

III MULTIMODAL CORPORA

Jens Edlund, David House and Jonas Beskow
Gesture movement profiles in dialogues from a Swedish multimodal
database of spontaneous speech 265

Patrizia Paggio
Towards an empirically-based grammar of speech and gestures 281

Subject Index 315

Towards an Interactional Grammar

1. Interaction meets grammar theories

It has been widely shown that participants in conversation use lexico-grammatical structures specifically designed for the conditions and requirements of talk-in-interaction (e.g. Deppermann, Fiehler, and Spranz-Fogasy 2006; Ford, Fox, and Thompson 2002; Günthner and Imo 2006; Hakulinen and Selting 2005). Based on these insights, calls have recently been made for a “Grammar of Spoken Language” (e.g. Auer 2005; Günthner 2011) in research informed by Interactional Linguistics and related approaches. Specifically, Construction Grammar (Croft 2001), Cognitive Grammar (Langacker 1987, 1991), and Emergent Grammar (Bybee 2006; Hopper 1987) have been welcomed as grammatical theories that are particularly suited to modeling an Interactional Grammar (cf. e.g. Fried and Östman 2005).¹ These grammar models and interactionally informed approaches share the assumption that the linguistic knowledge of speakers is based on experiences from language use, thus adopting the view of grammar proposed by, e.g. Langacker (2001):

Cognitive Grammar takes the straightforward position that *any* aspect of a usage event, or even a sequence of usage events in a discourse, is capable of emerging as a linguistic unit, should it be a recurrent commonality. (Langacker 2001: 146, emphasis in the original)²

A similar view is taken by interactional approaches, which also assume that linguistic knowledge is not static and fixed, but is built, ratified, and modified in interaction. In this sense, grammar and usage are in a reciprocal relationship and cannot be treated as separate entities: On the one hand, grammar provides the basis for language use, but on the other hand, grammar is a flexible, permanently changing product emerging from language use. From this perspective, then, actual language use thus influences grammar. This assumption implies that the strict separation of competence and performance postulated by Generative Grammar approaches (e.g. Chomsky 1965) loses its validity. “Performance” gets revalorized and should serve as a starting point for an empirical analysis of grammar.

¹ In the present discussion, we will use the term Interactional Grammar to reflect the interactional, responsive nature of lexico-grammatical structures in embodied conversational encounters.
² Langacker (2001: 144) defines usage events as “actual instances of language use”.

Thus far, the major focus of interactionally informed research trying to bridge the gap between language use and linguistic knowledge has been the examination of lexico-syntactic structures, mostly within the framework of Construction Grammar (e.g. Deppermann 2006; Birkner 2008; Günthner 2011; Günthner and Imo 2006; Imo 2011). In comparison, only few in-depth studies have addressed the complexities of prosodic contextualization in grammar (but see Barth-Weingarten 2006; Couper-Kuhlen 2007; Gohl 2006). Also, despite a fast-growing body of interactional research on the visual resources used in embodied interaction such as gestures, gaze, facial expression, movements of the body and the head, body posture, and body position (e.g., Mondada and Markaki 2005; Stivers and Sidnell 2005; Streeck, Goodwin and LeBaron 2011), their role in a grammar of interaction has not yet been broadly discussed (but see Fricke 2012).

The contributions to the present volume demonstrate, from an Interactional Linguistics perspective, how prosody and embodiment form relevant parts of the linguistic and communicative knowledge of participants in interaction. In this sense, we argue, they are potentially relevant to an Interactional Grammar. While these contributions provide evidence for the notion of Interactional Grammar proposed here, the question of how to model this grammar within a theoretical framework must be left unanswered (cf. section 2). These interactionally informed contributions are followed by studies on (annotated) multi-modal corpora and instrumental approaches to the analysis of language use. They are intended to instigate a discussion of how such approaches might complement the study of multimodal meaning-making from a purely interactional perspective for reasons discussed below (cf. section 3).

In the following section, we turn to a discussion of prosody and embodiment and their relationship to grammar, pointing out some possible links and open questions. As research on these issues has just begun, this discussion might raise more questions than it can answer.

2. Prosody and embodiment in Interactional Grammar

Studies from the field of contextualization research and Interactional Linguistics underline the fact that prosody must not be treated as marginal, but as a crucial component in the description of linguistic structures (see e.g. Auer and Di Luzio 1992 for English and German; Bergmann 2008 for German; Couper-Kuhlen and Ono 2007 for English, German, and Japanese; Reber 2012 for English; Viscardi 2006 for Portuguese). These studies demonstrate that prosody is deployed as a resource for various turn-constructive, sequence-organizational, and interactional tasks.

In light of these and other findings, Selting (2010) identifies “two basic functions” of prosody in talk-in-interaction. First, it is “*always (a)constitutive*” of interactional meaning-making because “[t]here is no spoken language without prosody,” and for this reason, prosody always serves as a potential contextualization cue for the ongoing conversational project. Second, in some interactional activities, it may have a “*distinctive* function” (Selting 2010: 6, emphasis in the original). For example, Selting (1996) found that depending on the prosodic shape of the German open class repair initiator *was* (“what”), it was interpreted by the first speaker as a hearing problem, a problem of understanding, or a problem of expectations, i.e. as a display of astonishment.³

Interactionally informed studies which model the role of prosody in the construction of grammatical descriptions have recognized that prosody may make a relevant contribution to grammatical functions, such as the constitution of larger units, e.g. turn construction units, and information structuring (cf. Barth-Weingarten 2006). In our view, however, it is time to go one step further and, firstly, adopt as a basic assumption that prosodic devices may also be potentially pervasive in the construction of smaller lexico-grammatical units (with respect to size and/or frequency). For example, based on a corpus of television interviews, Berkenfield (2001) discovered that the phonetic design of American English *that* (with respect to the quantity and quality of the vowel) depends on whether *that* is serving as a demonstrative pronoun, demonstrative adjective, complementizer, or relative clause marker. Secondly, it should not be ignored that all prosodic devices are interactionally embedded and therefore are subject to the conditions that govern and constrain naturally-occurring interaction. For instance, similar to lexico-syntactic structures, prosodic devices unfold in time and – in terms of their forms and functions – relate both to prior and subsequent productions of turns and sequences.⁴ For these reasons, we take the position that prosody

³ Although the distinctive function of intonation has also been noted by both the so-called British and the American schools of intonation, these perspectives do not typically take into consideration the indexical nature of prosody and do not derive their findings from interactionally situated talk (cf. Batliner 1989; Gussenhoven 2004; Oppenrieder 1988; Pheby 1975). The same is true for the description of such thoroughly grammatical functions as information structuring and phrasing, both of which constitute well-established core functions of intonation in formal accounts of grammar (cf. Ladd 2008; Välimaa-Blum 2005).

⁴ Cf. Selting (1995: 366) “Es resultiert die weitergehende Perspektive einer Linguistik der Konversation, in der linguistische Strukturen und Systeme als Signalisierungsmittel und als Ressource der Organisation der konversationellen Interaktion beschrieben werden. In dieser Perspektive wären analog zur interaktionalen Pros-

should be regarded as a central component in the description of lexico-grammatical structures, and that an adequate model of grammar should be able to accommodate aspects of truly interactional language.

Turning to embodiment, the study of face-to-face interaction suggests that visual resources, similar to prosodic resources, also contribute to meaning-making in interaction. However, in contrast to prosodic devices, most linguists do not consider visual resources as belonging to the lexico-grammar of a language. Therefore, the question arises as to how embodiment can be incorporated into an Interactional Grammar, that is, into a grammar of interaction which models all communicative resources in a unified fashion. These resources, including lexico-grammatical, visual-spatial, and others, such as so-called paralinguistic resources (e.g. laughter or whistling) which are potentially relevant to the communicative construction of meaning, are viewed as intertwined in social action formation.

To illustrate this point, Reber 2012 finds that the sound pattern and context-specific use of affect-laden sound objects such as *oh* or *ah* in English talk-in-interaction are distinctive for their meaning-making. Additionally, sound objects may be accompanied by “visual-spatial resources which are (1) physiologically inherent to the articulation of sound objects and those which (2) (by convention) build part of an embodied gestalt in which the sound object is performed” (Reber 2012: 249). With regard to (1), Reber and Couper-Kuhlen (2010) suggest that producers of a whistled sound object must have pursed lips on production (Reber and Couper-Kuhlen 2010: 86). As to (2), the production of a “pained sound” by the rejectee in a rejection sequence may be accompanied by a conventionalized cluster of visual signals such as averted head plus lowered gaze (Reber and Couper-Kuhlen 2010: 84). Observations of this kind suggest that visual-spatial signals should be described as contextualization cues that intersect with linguistic lexico-grammatical cues and others in interactional processes of meaning-making. In this sense, they belong to a grammar of interaction as they form part of multimodal ges-

odie der Konversation auch segmental-phonologische, morphophonemische, syntaktische, lexikalisch-semantische u. a. Signalisierungssysteme als interpretativ relevante Kontextualisierungshinweise in der Alltagskommunikation zu untersuchen.” [This results in the further perspective of a Linguistics for conversation, in which linguistic structures and systems are described as signaling devices and as an organisational resource for conversational interaction. In this perspective, segmental-phonological, morphophonemic, syntactic, lexical-semantic etc. signaling systems would – by analogy to the interactional prosody of conversation – also have to be examined in terms of their role as interpretively relevant contextualization cues in everyday communication.] (our translation).

alts, be it for physiological reasons (as in (1) above) or by convention (as in (2) above, cf. also Mondada⁵).

Furthermore, visual resources play a central role in the constitution of units in interaction (cf. Ford, Fox, and Thompson 1996; Mondada 2007a). For example, in her study on repeated gestures which serve as a “tying technique to connect utterances over time” (Laursen 2005: 1), Laursen argues for an embodied grammar, claiming that gestures are an integral part of a turn-constructional unit (Laursen 2005: 19) and can ensure coherence in interaction. Moreover, visual signals can even form a turn (cf. e.g. Stivers 2008 on alternative recipient tasks performed by vocal continuers and nodding in story-telling) or form part of a turn-constructional unit on their own (cf. Ford, Thompson, and Drake 2012, Olsher 2004). This suggests that in addition to points (1) and (2) above, an Interactional Grammar must allow for (3) visual-spatial resources which form alternatives or are complementary to the use of lexico-grammatical structures in interactional turn construction. Furthermore, it is evident that prosodic and visual resources are closely related in the formation of interactional tasks, such as turn-taking (cf. e.g. De Stefani 2005; Iwasaki 2009; Mondada 2007b; Oloff; Streeck and Hartge 1992 for visual resources and e.g. the contributions in Couper-Kuhlen and Ford 2004 for prosody) and the affective framing of sequences (cf. Gülich and Lindemann 2010). In view of these findings, the research presented in our volume is meant to contribute to the discussion on how prosody and embodiment are relevant for Interactional Grammar.

In the following section, we briefly summarize some problem areas and lines of discussion found in the contributions of the present authors, proceeding from those directly relevant to the study of prosody and embodiment to those posing more general questions for the study of linguistic structures in interaction:

(i) The forms and functions of prosody and embodiment in interaction cannot usually be modeled in terms of simple form-meaning pairs, because these resources require indexical interpretation. They often do not have a fixed meaning, but refer in their specific context of occurrence, i.e. together with co-occurring contextualization cues, to the relevant interpretative framework which can be located outside of the utterance. For this reason, the Interactional Linguistic study of prosodic and visual resources emphasizes the need for a *holistic perspective on linguistic structure, and, more widely, interactional gestalts*, and underline that linguistic structures and co-produced visual signals can only be interpreted as *social and situated actions in their specific*

⁵ References without year refer to contributions to the present volume.

context (cf. Hörmeyer; Mondada). For this reason, the prospect of deepening of our understanding of multimodal structures as co-occurring, interacting cues for the formation of social actions conveyed linguistically (cf. Szczepek Reed) and of how such cues are embedded in the broader linguistic and non-linguistic context seems highly promising. In our view, such a holistic perspective on interaction is essential when dealing with questions about how linguistic knowledge is formed.

Turning to the issue of the interactional context, current debates about grammar often marginally consider or ignore altogether the primary site of language, i.e. face-to-face interaction. For instance, Välimaa-Blum (2005: 3) claims in her construction grammatical introduction to cognitive phonology that “the principal and only function of language as a system is the expression of meaning.” This point of view neglects the contextualization of linguistic actions in specific social situations, in which participants in conversation interact in time and space to pursue their communicative goals. This neglect is, therefore, difficult to reconcile with an Interactional Grammar. In the same vein, Günthner (2011) argues against such a position and criticizes the fact that a majority of studies within the construction-grammar framework, among others, still define constructions as “stable, homogenous and decontextualized units” [stabile, homogene und dekontextualisierte Einheiten] (Günthner 2011: 16, our translation) and do not consider the temporal, embodied, and interactional character of language.

(ii) The *online character* and the *temporal organization of interactional structures* must be accounted for in an Interactional Grammar. Studies exploring the syntax of spoken language have largely shown how the temporal unfolding of interaction shapes the form and the function of emerging syntactic structures (cf. e.g. Auer 2009; Günthner and Hopper 2010 on pseudocleft constructions in German). Prosodic projection (Auer 1996; Couper-Kuhlen 2007; Selting 1995: 73) interacts in a complex way with the emergent syntax of a turn (cf. Bergmann; Birkner; Brenning; Pfeiffer).

In face-to-face interactions, the temporality and interactional organization of visual resources also plays a crucial role and, therefore, must be taken into account. In fact, every resource (syntax, gesture, prosody) has its own temporal organization. For example, it has been shown that gestures precede their lexical affiliates (cf. Schegloff 1984) and that prosodic units do not always coincide with syntactic units (cf. Selting 2000). As Stivers and Sidnell (2005) claim, each modality thus has its own organization. The close coordination and timing of visual cues in relation to speech and the sequential position of the emerging gestalt has to be accounted for in an Interactional Grammar (Mondada; Oloff).

(iii) Correspondingly, the role of linguistic structures for the *management of conversational tasks* (e.g. the organization of turn-taking in conversation, the framing of activity types, and the organization of repair) has thus far been neglected in theories of grammar. With respect to repair, Fox, Hayashi and Jaspersen (1996) have shown that self-repair as a phenomenon specific to oral language has a reciprocal relationship with syntax: On the one hand, self-repair strategies in interaction are influenced by the underlying language-specific grammar; on the other hand, every grammatical system is designed in a way to allow for self-repair. In addition to the influence of language-specific grammatical features (cf. e.g. Birkner et al. 2010, 2012; Fox, Maschler, and Uhmman 2009), the syntactic and prosodic organization of self-repair seems to be shaped by various factors from interaction and cognition (cf. Pfeiffer 2010, this volume, in press). Unit construction in conversation also illustrates the fact that *conversational units* must be conceptualized by taking into account the contingencies of interaction (cf. Brenning; Bergmann; Hörmeyer; Szczepek Reed). Concerning the role of prosody for the constitution of units in interaction, both Bergmann and Birkner demonstrate that prosody allows for different degrees of prosodic phrasing. In her study on parentheses, Bergmann shows that prosodic means are used regularly to signal a break-off in the emergent syntactic structure. However, she finds considerable variation in the way different prosodic resources combine in order to accomplish this task. Similarly, Birkner demonstrates that the relationship between the semantics of relative clauses and its prosodic design is much more complex than past literature has suggested. These examples illustrate that an Interactional Grammar must provide for a contextualized description of the use of prosodic resources, recognizing that language is situated in interaction, and thus in time and context.

In conclusion, given the current state of research, we propose a view of prosody and embodiment in interaction in which they are considered contextualization cues (Gumperz 1982), intersecting with one another and with lexico-syntactic structures in interactional meaning-making. In this sense, lexico-grammatical structures, including prosody, and visual-spatial resources, together with paralinguistic and other potential cues, form part of what we wish to call an Interactional Grammar.

3. Multimodal corpora

In this section, we turn to the question of whether and how to use studio-recorded multi-modal corpora as a potentially complementary data base in interactionally oriented research.

Because of the radically empirical approach to the investigation of interaction, working with studio-recorded data is not common in Interactional Linguistics due to the unnatural character of the data gathered in such settings. However, with a focus on the online-production of language, on how the participants in interaction use prosodic and visual devices in time, and on the need for a profound understanding of these processes when modeling an Interactional Grammar, it seems worthwhile to consider the use of additional technologies, e.g. eye tracking or motion capture systems (cf. Edlund, House, and Beskow). Conceivably, the use of such technologies would allow deeper insight into the detailed timing and relations between prosodic and visual devices than would perhaps be possible on the basis of naturally occurring data. As regards gaze, for example, its description in naturally occurring interaction often poses problems for the analyst. To illustrate this point, Reinhold Schmitt (p.c.) observes:

Videoaufnahmen authentischer Situationen [...] sind [...] aufgrund aufnahme-technischer Kontingenzen (Vollständigkeitsorientierung, Lichtverhältnisse, Kameraperspektive(n) und Kameraführung) in der Regel für eine exakte Rekonstruktion der Komplexität und Dynamik von Blickorganisation nicht wirklich geeignet. Das Hauptproblem besteht dabei in der exakten Rekonstruktion des tatsächlichen Blickpunktes (dem Zielpunkt des Blicks). [Because of the contingencies of the recording situation (orientation to exhaustiveness, lighting conditions, camera perspective(s), and camera work), video recordings of authentic situations [...] usually do not lend themselves to an exact reconstruction of the complexity and dynamics of the organization of gaze. Here the main problem consists of the difficulty in exactly reconstructing the actual visual focus (the point of gaze).] (our translation)

As there is a large difference between a participant's gaze into the eyes of his/her co-participant or on the root of his/her nose, an analysis of video data may result in a description which does not give as much analytic detail of the point of fixation as may be required (Reinhold Schmitt, p.c.).

The development of studio-recorded multi-modal corpora such as the one presented in Edlund, House, and Beskow is, however, still in its beginnings (see also e.g. Kipp et al. 2009, Paggio for similar approaches). Nevertheless, previous findings on the interrelation of prosodic and visual resources underline the usefulness of such corpora. For instance, Loehr (2006) demonstrates that head movements, hand movements, and pitch accents (which he relates to movements of the larynx) are rhythmically coordinated with each other and "sometimes align on meeting points", i.e. they co-occur in time (Loehr 2006: 193). Similar observations have been made in interactionally informed research (Streeck and Kallmeyer 2000). It is true that researchers both within the interactional paradigm and in so-called multimodal

corpora deploy different methodologies and generally take different research interests. For example, Interactional Linguists may be primarily concerned with the situated multimodal organization and the accomplishment of social actions in their natural habitats, while those focused on multimodal corpora research may be focused on the coordinated production of speech and embodiment for applied issues such as multimodal user interfaces. However, we posit that a combination of the two may lead to synergies and supplementary benefits, at least for some research questions.

In addition to these efforts to meet the demand for multimodal *online*-corpora in a way that respects the temporal and multimodal organization of language, recent approaches to the representation of non-verbal behavior types may contribute to the development of a description of the relation between speech and visual resources in a multimodal grammar (Paggio). This approach does diverge from Interactional Linguistics theoretically and methodologically, by, for example, building grammar on the basis of annotations of a corpus that rely on the annotators' interpretations. However, it gives a clear prospect on what an abstract, holistic representation of multimodal signs within a certain grammatical framework (that of Head Driven Phrase Structure Grammar) may look like.

Based on these considerations, the interactionally informed contributions in this volume are complemented by contributions on annotated multimodal corpora. In this way, we hope to advance the discussion on how studies from these two domains may potentially draw from one another.

4. The contributions

The contributions to the present volume are organized into three major sections which reflect the major areas of interest of the volume: I) Prosody, II) Embodiment, and III) Multimodal corpora. Each study approaches the questions raised in section 1 on the basis of results from original case studies.

I Prosody

KARIN BIRKNER's study on the prosodic formats of relative clauses in spoken German takes a critical look at the commonly accepted assumption that restrictive and appositive relative clauses can be distinguished on the basis of their prosodic integration into the matrix clause. The analysis shows that the relationship between the semantics of relative constructions and their prosodic design is much more complex than previously suggested. While appositive relative clauses are usually non-integrated, as might be ex-

pected, the group of restrictive relative clauses is quite heterogeneous. As a result, Birkner concludes that the prosody of relative clauses is affected not only by semantic factors but also by various factors in conjunction with informational and interactional structure.

MARTIN PFEIFFER conducts a prosodic analysis of substitutions of the determiner in German prepositional phrases, showing that intonation is affected by changes in gender. More specifically, the intonation pattern of the repaired segment, i.e. on the preposition and the determiner prior to repair initiation, falls considerably lower when the gender of the determiner is subsequently altered, compared to alterations of definiteness, number, cliticization, and mere repetitions of the preposition and the determiner. Pfeiffer identifies a link between this falling intonation pattern and the cognitive processes involved in lemma substitutions, which must always be carried out in alterations of gender but not in other types of alteration, and discusses possible interactional implications of this finding. Given the relationship between the syntactic category of gender and intonation, he argues for an integration of this prosodic aspect into an Interactional Grammar.

Providing further evidence for the relevance of prosody for Interactional Grammar, the third contribution in this section addresses the grammar of syntactic co-constructions in spoken German. JANA BRENNING argues that intra-turn speaker change within *terminal item completion* in German can be systematically described by referring to the prosodic design of the co-constructed unit. It is shown how incoming speakers orient the beginning of their completion toward a projected possible position for the nucleus accent syllable to pre-empt another speaker's emerging syntactic gestalt. Brenning further discusses how the incoming speaker can anticipate this position by relying on the emergent syntax of the previous speaker's turn.

PIA BERGMANN's contribution on the prosody of parentheses in spoken German focuses on the marking of boundaries between different parts of the parenthesis. In a detailed prosodic analysis, Bergmann demonstrates that prosodic cues indicate upcoming syntactic breaks and contextualize the different parts of the parenthetical structure as being separated (host vs. parenthesis) or belonging together (multiple parenthesis). In other words, she demonstrates how prosodic cues are systematically exploited in the phrasing of units. Bergmann then discusses possible insights that might be gained from a combination of the concepts of Interactional Linguistics and Prosodic Phonology / Autosegmental-metrical Phonology.

The final chapter in the section on prosody also addresses the notion of units in conversation. BEATRICE SZCZEPK REED questions the common use of the intonation phrase as a unit of analysis for natural talk, asking

whether participants in interaction orient toward chunks (shaped like intonation units) which accomplish conversational actions. In her chapter, she suggests the term *action component* to refer to these units smaller than turn construction units in order to take into account their interactional relevance for participants as building-blocks for actions. She claims that we have to forgo a mere formal linguistic (syntactic and prosodic) conceptualization of these units and adopt a point of view which acknowledges their role in the formation of action.

II Embodiment

In her detailed analysis of a collection of instances of the French deictic *ici* (here) LORENZA MONDADA proposes a grammatical description of *ici* as a multimodal gestalt that is crucially based on the temporal unfolding of the turn, its sequential organization, and its context. She identifies two multimodal patterns surrounding *ici*: 1) *ici* + pointing gesture as introducing a new referent, and 2) *ici* as an attention getting device. She concludes that "[g]rounding grammar on use and users means [...] a focus on interaction, time and context" (Mondada: 202).

FLORENCE OLOFF examines the withdrawal from turns in overlap, demonstrating how the incorporation of a multimodal approach to this well-studied interactional phenomenon can shed new light on its organization. Providing evidence that a purely syntactic perspective does not explain the point in time when a speaker withdraws from a turn, she claims that speakers orient toward the (un)availability of other participants, as, for example, displayed by their body position or gaze.

Analyzing the role of gaze in the constitution of units in Augmentative and Alternative Communication, INA HÖRMEYER focuses on a kind of conversation in which essential interactive resources like prosody and syntax are missing. In her examination of interactions in which one conversation partner suffers from severe cerebral palsy, she demonstrates that interactions via an electronic communication aid require speakers to make explicit the boundaries of their units through the use of visual signals. As participants can be shown to regularly orient toward the aided speaker's shift in gaze to identify turn-constitutional units, Hörmeyer concludes that gaze must be seen as constitutive of a grammar of Augmentative and Alternative Communication.

III Multimodal corpora

JENS EDLUND, DAVID HOUSE, and JONAS BESKOW report on a method for using infra-red cameras and reflective markers to capture body and head gestures in which the data is used to automatically produce gesture movement profiles for spontaneous dialogues. Given the limitations of a fine-grained analysis of gesture on the basis of video recordings, the authors emphasize that using motion capture data in addition to audio and video data may aid in the analysis of the multimodality of language-in-interaction in more detail. For instance, motion capture can help us get a better grasp of the timing relationships between speech and facial and body gestures, and may lead to a better understanding of what aspects of gesture and motion should be considered as part of language and grammar.

Focusing on the analysis of head movements in video-recorded conversations in Danish, PATRICIA PAGGIO discusses how non-verbal behavior can be integrated into a theory of multimodal grammar. She presents a method of data annotation that allows for the representation of the gesture-speech relation, including aspects of information structure, and suggests modeling a multimodal grammar that is based on so-called feature structures that constitute multimodal signs. These multimodal signs represent the gesture (its shape and its communicative function) and the speech segment it is associated with.

Acknowledgments

We are grateful to Peter Auer, Elizabeth Couper-Kuhlen, and Florent Perek for their suggestions and critical comments on an earlier version of this chapter. Furthermore, we would like to thank the two anonymous reviewers of the advisory board of the School of Language and Literature, Freiburg Institute for Advanced Studies (FRIAS). Thanks are due to Elizabeth Tremmel for checking our English.

Some contributions to this volume were first presented at the FRIAS conference *Interaction and usage-based grammar theories. What about prosody and visual signals?* held in December 2009 in Freiburg. We are indebted to the FRIAS for supporting the organization of the conference and the publication of this volume. We would also like to thank Janaisa Martins Viscardi for her conceptual and organizational contribution to the conference.

References

- Auer, P. 1996 On the prosody and syntax of turn-continuations. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in Conversation. Interactional studies*, 57–98. Cambridge: Cambridge University Press.
- Auer, P. 2009 Online-Syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31: 1–13.
- Auer, P. and A. Di Luzio 1992 *The Contextualization of Language*. Amsterdam: John Benjamins.
- Barth-Weingarten, D. 2006 Parallel-opposition-Konstruktionen: Zur Realisierung eines spezifischen Ausdrucks der Kontrastrelation. In: S. Günthner and W. Imo (eds.), *Konstruktionen in der Interaktion*, 153–179. Berlin: de Gruyter.
- Batliner, A. 1989 Wieviel Halbtöne braucht die Frage? Merkmale, Dimensionen, Kategorien. In: H. Altmann, A. Batliner and W. Oppenrieder (eds.), *Zur Intonation von Modus und Fokus im Deutschen*, 111–162. Tübingen: Niemeyer.
- Bergmann, P. 2008 *Regionalspezifische Intonationsverläufe im Kölnischen. Formale und funktionale Analysen steigend-fallender Konturen*. Tübingen: Niemeyer.
- Berkenfield, C. 2001 The role of frequency in the realization of English that. In: J. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 281–308. Amsterdam: John Benjamins.
- Birkner, K. 2008 Was X betrifft: Textsortenspezifische Aspekte einer Redewendung. In: A. Stefanowitsch and K. Fischer (eds.), *Konstruktionsgrammatik II. Von der Konstruktion zur Grammatik*, 59–80. Tübingen: Stauffenburg.
- Birkner, K., S. Henricson, C. Lindholm, and M. Pfeiffer 2010 A contrastive study of the syntax of self-repair in German and Swedish prepositional phrases. *InLiSt – Interaction and Linguistic Structures*, No. 46.
- Birkner, K., S. Henricson, C. Lindholm, and M. Pfeiffer 2012 Grammar and self-repair: Retraction patterns in German and Swedish prepositional phrases. *Journal of Pragmatics* 44: 1413–1433.
- Bybee, J. 2006 From usage to grammar: the mind's response to repetition. *Language* 82: 711–733.
- Chomsky, N. 1965 *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The M.I.T. Press.
- Couper-Kuhlen, E. 2007 Prosodische Prospektion und Retrospektion im Gespräch. In: H. Hausendorf (ed.), *Gespräch als Prozess: Linguistische Aspekte der Zeitlichkeit verbaler Interaktion*, 69–94. Tübingen: Narr.
- Couper-Kuhlen, E. and C. Ford 2004 *Sound Patterns in Interaction: Cross-linguistic studies from conversation*. Amsterdam: John Benjamins.
- Couper-Kuhlen, E. and O. Tsuyoshi 2007 'Incrementing' in conversation. A comparison of practices in English, German and Japanese. *Pragmatics* 17: 513–552.
- Croft, W. 2001 *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Deppermann, A. 2006 Construction Grammar – Eine Grammatik für die Interaktion? In: A. Deppermann, R. Fiehler and T. Spranz-Fogasy (eds.), *Grammatik und Interaktion. Untersuchungen zum Zusammenhang zwischen grammatischen Strukturen und Interaktionsprozessen*, 43–65. Radolfzell: Verlag für Gesprächsforschung.
- Deppermann, A., R. Fiehler and T. Spranz-Fogasy (eds.) 2006 *Grammatik und Inter-*

- aktion. *Untersuchungen zum Zusammenhang zwischen grammatischen Strukturen und Interaktionsprozessen*. Radolfzell: Verlag für Gesprächsforschung.
- De Stefani, E. 2005 La suspension du geste comme ressource interactionnelle. In: L. Mondada and V. Markaki (eds.), *Interacting Bodies. Online proceedings of the 2d ISGS Conference*. [http://gesture-lyon2005.ens-lsh.fr/article.php3?id_article=259] (accessed April 10, 2011).
- Fillmore, C. J., P. Kay and C. O'Connor 1988 Regularity and idiomaticity in grammatical constructions. The case of let alone. *Language* 64: 501–538.
- Ford, C., B. Fox and S. Thompson 1996 Practices in the construction of turns: The 'TCU' revisited. *Pragmatics* 6: 427–454.
- Ford, C., B. Fox and S. Thompson 2002 *The Language of Turn and Sequence*. Oxford: Oxford University Press.
- Ford, C., S. Thompson and V. Drake 2012 Bodily-visual practices and turn continuation. *Discourse Processes* 49: 192–212.
- Fox, B. A., M. Hayashi and R. Jaspersen 1996 Resources and repair: A cross-linguistic study of the syntactic organization of repair. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 185–237. Cambridge: Cambridge University Press.
- Fox, B. A., Y. Maschler, and S. Uhmann 2009 A Cross-linguistic study of self-repair: Evidence from English, German, and Hebrew. *Gesprächsforschung – Online Zeitschrift zur verbalen Interaktion* 10: 245–291.
- Fricke, E. 2012 *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin: de Gruyter.
- Fried, M. and J.-O. Östman 2005 Construction Grammar and spoken language: The case of pragmatic particles. *Journal of Pragmatics* 37: 1752–1778.
- Gohl, C. 2006 *Das-Konstruktionen als Praktiken des Begründens*. In: S. Günthner and W. Imo (eds.), *Konstruktionen in der Interaktion*, 182–204. Berlin: de Gruyter.
- Goldberg, A. and R. Jackendoff 2004 The English resultative as a family of constructions. *Language* 80: 532–568.
- Gülich, E. and K. Lindemann 2010 Communicating emotion in doctor-patient interaction: a multidimensional single-case analysis. In: D. Barth-Weingarten, E. Reber and M. Selting (eds.), *Prosody in interaction*, 269–294. Amsterdam: John Benjamins.
- Günthner, S. and W. Imo 2006 *Konstruktionen in der Interaktion*. Berlin: de Gruyter.
- Günthner, S. and P. Hopper 2010 Zeitlichkeit und sprachliche Strukturen: Pseudoclefts im Englischen und Deutschen. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 11: 1–28.
- Günthner, S. 2011 Aspekte einer Theorie der gesprochenen Sprache – ein Plädoyer für eine praxisorientierte Grammatikbetrachtung, *Arbeitspapiere Reihe "Grammatik in der Interaktion" (GIDI)*, 32/2011.
- Gumperz, J. 1982 *Discourse Strategies*. Cambridge: Cambridge University Press.
- Gussenhoven, C. 2004 *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Hakulinen, A. and M. Selting 2005 *Syntax and Lexis in Conversation*. Amsterdam: John Benjamins.
- Hausendorf, H. and D. Wolf 1998 Erzählentwicklung und -didaktik. Kognitions- und interaktions-theoretische Perspektiven. *Der Deutschunterricht* 1: 38–52.
- Hopper, P. 1987 Emergent Grammar. *Berkeley Linguistics Society* 13: 139–157.

- Imo, W. 2011 Ad hoc-Produktion oder Konstruktion? – Verfestigungstendenzen bei Inkrement-Strukturen im gesprochenen Deutsch. *Arbeitspapiere Reihe "Grammatik in der Interaktion" (GIDI)*, 29/2011.
- Iwasaki, S. 2009 Initiating interactive turn spaces in Japanese conversation: Local projection and collaborative action. *Discourse Processes* 46: 226–246.
- Kehrein, R. 2002 *Prosodie und Emotionen*. Tübingen: Niemeyer.
- Kipp, M., J.-C. Martin, P. Paggio and D. Heylen 2009 *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications, Lecture Notes on Artificial Intelligence*. Berlin: Springer.
- Ladd, D. R. 2008 *Intonational Phonology*. Second Edition, Cambridge: Cambridge University Press.
- Langacker, R. W. 1987 *Foundations of Cognitive Grammar. Vol. 1, Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. 1991 *Foundations of Cognitive Grammar. Vol. 2, Descriptive Application*. Stanford: Stanford University Press.
- Langacker, R. W. 2001 Discourse in Cognitive Grammar. *Cognitive Linguistics* 12: 143–188.
- Langacker, R. W. 2008 *Cognitive grammar: a basic introduction*. Oxford: Oxford University Press.
- Laursen, L. 2005 Towards an embodied Grammar: Gesture in tying practices. Constructing obvious cohesion. In: L. Mondada and V. Markaki (eds.), *Interacting Bodies. Online proceedings of the 2d ISGS Conference*. [http://gesture-lyon2005.ens-lsh.fr/article.php3?id_article=259] (accessed December 10, 2010).
- Loehr, D. 2007 Aspects of rhythm and gesture in speech. *Gesture* 7: 179–214.
- Mondada, L. 2007 L'interprétation online par les co-participants de la structuration du tour in fieri en TCUs: évidences multimodales. *Travaux neuchâtelois de linguistique* 47: 7–38.
- Mondada, L. 2007 Multimodal resources for turn-taking: pointing and the emergence of possible next speaker. *Discourse Studies* 9: 194–225.
- Mondada, L. and V. Markaki (eds.) 2005 *Interacting Bodies. Online proceedings of the 2d ISGS Conference*. [http://gesture-lyon2005.ens-lyon.fr/article.php3?id_article=12] (accessed March 7, 2012).
- Müller, C. 1998 *Rede begleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berliner Wissenschaftsverlag.
- Olsher, D. 2004 Talk and gesture. The embodied completion of sequential actions in spoken interaction. R. Gardner and J. Wagner (eds.), *Second language conversations*, 221–245. London: Continuum.
- Oppenrieder, W. 1988 Intonatorische Kennzeichnung von Satzmodi. In: H. Altmann (ed.), *Intonationsforschungen*, 169–205. Tübingen: Niemeyer.
- Pfeiffer, M. 2010 Zur syntaktischen Struktur von Selbstreparaturen im Deutschen. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 11: 183–207.
- Pfeiffer, M. in press Formal vs. functional motivations for the structure of self-repair in German. In: B. MacWhinney, A. Malchukov and E. A. Moravcsik (eds.), *Competing motivations in grammar and cognition*. Oxford: Oxford University Press.
- Phoby, J. 1975 *Intonation und Grammatik im Deutschen*. Berlin: Akademischer Verlag.
- Reber, E. 2012 *Affectivity in Interaction. Sound objects in English*. Amsterdam: John Benjamins.

- Reber, E. and E. Couper-Kuhlen 2009 Interjektionen zwischen Lexikon und Vokalität: Lexem oder Lautobjekt? In: A. Deppermann and A. Linke (eds), *Sprache intermedial: Stimme und Schrift, Bild und Ton*, 69–96. Berlin: de Gruyter.
- Sacks, H., E. A. Schegloff and G. Jefferson 1974 A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735.
- Schegloff, E. A. 1984 On some gestures' relation to talk. In: J. Maxwell Atkinson and J. Heritage (eds.), *Structures of Social Action*, 266–298. Cambridge: Cambridge University Press.
- Selting, M. 1995 *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.
- Selting, M. 1996 Prosody as an activity-type distinctive cue in conversation: The case of so-called 'astonished' questions in repair initiation. In: E. Couper-Kuhlen and M. Selting (eds), *Prosody in Conversation: Interactional Studies*, 231–270. Cambridge: Cambridge University Press.
- Selting, M. 2000 The construction of units in conversational talk. *Language in Society* 29: 477–517.
- Selting, M. 2010 Prosody in Interaction: State of the art. In: D. Barth-Weingarten, E. Reber and M. Selting (eds), *Prosody in interaction*, 3–40. Amsterdam: John Benjamins.
- Sidnell, J. and T. Stivers (eds.) 2005 Multimodal Interaction. Special Issue of *Semiotica* 156.
- Stivers, T. 2008 Stance, alignment and affiliation during story telling: When nodding is a token of preliminary affiliation. *Research on Language in Social Interaction* 41: 29–55.
- Stivers, T. and J. Sidnell 2005 Introduction: Multimodal interaction. *Semiotica* 156: 1–20.
- Streeck, J. and U. Hartge 1992 Previews: Gestures at the Transition Place. In: P. Auer and A. Di Luzio (eds.) *The contextualization of language*, 135–158. Amsterdam: John Benjamins.
- Streeck, J. and W. Kallmeyer 2000 Interaction by inscription. *Journal of Pragmatics* 33: 465–490.
- Tummers, J., K. Heylen and D. Geeraerts 2005 Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1: 225–261.
- Välimaa-Blum, R. 2005 *Cognitive Phonology in Construction Grammar: Analytic Tools for Students of English*. Berlin: de Gruyter.
- Viscardi Martins, J. 2006 Prosódia e significação: considerações a partir da fala de um sujeito afásico. *Revista Virtual de Estudos da Linguagem – ReVEL* 4: 1–14.

I Prosody

linguae & litterae

Publications of the School of Language & Literature
Freiburg Institute for Advanced Studies

Edited by

Peter Auer · Gesa von Essen · Werner Frick

Editorial Board

Michel Espagne (Paris) · Marino Freschi (Rom)

Ekkehard König (Berlin)

Michael Lackner (Erlangen-Nürnberg)

Per Linell (Linköping) · Angelika Linke (Zürich)

Christine Maillard (Strasbourg) · Lorenza Mondada (Basel)

Pieter Muysken (Nijmegen) · Wolfgang Raible (Freiburg)

Monika Schmitz-Emans (Bochum)

18

De Gruyter

Prosody and Embodiment in Interactional Grammar

Edited by

Pia Bergmann, Jana Brenning,

Martin Pfeiffer and Elisabeth Reber

De Gruyter

ISBN 978-3-11-029504-7
e-ISBN 978-3-11-029510-8
ISSN 1869-7054

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

© 2012 Walter de Gruyter GmbH, Berlin/Boston
Druck: Hubert & Co. GmbH & Co. KG, Göttingen
∞ Gedruckt auf säurefreiem Papier
Printed in Germany
www.degruyter.com



Table of contents

<i>Pia Bergmann, Jana Brenning, Martin Pfeiffer, Elisabeth Reber</i> Towards an Interactional Grammar	1
--	---

I PROSODY

<i>Karin Birkner</i> Prosodic formats of relative clauses in spoken German	19
---	----

<i>Martin Pfeiffer</i> What prosody reveals about the speaker's cognition: Self-repair in German prepositional phrases	40
--	----

<i>Jana Brenning</i> Speakers' orientation to the nucleus accent in syntactic co-constructions	74
--	----

<i>Pia Bergmann</i> The prosodic design of parentheses in spontaneous speech	103
---	-----

<i>Beatrice Szczypek Reed</i> Prosody, syntax and action formation: Intonation phrases as »action components«	142
---	-----

II EMBODIMENT

<i>Lorenza Mondada</i> Deixis: an integrated interactional multimodal analysis	173
---	-----

<i>Florence Oloff</i> Withdrawal from turns in overlap and participation	207
---	-----

<i>Ina Hörmeyer</i> The importance of gaze in the constitution of units in Augmentative and Alternative Communication (AAC)	237
---	-----