candidate for being accompanied by a depictive gesture (Fricke 2012; Streeck 2002). Streeck (2002: 582) claims that *so* serves as “a ‘flag’ that alerts the interlocutor that there is extralinguistic meaning to be found and taken into account in making sense of what is being said”. Note that although *so* is indeed found to be gesture-attracting, it has the lowest RFR of all words that exceed chance level. The raw frequencies in the current data compromise Streeck’s (2002: 581) intuition that “when Germans depict the world with their hands as they talk […] they almost always utter *so* in the process.” Table 7.2 displays the words for which the RFR is lower than chance level.
Table 7.2 Gesture-repelling lemmas

<table>
<thead>
<tr>
<th>Lemma</th>
<th>English translation (most common senses)</th>
<th>N in gesture-accompanied sub-corpus</th>
<th>N in speech-only sub-corpus</th>
<th>Relative Frequency Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>wissen</td>
<td>know</td>
<td>25</td>
<td>70</td>
<td>0.29</td>
<td>5.42E-9</td>
</tr>
<tr>
<td>ja</td>
<td>yes / modal particle</td>
<td>141</td>
<td>348</td>
<td>0.33</td>
<td>2.3E-33</td>
</tr>
<tr>
<td>glauben</td>
<td>to believe</td>
<td>35</td>
<td>85</td>
<td>0.33</td>
<td>3.23E-9</td>
</tr>
<tr>
<td>Brunnen</td>
<td>fountain</td>
<td>28</td>
<td>57</td>
<td>0.40</td>
<td>3.48E-5</td>
</tr>
<tr>
<td>genau</td>
<td>exactly, I agree</td>
<td>94</td>
<td>168</td>
<td>0.45</td>
<td>2.24E-10</td>
</tr>
<tr>
<td>ah</td>
<td>uhm (filled pause)</td>
<td>329</td>
<td>567</td>
<td>0.47</td>
<td>2.1E-28</td>
</tr>
<tr>
<td>noch</td>
<td>still, yet</td>
<td>35</td>
<td>60</td>
<td>0.47</td>
<td>3.28E-4</td>
</tr>
<tr>
<td>nicht</td>
<td>not</td>
<td>78</td>
<td>133</td>
<td>0.47</td>
<td>8.79E-8</td>
</tr>
<tr>
<td>schon</td>
<td>already / discourse particle</td>
<td>31</td>
<td>49</td>
<td>0.51</td>
<td>0.0025</td>
</tr>
<tr>
<td>nee</td>
<td>no</td>
<td>42</td>
<td>64</td>
<td>0.53</td>
<td>0.0013</td>
</tr>
<tr>
<td>kommen</td>
<td>to come, to arrive</td>
<td>161</td>
<td>235</td>
<td>0.55</td>
<td>1.06E-8</td>
</tr>
<tr>
<td>müssen</td>
<td>must</td>
<td>67</td>
<td>94</td>
<td>0.57</td>
<td>3.62E-4</td>
</tr>
<tr>
<td>Skulptur es</td>
<td>sculpture</td>
<td>41</td>
<td>57</td>
<td>0.58</td>
<td>0.0071</td>
</tr>
<tr>
<td>(no article)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wir</td>
<td>we</td>
<td>40</td>
<td>52</td>
<td>0.62</td>
<td>0.019</td>
</tr>
<tr>
<td>Kapelle man</td>
<td>chapel, one (indef. pers. prn.)</td>
<td>69</td>
<td>90</td>
<td>0.62</td>
<td>0.0033</td>
</tr>
<tr>
<td>zu</td>
<td>to, for, at</td>
<td>81</td>
<td>95</td>
<td>0.68</td>
<td>0.0107</td>
</tr>
<tr>
<td>dann</td>
<td>then</td>
<td>445</td>
<td>522</td>
<td>0.68</td>
<td>1.44E-8</td>
</tr>
<tr>
<td>ich</td>
<td>I</td>
<td>678</td>
<td>771</td>
<td>0.71</td>
<td>4.99E-11</td>
</tr>
</tbody>
</table>

On the gesture-repelling side of the spectrum, there are twenty lemmas for which the RFR exceeds the baseline. The verbs *glauben* 'to believe' and *wissen* 'to know' are among the most gesture-repelling ones. These are both verbs of cognition and do not have clear spatial-perceptual features associated with them. Moreover, because these words take
propositional complements, one can expect some degree of structural distance to the elements of the utterance that are likely to receive gestural co-expression. The low ranking of the deontic modal müssen 'must' can also be accounted for by the first mentioned explanation, as it does not have clear spatial properties either.

More remarkable is the fact that kommen 'to come/to arrive' shows up as gesture-repelling. Like the gesture-attracting verb gehen 'to go', kommen expresses directed movement. The most salient semantic difference between these two verbs is that komen is associated with motion from a distal source to a proximal goal, whereas gehen refers to motion in the reverse direction. The finding that the former is substantially less often gesture-accompanied than the latter could be related to the fact that outward movements of the hands – congruent with the semantics of gehen – are more natural than effortful movements that start from a distal location and move toward the body. In addition, there are many instances of kommen in the corpus where the deictic center is not the speaker, but a location in the town (e.g. 'you come/arrive at a square'). It is in such cases not the entire path of movement, but the final segment of it that is in focus (cf. Langacker 1987: 69). Moreover, because gestures in direction-giving discourse typically take the speaker's body as the deictic center (Fricke 2007), performing a gesture that parallels the path described by this use of kommen would entail a radical viewpoint shift from the perspective of the route-follower to that of a, presumably inanimate, point of arrival. A third possible factor related to the low gesture-attraction of kommen is information structure, as will be discussed in Section 7.1.4.

According to the current data, the discourse particle ja is also unlikely to be gesture-accompanied. This is interesting given the finding by Schoonjans (2014) that ja tends to coincide with a specific head gesture. However, since the current results are based on hand gestures only, these findings cannot be compared directly. Moreover, the analysis above does not carefully distinguish between the uses of ja as a responsive particle (translating into 'yes') or as a modal particle (roughly translating into 'simply'; indicating that no contradiction is expected).
The particle äh ('uh'; a filled pause) also occurs relatively often in speech-only conditions. This appears to be at odds with the idea that gesture plays an important role in word retrieval (Krauss et al. 2000). Again, however, a direct comparison of these findings is not entirely appropriate. Filled pauses can be used for a range of functions other than concept search, such as marking the discourse structure (Swerts 1998), and these are not distinguished in the current analysis.

The adverbs among the gesture-repelling lemmas are quite different from the ones we have seen in the list of gesture-attracting words. None of the adverbs found to be gesture-repelling – dann 'then’, nicht ‘not’, noch ‘still’ and schon ‘already/just’ – have a clear visual-spatial nature (in contrast to the gesture-attracting adverbs rechts and links). With respect to nicht, there is again an ostensible conflict with previous literature, which has pointed out a close link between certain gestures and the verbal expression of negation (Harrison 2008, 2009; Kendon 2004). Although the current data do not challenge the existence of an association between negation and specific gestural patterns, they show that the German negation particle nicht is considerably more often expressed in a unimodal than in a multimodal context.

The only nouns in the list are Brunnen ‘fountain’ and Kapelle ‘chapel’. It is not obvious why these occur so infrequently in the temporal proximity of a gesture. A possible reason is that, contrary to the nouns that showed up as gesture-attracting, the words Brunnen and Kapelle refer to objects that have such specific spatial-perceptual characteristics that they are often followed by a spatial description (see e.g. (28)). These nouns, then, occur in the positions of the utterance which introduces the object as a discourse topic, but not during the subsequent part of the utterance, where its physical properties are described. For instance, when participants speak about the fountain in the SaGA-town, the word Brunnen is often used to introduce the object ('you then arrive at a fountain …'), but not during the subsequent description and gestural depiction of what the fountain looks like ('it looks like this and that'). This is exemplified in (28), where the speaker makes a sequence of gestures that depict the different parts of the fountain and their configuration, but the strokes are performed well after the (first mention of) the word Brunnen.
Based on these observations, one could hypothesize that when words with relatively unspecific lexical meanings (e.g. *Weg* 'road') occur with gestures, this co-occurrence tends to be immediate, rather than successive. Lexical items that refer to perceptually complex objects, by contrast, cannot always be sufficiently characterized in a single information unit, and are therefore sometimes followed by one or more descriptive utterances. As a consequence, the latter set of nouns only correlates with gesture performance if a wider time window is taken into account.

Finally, the list contains three personal pronouns: *es* 'it', *ich* 'I', and *wir* 'we'. These are typically unstressed words that occur in topic position. As gestures tend to occur together with newsworthy information (Levy & McNeill 1992; McNeill 1992) one can expect that pronouns are not the best candidates for gestural co-expression. In addition, since *ich* and *wir* are self-referencing, no depictive specification of the referent is expected. In order to more closely examine the relation between gesture performance and the grammatical properties of the co-occurring speech, the next section takes a closer look at the distribution of the parts of speech in the two sub-corpora.

7.1.3 **POS-level analysis**

To gain deeper insight into the relation between gesture performance and the grammatical categories of the co-expressed words, the analytical procedures described in the previous section were repeated as applied to the Part-Of-Speech (POS) tags in the corpus. That is, instead of looking
at lemma frequencies, the current section focuses on the frequencies of the 22 different POS labels in the speech-only and gesture-accompanied sections of the corpus. The POS-tags were automatically assigned, during the construction of the SaGA corpus, by the Weblicht plugin in ELAN (Hinrichs et al. 2010) and are roughly based on the Stuttgart-Tübingen-tagset (STTS; Schiller et al. 1995). Figure 7.2 shows the RFR values for each of the grammatical categories, with gesture-attracting POS-labels on the left, and gesture-repelling ones on the right.

![Relative frequency ratios for most common parts of speech in corpus](image)

Figure 7.2 Relative frequency ratios for the POS-corpus

From the visualized distribution, we see that five parts of speech exceed the baseline on the positive side, whereas seven parts of speech were found to be significantly gesture-repelling. I first examine the gesture-attracting parts of speech (Table 7.3).

---

39 The categories used in the current analysis correspond to the broader classes in which the more fine-grained tags were divided.
The gesture-attraction values are highest for pronominal adverbs. These are words that designate a spatial relationship with respect to a previously specified entity or location (e.g. *drauf* ‘on (top of) it/that’; *darin* ‘in(side) it/that’). The high RFR suggests that the multimodal expression of such spatial relations is a common phenomenon. In example (29) we see an example of how gestures and pronominal adverbs are co-expressed to describe the shape of a complex object (the same object as is described in (28) by another speaker). The speaker makes a series of iconic gestures that depict the different parts of the fountain and their spatial configuration, which is simultaneously described by the phrases that contain the pronominal adverbs *drin* ‘in it’ and *obendrauf* ‘on top of it’.

(29)  eine Säule dann eine sone Schale dann da drin nochmal eine kleinere Säule und noch eine kleinere Schale obendrauf

‘a column then one of those bowls then in it again a smaller column and yet another smaller bowl on top of it’
Prepositions are also found in the list of gesture-attracting parts of speech, although with a lower relative frequency than pronominal adverbs. As seen above, the prepositions *auf, von, nach* and *mit* have a particular tendency to be accompanied by gestures that depict the spatial configuration described by the speaker. The finding that gestures are likely to occur in the company of nouns and determiners is in line with the idea that the hands can function as an ‘attribute’ of a noun phrase (Fricke 2009; Ladewig 2012). Recall for instance example (30), seen in Chapter 3, repeated here for convenience.

(30)  *die Kirche hat halt ein einspitzes Dach und zwei Türme an diesen Seitenschiffen.*

In this utterance, the gesture depicts the shape of the referenced towers, bearing a functional analogy to the co-articulated noun phrase. In this light, the finding that adjectives show up as ‘gesture-neutral’ (RFR = 1.01, p = 0.95) is rather striking. A possible explanation is that adjectives and gestures fulfill similar roles and therefore cancel out each other’s necessity: when a depictive gesture is performed in concurrence with a noun phrase, an adjective with (roughly) the same meaning is no longer necessary, and vice versa. In the case of (30), for instance, it is imaginable that the shape-depicting gesture is performed because it is better suited for characterizing the contours of the towers than an adjective would be.

For verb phrases, the observed pattern is remarkably different from what we see for noun phrases. Gestures do have a significant tendency to co-occur with adverbs, but they are not correlated with any type of verb (e.g. for lexical verbs: RFR = 1.01, p = 0.74). This finding could allude to a differential contribution of gestures to noun phrases.
and verb phrases. Provided that immediate temporal coincidence is indicative of a functional analogy, it follows that when gestures co-occur with a verb phrase, they will tend to take the role of a modifier, not of the verb itself. For gestures performed in the vicinity of a noun phrase, by contrast, the closest functional analog of the gesture is the head noun. Given the limits of the current data set and discourse genre, claims like these surely remain quite speculative, but the statistical trends observed appear rather robust. Table 7.4 shows the parts of speech found on the lower end of the spectrum, which correspond to some of the linguistic categories that gestures are unlikely to co-occur with.

Table 7.4 Gesture-repelling parts of speech

<table>
<thead>
<tr>
<th>Lemma</th>
<th>N in gesture-accompanied sub-corpus</th>
<th>N in speech-only sub-corpus</th>
<th>Relative frequency ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total N</td>
<td>17,384</td>
<td>13,986</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filled Pause</td>
<td>330</td>
<td>567</td>
<td>.47</td>
<td>2.1E-28</td>
</tr>
<tr>
<td>Interjection</td>
<td>250</td>
<td>381</td>
<td>.53</td>
<td>3.77E-15</td>
</tr>
<tr>
<td>Refl. Pronoun</td>
<td>57</td>
<td>72</td>
<td>.64</td>
<td>0.012</td>
</tr>
<tr>
<td>Pers. Pronoun</td>
<td>841</td>
<td>971</td>
<td>.69</td>
<td>1.6E-14</td>
</tr>
<tr>
<td>Indef.</td>
<td>269</td>
<td>310</td>
<td>.70</td>
<td>2.29E-5</td>
</tr>
<tr>
<td>Particle</td>
<td>802</td>
<td>910</td>
<td>.71</td>
<td>1.23E-12</td>
</tr>
<tr>
<td>Wh-Pronoun</td>
<td>123</td>
<td>129</td>
<td>.77</td>
<td>0.038</td>
</tr>
</tbody>
</table>

The most obvious common denominator in the list of gesture-repelling parts of speech is that all are generally short words: there are four different types of pronouns, filled pauses, interjections and particles. As mentioned in 7.1.2, it can be assumed that pronouns are gesture-repelling because they are likely to occur in positions with given, rather than new information. The low gesture-attraction for the other word types – filled pauses, interjections and discourse particles – are understandable for the same reason. An additional explanation might be that the latter set of words do not have clear iconic or indexical properties. Some of their pragmatic functions could be co-expressed by facial gestures (e.g. nodding, shoulder shrugging), but the current data
suggest that the functions of interjections and particles are not systematically associated with hand movements.

7.1.4 Bigram-level analysis

Advances in construction-oriented theories of grammar have suggested that single words are not always the most appropriate units of analysis for the study of language (e.g. Fillmore et al. 1988). Some multi-word units (e.g. idioms such as all of a sudden or the Xer the Yer) may be equally fundamental units of natural language. Here, I investigate one type of verbal structure that transcends the level of single words, namely bigrams (adjacent pairs of words). Applying the same method as above, two lists of bigrams were extracted from the SaGA corpus. All chronologically successive pairs of words were included, except those between which there was a temporal interval of more than two seconds. This procedure resulted in 17,680 bigrams in the speech-only corpus and 12,152 bigrams in the gesture-accompanied part. Figure 7.3 plots the ratios of the relative frequencies of these bigrams in the two corpus sections.

We see that only two bigrams exceed the chance baseline on the gesture-attracting side, while five bigrams have scores that are significantly lower than what one might expect by chance. Table 7.5 and Table 7.6 list these bigrams and the relevant statistics.

The highest RFR is for auf der 'on (top of) the'. This is somewhat surprising because the definite article der (as a lemma; including all genders) did not show up as gesture-attracting by itself in the lemma-based analysis. It is, however, consonant with the observed high rating for pronominal adverbs, which also combine prepositional meaning with definite reference. Although auf der cannot sensibly be considered a holistic unit or idiom, this result suggests that there are more subtle correlations between segments of speech and gesture use that are not visible when focusing on individual occurrences of a lemma or part-of-speech tag.

40 The construction of the bigram-corpus is based on the original words, not lemmas. This makes it easier to deduce the original patterns in the data (e.g. kommst du). The lemmatized forms of such patterns (e.g. kommen ich), are typically less faithful to the original data.
The second gesture-attracting bigram, *so ein* 'such a/a X like this' does plausibly qualify as a self-contained linguistic construction. It is (generally) equivalent to the word *son*, which we have seen above in the list of gesture-attracting lemmas (Table 7.1). Like *son*, it combines a qualitative deictic component with an indefinite article. This combination of linguistic functions, as argued before, makes it particularly susceptible to gestural co-expression. Next, let us look at the bigrams on the repelling side (Table 7.6).

### Table 7.5 Gesture-attracting bigrams

<table>
<thead>
<tr>
<th>Lemma</th>
<th>English translation</th>
<th>N in gesture-accompanied sub-corpus</th>
<th>N in speech-only sub-corpus</th>
<th>Relative Frequency Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>auf der</td>
<td>On the</td>
<td>70</td>
<td>18</td>
<td>3.03</td>
<td>9.57e-6</td>
</tr>
<tr>
<td>so ein</td>
<td>Such a/a X like this</td>
<td>48</td>
<td>18</td>
<td>2.18</td>
<td>0.021</td>
</tr>
</tbody>
</table>

**Figure 7.3** Relative frequency ratios for the bigram-corpus

![Relative frequency ratios for most common bigrams in corpus](image)

---

**Table 7.5 Gesture-attracting bigrams**

<table>
<thead>
<tr>
<th>Lemma</th>
<th>English translation</th>
<th>N in gesture-accompanied sub-corpus</th>
<th>N in speech-only sub-corpus</th>
<th>Relative Frequency Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>auf der</td>
<td>On the</td>
<td>70</td>
<td>18</td>
<td>3.03</td>
<td>9.57e-6</td>
</tr>
<tr>
<td>so ein</td>
<td>Such a/a X like this</td>
<td>48</td>
<td>18</td>
<td>2.18</td>
<td>0.021</td>
</tr>
</tbody>
</table>
The most striking pattern in the table concerns the three occurrences of the verb *kommen* 'to come/arrive': *dann kommst* 'then come (2\textsuperscript{nd}prs. sing.)', *kommst du* 'you come', and *dann kommt* 'then comes'. Possible explanations for the gesture-repelling character of *kommen* have already been mentioned above: the semantics of the verb *kommen* involves the conception of movement from a distal source to a proximal goal, which corresponds to a relatively unnatural manual movement (moreover, the deictic center of *kommen* in route direction is often displaced). Another possibility can be derived from the fact that the bigrams observed here seem to form part of the phrase *und dann kommst du* 'and then you come'. This phrase, or a shortened version of it, typically occurs at the beginning of a new phase of a route description. In example (28) above, for instance, we have seen that the phrase *und kommst* precedes the introduction of the object *Brunnen* 'fountain', which is characterized in speech and gesture during a later segment of the utterance. Because the word *kommen* in such phrases allocates focus to the end point, rather than to a path of motion, no high gesture rates are to be expected in its direct proximity.

Further, the bigram *und åh* 'and uh' as well as *und dann* 'and then' have a significantly gesture-repelling RFR. Both of these bigrams could be used as placeholders, to delay the discourse while planning the
upcoming speech. In line with the negative RFR values for filled pauses, this activity appears unlikely to be co-performed by the hands.

Despite the fact that the bigram-based analyses appear more noisy than those seen for lemmas and POS-tags, some insights can be gained from the results presented here. It appears that certain specific multi-word expressions have a particular (dis)affinity with manual gestures (e.g. *so ein* on the attractive side, *und äh* on the repelling side). In addition, we have seen possible evidence for even larger recurrent verbal patterns that correlate negatively with gestural occurrence (e.g. *und dann kommst du*). Thus, although bigrams are obviously not the most cognitively realistic units of linguistic organization, including them in the analysis can help to reveal the existence of gesture-affiliated structures that remain unexposed when focusing on individual words only. A possible next step is to look at longer word sequences (e.g. trigrams) or semi-filled constructions (e.g. *so ein [NP]*). Taking into account more complex structures, however, results in relatively low token frequencies and therefore less reliable data. Instead of pursuing this line of research further, the next section addresses a question that is fundamental to assessing the robustness of the results presented here and above: the effect of the choice of the time window that defines speech-gesture 'co-occurrence'.

### 7.2 Temporal dynamics

The findings presented so far are based on a somewhat arbitrarily chosen definition of temporal co-occurrence – linguistic units were considered to be gesture-accompanied if they were performed no more than one second before or after the stroke phase of a gesture. Although this decision was motivated by previous literature (Leonard & Cummins 2011; Loehr 2004; McNeill 1992), it is imaginable that the results vary when using a time window of a different size. Here I explore how modifying the operational definition of co-occurrence influences the results of the above analysis, and how this informs the temporal dynamics of spoken-gestured language use. Before discussing the methodology employed for addressing this question, I briefly consider some previous literature on the relative timing between speech and gesture.
Based on a large body of experimental research, McNeill (1992; Ch.1) formulated three guiding principles for speech-gesture synchrony. The *phonological synchrony* principle states that gesture strokes appear either simultaneously with or right before the nearest stressed syllables in speech. This hypothesis has been copiously tested in experimental and more natural settings (Ekman et al. 1976; Loehr 2004; McClave 1991). Loehr (2004) found that gesture strokes and gesture apexes (the most salient energy pulse within the stroke) line up with peak syllables as well as with pitch accents. The vast majority of the gesture apexes in his data occurred within 300ms of the closest stressed syllable. Leonard and Cummins (2011) examined whether artificial modulations of the speech-gesture interval are noticeable to attenders of multimodal utterances. The subjects in their experiments easily detected altered timing when beat gestures were presented later than in the original video, but they were rather slow to detect such modifications when the videos were altered so that the gestures occurred earlier. The most stable temporal relationship found by the authors was between the pitch accent in speech and the moment where the gesture reaches maximal physical extension (in comparison to combinations of other characteristics of the two channels).

The other two principles formulated by McNeill are *semantic synchrony* and *pragmatic synchrony*. As the names suggest, these hold that if speech and gesture are expressed together, they cover the same semantic and pragmatic import. Various studies have reported that gestures occur simultaneously with, or right before their ‘lexical affiliates’ (i.e. the word or phrase they were judged to relate to in meaning; Chui 2005 for Chinese; Ferré 2010 for French; Habets et al. 2011; Morrel-Samuels & Krauss 1992 for English). Some, however, have reported instances of gestures following speech, thus contradicting McNeill’s phonological synchrony rule (e.g. Chui 2005).

A relevant finding in the light of the current research interest is that phonological synchrony and semantic synchrony are interrelated. Morrel-Samuels and Krauss (1992) found that the onset latency between gestures and their lexical affiliates is inversely correlated with the familiarity of these words; less familiar words occur with more temporal distance to the co-expressed gestures than more familiar
words. Bergmann et al. (2011) also reported on an interaction between timing and semantics. They showed that speech and gesture are produced in closer temporal proximity when they are semantically redundant (i.e., when they express more or less the same information) than when they are complementary in meaning.

These studies have helped to elucidate certain aspects of the temporal dynamics of spoken-gestured language production. Strikingly, however, almost all current work is based on the notion of lexical affiliate. As argued previously, this notion is problematic because it is biased toward clearly delineable, semantic units of speech. As a result, most current studies have been focused on low-level semantic interactions between speech and gesture and have neglected other levels of linguistic organization on which the two channels can intersect. In fact, McNeill’s notion of pragmatic synchrony has been largely ignored in current work on speech-gesture timing. This is a significant pitfall because, as the data from the previous section suggests, the various words that can be considered ‘affiliated’ to gestural expression are not only those with referential meanings, but also include words with relational meanings (e.g. prepositions) and textual or interpersonal meanings (e.g. discourse particles). An additional and related shortcoming is that the annotation of lexical affiliates unavoidably involves some degree of subjective interpretation.

Here, I pursue a more objective approach, focusing on the temporal aspects of the correlations between gestures and the gesture-attracting elements of speech discussed above. The temporal characteristics of these linguistic items are studied by repeating the above procedure with amended criteria for dividing the corpus into speech-only and gesture-accompanied parts. The next section examines the effect of decreasing and increasing the time window that determines which words are considered to co-occur with gestures. Then I look more specifically into the question of when the gesture-related words occur with respect to the gestures in the corpus.

7.2.1 The effect of the choice of time window
Repeating the above procedures with different time windows can shed further light on the interaction between phonological synchrony and the
semantic and pragmatic characteristics of the verbal channel. It also has methodological relevance: many studies so far, including the one presented above, employ somewhat arbitrary definitions of speech-gesture co-occurrence (e.g. direct overlap between the two channels, or a temporal tolerance of 1 or 2 seconds). Here, I compare the gesture-accompanied and speech-only segments of the corpus, under a range of different operational criteria for considering a word ‘gesture-accompanied’. These include a zero-lag condition, where only those words are regarded as gesture-accompanied that overlap directly with a gesture stroke, as well as conditions with a temporal tolerance of up to four seconds (Figure 7.4).

Apart from the modified time window, the analyses carried out here follow the exact same procedures as above. To avoid data abundance, a set of eight lemmas and a set of seven parts of speech were selected to serve as the case in point. The selection was based on the RFR scores (all significantly above chance according to the previous analysis) and on functional diversity. Figure 7.5 shows the RFR scores of the eight selected lemmas as a function of the size of the time window, ranging between zero and four seconds. The dashed lines represent the chance baseline (upper limits of the 95% confidence intervals), computed separately for each corpus division. Note that the plots are scaled to fit the window, so that the contours are most visible. As a consequence, different scales are used on the y-axes for each of the lemmas.

From visual inspection of the plot, it is evident that the choice of time window has a differential impact for the various lemmas. Some lines, in particular those for *dies* and *son*, are relatively flat. Both of these
words are determiners with a deictic component, which can allocate the interlocutor’s attention to some entity or quality depicted gesturally. The correlation of these words with gestural expression, however, appears not to be limited to direct co-occurrence; RFR scores are of the same order of magnitude when considering wider time windows. This suggests that *son* and *dies* potentially (re)allocate the interlocutor’s attention not just directly after these words are uttered, but possibly up until multiple seconds thereafter. The relationship between the word *so* and gesture performance also plays out on a rather wide time scale. However, the preferred time window seems to be more restricted: the RFR is at chance with a zero-lag window, and peaks for windows of around 3 seconds. A similar type of pattern is found for *hier*. Its gesture-attractiveness holds for any time window, but the signal-to-noise ratio appears highest when the temporal tolerance is defined at one second.

![Graph showing relative frequency ratios of gesture-attracting words for different time windows](image)

**Figure 7.5** Relative frequency ratios of gesture-attracting words for different time windows
One of the words in Figure 7.5 has a maximum RFR for a time window of zero seconds: *rechts* 'to the right'. This suggests that when *rechts* is expressed together with a manual gesture, there tends to be a very short lag or no lag at all. The opposite is true for the word *gehen* 'to go'. We see that *gehen* has an RFR that is close to the chance baseline when looking only at immediate temporal overlap of the verbal and gestural channels. For all larger time windows, however, the gesture-attraction value remains well above chance. The inverse relationship between the contours of *rechts* and *gehen* is striking as it may be expected that these words often go together in route directions, for instance in phrases like *du gehst rechts* 'you go to the right'. The current data suggest that when such phrases are accompanied by a gesture, the gesture is more likely to temporally coincide with the adverb than with the verb.

For two of the words inspected, we see a rather stable increase of the RFR as a function of the temporal tolerance: *quasi* and *halt*. The observed correlation of these words with gesture occurrence becomes stronger when larger time windows are taken into account. As discussed above, *halt* and *quasi* are both discourse particles that have a relatively indirect relation to gestural expression. A hypothesis that can be derived from the current data is that gestures that perform meta-discursive functions are subject to more flexible time constraints than those that directly pertain to semantic units (e.g. gestures that modify to nouns).

Next, we see how the choice of time window impacts the results on the level of grammatical categories (Figure 7.6). This analysis is based on the entire data set. It concerns all gesture-attractive parts of speech, plus adjectives and lexical verbs (which are very common in the corpus, but were not found to be gesture-attractive when assuming a one-second window). On the level of grammatical categories, we also see a diversity of temporal patterns. Determiners have a relatively flat contour, with slightly higher scores for smaller time windows. This contour plausibly results from collapsing over definite articles, indefinite articles and demonstratives, which have somewhat diverse
dynamics, as seen above. Nouns and adjectives have homologous patterns, with higher RFR values for direct co-occurrence than for larger time windows. The line for adjectives takes the steepest descent, dropping below the chance baseline for all time windows other than the zero-lag one. This presents an important qualification to the findings in the previous section, where no positive values for adjectives were reported: adjectives are apparently correlated with gesture use only when looking at immediate coincidence. In contrast, the RFR values for the nouns stay above chance for all time windows.

For prepositions and pronominal adverbs, the observed relation is somewhat inverse to what we see for nouns. When looking at direct

---

41 A more fine-grained distinction is not possible here because of the limits of the automatic part of speech tagger, which was not always accurate in distinguishing the uses of *der* as definite article or as demonstrative pronoun.
temporal simultaneity, no significant gesture attraction is observed for prepositions, and the RFR only marginally exceeds the baseline for pronominal adverbs. The gesture-attraction of these grammatical classes shows up more clearly, however, with any larger time window. A possible interpretation of this finding is that the semantic relationship between gestures and prepositional meaning is indirect in nature. Since prepositions and pronominal adverbs often contribute to descriptions of spatial configuration of two or more objects, they relate more strongly to the relative temporal and spatial positioning of successively performed gestures than to individual gesture strokes.

Regarding lexical verbs and adverbs, the current data show that the findings reported in Section 7.1.3 are relatively independent of timing. For almost all time windows, there is a discrepancy between the high RFR rates for adverbs, and the low ones for verbs. Unlike what we have seen for adjectives, the low values for verbs hold for any choice of time window, although the chance baseline is approximated for larger windows. As far as adverbs are concerned, we see that the RFR values peak at one second and drop below chance at three seconds. This seems to suggest that gestures that modify verb phrases are typically performed within three seconds of the articulation of the adverb(s) they relate to. When interpreting these results, however, it should again be borne in mind that subtle patterns in the data could be masked as a result of averaging across all adverbs in the corpus, including words such as links and rechts (with short articulation-gesture lags) and so (with larger lags).

7.2.2 Relative timing between words and strokes
Investigating the influence of the choice of time window as done so far only informs us about one aspect of speech-gesture timing: it shows how the signal-to-noise ratio for detecting functional relations between speech and gesture changes when adopting different operational definitions of temporal co-occurrence. Given that the time window was always expanded to an equal extent on two sides, however, the data above leave us with open questions as to the timing of the examined linguistic units relative to the gestures. Here I address this question in more detail, by shifting the time window in time relative to the
occurrence of the gestures in the corpus, while keeping its duration constant. That is, the division of the corpus is now performed with a range of different intervals between speech and gesture, as depicted in Figure 7.7.

![Figure 7.7 Operationalization of the corpus sub-division for examining patterns in the timing of the lemmas and parts of speech relative to the gestures in the corpus. All intervals between -3 and +3 seconds are taken into account.](image)

Apart from the modification of the time window, the procedures for the analysis are exactly as above. With respect to the interpretation of the data, it is important to keep in mind that the two sub-corpora do not directly represent gesture-accompanied versus gesture-unaccompanied words (at least not in the sense that ‘company’ refers to immediate temporal proximity). The RFR here instead gives an indication of how likely the verbal units are to occur within a certain time frame before or after the gestures. Figure 7.8 presents the data for the same eight lemmas as analyzed above. The dotted vertical line in the middle of the plot corresponds to the zero-lag condition. Values on the left of this line correspond to the windows that were shifted backwards (i.e. windows that include words articulated before the gestures), whereas values on the right correspond to the relative frequencies of the words in the conditions where the time windows were shifted forward (articulated after the gestures).
For most of the lemmas, we see a clear preference for gestures to occur in close proximity of the gesture strokes. *Quasi* is the only word for which the RFR does not reach chance level for larger relative to smaller latencies. This suggests that *quasi* occurs just as often before as after a gesture, and equally likely with small and larger temporal intervals. Of the other lemmas under investigation, *gehen* is most closely centered around the zero-lag condition, i.e. it has a balanced tendency to precede and follow a gesture. For the remaining words, we see subtle asymmetries as to the location of the RFR peaks relative to the zero-lag condition. For *rechts*, the peak is on the right side of the plot, suggesting that gestures generally precede the articulation of this adverb. For *son*, *hier*, *halt* and *dies*, by contrast, we see that most of the weight is on the left side. This indicates that these words more often occur before than after the gestures in the corpus on average. This is a noteworthy finding given the claims by McNeill (1992) and others that gestures typically occur before the words they relate to. It is possible that the words

Figure 7.8 Relative frequency ratios of gesture-attracting lemmas for different lags between speech and gesture

![Graph showing the relative frequency ratios of gesture-attracting lemmas for different lags between speech and gesture](image)
considered here are exceptions to that rule, which has largely been inspired by research on the relation of gestures to noun and verb phrases. The set of gesture-following words observed here seems to have a common denominator that sets them apart from previous research: most of them (in particular son, hier and dies) have a deictic and/or demonstrative component – they can all be used to overtly direct attention to a gesture. The fact that doing so before any gesture is performed can be infelicitous could be a partial explanation of the trend observed here.

The observed dynamics of halt, for which the bias toward gesture precedence is greatest, are not covered by this account. As we have already seen in Figure 7.5 above, halt does not correlate with gesture occurrence when looking at immediate coincidence. Instead, according to the current data, halt tends to precede gestures within a rather specific time window of 0.5 to 1.5 seconds. This pattern could reflect the fact that halt typically occurs in relatively early positions in a sentence. Furthermore, it is consonant with one of the discursive functions of halt mentioned in Section 7.1.2 – its role as a placeholder. The observed gesture precedence can be expected on the basis of the potential role of halt in delaying the discourse for the purpose of planning upcoming utterances. If it is consistently used in situations where speech planning is difficult, and if speakers compensate for this difficulty through gestural co-expression of the planned utterance, one would expect to find a pattern like the one observed in Figure 7.8.

The most conspicuous temporal contour is that of the qualitative deictic so: the RFR values have two peaks – one right before and one right after the stroke. Although it must be conceded that these peaks only barely exceed the chance baseline and might partially reflect random variation, a linguistic explanation of this finding can be given as well. The finding is consonant with the dual use of the so + gesture construction described by Streeck (2002). Streeck notes that so can have two possible relations to gestural expression; it can mark the gesture as being part of a noun phrase or as part of a verb phrase. He illustrates this with the sentence in (31), where G1 and G2 are gestures that depict the action of holding a microphone. Streeck notices that the first so in combination with the first gesture modifies the noun phrase zwei
Mikrophone, whereas the second places the focus on the very action of holding the microphone and is therefore more closely related to the verb phrase.

     ‘And (she) has *these* two microphones, in her hand *like this*, right?’

The SaGA corpus contains various instances of both uses of *so*. The NP-related use sometimes even directly precedes the noun, as in the first part of (32), where the performed gesture traces the outline of the lamps referred to. Example (33), where the phrase *geht so* ‘goes like this’ coincides with a gesture that traces the outline of a sculpture, demonstrates the adverbial use of *so*.

(32)  *vor der Kapelle stehen zwei *so* Kugelleuchten also *so* Leuchten wie man sie

| ~~~ | *****|**************|
| prep. | str. | hold

*aus Parks und so kennt*

[in front of the chapel there are two of *these* spherical lamps, so *these* lamps like those you know from parks for instance’

(33)  *man muss sich das vorstellen die geht *so* aber macht dann nochmal *so*

| ~~~~~~~~~~~~~~~~|**************|
| prep | stroke

*eine Schleife drum herum*

| ******************|**********|
| stroke | hold

‘you have to imagine it goes *like this* but then makes another loop *like this* around it’

The quantitative data presented in Figure 7.8 are in line with the hypothesis that the use of *so* in combination with a gesture can be divided into two sub-constructions: the first peak in the bottom left chart possibly corresponds to the use of *so* as G2 in Streeck’s example
(31), where it modifies a verb phrase and is articulated before the gesture, whereas the second peak corresponds to the use of so as relating to the noun phrase, where it occurs after the gesture. However, as seen from the additional examples from the SaGA corpus, this correspondence is not without exceptions and both uses of so can be timed differently as well. Moreover, the current analysis does not distinguish between cases where so is used in expressions like so ein ‘such a; a X like this’ or und so ‘and the like’.

The final analysis in this chapter investigates the temporal dynamics of the gestures in relation to the POS layer of the corpus (Figure 7.9). The same procedures as above were performed, applied to the selection of grammatical categories that was motivated in the previous section.

![Figure 7.9 Relative frequency ratios of gesture-attracting parts of speech for different lags between speech and gesture](image)

The patterns seen in Figure 7.9 are generally compatible with the results of the lemma-level analyses. Prepositions and determiners have the
most symmetrical RFR contours, with peaks for the zero-lag condition. Adverbs have a general tendency to occur right before the gestures, although we also see a minor peak for +1.5 seconds. This mixed pattern possibly reflects the diversity of grammatical roles that adverbs can have (e.g. modifying verbs, adjectives or other adverbs) and the various ways in which they can connect with gestures (recall the difference between so and rechts in combination with a gesture). Lexical verbs, consistent with what we have seen in Figure 7.6, are not classified as gesture-attracting regardless of the temporal interval assumed.

For adjectives, nouns and pronominal adverbs, there is a slight tendency toward post-gestural occurrence. A striking aspect of the current analysis is that the peak for adjectives is at half a second, whereas the previous analysis suggested a tendency toward immediate coincidence. This difference is possibly related to operational choices: the data in Figure 7.8 were based on a corpus division with only directly co-occurring words, which yields a more skewed division than in the current analysis, where a tolerance of 0.5 seconds is employed for all analyses. The observed discrepancy seems to suggest that the most canonical timing for gesture-related adjectives is between zero and half a second after the gesture. For pronominal adverbs, the tendency to follow gestures is clearest. Whereas RFR values are at chance level for all time windows that were shifted backwards, most of the forward-shifted windows result in values well above chance. This is understandable, given that pronominal adverbs denote a relationship – often spatial character – with regard to an entity that has already been mentioned earlier in the discourse. The current data suggest that when this relationship is also depicted gesturally, the phrase that includes the pronominal adverb tends to be uttered after the first gesture has been performed (see e.g. example (29) above). In general, the temporal contour for pronominal adverbs, as observed in Figure 7.9, is likely to be driven by gesture sequences whereby spatial relations between objects are successively described and depicted.

7.3 Summary and conclusion
The current chapter has shown how a multimodal corpus can be used to gain relatively objective insights into the functional and temporal
relations between verbal and gestured components of expression. Through a bottom-up method, it has investigated the tendencies for gestures to be co-expressed with particular words and parts of speech. A small set of words was found to be positively correlated with gesture performance, including the deictics hier, dies, so and son, the discourse particles halt and quasi and a number of highly spatial lexemes. Other lemmas were less often gesture-accompanied than might be expected by chance, including words without clear spatial features (e.g. verbs of cognition such as glauben ‘to believe’) and words that typically have topical status in an utterance (e.g. pronouns). A comparable analysis applied to parts of speech tags corroborated the finding that certain word classes are more ‘gesture-friendly’ than others. Pronominal adverbs, nouns, determiners, prepositions and adverbs are particularly often accompanied by gestures. This is in line with the view that gestures can take over some of the functions that these words have, such as making reference to entities and ascribing static and dynamic properties to them. However, the grammatical potential of gestures appears more complex: neither adjectives nor verbs were found to be on the gesture-attracting side of the spectrum. When applying the same analysis to a version of the corpus reorganized as a set of bigrams, a number of (negative and positive) correlations were observed with specific verbal patterns beyond the level of single words. In particular, the gesture-attracting construction so ein can be thought of as a (semi-)holistic unit that is closely related with gestural depiction.

A subsequent analysis examined the relative timing of some of the most gesture-attracting words relative to gesture onsets and offsets. The degree of gesture-attraction of the linguistic units inspected was found to vary substantially with the choice of time window that was used to define co-occurrence. On the word level, the most important finding is that certain lemmas are most strongly gesture-attracting when considering a small window (e.g. gehen ‘to go’), whereas other gesture-correlated lemmas seem to be in a much looser temporal connection with gesture performance (e.g. quasi ‘so to speak’). Further analysis focused in more detail on the temporal characteristics of the linguistic units of interest relative to the gestures. In contrast to dominant assumptions in the literature, various lemmas were found to have a
tendency to appear after a gesture (mostly the ones with a deictic component). Of these two sets of findings, the former has clearer methodological implications. It shows that the correlational results one obtains are strongly dependent on one's criteria for considering a word 'gesture-accompanied'. Thus, McNeill's phonological synchrony rule is not equally tenable for all types of words. The relative timing between speech and gesture appears to vary with the linguistic functions gestures serve in the context of the utterance.

Several extensions of this research are possible. One of the most urgent ones would be to validate the methods and the results across different discourse contexts, so that more insight could be gained about the extent to which the outcomes are dependent on the current experimental setting. Another avenue of future research would be to take more complex verbal units into account, such as trigrams and semi-filled word sequences (e.g. 'VP + like' in English). A further refined categorization of the gestural behaviors could also be worthwhile. For instance, individual analyses could be conducted for iconic, indexical and discourse-related gestures. However, the pervasive multifunctionality of gestural expression renders the notion of gestural category somewhat problematic (see Chapter 4; Kok et al. 2016). A perhaps more fruitful avenue of further research would be to concentrate on a set of specific gestural patterns, and examine the linguistic characteristics of the verbal contexts in which these are performed. With a few modifications, the current method can be applied to arrive at a detailed characterization of the 'linguistic profiles' of emblems, recurrent gestures and more subtle patterns in gestural expression. These profiles would not only include the lexical-grammatical characteristics of the contexts in which they occur, but also temporal aspects. Given the numerous potential ways for validating the methods and results, the contents of this chapter are surely just the tip of the iceberg when it comes to seeking convergence between gesture studies and corpus linguistics.

In light of the development of multimodal models of grammar, the current findings are informative about what kind of verbal structures can constitute points of connection between speech and gesture. In general, we can expect that the elements found here to be gesture-attractive will have a prominent role in spoken-gestured
utterances. One of the most relevant conclusions of this chapter, in this respect, is that the group of gesture-attractive words is rather diverse – it contains various forms of locative and spatial terms, demonstratives, discourse markers, and other types of words. To gain a better understanding of how each of these elements relates to gestural expression, the next chapter presents an number of detailed grammatical analyses, applied to a diverse range of example utterances.
Chapter 8. Multimodal grammars at work

8.1 Introduction
In this chapter, FDG and CG are put to work: I apply the models to a range of gestural phenomena, based on video data from the SaGA corpus. The analyses build on Chapters 2 and 3, which discussed the theoretical foundations for incorporating gestural phenomena into each of the two frameworks. They furthermore draw on the empirical insights gained from Chapters 4-7, which addressed some of the challenges raised in the theoretical discussion. Although these challenges may not have been fully resolved, the theoretical and empirical insights reported in this dissertation motivate a number of key assumptions relevant to applying FDG and CG to spoken-gestured data. To recap, these are some of the most important assumptions that can be derived from the previous chapters:

1. Speech and gesture form part of a unified communicative message; they are components of a single utterance (Chapter 3).

---

42 Some of the material presented in this chapter also appears in adjusted format in Kok (2016) and Kok and Cienki (2016).
2. Functional analogies between elements of vocal and gestural expression exist on various levels of linguistic organization (Chapter 3 and 4).
3. Gestures are typically multifunctional: a single gesture may contribute to multiple layers of the discourse simultaneously (Chapter 4).
4. Functionally relevant formal features of gestures are sufficiently contrastive for gestures to be interpretable when access to speech is denied (Chapter 5).
5. Although gestures are undeniably sensitive to contextual factors, some pairings between aspects of form and function are stable across usage instances (Chapter 6).
6. Certain linguistic units (e.g. words, constructions, grammatical categories) have a special role in the composition of spoken-gestured utterances. In German, these include words with a deictic component and those used to allocate attention to extra-linguistic context (Chapter 7).
7. Temporal coordination plays a role in establishing a functional connection between verbal and gestural elements of expression, but the preferred relative timing between the channels is variable across different types of linguistic elements (Chapter 7).

On the basis of these assumptions, the current chapter deploys FDG and CG to analyze a number of multimodal utterances, taken from the SaGA corpus. The example utterances were selected such that they cover not only a wide range of gesture types, but also various ways in which speech and gesture can be connected in the context of an utterance. Thus, although the set of examples is surely not exhaustive, the current chapter offers a test case for the flexibility of FDG and CG in accounting for a broad spectrum of multimodal expressions.

The structure of the chapter is as follows. For each example utterance, I first provide a characterization, from both FDG and CG perspectives, of the most relevant form-meaning pairing(s). This step pertains to the type-level meaning of the gesture; it concerns the functional potential of a gestural category, independent of the way it is
instantiated. Next, I analyze the gesture’s role in interaction with the verbal tier. This step concerns the token-level meaning of the gesture (situated, enriched by context). The analyses below draw, as much as possible, on the analytical tools that are currently available in CG and FDG, introduced in Chapters 2 and 3. Where appropriate, I introduce small amendments to each of the models.

It should be noted that in the remainder of this chapter, I selectively adopt the rationale and vocabulary of the two models under investigation. Certain technical terms might be used in different ways, depending on whether they occur in an FDG-based or CG-based discussion. For instance, from an FDG perspective, I might describe a given linguistic unit as ‘devoid of semantic value’, whereas the same element can be a subject of semantic analysis when adopting CG as the starting point. Up until the discussion section, I remain as neutral as possible as to the adequacy of the vocabulary used by either framework.

8.2 Example analyses

The following sections cover a total of eight example utterances. The first four contain tracing and pointing gestures with reference and depiction as their primary functions. Next, I turn to more holistic patterns, which often involve fixed combinations of form features. These include gestures used for signaling quantity and allocating focus. The final two examples are somewhat further removed from mainstream linguistic theory; they concern the roles of gestures in discourse management and meta-communicative signaling.

8.2.1 Shape tracing for attribution

The SaGA corpus contains numerous instances of gestures that draw the physical contours of an object in the air. A typical example is the gesture displayed in Figure 8.1, which co-occurs with the description of an entrance gate – one of the landmarks in the corpus (34). While saying *blaues Eingangstor* ‘blue entrance gate’ the speaker moves both his hands, with the index fingers stretched away from his body, first outwards and then downwards, thus tracing the rectangular shape of the entrance gate in the air.
und hinten in der Mitte ist noch ein blaues Eingangstor.

Following the proposal that modes of gestural representation can be interpreted as linguistic categories (see Chapter 3), this gesture can be classified as an instance of the category of tracing gestures. It satisfies the defining characteristic of this category, that the speaker acts as if an edge of his hand (the index finger in this case) were a writing utensil (Müller 1998a: 325). As discussed in Chapters 5 and 6, tracing gestures are semantically ambiguous: they can either be used to depict an object, by outlining its shape, or to mimic the trajectory of a movement. The inherent, type-level meaning of the category of tracing gestures therefore can only be described in rather abstract terms. Table 8.1 shows possible representations, adopting the notation conventions of FDG and CG.
Table 8.1 Possible representations in FDG and CG of the type-level meaning of attributive tracing gestures

<table>
<thead>
<tr>
<th>Prototypical form</th>
<th>FDG representation</th>
<th>CG representation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IL:</strong></td>
<td>(C₁: [(R₁) (T₁)])</td>
<td>![Diagram of shape-contour-trajectory space]</td>
</tr>
<tr>
<td><strong>RL:</strong></td>
<td>(v₁: gestured shape/trajectory)</td>
<td>![Diagram of shape-contour-trajectory space]</td>
</tr>
</tbody>
</table>

The representations draw on the assumption that gestures that depict the contours of an object have not only a descriptive, but also a referential function. This is motivated by the finding from Chapter 4 that these functions are strongly correlated (in the eyes of naive observers). In FDG terms, tracing gestures of this type perform a Subact of Reference (R) and a Subact of Ascription (T), when analyzed on the Interpersonal Level. On the Representational Level, the Referential Subact corresponds to a semantic unit that is most likely an Individual (e.g. the object of which the contour or trajectory is being traced), although it may in principle also be a higher level semantic unit (one can imagine a circle-shaped tracing gesture produced while saying *this idea is circular*). The placeholder v is used throughout this chapter as a placeholder for semantically underspecified units.

The movement itself is simply represented orthographically: *gestured shape/trajectory*. Given FDG’s focus on ‘digital’ aspects of language, there is no interest in specifying iconic features of the hands in greater detail. Overall, according to the analysis in Table 8.1, the gesture has the properties of a head-modifier pair on some layer of semantic organization. The proposed representation could be supplemented with a subscript G that marks it as a candidate for gestural formulation: (f: *gestured shape*), but I choose not to adopt this notation here; it would

---

43 This analysis applies to attributively used tracing gestures. Those used for predication only involve a Ascriptive Subact, as argued in the following section.
falsely imply an analogy between gestures and grammatical categories, for which such subscripts are normally used.

The CG representation of the category of tracing gestures is akin to a grammatical class that overlaps with adjectives as well as adverbs: it profiles a relationship between some physical characteristic (a region of shape, contour or trajectory space; analogous to Langacker 2008a: 102) and some THING. In line with this interpretation, the diagram in Table 8.1 contains two elaboration sites. The site on the right corresponds to the shape or contour that is drawn by the hand; the one on the left corresponds to the entity to which this shape or contour is to be attributed (presumably elaborated in the verbal channel). Thus, the diagram indicates that tracing gestures – as a category – convey as much as ‘there is some entity that has some spatial property – presumably a path, motion, or contour’.

Based on these representations of the elementary structure of this gesture type, I now turn to an analysis of the utterance as a whole. Here and below, the representations are limited to the parts of the utterance that are relevant to the situated function of the gesture, and that are not redundant given previous analyses. I first provide an FDG analysis of the gesture in Figure 8.1 in combination with the part of the utterance marked in boldface: *son blaues Eingangstor*. The analyses in (35) and beyond pertain to the Interpersonal and Representational level only, and provide a threefold representation for each: the contributions of the verbal and gestural components of the utterance are first analyzed separately, and then integrated into a combined representation of the utterance segment.

<table>
<thead>
<tr>
<th>(35)</th>
<th>(ILspeech)</th>
<th>(ILgesture)</th>
<th>(ILcombined)</th>
<th>(RLspeech)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(C₁: [(-id+s R₁) (T₁) (T₂)] c)</td>
<td>(C₂: [(R₂) (T₃)] c)</td>
<td>(C₁₂: [(-id+s R₁) (T₁) (T₂) (R₂) (T₃)] c)</td>
<td>(1x1:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(recog f₁: Eingangstor):</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(f₂: blau)</td>
</tr>
</tbody>
</table>

x)
Multimodal grammars at work | 255

\( \text{RL}_{\text{gesture}} \) \( (v:\) \\
\( (f_3: \text{gestured shape}) \)

\( \text{RL}_{\text{combined}} \) \( (1x_{12}: [ \) \\
\( (\text{recog } f_1: \text{Eingangstor}): \) \\
\( (f_2: \text{blau}) \)

\( ]^3 \) \\
\( (f_3: \text{gestured shape}) \)

The spoken part of the utterance segment performs a Referential Subact and two Ascriptive Subacts. The Ascriptive Subacts correspond to the Properties \textit{Eingangstor} ‘entrance gate’ and \textit{blau} ‘blue’ on the Representational Level. The ‘definite-indefinite’ article \textit{son} indicates that the object referred to is presumed to be unidentifiable to the addressee, while it belongs to a class of entities that is familiar (Hole & Klumpp 2000). The unidentifiability of the referent is represented by the -id operator on the Referential Subact (identifiability is analyzed on the Interpersonal Level in FDG because it pertains to the interlocutors’ shared knowledge about the context of communication). The fact that \textit{son} simultaneously marks entity type as definite is represented by the recognizability operator (recog) on the Property \( f_1 \) (\textit{Eingangstor}).

Following Table 8.1, the gestural component of the utterance combines a Referential Subact and an Ascriptive Subact. I have refrained from making any assumptions as to whether the gesturally expressed Subact \( R_2 \) has a specificity operator (by itself, independent of the verbal channel). Although the tracing gesture can be interpreted as referring to a specific instance of a gate (+s), it could also be seen as exemplifying the type (-s) of object that the speaker refers to (Fricke 2012).

Next, I turn to the functional integration of the two tiers. The combined representation of the Interpersonal Level simply lists all Subacts expressed. This may appear redundant, in the sense that the combined representation is fully predictable from the two component tiers, but later examples show that the establishment of the combined Interpersonal Level is not always obvious (see, e.g. 8.2.8). Because the verbally and gesturally expressed Referential Subacts clearly have a
shared referent, the integrated semantic representation contains only a single Individual \(x_{12}\). In contrast, the Property corresponding to the gestural Ascriptive Subact is complementary with the verbal tier. Akin to the phrase *blaues Eingangstor*, the gesture acts as a restrictor on the Individual \(x_{12}\). The Properties expressed verbally and gesturally are represented as equipollent, as there is little reason to assume a hierarchical relation between them (they are not in a linear relation, but expressed simultaneously).

An analysis of the same utterance segment from a CG point of view is given in Figure 8.2. As argued in Chapter 3, CG can be useful not only for analyzing the interaction between gesture and speech, but also for unraveling the internal structure of certain gestures. I first demonstrate the latter utility of CG. Figure 8.2 shows an analysis of the gesture in Figure 8.1 as composed of a conventionalized component and an ad hoc (self-symbolized) component.

The diagram in the bottom left segment of Figure 8.2 represents the assumption that the handshape instantiates a category of tracing gestures (see Table 8.2). The ad hoc component of the gesture signifies, through self-symbolization, a shape or path that is homologous to the tracing movement that is performed.\(^{44}\) The symbolic unit corresponding to the path itself is shown on the bottom right part of the figure, where the label 'self-symbolized' is added to emphasize that this dimension of the gestural form is not a direct manifestation of an entrenched/conventionalized mental structure. As seen in the upper part of the diagram, the unification of the two symbolic structures simply entails recognition that the ad hoc component of the gesture (the traced line) elaborates one of the e-sites invoked by the handshape: it restricts the region of shape-contour-trajectory space that is being

\(^{44}\) The iconicity that governs the self-symbolized tracing movement is not as straightforward as suggested by the diagram in Figure 4. Because iconicity rarely involves full overlap between form and meaning, some degree of arbitrariness remains as to what aspects of the referent are profiled by the gesture. With the rare exception of cases where a speaker’s hand actually represents her own hand at the moment of speaking, iconic reference involves some degree of schematization. The challenge of capturing the systematicity that governs iconic mappings in a CG-based analysis remains outside the scope of this dissertation.
attributed. What results from the integration of the components is a construct that functionally expresses the idea that ‘there is some entity that has a spatial feature resembling the trace of the hand.’ In Figure 8.3, I analyze how the gesture interacts with the verbal component of the utterance *son blaues Eingangstor* ‘a blue entrance gate like this’ (inspired by Langacker 2008a: 345).

Figure 8.2 A CG representation of the internal structure of the gesture in Figure 8.1, as composed of conventional and self-symbolized parts
The representation of *son* in the bottom left part of the figure sketches its contribution to the grounding of the noun phrase *blaues Eingangstor*: the definite-indefinite article *son* marks a transition from a type conception (any entrance gate) toward an instance conception of the referent (a specific entrance gate). The noun phrase *son blaues Eingangstor* restricts the search space of candidate referents to a subset that is not only refined by the verbal type specification, but also by some salient quality in the immediate context. When combined with the
gesture, it can be presumed that the search space is limited to objects with shapes that are iconically related to the kinesic features of the movement of the hand(s).

As shown in the diagram, one of the elaboration sites set up by son is in correspondence with the noun phrase blaues Eingangstor, while the other corresponds to the gestural trace. The emergent meaning can be represented as in the upper part of the figure: the full multimodal utterance profiles a region of conceptual space that has the properties of a THING and holds the middle ground between a type conception and instance conception. It is furthermore characterized by two NON-PROCESSUAL RELATIONSHIPS: one with a region of color space and one with a conceived shape.

8.2.2 Shape tracing for predication

A second instance of a tracing gesture is shown in Figure 8.4. Its inherent meaning is similar to the example above, but it has a different role in utterance (36). Given that the gesture coincides with the phrase ist so helixförmig ‘is helix-shaped like this’, it can be assumed to have a predicative, rather than attributive function.
ja [...] und die sind halt so helixförmig

| prepositional phrase | stroke | recovery |

'yes [...] and they are well helix-shaped like this'

From an FDG perspective, the difference between the referential and predicative use of a tracing gesture is apparent on both the Interpersonal Level and the Representational Level. In (37), analyses are given of the gesture in combination with the phrase *die sind so helixförmig* 'they are helix-shaped like this'.

(36) \( (A) \ (B) \ (C) \)

\( ja [...] und die sind halt so helixförmig \)

\( \text{prepositional phrase} \ | \text{stroke} \ | \text{recovery} \)

The gesture does not perform an act of reference, as was the case in the previous example, but it contributes to the predication of a Property. On the Interpersonal Level, its pragmatic import is therefore represented as a single Ascriptive Subact. This subact corresponds to the Property \( (f_2) \) on the Representational Level, which is comparable to the gesturally
evoked property seen in the previous example. The copular clause in the verbal channel is conditioned by a one-place Property \( f_1: \text{helixförmig} \), which is modified by a Manner variable \( m_1: \text{so} \). In line with Hengeveld & Mackenzie’s (2008: 267) treatment of the English *like this*, the manner specified by *so* is represented as restricted by a proximity operator. The dependent Individual \( x_1 \) has an absent head – pronouns are introduced on the Interpersonal Level in FDG – and an operator for plurality (m). In the combined representation, the gesturally evoked Property is analyzed as a modifier of the Manner variable \( m_1 \). Thus, in line with the results from Chapter 7, the pro-adverb *so* has an important role; it provides an interface between the verbal and the gestural tier. Overall, we see that the gesture in this utterance has a different role when compared to the previous example. It does not act as a specifier that operates in parallel to the adjective, but it is embedded in a complex predicate. A CG representation of the utterance fragment in (36) is given in Figure 8.5.

This analysis of *die sind so helixförmig* ‘they are helix-shaped like this’ combines the simplified representation of a copular clause, introduced in Chapter 2, with a representation of the pro-adverb *so*. It incorporates the fact that *so*, when used in pre-adjectival position, schematically refers to a relationship between some entity and some aspect of the immediate context. The rounded rectangles in the bottom left diagram represent attentional frames, which are postulated in CG to analyze the linear-sequential dimension of discourse organization (see Chapter 2). The semantics of the full verbal expression involve an anticipated shift of attention from the first frame, which includes the relationship described by *die sind […] helixförmig* ‘they are […] helix-shaped’, to a subsequent frame, where some aspect of the situational context is designated.

The extra-linguistic material projected by *so* is elaborated by the gesture. In and of itself, the semantic pole of the gesture is similar to the example above; it profiles a region of shape-contour-trajectory space that is abstracted from the physical contours of the tracing movement. In this case, the trace of the hand specifies the meaning of the co-articulated adjective *helixförmig* ‘helix-shaped’: it narrows down the set of candidate shapes to those that are helix-like and, more specifically, resemble the traced line. Thus, the analysis accords with Streeck’s
description of so as providing a "structured frame [...] within which symbolic body action can be produced and understood".

A considerable difference with respect to the previous example is that the predicates expressed verbally and gesturally relate to the same conceptual domain, namely that of spatial experience (in example (34), both space and color conception were evoked as background frames; the gesture did not interact with the latter). For the unified meaning, a single experiential domain figures as the conceptual base, as shown in the top diagram. An additional difference, crucial to the difference between attribution and predication, pertains to the profile of the utterance as a whole. It is not just the designated THING(s) that are

Figure 8.5 A CG analysis of the interaction between the spoken and gestured component in example (36)
professed in the unified representation, but also the relationship that holds between them.

8.2.3 **Repeated pointing in combination with spatial description**

In the following example, two subsequent pointing gestures are performed (38; Figure 8.6). For both of these, the speaker's physical environment serves as reference frame. The speaker first refers to a spiral staircase and says it looks like the staircase *hier in der Halle* 'here in the hallway'. While saying *hier* 'here', she points to the side, ostensibly in the direction of the actual hallway. Shortly after, she continues her utterance, saying *in der Halle* 'in the hallway' and she points with her other hand in the same direction.

![Figure 8.6 Two consecutive pointing gestures](image-url)
As seen from the transcript, the first of the two consecutive pointing gestures is performed right before the noun phrase *die Wendeltreppe* 'the spiral staircase' is vocalized. In the part of the utterance that follows, another pointing gesture is performed, concurrent with the phrase *hier in der Halle* 'here in the hallway'. I analyze these two fragments of the utterance separately, under the assumption that the meaning of the second gesture is anchored in the first one.

For pointing gestures, the association between the handshape and the referent is based on an indexical, rather than an iconic relationship. The handshape itself bears no physical resemblance to the object referred to, but is meaningful through a relation of spatial contingency: its interpretation relies upon "the understanding that a body part projects a vector toward a particular direction" (Kita 2003: 5). From a grammatical perspective, referentially used pointing gestures are in some ways comparable to iconic gestures. Table 8.2 shows possible analyses from the perspectives of FDG and CG.

From an FDG point of view, the pragmatic import of referential pointing gestures is similar to the gestures analyzed previously in this chapter. In interpersonal terms, gestures of this type combine a Referential Subact with an Ascriptive Subact. As represented by the specificity operator (+s), it can be assumed that the entity referred to is identifiable to the speaker; pointing gestures are typically used to single out a referent from among a set of possible candidates in the immediate environment. The analysis on the Representational Level acknowledges
that the semantic unit referred to as \((v)\) is restricted by a Location variable \((l)\).

Table 8.2 Possible representations in FDG and CG of the type-level meaning of referential pointing gestures

<table>
<thead>
<tr>
<th>Prototypical form</th>
<th>FDG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>((C_1: [(+s R_1) (T_1)]) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>((v_1:\ (l_1: \text{indexed location})) v))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The CG analysis in Table 8.2 represents these characteristics of pointing gestures graphically. It incorporates the fact that the region of conceptual space relevant for interpreting pointing handshape partially overlaps with the actual, physical space in which the interaction takes place. As shown in the diagram, the location of the speaker (S) constitutes the deictic center relative to which the conceived location is construed. The semantic pole of the pointing gesture furthermore includes a profiled instance of a \textit{THING} that is selected from a set of candidates (Langacker 2008a: 283). The semantic category of the profiled referent, however, remains unspecified by the gesture (when analyzed on the type level).

Because the utterance in (38) contains two pointing gestures that are not fully identical in their function, I provide separate analyses for the first and second part of the utterance. In (39), I focus on the semantic integration of the first pointing gesture and the noun phrase \textit{die Wendeltreppe} ‘the spiral staircase’. The gesture, which is performed right before the verbal expression, characterizes the object evoked by the noun phrase by specifying a region of space where it is located. The following FDG analysis represents the contributions of, and interaction between the verbal and gestural components.
The spoken part of the utterance segment (die Wendeltreppe 'the spiral staircase') performs a Referential Subact (R₁) and an Ascriptive Subact (T₁). The operators +s and +id on the Referential Subact represent the definite article, which marks the referent as identifiable to the speaker and addressee. Given that the noun phrase is immediately preceded by a pointing gesture, it can be presumed that the identification of the object referred to – a specific spiral staircase – is at least to some extent facilitated by the gesture. However, FDG offers no obvious means for indicating the contribution of each channel to the common ground (cf. the CG-based example below). The combined representation contains a single Individual (x₁₂), which is co-denoted by the noun phrase and the gesture. The Location variable l₁ evoked by the gesture is analyzed as parallel to the Property f₁; the Property and Location variables together restrict the set of available referents. Overall, the most significant difference with respect to the tracing gesture analyzed in 8.2.1 concerns the Interpersonal Level; in contrast to depictive gestures, pointing gestures are inherently associated with identifiable instances, rather than categories of entities. A CG-based representation of the same utterance fragment is given in Figure 8.7.
CG’s analytical devices allow for a detailed analysis of the interaction of the gesture and the definite article *die* (Figure 8.7). Langacker (2008a: 285) represents the grammatical category of definite articles, in their most basic meaning, roughly as in the bottom right diagram. The definiteness is characterized by the assumption that the referent of the

Figure 8.7 A CG representation of the interaction between the noun phrase *die Wendeltreppe ‘the spiral staircase’* and the first gesture in example (38)
(projected) noun phrase has been attended to in some previous attentional frame. Because of the close temporal succession between the gesture and the noun phrase (the speaker says die Wendeltreppe ‘the spiral staircase’ about half a second after having performed the gesture), one can assume that the attentional frame that is schematically referred to by the definite article aligns with the one in which the gesture was performed.

Thus, the referential content of the gesture elaborates one of the e-sites set up by the definite article; a correspondence is established between the referent of the gesture and that of the noun phrase. The subsequently articulated noun Wendeltreppe ‘spiral staircase’ further specifies the type of entity that the speaker refers to: it elaborates the relatively schematic referential content of the article-gesture combination, as shown in the middle and upper segments of the diagram. The verbal and gestural components of the utterance segment are mutually informative: the noun phrase elaborates the e-site that is created by the gesture (it classifies the referent that is being singled out) and conversely, the gesture contributes to grounding the utterance in the immediate physical context.

During the second part of the utterance, which consists of the locative phrase hier in der Halle ‘here in the hallway’, the speaker performs a second pointing gesture in the same direction. In the analysis below, it is assumed that the meaning of this gesture is anchored in the preceding one; it profiles the same spiral staircase that was brought to focus in the first part of the utterance. The representation of the pointing gesture in Table 8.3 includes the interpretation that it is co-referential with the preceding noun phrase die Wendeltreppe ‘the spiral staircase’. The type of gestural pattern seen here is not specific to pointing, but can be thought of as a case of the more general phenomenon of gestural repetition (Bressem 2012; McNeill 2000a). By repeatedly using a specific gesture form or region of space to refer to the

---

45 Alternatively, it may be contextually salient for another reason. The current analysis is limited to the anaphoric use of the definite article.

46 The semantics of the gestural component, in the preceding discourse frame, are also elaborated by the noun. To maintain clarity in the diagrams, this correspondence is not depicted here.
same entity, a form of ‘anaphoric’ reference is established. A schematic representation of the phenomenon of gestural repetition, applied to pointing, is given in Table 8.3.

Table 8.3 Possible representations in FDG and CG of the type-level meaning of repeated pointing gestures

<table>
<thead>
<tr>
<th>Prototypical form</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FDG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL: (C1: [(+s R1) (T1)] c)</td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
<tr>
<td>RL: (vco-indexed: gestured location)</td>
<td>Previous attentional frame</td>
</tr>
<tr>
<td>v)</td>
<td>Current attentional frame</td>
</tr>
</tbody>
</table>

In FDG, anaphoric reference is handled through co-indexing. In the case of repeated pointing, the indices of the v- and l-variables that correspond to the two gestures strokes are aligned. Other than this, the semantic representation of the second pointing gesture does not differ significantly from that of the first one. In the CG analysis, the anaphoric relation is represented through a correspondence between two elements of attentional frames that are accessed in (not necessarily immediate) succession. As shown in the diagram, repeated pointing entails a conceived correspondence between the THING that is designated currently and the referent of the gesture that was performed and attended to in a recent moment.
In (40), I propose an FDG analysis of the semantics of the second pointing gesture in combination with *hier in der Halle* 'here in the hallway'.

(40)

\[
\begin{align*}
\text{RL}_{\text{speech}} & \quad \text{(prox } l_2: [ \\
& \quad (f_2: \text{in}) \\
& \quad (x_3: (f_3: \text{Halle})) \\
& \quad ] ) \\
\text{RL}_{\text{gesture}} & \quad (x_{12}: \\
& \quad (l_1: \text{gestured location}) \\
& \quad ] ) \\
\text{RL}_{\text{combined}} & \quad (x_{12}: [ \\
& \quad \text{(prox } l_2: [ \\
& \quad (f_2: \text{in}) \\
& \quad (x_3: (f_3: \text{Halle})) \\
& \quad ] ) \\
& \quad (l_1: \text{gestured location}) \\
& \quad ] )
\end{align*}
\]

The phrase *hier in de Halle* 'here in the hallway' corresponds to a Location variable \(l_2\) that is proximate (prox) and has a configurational head \((\text{in der Halle})\).\(^{47}\) Because it cannot be assumed that this Location is identical to the one designated by the gesture \(l_1\), it receives a separate index. The location referred to by the gesture is analyzed as a restrictor that operates in parallel to the locative phrase *hier in der Halle*. This representation is without doubt somewhat rudimentary, as it contains no assumptions about the spatial relation between the locations referred to by each channel (e.g. whether one is contained within the other). Because the gesture is co-indexical with the first gesture as well

\(^{47}\) Note that the subscript \(2\) is used here, because the indexing of variables continues from the representation of the first part of the utterance (39).
as with *die Wendeltreppe* 'the spiral staircase', the combined representation has a single Individual ($x_{12}$) as its head – it is anchored in the multimodal noun phrase analyzed in (39). Altogether, the location of the referenced Individual is specified by lexical, grammatical and gestural elements of the utterance.

Figure 8.8 shows a CG-based analysis of the interaction of the gesture with the phrase *hier in der Halle* 'here in the hallway'.

![Diagram](image-url)

Figure 8.8 A CG analysis of the adverbial phrase *hier in der Halle* in combination with the pointing gesture in example (38)

This diagram shows that the combined meaning of *hier in der Halle* 'here in the hallway' and the pointing gesture derives from the overlap
between the three spatial frames evoked. The locative expression *hier* profiles a region of space that is proximal with respect to the position of the interlocutors. The combination of the gesture with *hier* results in the profiling of the spiral staircase (given its anchoring in the first gesture) and includes a proximal region of space as a salient aspect of the conceptual base (see Fricke 2007 section 2.3.2 for a similar proposal). Thus, according to this analysis, the pointing gesture serves as the profile determinant for this part of the utterance. The search domain for the referenced object is further restricted by the prepositional phrase *in der Halle*, which profiles a landmark (*Halle* 'hallway') and a relation of inclusion with an unspecified trajector. The upper diagram in Figure 8.8 incorporates the interpretation that the landmark is included within the scope of *hier* 'here'. Notwithstanding the fact that alternative interpretations of the semantics of pointing gestures in combination with locative terms are indeed possible (see e.g. Fricke 2007 for a comprehensive discussion), the current analysis demonstrates the potential utility of CG's diagrams in detailing the interaction between gestural and verbal means of spatial reference.

8.2.4 **Relative movement tracing**

Figure 8.9 shows an example of a gesture that enacts the movement of one entity relative to another. The gesture is performed in concomitance with the verb phrase *geht quasi drei Viertel um den Teich herum* '[you] let's say go three quarters around the pond' (41). There is a clear asymmetry in the dominance of the hands. The right hand traces roughly three-quarters of a circle, with the index finger extended and the other fingers curved. The left hand also has an extended index finger, but remains close to rest position and does not undergo effortful movement.
At least two semantic structures play a role in the interpretation of this gesture. First, as seen in Chapter 6, the use of two hands with Index-only handshapes is typically interpreted as referring to two separate entities. Second, effortful movement of a gesture with this handshape is often perceived as representing the movement of some object or person. The pragmatics and semantics of gestures that combine these two characteristics can be represented as in Table 8.4. Note that although the image depicts a gesture with an extended index finger, the pattern of interest is relatively independent of handshape; it is for instance also possible to trace movement with a flat hand.

As seen in Chapter 5, gestural movement can be ambiguous between movement depiction and shape depiction. The current analysis only pertains to the latter function: the use of gestural movement to depict a process. Other handshapes might also be used to represent movement, for instance when moving a clenched fist downward while describing a rock falling down a cliff. However, flat hands and index-only handshapes seem to be the most preferred handshapes for this purpose, as seen in Chapter 6.
Table 8.4 Possible representations in FDG and CG of the type-level meaning of gestures that represent relative movement

Example of form

<table>
<thead>
<tr>
<th>FDG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL:</td>
<td><img src="image1" alt="Diagram" /></td>
</tr>
<tr>
<td>(C1: [(T1) (R1) (R2)])</td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
<tr>
<td>RL:</td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td>(e1: [</td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
<tr>
<td>(f1: gestural movement)</td>
<td><img src="image5" alt="Diagram" /></td>
</tr>
<tr>
<td>(v1: (l1: position of dominant hand), (v2: (l2: position of non-dominant hand)))</td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
<tr>
<td>])</td>
<td></td>
</tr>
</tbody>
</table>

Following the same line of argumentation as in 8.2.1, the gesture is not analyzed as an isolated predicate, but as a predicate-argument combination. In this case, the predicate is relational in nature and has two arguments: one for each hand involved. The FDG representation of the Interpersonal Level accordingly includes one Subact of Ascription and two Subacts of Reference. On the Representational Level, these correspond to a State of Affairs (e1) with a two-place Property (f1). The first participant (v1) is evoked by the dominant hand, while the second participant (v2) corresponds to the non-dominant hand. Because the relative positioning of the hands provides further characterization of
these participants, both are analyzed as restricted by a Location variable. It furthermore seems appropriate to assign the function of Actor (A) to the participant associated with the dominant hand, which performs the action depicted. Likewise, the non-dominant hand functions as a Locative (L); it designates a landmark relative to which the movement takes place (in cases where the effort of the dominant hand directly impinges upon the other hand, labeling the former as Undergoer (U) might be more appropriate).

The CG representation shows that the gesture represents a PROCESSUAL RELATIONSHIP between two THINGS. One of these has the role of trajector (primary participant), the other of landmark (secondary participant). As time unfolds, the relative position of the two participants changes in a way that is analogous to the positioning of the hands in space. In (42) we see an FDG-based representation of the combination of the gesture with the boldfaced part of example (41): [du] geht drei Viertel um den Teich herum ‘[you] go three-quarters around the pond’.50

(42)

IL speech  
(C₁: [(T₁, (R₁: [-S+A]))] (R₂, L))

IL gesture  
(C₂: [(T₂, (R₃, (R₄, L))])]

IL combined  
(C₂: [(T₁, (R₁: [-S+A]))] (R₂) (T₂) (R₃) (R₄))

RL speech  
(e₁: [(f₁: herumgehen) (x₁): (l₁: [(f₂: um: (q₁: drei Viertel))])])

50 The particle quasi is ignored for now, but will be discussed in the next subsection.
The clause denotes a State-of-Affairs which combines a two-place Property (herumgehen), an Individual (du) and a Location (drei Viertel um den Teich). The dominant hand of the speaker can be interpreted as co-indexical with the first participant (x₁), as it embodies the agent that carries out the movement. There is also a correspondence between the non-dominant hand and the second participant, but this relation is more
intricate; two separate points of connection can be identified between modifiers of the Location variable \( I_2 \) and aspects of the gesture. First, the position of the non-dominant hand characterizes the Individual \( x_2 \) as being located in a specific (relative) location. It is therefore analyzed as a Location variable that operates in parallel with the lexical modifier \( f_3 \) (Teich). Second, the movement dimension of the gesture restricts the prepositional predicate *um* ‘around’. Together with the verbal quantifier *drei Viertel*, it specifies the spatial relationship denoted by the prepositional phrase. Whereas *drei Viertel* is represented as a Quantity variable \( q \) – it indicates the extent to which the preposition *um* (‘around’) applies – the gesture is analyzed as a Manner variable \( m \), because it provides additional qualitative information about the spatial relationship described (e.g. the direction relative to the starting position).

In terms of Cognitive Grammar, an analysis of the gesture can be given as evoking two **things** and a **processual relationship** between them. Some of the spatial aspects of this relationship can be represented diagrammatically, exploiting the heuristic illustrations that CG provides. To avoid redundancy with the analyses presented earlier in this chapter, Figure 8.10 displays the semantics of the two tiers in a simplified manner, leaving out the first stages of the compositional path.

The CG representation shows that the conceptual substrates of the verbal and gestural components of the utterance are much alike. The verbal channel evokes a scene where the primary participant (the addressee) traverses in a circular fashion around a landmark, for three-quarters of the way. The path followed by the trajector is represented here in dashed lines because it is relatively unspecific: the verb phrase does for instance not specify the direction of the movement (e.g. as clockwise or counterclockwise). The gestural tier also involves two participants; the dominant hand has the role of trajector, the non-dominant hand serves as the landmark. The tracing movement, moreover, evokes a conception of a path of movement by the former around the latter. The trajector and landmark are both represented as e-sites here because they are inherently schematic: the handshapes provide little or no information as to the type of semantic entity they refer to. As represented by the arrows, both e-sites are elaborated by
elements of the verbal channel. The fact that the verbally evoked trajector corresponds to one of the interlocutors (the hearer, H) is represented by a correspondence link in the bottom left diagram.

As seen on the right, the gestural tier is more objectively construed than the verbal expression; the hands make no direct reference to either interlocutor. In line with the FDG representation above, the gesture is analyzed as adding specificity to the PROCESSUAL RELATIONSHIP that is profiled verbally. It characterizes the path that the speaker describes in a more granular way.\(^{51}\) Thus, this example presents

\(^{51}\) This interpretation could be called into question: it is not entirely obvious that the gesturally traced path adds specificity to the utterance. The hands make a rather subtle movement that is not necessarily more elaborate than what can already be presumed on the basis of the verbal description. Even the direction of the movement as counterclockwise can already be presumed, given
a case where the notion of autonomy-dependence is not one-sided: there is some degree of mutual elaboration between the two channels.

A point worth noting here is that CG’s trajector-landmark distinction does not always map onto the hands in the case of asymmetric, two-handed gestures. The SaGA corpus also contains utterances where the dominant hand (the gestural trajector) aligns with the secondary participant in speech (the verbal landmark). The integrated representation of such utterances is not as straightforward as in the current example, as the roles of primary and secondary participant appear to be opposite for the two channels.

8.2.5 **Quantity specification and approximation**

In example (43), the speaker performs a gesture that combines a number-signaling handshape and a movement pattern that is characteristic of hedging (Figure 8.11). The timing of the gesture is aligned with the articulation of *zehn Bäumen quasi* ‘ten trees so to speak’. The speaker raises both hands with the palms facing towards the addressee and all fingers extended vertically. Thus, her hands co-express the quantity of the trees she is talking about (ten). Meanwhile, she repeatedly rotates her hands sideways at the wrist, ostensibly signaling uncertainty as to whether the information she provides is accurate.

The handshape and the oscillating movement can be analyzed as separate expressive elements. The vertical raising of one or more fingers is of course strongly conventionalized as a way of signaling a quantity, but it is nevertheless not self-sufficient: its interpretation relies on the presence of a verbal or contextual element that specifies the type of entity referred to. Grammatical representations of such quantificational handshapes in FDG and CG are given in Table 8.5.

the driving conventions in Germany. An alternative interpretation, therefore, is that both channels express the general concept of a three-quarter circular movement, not that of a specific path (cf. Fricke's 2014a notion of object-related versus interpretant-related gestures).
‘You well have to [go] between through between those ten trees so to speak’

Table 8.5 Possible representations of the type-level meaning of quantity-signaling gestures

<table>
<thead>
<tr>
<th>Example of form</th>
<th>FDG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td>IL: (C₁: (R₁) ( v ))&lt;br&gt;RL: (gestured number( v ))&lt;br&gt;or&lt;br&gt;IL: (C₁: [(R₁) (T₁)]) ( v ))&lt;br&gt;RL: (( v ))&lt;br&gt;(q₁: gestured number)</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
</tbody>
</table>

From an FDG perspective, the grammatical import of number-signaling gestures is akin to a verbal quantifier: they inherently evoke some entity and ascribe a number to it. A possible difficulty here is that FDG enforces a classification of the gestural meaning as either a modifier or an
operator. This requires the analyst to decide whether gestures of this kind are to be seen as lexical or grammatical items. Given that no clear standards exist for regarding a co-speech gesture as grammaticalized (other than a preliminary proposal by Schoonjans 2014a; see also Wilcox 2007 for a discussion with respect to sign languages), the former option seems theoretically more viable. A lexeme-like representation in the form of \( (v: (q: \text{gestured number})) \) – where \( q \) is a Quantity variable – aligns best with the analyses presented earlier in this chapter, where all gestures were regarded as Lexical Properties. The Interpersonal Level would in this case involve not only a Referential, but also an Ascriptive Subact. However, since the verbal numeral \( \text{zehn} \) ‘ten’ is seen as an operator in FDG, this option makes it difficult to acknowledge the co-expressiveness of the quantificational elements in the two channels. If the gesture were interpreted as lexical, a unified representation could get somewhat cumbersome, for instance in the form of \( (10x: (q: 10)) \).

An alternative option, pursued in Kok (2016), is to analyze the gesture as an operator. This does justice to the fact that quantificational gestures are not ‘descriptive’ in the same way as, for instance, the shape-tracing gestures seen above. That is, they do not help to characterize the type of entity referred to. Notably, the problem that arises here is not specific to gestural expression: the fact that certain words (including numerals, demonstratives and prepositions) combine lexical and grammatical properties presents a general problem for FDG (Keizer 2009). In the discussion section of this chapter, I briefly return to this issue.

The CG representation in Table 8.5 assumes that number-signaling gestures are meaningful against a conceptual base that includes a mental number line. As in the case of verbal numerals, the onstage region of conceptualization includes a certain quantity of tokens that all belong to a the same type \( (t) \) (Langacker 1991: 166). In CG, the analyst is not forced to qualify the gesture as either lexical or grammatical.

A second meaningful pattern in (43) is the oscillating movement of the hands. This movement type is often observed in combination with the expression of indeterminacy or reduced commitment (Bressem & Müller 2014; Calbris 1990). Assuming that this form-function mapping is consistent, some analogy can be assumed to exist between oscillating
gestures and verbal elements that express approximation (e.g. blue-ish, kinda in English). In terms of their form, hedging gestures are relatively flexible: although they have a clear formational core, primarily defined by the repeated swaying or rotating of the hand(s), they occur with a variety of handshapes. In the SaGA corpus, for instance, oscillating movement patterns can be found in combination with unmarked, lax hands, but also in combination with handshapes that express information about the shape or location of the referent. Thus, although Table 8.6 provides a schematic representation of this gesture type combined with loose, open palms, it should be borne in mind that it is the movement type, not the handshape, that is of interest here.

Table 8.6 Possible representations in FDG and CG of the type-level meaning of gestures that signal uncertainty

<table>
<thead>
<tr>
<th>Prototypical form</th>
<th>FDG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(approx V₁)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ˉ)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In FDG, the meaning of this gesture type can be analyzed as analogous to an approximation operator. Given that approximation can be applied to multiple layers of pragmatic organization (Hengeveld & Keizer 2011), the placeholder V is used here instead of a specific interpersonal variable. No representation is given at the Representational Level, since the oscillating movement by itself does not have semantic import.

Providing a CG analysis of this pattern is somewhat less straightforward. Since CG is primarily geared towards semantic correlates of clause structure, no widely used notation conventions exist for phenomena like hedging. However, Langacker (2008a: 431) does
discuss epistemic stance constructions (e.g. *I am sure that...*), which also relate to the expression of the speaker’s commitment. An essential part of the conceptual base for such constructions, according to Langacker’s analysis, is a continuous scale of commitment of a conceptualizer (C) toward the proposition expressed. The representation in Table 8.6 follows the assumption that oscillating gestures are functionally related to expressions of epistemic stance. In the diagram, the semantic pole of the oscillation gesture is analyzed as profiling a relation between the conceptualizer and a region on the uncommitted side of the scale. In essence, this representation is not very different from the FDG representation; the most salient difference is that CG representations emphasize that epistemic commitment is a matter of degree.

Turning to the utterance in (43) (*zehn Bäumen quasi* ‘ten trees so to speak’), it can be noticed that both the numerical and pragmatic facets of the gesture receive expression in speech as well. The quantity of ten is made explicit with a numeral (*zehn*), whereas the reduced commitment is expressed by the hedging particle *quasi*. An FDG analysis of this utterance segment is given in (44).\(^5^2\)

\[
\begin{align*}
\text{IL} &\quad \text{speech} & (\text{approx } C_{1}: & [(R_{1}) \quad [-] \quad (T_{1})])^c \\
\text{IL} &\quad \text{gesture} & (\text{approx } C_{2}: & (R_{2})^c) \\
\text{IL} &\quad \text{combined} & (\text{approx } C_{12}: & [(R_{1}) \quad [-] \quad (T_{1}) \quad (R_{2})])^c \\
\text{RL} &\quad \text{speech} & (10 \times_{1}: & (f_{1}: \text{Baum}) \\
& & & x) \\
\text{RL} &\quad \text{gesture} & (10 \times_{2}) \\
\text{RL} &\quad \text{combined} & (10 \times_{12}: & (f_{1}: \text{Baum}) \\
& & & x)
\end{align*}
\]

\(^5^2\) An analysis of the interaction between the gesture and the demonstrative *diese* ‘this/that’, is not given here, but can be derived from the discussion of definiteness and locatives above.
The Communicated Content $C_1$ on the spoken tier represents the action of informing the addressee ‘that she has to go through those ten trees’. The speaker’s low commitment to the accuracy of this information (or of the way she formulates it) is signaled by the hedging particle quasi ‘so to speak’, as represented by the approximation operator $\text{approx}$ on the Communicated Content (cf. Hengeveld & Keizer 2011). Note that the subacts that are not directly relevant here have been omitted from the representation. The gesture tier involves a combination of the structures analyzed in Table 8.5 and Table 8.6. On the Interpersonal Level, it includes a Communicated Content that carries an approximation operator and contains a single Referential Subact. Although it is possible that oscillating gestures pertain to a single act of reference or ascription, there is good reason to believe that the gestural hedging has scope over the entire Communicated Content in this case. It is unlikely that the gesture only relates to the quantification, as it is clear from the previous discourse that the speaker is aware of the number of trees. Moreover, the gesture is roughly aligned in time with the articulation of quasi, which typically has scope over larger stretches of discourse than a single word or phrase.

In addition, the gesture makes a rather specific contribution to the Representational Level: the vertical raising of all fingers on both hands evokes a semantic variable and a quantity of ten. The combined analysis on the Representational Level in (44) elucidates how this information relates to the semantics of the verbal channel. The gesturally evoked entity is clearly co-indexical with the Individual that is semantically restricted by the Property Baum ‘tree’ and the quantificational operator $\text{zehn} ‘ten’$ (i.e. the same set of trees is being referred to in speech and gesture). Because of this semantic correspondence, the combined Representational Level contains just a single Individual ($x_{12}$) and a single quantificational operator. In sum, the gesture interacts with speech on two different levels of representation. It expresses approximation with respect to the Communicated Content – an operation that is relevant at the Interpersonal Level – and it performs a low-level semantic operation: the quantification of an Individual.

A CG analysis of the same utterance fragment, in simplified form, is given in Figure 8.12.
The diagram on the left represents the combination of the noun phrase *zehn Bäumen* ‘ten trees’ and the quantity-signaling dimension of the gesture. Most relevant here is that the noun phrase characterizes the type of entity (*Baum*; B), of which ten instances are referred to by the gesture. The diagram on the right displays the basic representation of the oscillating dimension of the gesture in combination with the particle *quasi*. The proposition toward which reduced commitment is expressed is displayed as an e-site, as the movement pattern itself provides no information about its content. The unification of the two tiers entails the elaboration of this e-site by the proposition *du musst zwischen durch zwischen diesen zehn Bäumen* ‘you have to go between through between those ten trees’, the semantics of which are partially represented in the diagram on the left. To avoid redundancy, I do not display the unification of these two utterance segments in a separate diagram.

### 8.2.6 Gestural focus allocation

In (45), the speaker first asks *weißt du was es noch geben sollte* ‘do you know what else there should be’, then pauses her speech for about 600 milliseconds, and subsequently answers her own question by saying *ach*
diesen Kirchturm 'oh the church steeple'. During the second part of the utterance, the intensity of her voice increases and the first syllable of Kirchturm has a pitch accent. The answer coincides with a gesture whereby the left hand is moved forward toward the addressee, with the palm facing up and the index finger slightly extended (Figure 8.13).

Weißt du was es noch geben sollte [...] ach diesen Kirchturm

‘Do you know what else there should be [...] oh that church steeple’

The gesture combines features of two well-documented gesture types: the palm-up-open-hand (Ferré 2011; Kendon 2004; Müller 2004) and the ‘abstract deictic’ (Bavelas et al. 1992). The analyses below concentrate on a function that is shared by both of these gesture types and appears most prevalent in the current context: bringing a discourse referent into focus. In grammatical terms, gestures like the one in Figure 8.13 can be thought of as marking an element of the utterance as having a prominent status in the discourse. In Table 8.7, we see how this interpretation can be represented in terms of the notation conventions available in FDG and CG.
Multimodal grammars at work

Table 8.7 Possible representations in FDG and CG of the type-level meaning of gestures that allocate focus

<table>
<thead>
<tr>
<th>Prototypical form</th>
<th>FDG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL: ((V_1)_{\text{Foc}})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL: -</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Being a discourse-oriented framework, FDG offers an elaborate treatment of information structural phenomena. As discussed in Chapter 2, it distinguishes Focus-Background and Topic-Comment relations as separate dimensions of discourse structure. The Focus function (Foc) is assigned to interpersonal variables (Subacts or Communicated Contents) that are marked as newsworthy. The Topic function (Top) is reserved for elements that are marked as the starting point for an utterance. The former distinction seems most relevant to the gesture in the current example. Palm-up-open-hand gestures are found to accompany emphatic stress in speech (Ferré 2011), which suggests they can play a role in discursive foregrounding. Because Focus can relate to multiple layers in FDG, the representation in Table 8.7 includes the variable V as a placeholder for any interpersonal variable. Since the palm-up-open-hand gesture, in and of itself, does not contribute to the semantics of the utterance, the Representational Level is left empty.53

In CG, too, the notion of prominence is seen as a pivotal determinant of grammatical organization. Various semantic notions, including profiling and trajectory-landmark alignment, are defined in terms of the relative prominence of elements of conceptualization.

53 Some gesture researchers might take issue with this analysis, as it can be argued that all gestures have some semantic value. The gesture in (45) could, for instance, be interpreted as enacting the presentation of some objectified discursive content to the addressee. Although the roots of the gesture might indeed be iconic, the gesture under investigation here can be regarded as semantically bleached (Schoonjans 2014a); the iconic aspects of the gesture do not directly contribute to the semantics of the utterance.
Discursive prominence is accounted for in CG with recourse to the conceptualizer’s focus of attention (Langacker 2001, 2013). Langacker follows Chafe’s (1994) hypothesis that language users allocate attention to a single, typically clause-sized unit of discourse at a time: “Dwelling on each clause individually, in a separate attentional gesture, enhances its cognitive salience and that of the elements it contains, if only by according them more processing time and thus a fuller realization” (Langacker 2001: 157). Focus markers, accordingly, can be thought of as refining the conceptualizer’s scope of attention to a specific portion of the profiled region of conceptual space, thus marking it as new or significant. Table 8.7 sketches how this principle may be represented in a CG-based analysis of referential focus. The diagram shows that the semantic pole of the palm-up-open-hand gesture, when used for emphasis, involves narrowing down the conceptualizer’s window of attention to include only a specific segment of the conceptualized material. The process of attention narrowing is represented by the relative heaviness and size of the rounded rectangles that represent the attentional frames. Whereas the attentional window that precedes the performance of the gesture includes a semantic structure composed of multiple elements (e.g. the full content of a clause), the subsequent one is scoped down to just a single THING – the other elements are backgrounded.

The following FDG analysis can be given of the second part utterance in (45): diesen Kirchturm ‘this/that church steeple’.

(46)

ILspeech (C1: [ (R₁)TopFoc(T₁) ] )

ILgesture (V₂)Foc

ILcombined (C1: [ (R₁₂)TopFoc(T₁) ] )

RLspeech (dem x₁:

(f₁: Kirchturm)

x)

RLgesture -
Because the Referential Subact presented in speech ($R_1$) is articulated with prosodic stress, it is analyzed as carrying Focus. It is, moreover, interpreted as having a Topic function, as the isolated presentation of this noun phrase in a single intonation unit marks it as a likely starting point for the subsequent utterance. In line with Table 8.7, the gesture is analyzed as evoking an underspecified interpersonal variable $V_2$, to which it assigns focus. Given the simultaneity of the verbal and gestural components of the utterance, a direct correspondence can be assumed between the focal Subact $R_1$ and the interpersonal unit $V_2$ expressed by the gesture. This correspondence is represented in the combined interpersonal representation, which inherits the Topic function from the verbal channel and the Focus function from both tiers. Given that the gesture does not denote any semantic material, it does not make a contribution to the combined Representational Level. A (simplified) analysis of the utterance fragment in CG is provided in Figure 8.14.

As indicated by the relative heaviness of the rounded rectangles, the gesture narrows down the conceptualizer’s window attention to a single element. The element brought into focus is elaborated by the noun Kirchturm ‘church steeple’ that is expressed verbally. As shown in the unified representation in the upper diagram, it is not only the profiled referent that is relevant to the interpretation of the speech-gesture pair, but also its relative prominence with respect to the preceding discourse frame(s). According to this analysis, the main function of the gesture is to contribute an additional layer of prominence to the semantics of the utterance.

Obviously, the representations given here are strongly simplified. In addition to a background-foreground contrast, the utterance also involves a specific discursive pattern in the form of a (self-directed) question-answer schema. The noun phrase diesen Kirchturm, which constitutes the answer component of the schema, is interpreted against the background of the question that precedes it. A detailed analysis of how such interactional structures may be accounted for in FDG and CG.
remains outside the scope of this chapter. A further note is that focus as expressed through words or intonation is not necessarily of the same nature as focus expressed gesturally. In addition to directing attention, the co-verbal behaviors of the speaker might contribute to fostering interactional engagement and/or common ground among the interlocutors in a way that is specific to gestural expression. For a more detailed account of the various ways in which gestures can add prominence to elements of discourse, analytical notions might be required that go beyond those currently offered by FDG and CG.

8.2.7 Temporal scenario construction
Most of the examples discussed so far pertain to linguistic units that comprise a single word or phrase. Here I consider a phenomenon that is further removed from the traditional scope of grammatical theory: the expression of coherence between events. As discussed in Chapter 6, the
consecutive performance of multiple gesture strokes can mark the events described as temporally or spatially coherent (see also Müller et al. 2013).

Figure 8.15 and example (47) show an example of this phenomenon. The speaker describes three consecutive situations and performs three individual strokes in concurrence with the three spoken clauses. The first gesture (still B) mimics the interlocutor’s orientation at the starting point of the route he is instructed to follow, the second one (stills C and D) traces the orientation of the street that the speaker refers to, and the third (still E) depicts the shape of the roundabout that he will encounter. The three strokes are performed in rapid succession and the hands are not returned to rest position in between.

Figure 8.15 A gesture that depicts a temporally coherent scenario
Semantic structures beyond the clause level have received somewhat limited attention in traditional grammatical theories. FDG is one of the few frameworks that explicitly incorporates such structures into its analytical architecture. CG, although more typically applied to clause-level phenomena, also provides tools for analyzing the incremental construction of larger semantic structures (Langacker 2001). Possible representations of sequences of connected gestures are given in Table 8.8.

Table 8.8 Possible representations in FDG and CG of the type-level meaning of gestures that signal temporal coherence

<table>
<thead>
<tr>
<th>Prototypical form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke 1</td>
</tr>
</tbody>
</table>

Rest position

Rest position
FDG reserves a special semantic variable for sets of semantic units that encompass spatially or temporally connected situations: the Episode (ep). This variable is considered relevant for grammatical organization because in certain languages, groups of clauses that represent connected events receive specialized morphological marking (Hengeveld & Mackenzie 2008: 157). Given the observed functionality of gesture sequencing, discussed above and in Chapter 6, the notion Episode seems an appropriate unit of analysis for gesture sequences such as the one seen in (47). The individual gesture strokes that comprise the sequence, accordingly, map onto States of Affairs (e), which are connected through operators of subsequence (subs). In CG, one can imagine a treatment of gesture sequences in terms of a specific arrangement of time points on a conceived mental time line. As displayed schematically in Table 8.8, gesture sequences that mark temporal coherence can be analyzed in terms of clusters of time points on a mental time line. This is diagrammed by dividing the time point $t_2$ into sub-points $t_{2a}$, $t_{2b}$, and $t_{2c}$.

Turning to the example utterance in (47), a (strongly simplified) FDG analysis is given in (48).

---

54 The current analysis only concerns the temporal connection between the events. The spatial coherence that exists between them can be analyzed along similar lines, with a conceived spatial environment as the most relevant background frame.
Both the verbal and the gestural tier involve three core units, together constituting a larger semantic structure. On the verbal tier, the continuity of the three States of Affairs is marked by the connectors \textit{dann} ‘then’ between the first and second clause and \textit{und dann} ‘and then’ between the second and third. These are represented as operators of subsequence on the States of Affairs. As the coherence between the depicted events is also marked by the kinesic continuity of the movement of the hands, the gestural tier also contains operators of subsequence. Because the three verbally expressed clauses are temporally aligned with the gestures, a one-to-one projection can be assumed between the elements of the two channels. In (48), the representation of the lower levels is represented in a simplified manner. Instead of explicating the semantic relations between the verbal and gestural components (as I have already done in detail for various examples above), I here simply use the representation \textit{verbal content + stroke} to indicate that the Communicated Content comprises both spoken and gestured elements.

A CG analysis of the interplay between speech and gesture is given in Figure 8.16. The CG diagram is in essence similar to the FDG representation; it represents a tight alignment between the three clauses in speech on the one hand and the three gesture strokes on the
other. The temporal grouping of the described events is represented through correspondence links between the elements of expression (the three clauses and the three gesture strokes) and the clustered time points on the conceived time line. The frame around the connected time points indicates that the gesture sequence has the effect of bringing the spatial-temporal coherence between the events described into attentional focus. Thus, the temporal organization of the gesture can be analyzed as a symbolic structure that has semantic value above and beyond the meaning of the individual gesture strokes.

8.2.8 Meta-communicative signaling
The utterance in (49), which we have already seen in Chapter 5, exemplifies an instance of a gesture that has a meta-communicative role. The speaker signals that he is not entirely satisfied with the choice of the word Überführung to describe the landmark he refers to (49). He then continues to look for a better term: *sone ja eine Überführung ist es nicht*
aber da sind so zwei Gebäude äh baulich äh verbunden also so eine son son Übergang da 'Such a well ... it is not an Überführung ... but there are two buildings uh structurally uh connected so one of those those ... overpasses there'. While uttering so eine son son 'one of those those those (or: such a such a such a)', he moves his hands upwards, swaying them back and forth in an energetic manner and with substantial tension in both hands (Figure 8.17).

Figure 8.17 A gesture that signals speech or formulation difficulty

(49) sone ja ... eine Überführung ist es nicht ... aber da sind so zwei Gebäude

\[ \text{(A)} \quad \text{(B)} \quad \text{(C)} \]

\[ \text{äh baulich äh verbunden also so eine son son son ... Übergang da} \]

\[ |~~~|*************|..,.|..,.|] \]

prep. stroke recovery

'such a well ... it is not an Überführung ... but there are two of those buildings uhm structurally uhm connected so one of those those those ... overpasses there'

Signaling that one is having trouble finding the right words or formulation ('own communication management'; Allwood et al. 2006) is even further removed from canonical linguistic theory than the previous example. A possible reason for the general neglect of this activity type in linguistic analysis is that it is not obvious whether it is to be considered as communicative, or merely self-oriented. Although there might not be a conclusive answer to this question, compelling evidence exists that markers of own communication management, such as uh and uhm in
English, are at least to some extent addressee-oriented (Clark 1996; Clark & Fox Tree 2002). According to Clark and colleagues, self-interjections of this kind are often aimed at ‘holding the floor’ in a conversation. Gestures of the type seen in Figure 8.17, likewise, may to some extent be speaker-oriented (Beattie & Coughlan 1999; Krauss et al. 2000), but can nevertheless serve a communicative purpose in spoken interaction. This is corroborated by the finding that substantial systematicity exists in the forms of gestures performed during speech management; they often take the form of circular patterns, finger snapping or rapid back and forth movement (Allwood et al. 2006; Ladewig 2011). If gestures like the one in Figure 8.17 are indeed communicative in character, it is worthwhile considering how this category of gestures can be analyzed in models of grammar. A sketch of the prototypical form of gestures that signal own communication management is given in Figure 8.18.

As this category of gestures typically involves some form of back-and-forth swaying motion, there is some overlap in formal characteristics with the approximation-related gestures discussed in 8.2.5. However, compared to gestures used for hedging, hand movements performed during own communication management generally involve greater tension in the fingers and more energetic movement (provided that the observations from Chapter 6 can be generalized). They may also involve iconic features of the object that the speaker attempts to describe. The

---

55 I do not provide a separate type-level representation here, because there would be no significant difference from the token-level representation.
gesture displayed in Figure 8.17, for instance, starts from the end position of a previous gesture, which represented the location of the two buildings referred to.

From an FDG perspective, it is clear that the holding-the-floor function of the gesture is a concern for the Interpersonal Level – it pertains to the linguistic interaction itself. The integration between the gesture and the verbal component of the utterance takes place at the highest level(s) in the hierarchy of interpersonal layers; the gesture does not intersect with the Communicated Content, but signals that the performance of upcoming linguistic actions (e.g. Discourse Acts, Moves) is delayed. That is, the gesture in (49) is performed in the middle of a Move, but does not have an impact on its internal grammatical organization. For this reason, it can be analyzed as a separate, self-contained Move (50).

\[(50)\]

IL\text{speech} \quad (M_1:\ A_{1:a}::[
\quad \quad \quad (F_1:\ DECL)
\quad \quad \quad (P_1)s
\quad \quad \quad (P_2)_A
\quad \quad \quad (C_1:\ sone\ ja\ ...\ Übergang\ da\ -)
\quad \quad \quad ]^A)
\quad ]^M)\]

IL\text{gesture} \quad (M_2:\ A_{2}::[
\quad \quad \quad (F_2:\ HOLDING\ FLOOR)
\quad \quad \quad (P_1)s
\quad \quad \quad (P_2)_A
\quad \quad \quad ]^A)
\quad ]^M)\]

IL\text{combined} \quad (M_{1:a}:: (F_1:\ DECL) \ldots ^M)
\quad (M_{2}:: (F_2:\ HOLDING\ FLOOR) \ldots ^M)
\quad (M_{1:b}:: (F_1:\ DECL) \ldots ^M)
The utterance in (49) is analyzed as comprising a primary move $M_1$ that is interrupted by a secondary move $M_2$. The primary Move includes the totality of the semantic content expressed by the utterance. The gesture performs an independent contribution to the interaction: it has a separate Illocution, which I represent here as HOLDING FLOOR. The Move performed by the gesture is functionally comparable to an Interactive (cf. Hengeveld & Mackenzie 2008: 76-77), but it is specific to situations where it is surrounded by parts of another Move. In the combined representation, the primary move is divided into two sub-moves: $M_{1<a>$ and $M_{1<b>$. This aspect of the analysis resembles Hengeveld & Mackenzie’s (2008: 51) illustration of a move that runs across multiple turns in a conversation.

As for the CG treatment, Langacker’s (2001: 148) discussion of discourse expectations is of interest here. To account for the meaning of expressions like uh and uhm in English, Langacker postulates ‘speech management’ as one of the channels of conceptualization that constitute the semantic pole of a symbolic unit. However, Langacker does not elaborate in depth on the conceptual underpinnings of particles or gestures that contribute to holding the floor or have a related function. Nonetheless, CG provides useful analytical tools for representing aspects of the conceptual substrate of expressions that signal a delay of the discourse. Figure 8.19 proposes a CG analysis that captures some of the relevant conceptual structures.

The diagram shows that the temporal dimension of the ground is a salient aspect of the conceptual base for the interpretation of holding-the-floor gestures. Whereas time conception already proved relevant in earlier examples, it is in this case pertinent to the organization of the communicative actions conducted by the speaker, not to the referential content of the utterance. The gesture devotes attention to a discrepancy between the timing of the subsequent discourse frame ($t_2$) as anticipated by the hearer ($H$) and as projected by the speaker ($S$). Thus,

---

56 The verbal indicators of difficulty in word retrieval (pausing and repetitive uttering of the word son) can be seen as verbal markers of the holding-the-floor Illocution.
as is indicated by the position of the heavy frame in the diagram, the gesture plays out primarily on the level of interaction management.

8.3 Discussion: merits and drawbacks of FDG and CG

In this chapter, we have seen that the grammar models under investigation lend themselves to the analysis of speech-gesture assemblies across the board of linguistic organization. Both frameworks can be applied to analyze a wide variety of functional relations between the elements of multimodal utterances. In addition, they prove sufficiently flexible to incorporate patterns of gestural expression that are beyond the traditional purview of grammatical theory (e.g. discourse structure, speech management). In the final sections of this chapter, I summarize some of the most important observations and compare the strengths and weaknesses of the models.

Before going into the details of the evaluation, it is important to repeat a few notes on the comparability of the two frameworks. As put forward in Chapter 2, FDG and CG are not fully comparable on all counts. They derive from different research traditions and do not subscribe to the exact same empirical agenda. Consequently, the notation systems
deployed by the two frameworks are different in nature. Whereas FDG employs strict, explicit formalisms, CG’s notations are merely heuristic and, in some ways, closer to a sketchbook than to a formalism. When comparing the models, it should also be taken into account that both are continuously under development. If a particular phenomenon has not received attention in the FDG or CG literature so far, this need not indicate a fundamental flaw of the model; neither framework purports to claim that the set of notions it recognizes covers all that is potentially relevant to linguistic analysis.

8.3.1 **Strengths of FDG over CG**

FDG has a strong focus on the hierarchical relations that hold between elements of linguistic utterances. It therefore provides a rich vocabulary for specifying on what level of linguistic organization elements of speech and gesture can be connected. A clear benefit of FDG’s architecture, as became clear in Section 8.2.7, is that it reaches beyond individual clauses. FDG’s built-in notions for analyzing functional and structural relations between clauses make it possible to analyze gestural functions that play out on higher levels of discourse. We have seen that the notion of Episode, for instance, is useful when accounting for gesture sequences that mark spatial-temporal coherence between events. Although CG certainly also recognizes the importance of discourse, it has a smaller set of built-in tools for this purpose. Due to its primary focus on clause-level phenomena, and perhaps because its development has primarily been based on analyses of English, structures such as FDG’s Episode have not been explicitly recognized in CG as fundamental units of grammar.

A second benefit of FDG is that its high degree of explicitness provides a potential to detect patterns in the data. One pattern observed in the current chapter is that the semantic-pragmatic material provided by gestures often pertains to a layer of linguistic structure that also receives verbal expression. Another pattern is that gestures most often take the role of modifier of Individuals, Locations and Manners, compared to other variables. Both of these observations are based on small data samples, however, and call for investigation on a larger scale and across languages. The formulation of testable hypotheses can shed
new light on prevalent questions in gesture research (e.g., whether gestures are redundant or complementary with respect to speech, see Chapter 9). What kind of hypotheses could be derived from the CG analyses is less obvious. Empirical predictions based on CG are certainly imaginable (Langacker 2008a: 9), but not as easy to operationalize without profoundly simplifying the theoretical assumptions that underlie the model. That is, since CG generally denies the existence of rigid, all-or-nothing categories, it is often not straightforward how patterns observed through CG-based analyses can be translated into a (quantitative) research design.

Another, specific advantage of FDG is that it yields more parsimonious representations for certain phenomena. Whereas CG’s diagrammatic representations undoubtedly have great diagnostic utility for gesture studies (see 8.3.2 below), the value of CG’s illustrations over FDG’s symbolic formalisms is not evident in all cases. Especially with regard to certain pragmatic phenomena, CG’s diagrams sometimes have a somewhat high ink-to-information ratio. When we for instance compare the FDG-based and CG-based representations of the expression of reduced commitment (Table 8.6), the question rises whether the use of a diagrammatic representation is more informative than FDG’s orthographic one (unless the purpose is to remind the reader that commitment is essentially a continuous phenomenon). Another example of a difference in representational efficiency concerns anaphoric relations. Whereas FDG represents these simply by means of co-indexical subscripts, a CG-based treatment of anaphors involves a rather elaborate supplement to the diagram (see e.g. Table 8.3). As mentioned before, however, the notation systems do not have the same purpose and a direct comparison between them is not entirely appropriate.

8.3.2 **Strengths of CG over FDG**
The analyses presented in this chapter make clear that CG’s notation conventions have many interesting features in terms of analyzing the connection between speech and gesture. Compared to FDG, CG’s diagrams enable a great deal more detail as to the role of iconicity and indexicality in gestural expression. As seen above, CG-based diagrams can sometimes directly incorporate the relevant physical characteristics
of a gesture, and can help to show how these map onto the conceptual schemas evoked verbally. In Figure 8.9, for instance, the pictorial representations proved useful to show how the movement of the speaker’s hands was isomorphic to certain aspects of semantic pole of the co-expressed verb *herumgehen* ‘to go around’. In addition, CG’s diagrams can elucidate the cognitive principles that allow speech-gesture combinations to be meaningful. As seen in Figure 8.5, CG can help to clarify to what extent speech and gesture draw on shared conceptual domains. With respect to example (41), for instance, the CG analysis shows how the domain of spatial experience underlies the semantic connection between the verbal and gestural components of the utterance (Figure 8.10). Such insights do not follow from the FDG analyses, where the description of semantic predicates is limited to an orthographic representation. Thus, the fact that FDG dismisses conceptual semantics from the grammar proper results in a relatively shallow account of the semantic overlap between verbal and gestural forms of expression.

In addition, CG’s notion of construal proved particularly fruitful for the analysis of speech-gesture ensembles. As discussed in Chapters 3, and empirically supported by the data in Chapter 5, most gestures are inherently underspecified and to a large extent dependent on speech. CG’s notions of schematicity, specificity, dependence and elaboration are of great utility when analyzing how the schematic meanings of gestures interact with more specific elements of speech. Given the proposed parallel between gestural structures and CG’s basic characterization of grammatical categories (*THINGS*, *PROCESSES*, etc.), CG moreover enables a semantic description of the inherent meaning of gestures and their relation to elements of the verbal grammar. FDG is less appropriate for representing the various degrees of specificity that gestural meanings can have. The solution chosen above is the use of the \( v \) and \( V \) placeholders whenever a gesture could not be classified in terms of FDG’s pre-defined semantic units. As argued in more detail in 8.4, however, the classes represented by these placeholders might be too broad to adequately capture the semantic potential of specific gesture types.
A related, more general benefit of CG is that it draws on categories and distinctions that are defined with respect to domain-general cognitive principles. The fact that speech and gesture rely on an overlapping set of cognitive operations can help to motivate novel analyses of phenomena that have not received previous discussion in the literature. In example (25), for instance, I combined existing CG notions (e.g. attentional framing, grounding, profile-base) to arrive at a representation of gestures with a discourse-delay function. The resulting analysis provided in Figure 8.19 is motivated by the fact that cognitive operations such as attentional framing and time conception are equally applicable to both modalities. In FDG, the introduction of new categories can be more difficult to motivate. The categories and distinctions in FDG are in principle only valid when they are systematically encoded in the grammar of a given language. Strictly speaking, the extensions of FDG proposed above (e.g. HOLD FLOOR as a special type of illocution) can only be adequately validated through systematic inquiry within and across languages, where speech is analyzed in combination with gesture.

8.4 Possible modifications of FDG

Fully resolving the drawbacks of FDG mentioned above would entail a radical transformation of the model. To achieve more insight into the conceptual overlap between verbal and gestural meaning, for instance, the separation of conceptual and semantic components would need to be reconsidered, or even eliminated (cf. Butler & Taverniers 2008; Jackendoff 2002). Doing so, however, would go against FDG’s principled agnostic stance on the organization of the human conceptual system. Moreover, it is imaginable that the postulation of a separate semantic component has its benefits: it enforces explicitness in designating the formal and functional categories that underlie gestural forms. Instead of fundamentally redesigning the model, I propose three smaller amendments that could help to improve FDG’s potential to include gestural behaviors as a component of grammar.

My first suggestion is related to the proposals made by Anstey (2008) and Keizer (2009), to redefine and/or loosen the distinction between modifiers and operators. We have seen that the inclusion of gestures in FDG blurs the modifier-operator distinction, which has a
somewhat fuzzy border anyway (Keizer 2009). The categorization of the number-signaling gesture in (44), for instance, entailed a decision as to whether it was to be treated as lexical or grammatical. This can be problematic: most gestures are not evidently grammaticalized, yet some are functionally analogous to an operator under FDG’s criteria (for example because they are non-descriptive). A proposal for a less rigid distinction is made by Keizer (2009), who postulates a ‘lexical operator’ for elements of speech that have lexical as well as grammatical features (e.g. demonstratives, numerals). This proposal would, however, not solve a more fundamental issue pointed out by Anstey (2008): modifiers and operators are defined primarily by morphosyntactic criteria, but they are already introduced in the formulation component. In other words, they are conceived as elements of the semantic-pragmatic structure of an utterance, but they do not represent classes of entities that share semantic or pragmatic features. Because gestures are bound to a different set of ‘morphosyntactic’ criteria than verbal expressions are, the current definitions of modifier and operator cannot be seamlessly extended to co-verbal expression. In line with Anstey’s (2008) proposal, a possible solution is to abolish the current view on the modifier/operator duality and replace it by a semantically motivated alternative. Anstey (2002: 5) advocates the introduction of ‘specifiers’ in FDG, which are defined “not in terms of the layer they modify or the grammatical categories they express, but in terms of the semantic domain to which their specification relates.” According to this proposal, there should be as many types of specifiers as there are ontological classes. Given that speech and gesture play out in partially overlapping ontological domains (space, manners, things, events etc.), the proposed reconsideration might yield a greater potential for capturing the semantic interfaces that exist between the two channels.

A second, more specific suggestion is to include a wider range of possible placeholders. The current set of variables in FDG, which

---

57 One of the reasons FDG takes this approach is that it brings out the fact that languages differ in what requirements they impose upon speakers when it comes to the formulation of certain grammatical distinctions (e.g. tense, evidentiality). My claim here is not that this is necessarily a bad decision, but that applying the same criteria to gesture further problematizes the issue.
includes Lyons’ (1977) classification of semantic units, is based on the properties of written (and spoken) elements of language. Gestures do not necessarily respect the same distinctions, however; they may convey meaning on a more abstract level. To cope with gestural underspecificity, the maximally general placeholders $v$ and $V$ were used in this chapter. A possible drawback of this approach is that it treats all variables as equally susceptible to gestural expression. This might not always be justified. Pointing gestures, for instance, typically refer to Individuals or Locations, but are far less likely to refer to Propositions. The introduction of more specific placeholders for a subset of semantic variables could help specify the expressive potential of certain gesture types. For pointing gestures, for instance, a more appropriate placeholder might be represented as $v_{\text{loc}}$ where the subscripts define the types of pragmatic or semantic variables within their expressive potential.

A third suggestion is to introduce a special notation for constructions that serve as an interface between the two channels of expression. We have seen that certain verbal elements come with implicit ‘slots’ for modifiers that are likely to be filled by some form of extra-linguistic expression. The German word *so*, as discussed in 8.2.1, evokes the expectation that the Manner variable denoted will be modified by some contextual element, possibly a gesture. The representation of *so* as ‘prox m’ in example (36), however, does not include this property. Given that the tight relation between *so* and gestural expression has systematic impact on the (multimodal) output of the grammar, one could argue for a notation that foreshadows the potential gestural modification of the Manner variable. In the case of *so*, a notation like $m^G$ could be introduced to signal that certain variables will be passed on to the gestural formulation component.\footnote{A possible complication here is that gestural modification is, although common, not obligatory.}

### 8.5 Possible modifications of CG

In Chapter 3, I have discussed a number of theoretical points of connection between speech and gesture research. However, as seen in the current chapter, not all of these points can be equally well
represented in CG-based analyses. In order to further develop CG’s potential as a tool for analyzing multimodal data, further convergence is required between theoretical and representational aspects of the model.

One step in this direction was already made in Figure 8.2, where I showed how CG’s notion of self-symbolization can be employed to deal with iconic features of creative gesticulation. As argued in that section, accommodating gesture in CG requires a way of marking which symbolic units rely on conventionalization, and which rest on ad hoc mappings.

Another discrepancy between theory and notation in CG concerns the notions of gradience. CG acknowledges, at least in theory, that the entrenchment and conventionalization of symbolic structures as units of grammar is a matter of degree. However, CG’s notations do not represent this continuum. To better capture the different degrees to which different gesture types have undergone conventionalization (e.g. Kendon’s continuum; Kendon 1980), an augmented visual representation would be appropriate. The same applies to the representation of the various degrees of dependence that may exist between verbal and gestural forms of expression. In contrast to its underlying theoretical convictions, current CG diagrams make it seem like autonomy-dependence is a binary (all-or-nothing) distinction. A diagrammatic notation of the gradient nature of conventionalization and autonomy-dependence could, for instance, follow the lines of Langacker’s (2013) treatment of the variable degrees of activation of discourse elements.

8.6 Summary and outlook
The comparative data analysis in this chapter complements the theoretical discussions in Chapters 2 and 3. The most salient differences between FDG and CG seen in this chapter are directly related to the divergence in theoretical underpinnings and foci of the models. FDG’s architecture is primarily designed to account for the various types of structural hierarchy that play a role in linguistic organization; it is capable of accounting for the scope relations between verbal and gestural elements of utterances in a very explicit manner. CG, by contrast, was designed with special attention to the unification of grammatical theory with lexical semantics and cognitive science; it
Chapter 8 offers an elaborate framework for studying the relation between gesture and grammar from a cognitive linguistic point of view. Overall, as pointed out in this chapter, and as already alluded to in the theoretical discussion in Chapter 3, FDG and CG have quite complementary strengths in their potential to handle gestural phenomena.

Various aspects of the research presented in this chapter can be improved, regardless of the framework chosen. For one, the question should be addressed to what extent the current approach is biased toward pre-defined linguistic categories. Meaningful elements of the gestures in the data might have been overlooked as a consequence of taking models of written/spoken language as a starting point. To assess the appropriateness of current linguistic terminology for modeling multimodal utterances, there is need for a clearer understanding of the functional categories that are specifically relevant to gestural expression. Although research that takes the grammar of written/spoken language as a starting point has been fruitful (see Chapter 3), researchers need to be open to the possibility that some gestural patterns are best captured by a set of categories that is different from (or overlaps only partially with) those assumed by current grammatical theories.

A second possible improvement is to extend the analyses beyond the exclusive focus on manual gesture. Hand gestures are not a self-contained medium, but tightly connected to movement of other bodily articulators and eye gaze. The manual gesture in example (45), for instance, clearly went along with eye contact between the interlocutors. This strengthens the impression that the co-verbal components of this utterance did not only express focus, but also contributed to interactional engagement. Likewise, the fact that the word-search gesture in (25) went along with downward eye gaze could be an additional cue to its function of facilitating memory retrieval. These qualitative observations suggest that the incorporation of manual gesture in models of grammar cannot be fully adequate without considering the ways in which they intertwine with other bodily behaviors.

A third point of enhancement is to incorporate (multimodal) phonological frames in each of the models that are compatible with
patterns of preferred timing. In Chapter 7, I found substantial regularity in the timing of gesture strokes relative to certain words and constructions. Some of the observed patterns may be interpretable as grammatically relevant ‘templates’ for spoken-gestured expression. For instance, I speculated on a dual functionality of the word so in combination with a gesture, depending on whether its articulation precedes or follows the gesture. This pattern might be captured by a relatively simple rule, specifying two different phonological frames for the so + [gesture] construction, with (arguably) different functional properties. Although, as discussed in Chapter 3, a dynamic representation type might be favorable for representing such frames, their incorporation within the notation systems offered currently FDG and CG is also conceivable.
Chapter 9. Conclusion

In this final chapter, I summarize the findings of the dissertation, discuss its implications for linguistics and gesture studies, and touch upon future directions for research on multimodal grammar.

9.1 Summary of the findings
The first part of this dissertation considered the inclusion of gestures in grammar from a theoretical perspective, taking FDG and CG as points of departure. Chapter 2 introduced these frameworks to the extent relevant. A comparative evaluation pointed out that even though both models belong to the functional-cognitive strand of linguistics, they diverge in many respects. Generally speaking, CG is geared towards compliance with principles from cognitive psychology and takes the notion of conceptualization as integral to grammar, whereas FDG takes typology as a primary source of inspiration and is designed to study grammar as situated in discourse. Further points of divergence concern their assumptions on the modularity of linguistic processes, their approach to semantics and the nature of the notation systems employed.

In Chapter 3, I evaluated FDG and CG with respect to their potential to accommodate speakers' gestures as elements of the grammar proper. I highlighted a number of features of both theories that can contribute to understanding certain aspects of the grammar-gesture connection in greater detail. FDG's layered architecture encourages balanced attention to the semantic and pragmatic dimensions of
gestural behaviors. It also offers a rich analytical tool for analyzing scope relations between spoken and gestured forms of expression in the context of the utterance as a whole. As for CG, it too provides useful analytical notions in view of including gestures in its analyses. The grammatical potential of gesture is inherently recognized by CG’s conception of grammar as a prototype category that emerges from contextualized language use. Many of the construal operations that are seen in CG as central to grammatical organization have analogies in gestural expression. The general conclusion is that FDG and CG offer complementary strengths in view of advancing the incorporation of gesture into grammatical theory.

Despite the promising potential of each framework to adopt a multimodal view on language, a number of important challenges remain. Most importantly, the accommodation of gesture in FDG and CG requires a better understanding of (1) the extent to which, and ways in which, gestures are multifunctional; (2) the existence of speech-independent mappings between gestural function and form; and (3) the dynamic dimension of spoken-gestured expression. These challenges were empirically addressed in the second part of the dissertation (Chapters 4-7).

Chapter 4 advanced the view that gestures, when analyzed from a functional grammatical perspective, are inherently multifunctional. I reported on a perception study that was designed to be open to the potential of gestures to express more than one (linguistic) function at a time, with different degrees of prominence. A large set of gesture stimuli, drawn from a multimodal direction-giving corpus, was rated by a set of independent judgers. These indicated, on Likert scales, whether they thought the gesture performed each of nine functions. The results suggest that gestures cluster along three, relatively independent functional dimensions: reference and representation of objects, representation of space and movement, and meta-communicative signaling. These findings are at odds with the view that gestures can be classified as belonging to a single functional category. Instead, the picture emerged that any gesture occupies a location in a multidimensional ‘functional space’.
Building on the methods of Chapter 4, Chapter 5 examined which aspects of gestural expression can be understood when access to speech is denied. By extending the perception study with an audio-muted condition, I examined to what extent access to the verbal channel is necessary to comprehend what type of function a gesture carries out. Overall, the interpretation of the gesture stimuli was very similar in the conditions where the sound was on or off. For all functions investigated, responses in the two conditions were strongly correlated and mean ratings differed only modestly. The most consistent difference between conditions was that the participants were more assured in their judgments when they could hear the speech. These findings support earlier evidence that gestures carry speech-independent meaning on some level of abstraction. It moreover suggests that a functional level of description is appropriate for capturing their inherently meaningful qualities.

Chapter 6 aimed to gain a better understanding of gestural functions in relation to their formal properties. I investigated whether the functional perception ratings from Chapter 4 can be predicted on the basis on the formal properties of the gestures (e.g. handshape, orientation, direction of movement). On the basis of a series of regression models, an estimate was made for each function of what form features are most predictive of the Likert ratings assigned in the online experiment. This procedure resulted in a (very tentative) list of form-function associations. Together with the results from Chapter 5, the findings compromise the view that gestures are fully idiosyncratic.

Chapter 7 concentrated on the spoken tier of the corpus. It compared the distribution of linguistic elements (lemmas, parts of speech, bigrams) in the gesture-accompanied and speech-only segments. Some elements were found to be expressed more often in the temporal proximity of a manual gesture than can be expected by chance, whereas others occurred conspicuously often in unimodal conditions. The most salient ‘gesture-attracting’ lemmas include certain locative terms, demonstratives and discourse particles related to hedging and plausibility. On the level of grammatical categories, I found a tendency for (pronominal) adverbs, nouns and prepositions to be expressed in the company of a gesture. Other linguistic items were observed to be
'gesture-repelling', i.e. they occurred less often in the company of a manual gesture than can be expected by chance. These included verbs of cognition and words that typically occur in topic position (e.g. pronouns). The second part of the chapter investigated whether these results are contingent upon the choice of time window that defines whether a word is to be considered as gesture-accompanied. The results confirm that the definition of co-occurrence has an impact on the results: some linguistic units were only found to correlate with gesture performance when considering direct overlap in time, whereas for others, the relative timing appeared to be rather flexible.

The outcomes of the theoretical and empirical parts of the dissertation motivated the assumptions drawn on in Chapter 8, where FDG and CG were applied to a range of multimodal utterances. Both models were found to be capable of accounting for gestures with a variety of different linguistic functions. The observed strengths and weaknesses of the models mirrored their theoretical and empirical points of focus. FDG proved to be suitable for pinpointing on what level of linguistic organization a gesture operates, and what verbal elements of an utterance it has scope over. CG analyses provided more depth with respect to the conceptual frames and construal operations that underlie the semantic connections between the channels. In line with the theoretical evaluation presented in Chapter 3, the data analyses showed that FDG and CG give emphasis to different aspects of the grammatical import of gestural behaviors.

9.2 Implications for research on grammar
A central motivation behind this dissertation is that many current models of grammar do insufficient justice to the realities of spoken language. The incorporation of gesture in specific grammatical frameworks, as advanced in this dissertation, addresses some facets of the issue at stake. However, the theoretical and empirical insights put forward raise the bigger question of how research on grammar can develop to fully overcome the bias toward written language. When accepting gestures as elements of language, one faces a conundrum that also exists in the field of sign language studies: should one aim to interpret manual behaviors in terms of existing models of (verbal)
Grammar, or should these models be adapted to fit the characteristics of manual expression? Slobin (2008) describes these strategies as 'theories in search of data' versus 'data in search of theories'.

The approach taken in this dissertation is much along the lines of the 'theories in search of data' approach. I started out from existing models and considered how these could accommodate gestural forms of expression. As argued in Chapter 8, some linguistic notions seem applicable to aspects of gestural expression simply because their definitions are compatible with the functions that gestures can perform. For instance, certain functions of gestures seem to map onto the semantic variables (e.g. Individual, Episode) or functions (e.g. Actor, Locative) recognized in FDG, or onto CG's basic semantic categories (e.g. THINGS and PROCESSES). Analyzing gestures in such terms helps to identify the various functional analogies that exist between elements of the two channels. Thus, adopting a 'theory in search of data' approach can be helpful as a test case, highlighting functional commonalities between verbal and gestural modes of expression (Bressem & Ladewig 2011). However, the imminent danger of this approach is that it motivates the analyst to impose a set of pre-defined categories upon a semiotic channel for which these were not originally designed. It is by no means evident that current linguistic categories yield the most appropriate analyses for understanding gestures' role in communication. Turning back to the parallel with sign language studies, Slobin's (2008: 117-118) skepticism about the theory-first approach can be assumed to apply at least to some extent to the study of the gestural component of spoken language:

I maintain that the first approach [theory in search of data], which begins with an established theory based on spoken languages, keeps us from seeing deeper into the nature of signed languages and ultimately into the nature of human languages in general. The second approach – data in search of theories – is the one that is leading us to new insights.

On this account, the implications of a multimodal view of spoken language go beyond the incorporation of gestures in existing models of
grammar. It calls for a radical makeover of linguistic models, where the gestural component will need to be designed from scratch. The question arises, then, what categories and distinctions one would end up with when studying gestures on their own terms. To some extent, recent developments in sign language studies can be a source of inspiration:

Sign language linguists have begun to create tools with which to arrive at a deeper and more comprehensive understanding of human language, human cognition, and human social interaction. Indeed, old paradigms are fading and revolutionary new ideas are growing up across disciplines, languages, and countries. The habitual molds are being broken and new molds are beginning to be formed. (Slobin 2008: 129)

Some of the newly developed analytical notions can be a source of inspiration for gesture analysis. For instance, the notion of ‘buoy’, which refers to the phenomenon where one hand is held up to sustain a discourse topic while the other hand keeps expressing new information (Liddell 2003), seems to have parallels in co-speech gesturing (Engberg-Pedersen 2011). Notions like these, in addition to those that have been used for specific types of gestural behaviors (e.g. catchments, gestural reiteration) could be seen as elements of a linguistic repertoire that is specific to manual expression.

One of the most significant challenges for a model that analyzes speech and gesture as (partially) different systems is that it needs to remain informative with respect to the interface between the different channels. That is, while avoiding the assumption that systems designed for verbal grammar are entirely appropriate for analyzing gestural behaviors, one should also be careful not to assume that speech and gesture are entirely separable (see McCarthy & Carter 2006 for a discussion of the same dilemma in the debate on written versus spoken grammar). For reasons described in Chapter 8, models that describe linguistic structure in terms of general conceptual-communicative categories, rather than pre-defined linguistic notions, offer the greatest potential for striking such a balance.
A further challenge is that language is not expressed through multiple modalities under all circumstances. Models of grammar should be able to deal with the fact that language is ‘variably multimodal’: the semiotic channels that are relevant to linguistic expression may differ across communicative settings and even from moment to moment within a conversation (Cienki 2015). As discussed in Chapter 7, language can be seen as a system that has different possible ‘modes’ (e.g. spoken-only, spoken-gestured, etc.), each of which has different linguistic characteristics (cf. Cienki 2013a; Wilcox & Xavier 2013). Therefore, it would be sensible to commence every linguistic analysis with a specification of the semiotic channel(s) of which the utterance is composed (e.g. text, image, speech, intonation, manual gesturing, facial gesturing). Accordingly, the gestural component of a model of grammar only comes into play when it is applicable to the data analyzed.

9.3 Implications for gesture studies
The content of this dissertation accentuates the significance of studying gesture in relation to linguistic structure. However, it also problematizes the adoption of a linguistic perspective: even within the field of language studies, there is little agreement as to how language should be conceived of and analyzed. I have aimed to add depth to the linguistic branch of gesture studies by focusing on two specific models of grammar with largely complementary assumptions. By simultaneously adopting the perspective of a cognitive linguistic model and a structural-functional model, this dissertation emphasizes the importance of seeing gestures as constrained by both cognitive and discursive-communicative factors. Although these two dimensions are often divorced from each other in current research on gesture, their interaction is essential for understanding the connection between verbal and gestural forms of expression.

Both models taken into consideration present novel ways of understanding gestural meanings and functions. The layered structure of FDG, as argued in Chapter 3, presents a rich model for framing the numerous ways in which speech and gesture can relate to each other. It invites a close account of the possible linguistic functions that gestures can carry out and of the scope relations that can exist between gestures
and elements of speech. This perspective adds significant nuance to one of the most longstanding debates in gesture studies: the question of whether gestures are ‘redundant’ or ‘complementary’ with regard to speech. According to the FDG analyses discussed in the previous chapter, gestures are typically associated with multiple semantic and pragmatic variables simultaneously, only some of which also received verbal expression. These analyses suggest that the functional overlap between gesture and speech is typically partial – the redundancy-complementarity issue is not a black and white matter. The CG perspective casts yet a different light on this debate. CG’s reliance on cognitive semantics helps to understand how gestures tie in with the conceptual frames that are evoked by verbal structures. On this view, speech and gesture can never be considered fully redundant with each other; by virtue of the intrinsic differences between spoken and gestured behaviors, the conceptual frames that underlie their semantic poles will never be fully overlapping.

The dissertation also stimulates methodological innovation. Chapter 4 introduced a methodology that is open to the potential of gestures for having multiple functional dimensions with different degrees of prominence. This conception of gestural functionality suggests that a multi-layered annotation system is needed to adequately capture the contribution of a gesture to an utterance (consistent with recent systems such as LASG and NEUROGES). In addition, Chapter 7 introduced a corpus-based methodology for revealing statistical patterns in the co-occurrence of speech and gestures in a way that bypasses subjective interpretation. The results suggest that gestures are not linked to a single word or phrase (as supposed by the notion of ‘lexical affiliate’, used in many previous studies). Instead, there can be various types of relations at once between a gesture and the semantic and pragmatic layers of the verbal channel.

9.4 Future directions of research on the multimodality of grammar

Although research on the multimodality of grammar already has a history dating back more than four decades (e.g. Slama-Cazacu 1976), it has largely remained exploratory in character. Along with increasing
attention and recent developments in research on grammar and gesture, the question arises as to how this line of inquiry can develop into a more mature field of study. In my current view, which rests upon the contents of this dissertation, the following two directions deserve a place on the agenda.

9.4.1 Definitions and goals
As seen in Chapter 2, grammatical theories assume different definitions of grammar and pursue different empirical goals. Whereas formal models of grammar were developed to make predictions about syntactic well-formedness, other models have been developed with typological, discourse-analytical, cognitive, developmental or computational purposes in mind. Given this diversity in empirical orientations, it is important for any account of multimodal grammar to formulate what it wants to achieve. The goals behind research on multimodal grammar can be theoretical – understanding of the status of co-verbal behaviors in a specific model of grammar – but they can also be directed toward more practical applications. One can think of gearing the way one thinks about the gesture-grammar nexus toward therapeutic applications in relation to speech deficits. This could ultimately benefit those seeking solutions for communicative disorders in assessing if a disorder is a linguistic one or something broader. Further applications could involve automatic speech processing on the basis of video recordings, enabling computers to simultaneously process auditory and visual signals. More generally, it is imaginable that models of multimodal grammar can inspire matters concerning multimodal communication via film, television, or theater. In any case, a clear definition of the empirical goals can motivate the selection of an appropriate model and guide the research toward a specific outcome, within or outside of the realms of linguistics and gesture studies.

9.4.2 Proofs of concept
The theoretical claim that grammar is multimodal has been made on numerous occasions. To make this claim substantive, a ‘deeds, not words’ approach is needed. As discussed in Chapter 3, the actual
application of models of grammar to gesture data has so far only been attempted for a few models and for a few types of gestures. Chapter 8 of this dissertation aimed to provide more depth by presenting detailed analyses of a wide range of gestural phenomena. Such an approach can help to make theoretical claims more explicit and enforces the formulation of specific assumptions about the appropriateness of grammar terminology for analyzing gestural behaviors. In addition, because actual data is never as ‘clean’ as the categories of a theoretical framework, it helps to pinpoint the weaknesses of the approach adopted. Thus, by putting a model to the test against multimodal data, its quality can be iteratively assessed and adapted.

In the long run, testing a model by its ability to account for selected samples of data will not be enough. To arrive at a more solid proof of concept, a transition might be needed from analytical to generative models.\(^{59}\) That is, in order to test whether a model of multimodal grammar is sufficiently generalizable, it should make predictions about the circumstances under which certain gesture types are appropriate. With recent advances in the development of computer avatars and robots, opportunities have opened up to make such predictions testable. Individual features of gestures can be isolated to see under what parameter settings humanoid avatars appear most natural. The field of multimodal grammar can thus converge with work on human-computer interaction (Alahverdzhieva 2013; Kipp 2004; Kopp et al. 2008; Kopp & Wachsmuth 2010). This would also enable formulating and testing assumptions about the relation between manual gestures and movements of the shoulders, head and face, helping to move beyond the current bias toward just manual gestures. Solid theoretical and empirical foundations are essential, but the ultimate way to validate a theory of multimodal grammar is to see it in action.

\(^{59}\) The word *generative* is used in the general sense here, not as a specific reference to Chomsky’s generative grammar.


Bibliography


Cienki, A. (2012). Usage events of spoken language and the symbolic units we (may) abstract from them. In K. Kosecki & J. Badio (Eds.), *Cognitive Processes in Language* (pp. 149 - 158). Frankfurt am Main: Peter Lang.


Lücking, A, Bergmann, K., Hahn, F., Kopp, S., & Rieser, H. (2010). The Bielefeld speech and gesture alignment corpus (SaGA). In M. Kipp, J. C. Martin, P. Paggio & D. Heylen (Eds.), Proceedings of the 7th International
Conference for Language Resources and Evaluation (pp. 92-98). Valetta, Malta.


Mackenzie, J. L. (2013). Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. Journal on Multimodal User Interfaces, 7(1-2), 5-18.


Okrent, A. (2002). A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics (or at least give us some criteria to work with). In R. P. Meier, K. Cormier & D. Quinto-Pozos (Eds.), Modality and Structure in Signed and Spoken Languages (pp. 175-198). Cambridge: Cambridge University Press.


Summary of the dissertation


The first part of the dissertation reviews a number of theoretical points of convergence between these theories and gesture research. In Chapter 1, the choice of the models of interest is motivated and both models are introduced. A systematic comparison points out that although both models belong to the functional-cognitive branch of linguistics, they differ considerably in terms of their underlying assumptions and principles. Generally speaking, CG accords a central role to principles from cognitive psychology and the notion of conceptualization, whereas FDG takes typology as a primary source of inspiration and provides a more formal, interaction-oriented model of grammar. The theoretical review continues with a literature-based discussion of the potential of each of the models to incorporate gestural expression. A general conclusion from this assessment is that both models provide novel ways of understanding the relation between gesture and speech. FDG’s explicit distinction between representational and interpersonal dimensions of language promotes equal attention to semantic and pragmatic functions of gestural behaviors, and offers a rich framework for analyzing scope relations between spoken and gestured elements of multimodal utterances. The analytical tools offered by CG, on the other hand, shed light on the different construal operations that
underlie gestural expression, and lend themselves better to describing the schematicity of gestural meanings.

Importantly, a number of challenges emerge from this theoretical discussion, relevant to each of the two models. A first challenge is that the pervasive multifunctionality of gestural behaviors calls for an approach that assumes relatively complex form-meaning mappings. Second, it remains unclear how the units of analysis are to be defined, i.e. whether and how gestural behaviors can be described in terms of ‘primitive’ linguistic elements. Third, the temporal dimension of spoken-gestured expression adds substantial complexity to the notion of linguistic structure. The second part of the dissertation address these issues through a number of empirical studies. All of these are based on a large-scale video corpus, which is introduced in Chapter 4. Using excerpts from this corpus as stimuli, Chapters 4 and 5 present studies that investigate the perception of manual gestures by laypeople. In these studies, participants indicated how certain they were that the gestures they observed performed one or more of a total of nine given functions. Thus, the design was open to the potential of gestures to express more than one (linguistic) function at a time, with different degrees of prominence. The results suggest that most gestures combine multiple functional dimensions, some of which are more salient than others, and should therefore not be thought of as belonging to a single class or category.

In a subsequent study, the experiment was repeated in a condition where participants had no access to speech. Overall, no fundamental differences were observed in the perception of the gesture stimuli in the conditions where the sound was on or off. Participants were, however, more assured in their judgments when they could hear the speech. These findings support the view that gestures carry inherent, speech-independent meaning on some level of abstraction.

Chapter 6 aims to gain a better understanding of the functions of gestures in relation to their forms. It reports on a number of correlations between the formal parameters of the gestures in the stimulus set (e.g. handshape, palm orientation, direction of movement) and the functional perception ratings from Chapter 4. For each function, a statistically based estimate was made of which form features were the most
characteristic (i.e. most strongly predictive of a high rating). Although the regression models used for this assessment only explained a modest portion of the variance in the data, their predictive ability was statistically significant for all functions examined. On the basis of these results, a tentative list of form-function pairings was devised.

The final empirical chapter examines the linguistic side of the corpus, asking which words and grammatical categories are likely or unlikely to coincide with manual expression. It compares the distribution of verbal structures (lemmas, parts of speech, bigrams) that occur in the temporal proximity of the gestures in the corpus with those that occur in speech-only circumstances. Lemmas that were found to be ‘gesture-attracting’ (positively correlated with gesture performance) include a number of deictic words, nouns and adverbs with spatial-motoric features, as well as certain discourse particles. By contrast, verbs of cognition, interactives and pronouns were found to be ‘gesture-repelling’. To find out whether these results hinge upon the operational definition of ‘co-occurrence’, the procedures were repeated with different criteria for considering a word as gesture-accompanied. For most of the linguistic units examined, the choice of time window had a strong impact on the observed degree of gesture-attraction. The preferred timing between speech and gesture appeared to be more flexible for discourse-related lemmas than for descriptive words.

The third part of the dissertations brings together the theoretical and empirical insights from earlier chapters. It provides parallel analyses – using both FDG and CG as starting points – of a diverse range of spoken-gestured utterances. Both models were found to be capable of accounting for various ways in which speech and gesture can connect on both functional and structural levels. The perspectives provided by FDG and CG are in many ways complementary, as can be expected on the basis of the differences in their theoretical underpinnings. In accordance with the conclusions of the first part of the dissertation, the data analyses show that FDG and CG highlight different aspects of the grammatical potential of gestural expression. The discussion of these differences reveals a number of ingredients required for a model of multimodal grammar that is maximally informative of how speech and gesture interrelate during situated communication. In the final chapter,
the findings of the dissertation are synthesized and discussed in terms of their implications for the fields of linguistics and gesture studies.