



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 Szczepek 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
---	-----



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

Keywords 315





Pia Bergmann / Jana Brenning / Martin Pfeiffer / Elisabeth Reber

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 (co)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).



talts, 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 Jasperson (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 SZCZEPEK 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. Jaspersion 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. Uhmman 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 *Dass*-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, *Arbeitspapierreihe "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. *Arbeitspapierreihe "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 *Redebegleitende 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.
- Pheby, J. 1975 *Intonation und Grammatik im Deutschen*. Berlin: Akademischer Verlag.
- Reber, E. 2012 *Affectivity in Interaction. Sound objects in English*. Amsterdam: John Benjamins.





16 Pia Bergmann / Jana Brenning / Martin Pfeiffer / Elisabeth Reber

- 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







Karin Birkner
University of Bayreuth

Prosodic formats of relative clauses in spoken German

The prosodic design of a grammatical structure is normally not part of the description in grammar books. Relative clauses are an exception, however, as they are usually described as having two prosodic formats which differentiate restrictive and appositive relative clauses in spoken language. Restrictive relative clauses are assumed to be prosodically integrated into the matrix clause. Accordingly, the matrix clause and the relative clause form a single intonation phrase with one primary accent in the relative clause. Appositive (i.e., non-restrictive) relative clauses, on the other hand, are presented as forming a separate intonation phrase, with the matrix clause and the appositive relative clause each having their own primary accent (cf. also Becker 1978; Brandt 1990; Duden 2005; Eisenberg 1999; Fritsch 1990; Frosch 1996; Helbig and Buscha 2001; Hentschel and Weydt 2003; Holler 2005; Lehmann 1984, 1995; Weinrich 2005; Zifonun 2001; Zifonun et al. 1997; Greenbaum and Quirk 1992).

The semantic difference between restriction and apposition is mentioned by most grammarians, although the terminology often varies. It is generally understood that restrictive relative clauses modify the extension of the reference nominal¹ via referential restriction, whereas appositive relative clauses deliver additional information without restricting the extension of the reference nominal (Duden 1998; Eisenberg 1999; Helbig and Buscha 2001; Hentschel and Weydt 2003; Lehmann 1984; Weinrich 2005; Zifonun 2001; Zifonun et al. 1997; Holler 2007; Blühdorn 2007). This difference is rarely marked at the lexical level, thus ambiguous cases are common in which both a restrictive as well as an appositive reading is possible; the interpretation should then be clear from the respective context.

These assumptions regarding semantics, and especially those regarding prosodic design, need to be tested by a corpus-based approach to allow for an examination of these elements in authentic language use. In a study using a corpus of spoken German which comprised over 1,000 relative clauses, Birkner (2008) concluded that the semantic differentiation between restrict-

¹ I use *reference nominal* (RN) as the translation for German *Bezugsnominal*. The most commonly used term in English is “head-NP” which I avoid because of its theoretical implications in generative grammar approaches.



ive and appositive relative clauses based on extensional restriction alone is difficult, even though the corpus-based data provided the necessary context for disambiguation. Birkner also noted that the prosodic form of relative clause structures in spoken language is more heterogeneous than it is presented in grammar books, complying neither with dichotomous semantics nor with the two postulated prosodic formats (2008: 182ff.).

This article builds on Birkner's (2008) study by presenting findings about the correlation between the semantic features and the prosodic phrasing of relative clauses in spoken German. The corpus on which the present study is based is comprised of the following two corpora:

Table 1: Corpus

	Duration
First season of the reality series <i>Big Brother</i> (aired in 2000 on RTL 2)	22 h 40 min
Job interviews with college graduates for a trainee program at a bank (recorded 1995–1996)	10 h 29 min
<i>Total</i>	<i>33 h 09 min</i>

The language data from *Big Brother* is informal, including different types of discourse, such as mealtime conversations, arguments, and discussions, and it contains multi-party talk as well as one-to-one conversations and monologues. The job interviews, on the other hand, are homogenous, being primarily one-to-one conversations between applicant and interviewer.

The data analysis was carried out in several steps. Since the relative connector is obligatory in German, the relative clauses were collected by searching the transcripts for relative connectors.² Each individual example and its immediate context were copied as a text from the transcription files and as a sound from the audio/video files, which enabled analyses from syntactic, semantic and prosodic perspectives, taking into account the interactional embedding.

For the present study, only adjacent adnominal relative clauses were considered, and non-adjacent and free relative clauses were omitted.³ The study's total corpus includes 801 examples.

² The following forms were included: *das* (det/dat/des), *dem*, *den*, *der*, *deren*, *dessen*, *die*, *was* (wat), *welch*, *wem*, *wen*, *warum*, *weshalb*, *weswegen*, *wie*, *wo*, *wo(r)*+preposition.

³ Adjacent relative clauses directly follow the noun they modify. Since the analysis focuses on the prosodic features of the gap between noun and relative clause this is a prerequisite for the study. This is also the rationale for excluding free relative clauses: due to the missing nominal it is impossible to consider the junction between relative clause and nominal.





This chapter is organized as follows. In Section 1, the theoretical view of the semantic difference between appositive and restrictive relative clauses is explained (1.1), and the first presentation of results of this corpus-based study are introduced (1.2). Section 2 focuses on the prosody of relative clauses; after introducing what grammar books assume about the prosody of relative clauses (2.1), the results of the empirical analysis will be presented (2.2). In Section 3, the results of the semantic-prosodic analysis of the data are brought together. The article concludes in Section 4 with a summary of the findings on the prosodic design of relative clauses according to semantic type in spoken German.

1. Semantics of relative clauses

1.1. Assumptions about the semantic distinction between restrictive and appositive relative clauses

The semantic differentiation of the two types of relative clauses is based on the criterion of referential restriction and, hence, on an extensional notion of reference (Frawley 1992: 19; Löbner 2003: 354ff.; Blühdorn 2007). The restrictive relative clause causes an extensional limitation of the scope of reference of the reference nominal, while the appositive relative clause supplies additional information. Lehmann (1984) differentiates between restrictive and appositive relative clauses as follows: “The restriction operates on the basis of a given term by creating a new term with greater intension and less extension” (Lehmann 1984: 261, translation K.B.). He describes this operation with the example *Wir kennen einen Arzt, dem Anna vertraut* [We know a doctor whom Anna trusts], thus “The term ›doctor‹ is used as a basis. Its extension is limited to doctors that are characterized by their being in some way involved in another circumstance” (Lehmann 1995: 1200, translation K.B.). In the case of the appositive relative clause, the reference nominal is already sufficiently determined, and its reference is identifiable. As a result, the appositive relative clause does not contribute to the identification of the referent (Zifonun et al. 1997: 563), but delivers additional or background information (Lehmann 1984: 261ff.).

Quirk et al. (1972: 858) demonstrate this difference using the following two examples:

- i) The girl who stood in the corner is Mary Smith.
- ii) Mary Smith, who is in the corner, wants to meet you.



The relative clause *who stood in the corner* provides the necessary information to identify the girl in question in the first sentence. In the second sentence, the antecedent is a proper name, *Mary Smith*, which already ensures the identification of the person, so here, the relative clause is a non-restrictive post-modification providing additional information.

The establishment of reference is considered the prototypical function of relative clauses; it is assumed that relative clauses are part of referential acts in which the relative clause performs a set-theoretical operation on the reference nominal. The appositive relative clause is defined in relation to the restrictive clause (which is also reflected in the common use of the term *non-restrictive* for appositive), generally representing the marked type of relative clause formation.

Many researchers explicitly point out the difficulty of distinguishing between restrictive and appositive relative clauses (cf. e.g. Bache and Jakobson 1980: 243; Becker 1978: 1; Eisenberg 1999: 266; Eissenhauer 1999: 61; Lehmann 1984: 262f.; Tao and McCarthy 2001: 654; Weinrich 1993: 773). Context, world knowledge, and prosody can help to disambiguate them; the latter, in particular, is thought to play a central role. Birkner's (2008) corpus-based study showed, that the distinction between restriction and apposition is not always clear (despite using common tests, cf. Birkner 2008: 38ff., 111) and it concludes that most relative clauses are potentially semantically ambiguous. One reason for the ambiguity is the fact that restriction/apposition is not marked at the lexical level.

1.1 Empirical findings

In the following, we will see how the heterogeneity of restrictive relative clause structures, in addition to ambiguity, represents a problem for semantic identification. Several common relative clause structures will be presented from the corpus analysis.

1.2.1. Appositive relative clause structures

Let us look at a typical appositive relative clause, delivering additional information which does not influence the scope of reference of the nominal. (Transcripts follow GAT 2 conventions according to Selting et al. 2009. The symbols are listed in the appendix. The reference nominal is given in italics, and the relative clauses are in bold.)



Example (1) BB01–7414

- 01 Jrg: in äh malLoRca (--)
in uh Mallorca
02 bauen sich millioNÄre?
millionaires build themselves
03 (-)
04 **die gÄr nicht auf kosten des (.) dieses staates da**
LEben?
who are not even living there at the this State's expense
05 (-)
06 VILlen?
villas
07 (0.6)
08 und DIE werden da schon nicht gedULdet.
and even they are not tolerated there
09 Alx: ja
yes

The sequence stems from a discussion on racism. Jürgen (Jrg) uses the example of Mallorca to emphasize that xenophobia is not only used against potential “welfare freeloaders,” but also against well-off foreigners. The plural noun *millioNÄre* (l. 02) is a reference nominal which could be followed by either a restrictive or an appositive relative clause. The following relative clause *die gÄr nicht auf kosten des (.) dieses staates da LEben?* (who are not even living there at the this State’s expense) (l. 04) is an appositive relative clause because it characterizes the reference group of millionaires as a whole with its additional information and does not – like a restrictive relative clause would do – designate a subset of that group. The semantic independence of the two syntagmas can be proved via the main clause test, in which the subordinate structure can be transformed into two main clauses: *Millionaires build themselves villas in Mallorca. They are not living there at the State’s expense.*

1.2.2. Restrictive relative clause structures

Prototypical restrictive relative clauses establishing an extensional limitation are also found in the corpus. The following example, in which the relative clause limits the denotatum of the nominal *be we El studenten* (business students) to a subset, illustrates this type.



24

Karin Birkner

Example (2) BANK2–2294

((The interviewer in a job interview for a traineeship in a bank explains the job market condition.))

- 01 I: nun gibt es natürlich auch be we El (-) studenten (-) in
MASsen, (-)
now naturally there are tons of business students
- 02 eh nicht nur HIER;
uh not only here
- 03 sondern auch .h auch in den ALten bundesländern, (-)
but also also in the old West German states
- 04 eh sie können mir GLAUben.
uh you can believe me
- 05 es werden ja nIcht nur (-) eh: be we EL studenten
abgelehnt, (--)
not only uh business students are rejected
- 06 **die: eh (-) so wie SIE jetzt (.) n_ganz normales STUdium
gemacht [haben,]**
who uh like you now have done a totally normal study program
- 07 B: [<p>mhm;>]
- 08 I: sondern es werden auch be we el studenten ABgelehnt **die
vorher schon ne BANKlehre gemacht haben?**
but also business students are being rejected who have already done a bank apprenticeship



The Interviewer (I) makes a restriction here by first delimiting the group of rejected business students to those who have finished a study program, and then further to a subset of those who have also completed a bank apprenticeship. For this purpose he uses two relative clauses: *die: eh (-) so wie SIE jetzt (.) n_ganz normales STUdium gemacht haben* (l. 06) and *die vorher schon ne BANKlehre gemacht haben?* (l. 08).



These relative clauses are essential in order to accommodate the main clause proposition (cf. Blühdorn 2007). In other words, it is an act of reference in which the establishment of reference is implemented by means of a restrictive relative clause. Therefore, the function of this type of restrictive relative clause structure can be described as “identifying”.

1.2.3. Existence and presentative constructions

The next example is an existence construction. Similar to predicate nominative constructions, they are generally considered to be restrictive (Lehmann 1984: 266).





Example (3) BANK3–2424

((During a job interview, the interviewer comments on the housing situation in town.))

- 01 B: und anSONsten, (-)
and other than that
- 02 MAGdeburg,
Magdeburg
- 03 WOHNungssituation, ((räuspern)) (-)
apartment situation
- 04 I2: es GIBT (.) ne_ganz: (-)
there's a really
- 05 es gibt eine reihe von (.) NEUen wohnungen in magdeburg,
 (-)
there are a bunch of new apartments in Magdeburg
- 06 .hh mIEtniveau- (-) zwischen; (-) zwölf und achtzehn
 MARK <<p>pro quadratmeter,> (-)
price range between twelve and eighteen Marks per square meter
- 07 je nach LAge? (1.0)
depending on location
- 08 'h und eh (.) es gibt eine (-) ganze reihe an (-)
 saNIerten (.) WOHNungen? (-)
and uh there is a whole bunch of renovated apartments
- 09 **die sie so zwischen zehn und (--) vierzehn fuffzehn MARK**
[krie]gen.=
that you get for like between ten and fourteen fifteen Marks
- 10 B: [mhm]

In Example (3), the reference nominal *WOHNungen* used in the existence construction *es gibt* ... (l. 08) features the attribute *saNIerten*, which in turn is expanded with a complex quantifier *eine ganze reihe an*. Here, it is mainly the context which clarifies that the speaker is comparing new apartments and old apartments (i.e., renovated apartments) which are characterized by their respective prices. This is also reinforced by the *NEUen wohnungen* (l. 05) being attributed in the form of an apposition: *mietniVEAU- (-) zwischen; (-) zwölf und achtzehn MARK, <<p>pro quaDRATmeter.> je nach LAge?* (l. 06). In this example, a relative clause delivers additional, characterizing information, but does not delimit the scope of reference of the nominal. The semantic relationship is not based on the extensional limitation but on the intensional adding of features, and the function of the relative clause is not identifying, but rather descriptive.

Lehmann (1984) has already pointed out that a restrictive reading of existence and predicate nominal constructions is only possible in the case of a fully undetermined reference nominal (Lehmann 1984: 26). This can be confirmed or substantiated using these examples. If a reference nominal in an existence or predicate nominal construction has already been restrictively





delimited by other (non-relative-clause) attributes, an appositive relative clause is also possible. This is illustrated by the main clause test with the following example: *There are a whole bunch of renovated apartments. You get them for like between ten and fourteen fifteen Marks.* If the prenominal attribute (a whole bunch of renovated) is removed, the appositive reading of the remaining expression (there are apartments) does not make sense.

This is even more evident in the so-called “*Mensch*-construction” (cf. Birkner 2006a) that consists of a predicate nominal structure with the copula *sein* (to be) and a personal mass noun as well as a connecting attributive clause. Predicate nominal constructions with two full nouns as described by Lehmann (1984: 266) with the example *Herr Müller ist der Kandidat, der die besten Aussichten hat, gewählt zu werden* (Mr. Müller is the candidate who has the best chances of being elected) are notably rare in this corpus. Much more common are *pronoun – copula – full form + relative clause* structures. In the *Mensch*-construction, the full form consists of an unspecific term for humans (e.g., *Mensch* (person) or *Typ* (bloke)) which constitutes the first syntagma (the predicate nominal construction) and projects the personal attribution in the following relative clause. The second syntagma – normally considered the subordinate clause – provides the main predication of the construction. It predominantly draws on ways of acting, using an action-based typification for the positioning.

Example (4) BB84 509

((John and Andrea talk about John’s problems with his in-laws.))

- 01 Adr: <<all>na ich denk mal es hat de:nen’> (.)
 I think it was for them
- 02 DU hast denen dann erst mal GA:: nich da in den KRAM
 gepasst.>
 you didn’t fit in with them at all at first
- 03 Jhn: =nee überHAUPT nich. (--)
 no not at all
- 04 aber ick’ (.)
 but I
- 05 KENNST mich ja.
 you know me
- 06 <<all>ick bin ja dann o:_so_n tYp-> (0.5) .h der sich
 da ooch <<rhyth>kEin stÜck NACHgeben> will.
 I am the kind of bloke who won’t back down one bit
- 07 =wee [ßte?
 you know
- 08 Adr: [<<zustimmend>nee.>
 ((affirmative)) no



Like predicate nominal constructions in general, this example does not make a reference but rather a predication; the relative clause provides intensional-descriptive information. Unlike the first, the second syntagma cannot be left out; it is even projected by the cataphorical *so* (‘kind of’, l. 06).⁴

The reference nominal *typ* (‘bloke’) is lexically empty here; as a nominal placeholder it provides a link for a relative clause that supplies the descriptive meaning.⁵ Lehmann (1984: 293f.) mentions “relative clauses without a nucleus” which are similar to those described above, as the following example clearly shows: *Der Typ, den wir vorgestern bei Ede trafen, ist Botschafter* (‘The bloke whom we met the day before yesterday at Ede’s is an ambassador’) (Lehmann 1984: 294). He explicates that this kind of relative clause is “clearly identifying, i.e. in any case not appositive. Indeed, there is no reason to not call them restrictive [...] if one understands restriction in the wider sense as term specification and does not insist that the operation be applied to an already established term” (Lehmann 1984: 294; translation K.B.). Lehmann acknowledges that within restrictive relative clauses, this type – in which the reference nominal is specified without being extensionally delimited in the narrower sense – does exist.⁶

Cases such as those found in Examples (3) through (5) belong to a network of “presentative constructions.” Among other things, they use a typical verb in the matrix clause (e.g., often the copula *sein* (‘to be’), but also *es gibt*

⁴ The fact that the information-structural weight of the relative clause is quite heavy in the second syntagma often results in syntactic non-subordination (Birkner 2006a; cf. also Weinert 2004; Ravetto 2006). In German, subordination is marked by verb-final position of the finite verb, whereas in main clauses, the finite verb is in second-position. This kind of relative clause constructions often show verb-second-position, hence loosing the syntactic feature of subordination.

⁵ This is an important potential which is often overlooked in the discussion of relative clauses. Relative clause structures can be formed out of complex nominal phrases for which the language has no established nominal equivalent. A good example of this is made by an applicant in the bank corpus: *auf Lange sieht gesehen .b (-) eb: (-) bin ich: auch NICHT der typ der jedem kunden die DOPpelkarte; =damit er sein AUto anmelden kann; eb (-) kilometerweise hinterHERträgt* (‘In the long run, uh, I am not the sort of bloke who chases after every customer for many kilometers with their insurance ID card so that he can register his car.’). There is no noun in German whose scope of reference coincides exactly with these features.

⁶ As for the semantic identification of restrictive or non-restrictive relative clauses, Brandt comes to a similar conclusion: “Thus, these examples show that the common semantic denominator of restrictive relative clauses cannot be described as a limitation of the term that is described by the relative clause proposition, but rather must be sought elsewhere” (Brandt 1990: 42; translation K.B.).



(*there is/are*), *haben* (*to have*) or *kommen* (*to come*); cf. also Brandt 1990: 42f.) to present a referent in the rheme position which is then further specified with the relative clause. The category of presentative constructions subsumes various sub-constructions that share the feature of introducing a referent and subsequently specifying it by a relative clause, but which also differ considerably in their idiosyncratic features (cf. Birkner 2008: 389ff., Ch. 9).

1.2.4. Superlatives and relative clauses

Another common type of construction in the corpus is the superlative that connects a relative clause containing *es gibt* (*there is/are*). Here, too, we have a semantic type in which the narrow understanding of restriction as an extensional limitation does not seem appropriate. The following example illustrates this type:

Example (5) BANK4 4965

((The interviewer informs the applicant about the remuneration.))

- 01 I: 'h <<betont>tE gE SIEben? (-)
TG seven
02 taRIFgruppe soll das heißen? (-)
wage group is that supposed to be
03 te ge SIEben, (.)
TG seven
04 ELftes berufsjahr? (-)
eleventh year of work
05 **das is das: HÖCHste, (-)**
that is the highest
06 **was es GIBT?**
that there is'
07 (0.5)
08 B: mhm,
mbm
09 I1: 'h und das sind fünftausendzweihundertdreizehn DE mark.
and that's five thousand two hundred thirteen German Marks

The semantic relation between the reference nominal *das: HÖCHste* (*the highest*, l. 05) and the relative clause cannot be recognized as extensionally limiting (like e.g., by defining the highest thing that there is against one which does not exist). However, there is no appositive relationship, either. Even though the relative clause may be omitted without changing the scope of reference of the nominal, it does not provide an additional characterization (e.g., there is no semantic independence that could be tested by composing a main clause or by inserting *bekanntlich* (*as is well known*)). The relative





clause appears to be semantically redundant; in other words, it has a pragmatically motivated intensification function.

1.2.5. Possessive relative clauses

The final example presents a case in which an object relative clause with *haben* (to have) provides an anchor for a reference nominal (cf. also Birkner 2006b).

Example (6) BB01–5811

((Alex and John prepare to spend the night outdoors. John explains why he is wearing a heavy jacket.))

- 01 Jhn: ja weil ick kurz äh (2.0) äh (-) .h keene lange HOse
habe;
yes because I have short eh no long pants
- 02 Alx: ach so;
I see
- 03 Jhn: ick hab det nur UNTen rum. .h (--)
I only have them on the bottom
- 04 also ick hab SCHON ne LANGE, .h (---)
well I do have long ones
- 05 aber damit könnt ick überHAUPT ni schlafen.
but I couldn't sleep at all in them
- 06 weil die alle relativ ENG sind **die ick habe**. (2.0)
because the ones I have are relatively tight
- 07 Alx: die schUhe has_AU no an?
you are still wearing shoes too

John explains here that he is not wearing pants in his sleeping bag because he has only tight ones that are too uncomfortable to sleep in. Here, describing the relative clause *die ick habe* (the ones I have, l. 06) as an extensional limitation is clearly inadequate. An extensional limitation in the sense of pants which the speaker owns vs. pants that he does not own is not intended. It neither establishes a reference by means of a subset relation, nor does it provide an additional proposition which could be transformed into a self-contained main clause. The semantic relation which the relative clause establishes here is most accurately described as a possessive and temporal marking. It corresponds to the relative clause function described by Fox and Thompson (1990: 300) as “grounding”, which reflects the discursive necessity of anchoring a referent to, e.g., the speaker.





1.3. Summary

In sum, this analysis has shown that the definition of appositive relative clauses as deliverers of additional information is unproblematic. However, the so-called restrictive relative clauses, on the other hand, are strikingly heterogeneous, and the popular notion of restriction as an extensional limitation of the scope of reference of the reference nominal is adequate for only part of the data. Nevertheless, the examples of the latter type share the feature of further identifying the referent, either through extensional limitation or intensional enrichment. These cases will be labeled *non-appositive*, since they are easily distinguishable from appositive cases. Of the 801 relative clause structures in this corpus, this group comprises the majority with 670 tokens (84%), while the appositives are clearly in the minority with only 131 examples (16%). The following table illustrates the distribution.

Table 2: Distribution of relative clauses according to semantic type

Appositive	Non-appositive	Total
131 (16%)	670 (84%)	801 (100%)

We will now turn to the prosodic feature of relative clause constructions. Since Seiler (1960), a prosodic correlation has been associated with the above-described semantic difference. It has been assumed that appositive relative clauses occur in a parenthetical relation to the matrix structure, while restrictive relative clauses exhibit a clitic relation.

2. Prosodic features of relative clauses

2.1. Assumptions about the prosody of relative clauses

The claim of a systematic link between semantics and prosodic design is widespread (Bache and Jakobson 1980; Becker 1978; Brandt 1990; Buscha and Kempter 1983; Ebert 1973; Eisenberg 1999; Eissenhauer 1999; Fritsch 1990; Frosch 1996; Helbig and Buscha 2001; Lehmann 1984; Motsch 1965; Quirk et al. 1992; Zifonun 2001; Zifonun et al. 1997). It is assumed that appositive relative clauses possess accents both in the relative clause and on the reference nominal, and that the matrix construction and the relative clause constitute an independent prosodic unit. Restrictive relative clauses and their reference nominals, on the other hand, share one intonation unit, with the accent placed in the relative clause. In the typical appositive relative clause, an “intonation break” occurs between the reference nominal and the relative clause, while in restrictive relative clauses the progression is continuous



(Lehmann 1984: 263; 1995: 1204). Holler (2005: 28) assumes that the accent is obligatory in the appositive relative clause.

Quirk et al. report that in English, too, “restrictive modification tends to be given more prosodic emphasis than the head [the reference nominal, respectively, K.B.]; non-restrictive modification, on the other hand, tends to be unstressed in pre-head position, while in post-head position, its ›parenthetical‹ relation is endorsed by being given a separate tone unit [...], or – in writing – by being enclosed by a comma” (Quirk et al. 1992: 365). The independence of appositive relative clauses from the surrounding syntactic structure is often underlined by pauses, according to Becker (1978): “A relative clause can be unambiguous due to the intonation: appositive relative clauses are marked via a pause between the reference NP and the relative clause” (Becker 1978: 11; translation K.B.). Moreover, several authors refer to the role of stressed determiners in the reference nominal. The accentuation of determiners can accompany a restrictive relative clause or disambiguate the semantics as being restrictive (cf. e.g. Brandt 1990: 40; Duden 2005: 301; Eisenberg 1999: 482; Féry 1994: 103f.; Frosch 1996: 8; Lehmann 1995: 1205; Weinrich 1993: 784; Zifonun et al. 1997: 228). However, these assumptions of a direct correlation between prosodic design and semantics are not based on empirical studies. In the following we will present the prosodic analysis of the corpus in order to examine the prosodic-semantic relation in spoken language (cf. Birkner 2008).

2.2. Empirical findings

The features discussed in the above-cited studies for relative clauses are in accordance with the relevant features marking intonation phrases in spoken German (Birkner 2008: 84ff.). Appositive relative clauses represent a separate intonation phrase, while restrictive clauses form a phrasal unit together with their reference nominal. The central parameters for prosodic design are accent distribution and intonation phrasing, which are determined by the features of phrase-final pitch and boundary signals (e.g., pauses). In order to study the prosodic design of relative clauses, all examples in this corpus were analyzed prosodically. The analysis was conducted primarily aurally, though ambiguous cases underwent additional analysis with the acoustic analysis program PRAAT (cf. Boersma and Weenink 2006). Each example was examined with regard to primary accent and whether phrasal features could be detected at the junction between reference nominal and relative clause.

Results demonstrated that the degree of phrasing can be strong or weak (for a more in-depth discussion, cf. Birkner 2008: 131ff.). The analysis



yielded seven feature bundles (so called “formats”) of prosodic design in relative clauses in spoken German:

Format 1: Two intonation phrases and two accents: on the reference nominal and in the relative clause (distinct phrasing)

Format 1 includes two intonation phrases and two accents (on the reference nominal and in the relative clause), and the junction between reference nominal and relative clause features a clear phrasal boundary with boundary tones and a pause. This format corresponds to the assumed typical arrangement of appositive relative clauses.

Format 2: Two intonation phrases and two accents: on the reference nominal and in the relative clause (no distinct phrasing)

Format 2 also consists of two intonation phrases with two accents, but it is phrased less clearly than Format 1 (e.g., because there is no pause at the junction point).

Format 3: Two intonation phrases, accent in the relative clause, no accent on the reference nominal

In Format 3, there are two intonation phrases with an accent in each syntagma. The accent in the matrix structure is not on the reference nominal, however, but on another part of the syntagma.

Format 4: One intonation phrase and two accents on the reference nominal and in the relative clause

Format 4 differs from the previous formats in that only one intonation phrase exists, with one accent each on the reference noun and in the relative clause, together representing one intonation phrase.

Format 5: One intonation phrase and one main accent on the reference nominal

Format 5 consists of one intonation phrase with the main accent on the reference nominal. The relative clause is not (or only weakly) accentuated, and the prosodic integration can be strengthened (e.g., by latching).

Format 6: One intonation phrase and one main accent in the relative clause

Format 6 is distinguished by the reference noun and the relative clause comprising one intonation phrase in which the main accent lies in the relative clause. This format corresponds to the prototypical format of restrictive clauses assumed in the literature.

Format 7: One or two intonation phrases, with an accent neither on the reference nominal nor in the relative clause

Format 7 is a category of miscellaneous examples in which neither the reference nominal nor the relative clause are accented.

Table 3 summarizes the prosodic features of the seven formats.





Table 3: Prosodic formats of relative clause structures (parentheses indicate that the feature is facultative) (RN = reference nominal, RC = relative clause)

Format	Features	Phrase boundary	Prim. acc on RN	Prim. acc in RC	Pause
1	2 IP/2 acc (distinct)	+	+	+	+
2	2 IP/2 acc	+	+	+	–
3	2 IP/no acc on RN/acc in RC	+	–	+	(+)
4	1 IP/2 acc	–	+	+	(+)
5	1 IP/1 acc on RN	–	+	–	–
6	1 IP/1 acc in RC	–	–	+	(+)
7	1/2 IP/no acc on RN or RC	–/+	–	–	–

In a further step, the number of examples for each of the seven formats was determined in the relative clause corpus. Table 4 shows the quantitative distribution.

Table 4: Distribution of relative clause across prosodic formats

Format	Features	Percent	(number)
1	2 IP/2 acc (distinct)	26 %	(211)
2	2 IP/2 acc	13 %	(104)
3	2 IP/no acc on RN/acc in RC	3 %	(27)
4	1 IP/2 acc	23 %	(183)
5	1 IP/1 acc on RN	6 %	(45)
6	1 IP/1 acc in RC	27 %	(214)
7	1/2 IP/no acc on RN or RC	2 %	(17)
<i>Total</i>		<i>100 %</i>	<i>(801)</i>

Formats 1 and 6, which are considered in the literature to be the prototypical formats of appositive and restrictive relative clauses, respectively, appear with the same frequency, with each representing approximately one quarter of the total examples. Format 2, which differs from Format 1 only by a less distinct phrasing, constitutes 13 % of the examples. Format 4, with 23 % frequency, is roughly as common as Formats 1 and 6. Formats 3, 5 and the miscellaneous category 7 are only very sparsely represented.

When these results are compared with those of the semantic study, as presented in Table 2, it immediately becomes obvious through examining



the quantitative distribution that there is no simple correlation between semantics and prosody. In the next section, this will be explored in more detail.

3. The correlation between prosody and semantics

In order to explain the correlation between prosody and semantics, the results of the prosodic analysis presented in Table 4 will be merged with those of the semantic type study in Table 2. The results are shown in Table 5.

Table 5: The semantics and prosody of relative clause structures

Format	Appositive	Non-appositive
1	64 % (84)	19 % (127)
2	18 % (23)	12 % (81)
3	3 % (4)	3 % (23)
4	12 % (16)	25 % (167)
5	1 % (1)	7 % (44)
6	2 % (3)	31 % (211)
7	0 % (0)	3 % (17)
<i>Total</i>	<i>100 % (131)</i>	<i>100 % (670)</i>

Let us first examine the appositive relative clause structures. Sixty-four percent of these examples occur in Format 1. This, together with Format 2 (18%), results in a total percentage of 82%. In addition, a considerable number of examples is found in Format 4 (12%). Although this format does not include two intonation phrases, both the reference nominal and the relative clause are accented. Therefore, Holler's (2005: 28) assumption that the accent is obligatory in appositive relative clauses is confirmed.

In the case of non-appositive relative clause structures, the picture is altogether more heterogeneous. Although Format 6, the assumed prosodic design of restrictive relative clauses, is the most frequent one, it makes up only less than a third (31 %) of all the examples. Format 4 – two accents within one intonation phrase – is the second most frequent format at 25%. Formats 1 and 2, considered the typical formats of appositive relative clause structures, are also rather frequent (with 19% and 12%, respectively).

The important feature of the formats presented here and a central criterion for the prosodic design of relative clauses is the packaging of the relative clause and antecedent into two or one intonation phrase(s). This difference results in prosodic integration or disintegration of the syntagmas. In the following, therefore, Formats 1, 2, and 3 will be summarized as formats of prosodic disintegration, and Formats 4, 5, and 6 will be grouped together



as prosodic integration. Miscellaneous Format 7 and its small number of examples will be disregarded. Table 6 illustrates which percentages appositive and non-appositive examples represent for each of the formats of prosodic disintegration (Formats 1–3) and disintegration (Formats 4–6).

Table 6: Semantics in prosodic (dis-)integration formats

	Total	Appositive	Non-appositive
Formats 1, 2, 3 (prosodic disintegration)	42% (342)	32% (111)	68% (231)
Formats 4, 5, 6 (prosodic integration)	56% (442)	5% (20)	95% (422)

The results show that 42% of the total relative clauses take prosodic disintegration formats, while the prosodic integration formats make up the somewhat larger group with 56%. Analyzing the examples with regard to their semantics, results in the following: Only 32% of all examples of prosodic disintegration – the expected prosodic design in cases of apposition – actually contain appositive relative clauses, while 68% of prosodically-disintegrated examples are non-appositive. Even in cases of prosodic integration, appositive relative clauses constitute only 5%. However, 95% of all prosodically integrated examples are non-appositive. The link between semantics and prosody thus turns out to be much more complex than the literature would suggest. A systematic semantic disambiguation according to prosodic design cannot be confirmed.

The prosodic analysis of the corpus shows that, although appositive relative clauses occur more frequently in prosodically-disintegrated formats, restrictive relative clauses can also exhibit this format in a number of cases (68%). Presumably this is due to the fact that the prosodic design is affected by other factors, such as information structure and interactional features (Birkner 2008: 182ff).

From the recipient's point of view, determining the apposition of relative structures based on their prosodically-disintegrated design is not as promising an approach as it is assumed to be in the literature. For integration formats, in 95% of the examples, the prosodic integration is associated with non-appositive semantics. Nevertheless, a recipient cannot reliably deduce the restriction of a relative clause solely based on its prosodically-integrated format.





4. Conclusions

In the literature, it is generally assumed that relative clauses can be divided into two classes: appositive versus restrictive. This distinction is based on the criterion of referential restriction and, hence, on an extensional notion of reference. Prosodic design has always been considered an important means of determining semantic disambiguity between the two types of relative clauses; restrictive relative clauses have been assumed to be prosodically integrated, whereas appositive ones have been characterized by prosodic disintegration. Although this assumption is common in the literature, this empirical analysis of relative clauses in spoken German has shown as regards the semantic distinction that the generally accepted dichotomous distinction between appositive and restrictive relative clauses is oversimplified in many ways. While appositive relative clauses can be easily defined as deliverers of additional informational units and recognized (e.g., by means of the main clause test), the group of restrictive relative clauses is strikingly heterogeneous. For this reason, I suggest that the two groups be labeled as appositive and non-appositive relative clauses.

Further, it was shown that approaches based on an extensional definition cannot sufficiently explain many of the examples. The assumption that in restrictive relative clause structures, the relation between the reference nominal and relative clause represents an extensional limitation, applies to only a part of the data. Rather, analysis clearly shows that the modification of the relative clause often does not affect the extensional, but the intensional description. The reference-based definition of restriction implies that relative clause structures actually have a referential function. The corpus analysis shows, however, that in a high number of examples the matrix structures of relative clauses are predicative. Hence, instead of a referential restriction, an intensional description is made.

The correlation between the semantics of relative constructions and prosodic design has thus proved to be much more complex than the literature suggests. Therefore, a systematic and prospective semantic disambiguation based on prosodic design cannot be confirmed. The prosodic analysis of the corpus shows, for example, that although appositive relative clauses are more likely to occur in formats of prosodic disintegration, restrictive relative clauses also occur in this format in a great number of cases (cf. Table 5 and 6).

The reason behind this is presumably that the prosodic design is also affected by other factors in conjunction with the informational and interactional structure (Birkner 2008: 182ff.). Analyzing the examples within the categories in more detail, there are several starting points which help to ex-



plain the various prosodic formats. With regard to the arrangement in prosodic disintegration formats, factors such as stylization and emphasis may be relevant. Also, thematic progression has a significant effect on relative clause accentuation patterns. Relative clauses that, for example, show a semantic redundancy (like object relative clause with *haben*) are also characterized by deaccentuation. They are neither semantically independent like appositive relative clauses, nor do they have a restrictive effect on the reference nominal. Moreover, these relative clauses seem to be semantic postscripts that add given information. This semantic “redundancy” is also reflected in the impression of prosodic reduction.

Future research could explore the factors underlying the prosodic variation in non-appositive relative clauses in spoken language. Another interesting element for investigation is the relation between syntactic subordination and prosodic design. For this, the study of relative clauses is a promising field of research.

Appendix

Transcription conventions (GAT 2, Selting et al. 2009)

word [word]	
[word] word	overlapping text
(word)	uncertain transcription
((coughs))	commentary
wo'	truncation
=	latching
.h	inhalation
'h	exhalation
(.)	micropause
(-) (--) (---)	pauses up to one second
(2.5)	pause of indicated length
wo:rd	lengthened segment
NEver	primary accent
nEver	secondary accent (not always marked)
<<f> >, <<ff> >	loud/very loud, from <until>
<<p> >, <<pp> >	soft/very soft, from <until>
<<all> >	fast, from <until>
mhm	minimal feedback
eh, ehm	hesitation particles
words?	rising to final high
words,	rising to final mid





38

Karin Birkner

words; falling to final mid
words. falling to final low
words- level intonation

References

- Bache, C. and L. Kvistgaard Jakobson 1980 On the distinction between restrictive and non-restrictive relative clauses in modern English. *Lingua* 52: 243–267.
- Becker, R. 1978 Oberflächenstrukturelle Unterschiede zwischen restriktiven und nichtrestriktiven Relativsätzen im Deutschen. *Kölner Linguistische Arbeiten Germanistik* 4: 1–12.
- Behaghel, O. 1928 *Deutsche Syntax*. Bände I–IV, Heidelberg: Winter.
- Birkner, K. 2006a (Relativ-)Konstruktionen zur Personenattribuierung: 'ich bin n=mensch der ...'. In: S. Günthner and W. Imo (eds.), *Konstruktionen in der Interaktion*, 205–238. Berlin: de Gruyter.
- Birkner, K. 2006b Objektrelativsätze mit *haben*. In: A. Deppermann, T. Spranz-Fogasy and R. Fiehler (eds.), *Grammatik und Interaktion*, 147–177. Radolfzell: Verlag für Gesprächsforschung.
- Birkner, K. 2008 *Relativ(satz)konstruktionen im gesprochenen Deutsch: Syntaktische, prosodische, semantische und pragmatische Aspekte*. Berlin: de Gruyter.
- Blühdorn, H. 2007 Zur Struktur und Interpretation von Relativsätzen. *Deutsche Sprache* 4: 287–314.
- Boersma, P. and D. Weenink 2006 Praat: doing phonetics by computer, computer program, version 5.1.43, <http://www.fon.hum.uva.nl/praat/>, last access 08/28/2010.
- Brandt, M. 1990 *Weiterführende Nebensätze. Zu ihrer Syntax, Semantik und Pragmatik*. Stockholm: Almqvist & Wiksell International.
- Duden 1998 *Die Grammatik*, 6th edition, Mannheim, Dudenverlag.
- Duden 2005 *Die Grammatik*, 7th edition, Mannheim, Dudenverlag.
- Ebert, K. H. 1973 Functions of relative clauses in reference acts. *Linguistische Berichte* 23: 1–11.
- Eisenberg, P. 1999 *Grundriss der deutschen Grammatik* (Band 1: Das Wort; Band 2: Der Satz). Stuttgart/Weimar: Metzler.
- Eissenhauer, S. 1999 *Relativsätze im Vergleich: Deutsch – Arabisch*. München/Berlin: Waxmann.
- Fox, B. A. and S. A. Thompson 1990 A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66: 297–316.
- Frawley, W. 1992 *Linguistic Semantics*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Fritsch, W. J. 1990 *Gestalt und Bedeutung der deutschen Relativsätze*. München, Uni-Druck.
- Frosch, H. 1996 Appositive und restriktive Relativsätze. *Sprachtheorie und Germanistische Linguistik* 2: 7–19.
- Gärtner, H.-M. 2001 Are there V2 relative clauses in German? *The Journal of Comparative Germanic Linguistics*: 97–141.
- Greenbaum, S. and R. Quirk 1992 *A Student's Grammar of the English Language*. Harlow, Essex: Longman.





- Helbig, G. and J. Buscha 2001 *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Berlin: Langenscheidt.
- Hentschel, E. and H. Weydt 2003³ *Handbuch der deutschen Grammatik*. Berlin: de Gruyter.
- Holler, A. 2005 *Weiterführende Relativsätze. Empirische und theoretische Aspekte*. Berlin: Akademie Verlag.
- Holler, A. 2007 Uniform oder different? Zum syntaktischen Status nicht-restriktiver Relativsätze. *Deutsche Sprache* 3: 250–270.
- Lehmann, C. 1984 *Der Relativsatz. Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: Narr.
- Lehmann, C. 1995 Relativsätze. In: J. Jacobs (ed.), *Syntax. Ein internationales Handbuch zeitgenössischer Forschung*, 1199–1216. Berlin: de Gruyter.
- Löbner, S. 2003 *Semantik. Eine Einführung*. Berlin: de Gruyter (= de Gruyter Studienbuch).
- Motsch, W. 1965 Untersuchungen zur Apposition im Deutschen. *Studia Grammatica* V: 87–132.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1972 *A Grammar of Contemporary English*. London: Longman.
- Ravetto, M. 2006 *Es war einmal ein Königssohn, der bekam Lust in der Welt umher zu ziehen. Le false relative in tedesco*. Vercelli: Mercurio.
- Schaffranietz, B. 1997 Zur Unterscheidung und Funktion von restriktiven und appositiven Relativsätzen des Deutschen. *Linguistische Berichte* 169: 181–195.
- Schaffranietz, B. 1999 *Relativsätze in aufgabenorientierten Dialogen. Funktionale Aspekte ihrer Prosodie und Pragmatik in Sprachproduktion und Sprachrezeption*. Dissertation an der Universität Bielefeld, <http://bieson.ub.uni-bielefeld.de/volltexte/2003/180/>, last access 08/28/2010.
- Seiler, H. 1960 *Relativsatz, Attribut und Apposition*. Wiesbaden: Otto Harrassowitz.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzlufft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock and S. Uhmann 2009 Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.
- Tao, H. and M. J. McCarthy 2001 Understanding non-restrictive *which*-clauses in spoken English, which is not an easy thing. *Language Sciences* 23: 651–677.
- Weinert, R. 2006 Relative Clauses in Spoken English and German – Their structure and function. *Linguistische Berichte* 197: 3–51.
- Weinrich, H. 2005 *Textgrammatik der deutschen Sprache*. Hildesheim/Zürich: Olms.
- Zifonun, G. 2001 *Grammatik des Deutschen im europäischen Vergleich: Der Relativsatz*. Mannheim: Institut für Deutsche Sprache (= amades – Arbeitspapiere und Materialien zur deutschen Sprache 3).
- Zifonun, G., L. Hoffmann, B. Strecker et al. 1997 *Grammatik der deutschen Sprache*. 3 Bände. Berlin: de Gruyter.





Martin Pfeiffer
University of Freiburg

What prosody reveals about the speaker's cognition: Self-repair in German prepositional phrases

1. Prosody and self-repair

From the outset of the Interactional Linguistics enterprise, one of the main concerns has been the investigation of how participants deploy prosodic resources in conversation (e.g. Couper-Kuhlen and Selting 1996; Couper-Kuhlen and Selting 2001; Selting and Couper-Kuhlen 2001; Couper-Kuhlen and Ford 2004; Barth-Weingarten, Reber, and Selting 2010). Within this area, studies at the intersection of prosody and syntax have proved to be particularly fruitful. A significant body of research has demonstrated that prosodic resources interact with syntactic resources, such as in the constitution of units (e.g. Selting 1996; Auer 1996; Ford and Thompson 1996; Ford, Fox, and Thompson 1996; Selting 2000) and in the organization of turn-taking (e.g. Cutler and Pearson 1986; Selting 1995b; Wells and Peppé 1996).

However, there are other conversational domains in which the relationship between prosody and syntax still needs to be explored. A case in point is the phenomenon of self-initiated self-repair (cf. Schegloff, Jefferson, and Sacks 1977). In recent years, the investigation of self-repair has become a vibrant field of research in Interactional Linguistics (e.g. Fox and Jasperson 1995; Fox, Hayashi, and Jasperson 1996; Uhmann 2001, 2006; Rieger 2003; Wouk 2005; Fox, Maschler, and Uhmann 2009; Fox et al. 2009; Birkner et al. 2010, 2012; Pfeiffer 2010, in press). However, most of these studies have focused on the relationship between syntax and interaction, whereas the prosody of self-repair has not received much attention. Although there has been some research on cut-offs (e.g. Berg 1986; Jasperson 2002; Auer 2005a; Zhang and Luke 2010) and speech rate (e.g. Uhmann 1992; Plug 2006), the intonation of self-repair has been dealt with only marginally. Selting (1995a: 63), for instance, describes the function of intonation as preserving the cohesion of an utterance when a speaker continues to talk after a pause or an interruption. In her study on fragments, Selting (2001) discusses the intonational characteristics of several examples that can be analyzed as instances of repair. However, her analytical focus is not on how intonation can be deployed in repairing the fragmentariness of units, but rather on how intonation contributes to making fragmentary units recognizable as unfinished



for participants. Another investigation that touches on intonational aspects of self-repair is the study by Levelt and Cutler (1983) on prosodic marking in lexical repair in Dutch. Their main finding is that lexical-error repairs are often marked by accentuating the repairing lexical element and thereby establishing a clearly discernible contrast between the “old” and the “new” information. However, the study by Levelt and Cutler is restricted to lexical words – a prosodic analysis of function words in self-repair is still lacking. Therefore, to fill this gap, the present chapter explores the intonational characteristics of substitutions of determiners in German prepositional phrases. Consider the following example:

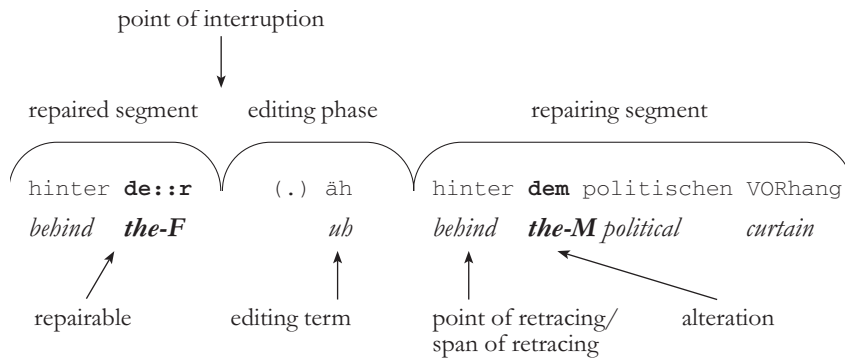


Figure 1: The three phases of self-repair.

Illustration and terminology (modified) are taken from Levelt (1983: 45). The terms *repaired segment* and *repairing segment* are taken from Fox and Jaspersen (1995: 81). See Appendix for transcription conventions (cf. Selting et al. 2009) and abbreviations. Boldface marks the substituted and the substituting elements.

In Figure 1, the speaker interrupts the emerging utterance after the lengthened determiner *de::r* (the-F). After the editing phase which is filled with the editing term *äh* (uh), she recycles to the preposition *hinter* (behind) and substitutes the “old” determiner with the “new” determiner *dem* (the-M). Substitutions of determiners within prepositional phrases are grammatically interesting in that they involve an alteration of at least one nominal category (i.e., gender, definiteness, or number).¹ Note that these grammatical changes concern not only the determiner itself, but also the form of the projected

¹ A modification of case is also possible, but involves the replacement of the preposition which determines the case of the nominal phrase. Such cases were not included in the corpus.





noun (see Section 2.2 on *projective repair*).² Alterations of gender (see Fig. 1) are particularly interesting because, in German, the selection of the gender form of the determiner is dependent on the noun, while the categories of definiteness and number are not; they are determined by the context or referent. This means that the substitution of a determiner with alteration of gender is, in fact, a substitution of the intended noun that the speaker already has in mind: The original lemma associated with the original determiner is substituted with the new lemma associated with the new determiner. Thus, substitutions of determiners with alteration of gender always involve a *lemma substitution*, whereas this need not be the case in alterations of the other grammatical categories or in repetitions of the determiner (see Section 5.1 for a detailed discussion).

This initial observation leads to the central concern of this chapter: Can we identify a link between alterations of syntax and prosody? More precisely, is the difference between alterations of gender and all other types of alteration and repetitions of the determiner reflected in the intonational design of self-repair? As lemma substitutions involve some kind of semantic readjustment, the production of this type of self-repair involves conceptualization and formulation processes that presumably occur early in the course of speech production (cf. Levelt 1989). Based on these considerations, it is reasonable to assume that the cognitive processes involved in lemma substitutions might start even before the initiation of self-repair is produced on the phonetic level. Therefore, the focus of the present chapter will be an intonation analysis of the initial phase of self-repair called the *repaired segment* (see Fig. 1). Combining a cognitive and an interactional perspective, the following research questions are pursued:

- (1) What is the relationship between intonation patterns on the repaired segment and subsequent alterations of syntax?
- (2) In what ways can these intonation patterns be relevant for interaction?

Self-repairs of determiners particularly lend themselves to an investigation of the relationship between prosody and syntax because determiners fulfill a grammatical function and, thus, their substitution involves a change in grammatical structure. The cognitive processes associated with these changes

² Of course, the modification of a nominal category influences the projected form of eventually occurring adjectives, too. However, as adjectives are facultative constituents in prepositional phrases and only occur rarely in my corpus, they will not be considered here.





must be executed in real-time, while the speaker must simultaneously manage other domains of speech production, including the prosodic design and the articulation of the ongoing utterance containing the repair. Such instances of self-repair involve syntactic reorganization on the fly, i.e., the processing of grammatical changes while talking (and therefore producing intonation patterns) at the same time. Therefore, the analysis of this type of self-repair might illuminate potentially different relationships between alterations of gender and intonation, on the one hand, and alterations of other grammatical categories and intonation, on the other.

Moreover, it is important to note that prepositions and determiners are unaccented in German, except for rarely occurring contrastive accents and stressed determiners in the reference nominal accompanied by a restrictive relative clause.³ Correspondingly, in my collection of 166 instances of self-repair there is only one stressed determiner (contrastive accent) and no stressed prepositions. The fact that determiners and prepositions lack stress assignment makes their intonation patterns more easily comparable. In nouns, by contrast, primary or secondary stress assignment would override other factors that could potentially influence the design of intonation.

In Section 2, the types of self-repair to be explored will be presented from a syntactic point of view. Section 3 describes the data and the method comprising instrumental and auditive analysis. In the subsequent empirical part of the chapter (Sections 4 and 5), the main results, in particular the finding that low falling pitch regularly occurs with alterations of gender, will be presented and discussed. Finally, Section 6 will conclude and make a suggestion for further research.

2. Presentation of the phenomena

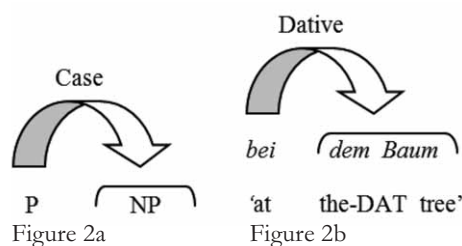
The present study examines two specific types of self-repair within prepositional phrases in German: (1) repetitions of the preposition and the determiner, and (2) substitutions of the determiner with repetition of the preposition. These two types of self-repair will be illustrated after a brief description of the main characteristics of prepositional phrases in German.

³ The assumption that stressed determiners occur in reference nominals of restrictive relative clauses still has to be tested empirically (cf. Birkner this volume).



2.1. The prepositional phrase in German

The preposition in German always requires a complement phrase, usually a nominal phrase, and governs the complement nominal phrase by determining its case. The cases determined by German prepositions are genitive, dative, and accusative, whereas nominative nominal phrases never occur as complements of a preposition (Gallmann 2006: 848). Figure 2a depicts a general schematization of case determination in German prepositional phrases; Figure 2b gives an example for the dative case:⁴



In Figure 2b, the preposition *bei* ›at‹ requires a dative nominal phrase. The case of the nominal phrase can be recognized by the dative marking of the (definite masculine singular) determiner *dem* ›the-DAT‹.

Another important characteristic of the prepositional phrase in German is the tendency of the preposition and the determiner to cliticize (e.g. *bei dem* ›at the‹ → *beim* ›at.the-CL‹).⁵ The process of cliticization makes evident the strong bond between the preposition and the determiner in German. This bond partly explains why there is a strong preference for speakers of German to begin with the repetition of the preposition when carrying out self-repair in prepositional phrases. If self-repair is carried out within a prepositional phrase containing a cliticized form of preposition and determiner, a retraction directly to the determiner is impossible, of course. However, even when the prepositional phrase contains non-cliticized forms (see Fig. 2b), speakers regularly ignore the determiner as a possible starting point, respect-

⁴ Due to the low frequency of occurrence of adjectives in my corpus (only 11 % of the prepositional phrases contain an adjective), no adjective is included in the constructed example in Figure 2b.

⁵ In abbreviations concerning the grammatical properties of the determiner (here: ›CL‹ for cliticization, see Appendix for an index of all abbreviations), only the feature which is important for the respective example under discussion will be specified. To ensure clarity, all the other grammatical features (here: case, number, gender, and definiteness) are left aside.





What prosody reveals about the speaker's cognition

45

ing the unity of the preposition and its complement (cf. Uhmann 2001, 2006; Pfeiffer 2008, in press; Birkner et al. 2010, 2012).

Within the nominal phrase complement, there is agreement according to case (genitive, dative, and accusative), number (singular and plural), and gender (masculine, feminine, and neuter). In addition, (in)definiteness influences the form the nominal phrase takes. The interplay of these grammatical categories in prepositional phrases is shown in Table 1 where the preposition *in* (<into>) requires an accusative complement:⁶

Table 1: The prepositional phrase in German.
The illustration is taken from Birkner et al. (2010: 10).

Feminine noun				
<i>Case</i>	<i>SG DEF</i>	<i>SG INDEF</i>	<i>PL DEF</i>	<i>PL INDEF</i>
ACC	in die Küche	in eine Küche	in die Küchen	in Küchen
(Ich gehe/ <i>I am going</i>)	<i>into the kitchen</i>	<i>into a kitchen</i>	<i>into the kitchens</i>	<i>into kitchens</i>
Masculine noun				
<i>Case</i>	<i>SG DEF</i>	<i>SG INDEF</i>	<i>PL DEF</i>	<i>PL INDEF</i>
ACC	in den Garten	in einen Garten	in die Gärten	in Gärten
(Ich gehe/ <i>I am going</i>)	<i>into the garden</i>	<i>into a garden</i>	<i>into the gardens</i>	<i>into gardens</i>
Neuter noun				
<i>Case</i>	<i>SG DEF</i>	<i>SG INDEF</i>	<i>PL DEF</i>	<i>PL INDEF</i>
ACC	in das Haus	in ein Haus	in die Häuser	in Häuser
(Ich gehe/ <i>I am going</i>)	<i>into the house</i>	<i>into a house</i>	<i>into the houses</i>	<i>into houses</i>

After this short overview of the grammar of the prepositional phrase in German, we now examine the phenomenon of self-repair within this syntactic unit.

⁶ The preposition *in* is one of the two-way prepositions in German that can take both dative and accusative complements, depending on the meaning of the utterance. However, as this is not crucial for the argument in this chapter, I restrict myself to an illustration of the accusative form.





2.2. Repetitions and substitutions of the determiner in prepositional phrases

This analysis focuses on two types of self-repair that are defined syntactically: (1) repetitions of the preposition and the determiner, and (2) substitutions of the determiner with repetition of the preposition. Both types of self-repair included in the corpus share two structural properties: first, the point of interruption is located after the determiner, and second, retracing goes back to the preposition. Cases of self-repair within the prepositional phrase in which the speaker interrupts the emerging utterance at another position (e.g. after the preposition or within the noun) were not included in the corpus. Likewise, instances of self-repair with different points of retracing (e.g. rare instances of retraction to the determiner) were excluded from this study.

The following example illustrates the first type of self-repair, the repetition of the preposition and the determiner:

Example (1)

01 hh04: gleich nach **de:r** (-) nach **der** GRENZöffnung;
right after the-F after the-F opening of the border

In Example (1), the speaker initiates repair within the production of the determiner through the lengthening of the vowel in the determiner *de:r* (>the-F<). The following pause after the determiner, which marks the point of interruption, functions as an additional resource to signal repair initiation. Subsequently, the speaker retraces to the beginning of the prepositional phrase in order to repeat the preposition and the determiner *nach der* (>after the-F<) before delivering the composed noun *GRENZöffnung* (>opening of the border<) that syntactically completes the prepositional phrase. To simplify matters, this type of self-repair will be referred to as *repetition*. In this type of self-repair, a part of the original utterance (preposition and determiner) is repeated without changing its morphosyntactic form.

The second type of self-repair we will explore here is the substitution of the determiner with prior repetition of the preposition. Consider the following example:

Example (2)

01 T: des war nach **de:r** nach **dem** KAIserschnitt;
that was after the-F after the-M Cesarean section





What prosody reveals about the speaker's cognition

47

In this example, just as in Example (1) above, self-repair is initiated directly after the determiner and retraction goes back to the preposition. However, contrary to instances of repetition, this type of repair involves a substitution of the determiner. In Example (2), the speaker replaces the feminine determiner *de:r* (the-F) with the masculine determiner *dem* (the-M). In addition to substitutions with alteration of gender, there can also be substitutions with alteration of definiteness or number (see Section 4 for the respective frequencies). Consider the following examples:

Example (3)

01 hh04: wenn jemand aus **de:r** (.) äh (-) aus **einer**
steuerberaterfamilie kommt.
if somebody from the uh from a
family of tax consultants comes
If somebody comes from a family of tax consultants

Example (4)

01 i-hh04: also da war ja auch vor einigen wochen dieses beispiel
mit **den:** °h mit **dem** KIND,
well there was PTCL PTCL some weeks ago this example
with the-PL with the-SG child

In Example (3), the speaker replaces the definite determiner *der* (the) with the indefinite determiner *einer* (a). In Example (4), the plural determiner *den:* (the-PL) is substituted with the singular determiner *dem* (the-SG). The instances of self-repair in the Examples (2), (3), and (4) have in common that they involve an alteration of a nominal category through substitution of their respective determiners. Given that every determiner in German projects a certain grammatical form of the noun to follow, these instances of substitution affect not only the determiner, but also the noun projected by the determiner.⁷

Such instances of self-repair cannot be adequately described by the distinction between prospective and retrospective self-repair (e.g. Papantoniou 2010), based on the distinction between prepositioned and postpositioned

⁷ Following Auer (2005b:8), I define projection as “the fact that an individual action or part of it foreshadows another”. Projections exist on both the interactional level and the grammatical (i.e., prosodic and syntactic) level. Grammatical projections help the listener anticipate the trajectory of an emerging linguistic structure and thus when the next transition-relevance place (cf. Sacks, Schegloff, and Jefferson 1974) will be reached.



self-initiations of repair made by Schegloff (1979: 273). In prospective self-repair, speakers delay the progression of the unfolding utterance by means of, for example, repetitions, lengthenings, and pauses, to treat a problem in parts of the utterance that have not yet been uttered. In retrospective self-repair, by contrast, speakers repair problems in parts of the utterance that have already been delivered and must be changed retrospectively. If we consider the replacement of the determiner in the Examples (2), (3) and (4), it seems obvious, at first glance, that these cases should be classified as instances of retrospective self-repair because they involve the retrospective substitution of an element within the emerging utterance. However, as the determiner projects syntactic information about the noun, the actual target of this type of self-repair is the grammatical form of the entire nominal phrase. As the substitution of the determiner is the consequence of the alteration of a nominal grammatical category, these instances of repair concern the determiner and the following noun in equal measure. Therefore, such cases of repair should be considered retrospective and prospective at the same time. In order to take into account both their retrospective and prospective character, which arises due to the substitution of an element that projects syntactic information about a constituent to follow, I will refer to these instances as *projective repair*.

Thus far, we have seen instances of two different basic types of self-repair. The first type is repetition (Ex. 1). The second type is the substitution of the determiner, that is, projective repair, with several subtypes: alteration of gender (Ex. 2), alteration of definiteness (Ex. 3), and alteration of number (Ex. 4). Additionally, the present study includes a type of self-repair that involves the process of (de)cliticization:

Example (5)

01 i-mu02: wenn_mer dorten warn **beim** (.) bei **dem**: SOMmerfest
 da-
 when we there were **at.the-CL** at **the** summer festival
 there
When we were at the summer festival

In this example, the speaker interrupts her ongoing utterance after the production of the cliticized form *beim* (<at.the-CL>), retraces to the beginning of the prepositional phrase, and replaces the cliticized form with its non-cliticized components, i.e., the preposition *bei* (<at>) and the determiner *dem*: (<the>). As (de)cliticization is a purely phonological process, nominal syntax



What prosody reveals about the speaker's cognition

49

is not changed.⁸ Thus, no change in syntactic projection of the determiner is involved. In this respect, this example is similar to repetition (see Ex. 1). Nevertheless, since a formal feature of the prepositional phrase being replaced in cases like Example (5), these instances of repair are classified as substitutions with alteration of cliticization.

In what follows, these types of self-repair, presented from a syntactic point of view in this section, will be subject to an intonation analysis.

3. Method

3.1. Data

For the present study, all the instances of self-initiated self-repair within prepositional phrases that exhibit the structural properties mentioned in Section 2.2 (i.e., point of interruption after the determiner and retraction to the preposition) were collected from three different corpora: everyday conversations, informal interviews and psychotherapeutic interaction. There were 166 instances of self-repair involving 37 native speakers of German within a total of approximately 45 hours of audio-taped data.⁹

3.2. Coding of the data

The instances of self-repair were classified according to the different types of self-repair: repetition (Ex. 1) and substitution (Ex. 2–5). Instances of substitution were further categorized according to the type of alteration: alteration of gender (Ex. 2), alteration of definiteness (Ex. 3), alteration of number (Ex. 4), or alteration of cliticization (Ex. 5). When examples could not be unambiguously categorized due to homomorphous forms of determiners (e.g. *die* can be definite singular feminine or definite plural for all genders) or due to the manipulation of more than one nominal category (e.g. alteration of gender and number), they were excluded from the corpus in order to ensure comparability among examples.

⁸ With regard to semantics, the meaning of cliticized forms is often different from non-cliticized forms. However, this will not affect the argumentation in this chapter.

⁹ I would like to thank Karin Birkner and Fabian Overlach for giving me permission to use their data from informal interviews and psychotherapeutic interaction. Furthermore, I am grateful to the members of the DFG-funded research project *Dialektintonation* (Universities of Freiburg and Potsdam, 1998–2005), Peter Auer and Margret Selting (principal investigators) as well as Peter Gilles and Jörg Peters (team members), for providing me with the informal interviews with urban dialect speakers recorded in this project.





3.3. Measurements

In addition to coding for type of repair and type of alteration,¹⁰ all instances of self-repair were subject to acoustic analysis with “Praat” (Boersma and Weenink 2010), a computer program for speech analysis. In order to obtain a value for the pitch movement on each repaired segment (i.e., the preposition and the determiner before the point of interruption, see Fig. 1), F_0 was measured in semitones at the beginning and at the end of the repaired segment, meaning on the center of the vowel of the preposition (X) and on the end of the last sonorant element of the determiner (Y). The value of the intonation pattern, that is, the range and the direction of the pitch movement for each repaired segment, was obtained through the subtraction of both measured values (X–Y).

Each repaired segment was also coded for lengthening or non-lengthening based on auditive analysis. This additional auditive analysis was necessary to distinguish between a slow-down in speech rate (i.e., syllables within the repaired segment evenly lengthened with respect to prior syllables), which was not coded as lengthening, and the lengthening of sounds (i.e., one syllable/phoneme being unevenly long compared to prior syllables/phonemes). This is an important distinction, since both phenomena occur with a longer duration of the repaired segment but may have different impacts on the intonation pattern of the segment in which they occur. The duration of (possibly occurring) editing terms (i.e., hesitation markers such as *äh* *uh*) and pauses between the repaired and the repairing segments was measured in milliseconds.

Instances of self-repair in which neither F_0 nor durations could be measured with Praat due to overlapping talk or creaky voice were excluded from the corpus.

4. Results

This section provides the main results of the data analysis. Since the observations of this study are based on naturally occurring interactions, the number of instances in each group of self-repair cannot be systematically controlled and balanced. Rather, the data reflect the natural frequency of occurrence of the phenomena under investigation. Some types of self-repair, especially substitutions of the determiner with alteration of definiteness, number and cliti-

¹⁰ The term *type of repair* refers to the basic types of repair (i.e., repetition and substitution), whereas *type of alteration* refers to a subtype of substitution (i.e., alteration of gender, number, definiteness, or cliticization).



What prosody reveals about the speaker's cognition

51

cization, are rare in spontaneous talk. In contrast, other types of self-repair (e.g. repetitions) occur much more frequently. This leads to quite pronounced differences between the numbers of instances of these types.¹¹ With regard to the differences in group size, this chapter reports on work in progress rather than on a completed project. However, I am confident that the corpus is substantial enough to allow for some interesting general conclusions.

To begin, I will give examples of intonation patterns that are representative of each type of repair. I will then present the statistical evaluation of the data, which reveals similarities and significant differences between these types of repair with respect to intonation patterns.

4.1. Intonation patterns

Let us now consider the different types of repair, previously described syntactically (see Section 2.2), from an intonational perspective. For each example, a transcript and translation are first provided, followed by acoustic analyses of the intonation pattern of the repaired segment.

While Examples (6)–(9) show only slightly falling intonation patterns, Examples (10) and (11) illustrate cases in which the pitch of the repaired segment falls significantly lower.

Example (6) and corresponding Figure 3 represent the repair type of repetition:

Example (6)

- 01 b07: aber ich glaub schon dass der rEAgan mit **seinem**: °hh äh:
 mit **seinem** proGRAMM,
 *but I do believe that Reagan with **his** uh*
 with **his** program
 02 de:r °h de:r äh (.) WELTraummilitarisierung,
 of of uh militarization of space
 03 i-b07: hm:,
 hm
 04 b07: dass der die kaPUTT jemacht hat.
 that he destroyed them

¹¹ As several studies on self-repair suggest (e.g. Rieger 2003; Fox, Maschler, and Uhmman 2009; Fox et al. 2009; Birkner et al. 2010, 2012), repetitions generally seem to occur more frequently than substitutions in the languages investigated so far.





52

Martin Pfeiffer

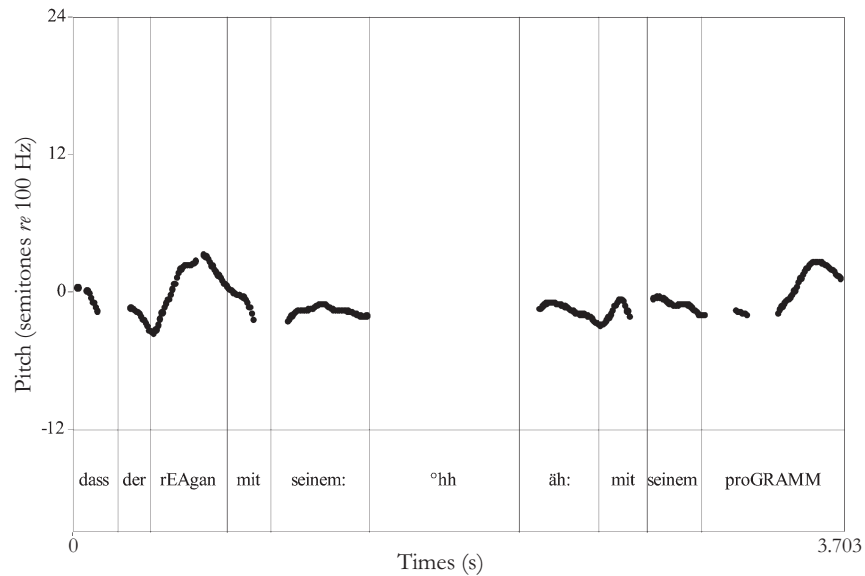


Figure 3: Intonation pattern in repetition

Figure 3 shows the intonation pattern (in semitones) associated with the repetition in Example (6). As can be seen, there is no strong pitch movement on the repaired segment (i.e., the first production of preposition and article) *mit seinem*: ⟨with his⟩. The measurement shows a slightly falling pitch of 0.9 semitones from the vowel of the preposition to the last sonorant element of the determiner. Although the repaired segments of sixteen repetitions show (slightly) rising intonation patterns, the intonation in Figure 3 is quite representative of this group ($n = 108$): The average value for intonation patterns on repaired segments of repetitions is a slightly falling pitch of 0.71 semitones. In addition to examples of repetitions with lengthening of the repaired segment and/or pauses between the repaired and the repairing segment (see Fig. 3), there is also a large subgroup of fluent repetitions ($n = 49$) that do not use lengthening or pauses (at the utmost a micropause of up to 0.2 seconds) for repair initiation. However, this does not seem to have an influence on the intonation pattern.





What prosody reveals about the speaker's cognition

53

Let us now turn to the group of substitutions with alteration of definiteness. Consider Example (7) and corresponding Figure 4:

Example (7)

01 hh04: dass dAs dann in **einem** in **dem** wirtschaftlichen bereich
KURklinik;
*that this in a in the economic sector
health clinic*
02 °hh (-) dazu führt (-) dass die sich also EINSchränken
müssen;=
leads to constrictions
03 =vielleicht auch persoNAL entlassen müssen;
maybe also to dismissals
04 i-hh04: ja
yes
05 DAran [habe]n die jetzt nich geDACHT.
they didn't anticipate that
06 hh04: [nech]
TAG

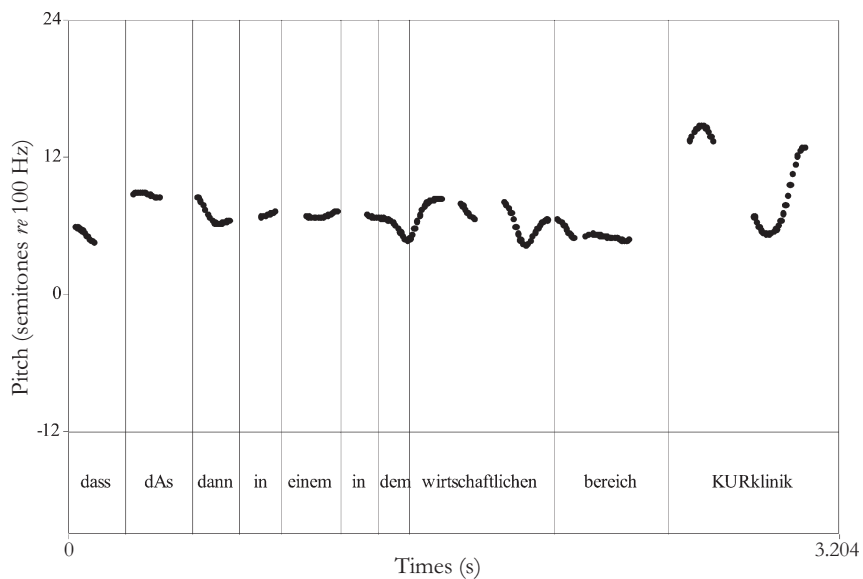


Figure 4: Intonation pattern in alteration of definiteness



In this example, the speaker initiates repair after the production of the indefinite determiner *einem* (ʔa), without pausing or using an editing term, by retracing to the beginning of the emerging prepositional phrase. The repairing segment substitutes the indefinite determiner with the definite determiner *dem* (ʔthe).

Similarly to Figure 3, there is practically no pitch movement on the repaired segment *in einem* (ʔin a) in Figure 4. The value for this pitch contour is 0.32, meaning there is minimal falling pitch on this segment. The average value for the group with alteration of definiteness ($n = 11$) is falling intonation of 1.04 semitones. Within this group, there are three cases of minimally rising pitch on the repaired segment, but also three examples with a falling intonation pattern of more than two semitones. In one, there is no pause between the repaired and the repairing segment, in the second there is a micro-pause, and in the third, a filled pause. None of these examples show a lengthening of the repaired segment.

The following example shows a substitution of the determiner with alteration of number:

Example (8)

01 b07: da wa:r gabs irgende veranstaltung in BAbelsberg-
there was some event in babelsberg
02 in **de:m** °hh in **den** ehemaligen FILMstudios
in **the-SG** in **the-PL** former film studios

In this instance of self-repair (Ex. 8 and Fig. 5), the singular determiner *de:m* (ʔthe-SG) is replaced with the plural determiner *den* (ʔthe-PL). On the repaired segment *in de:m* (ʔin the-SG), which is separated from the repairing segment by a pause used for breathing, we find a slightly falling intonation of 0.96 semitones. The average intonation pattern value of this group ($n = 6$) amounts to 1.09 and is thus similar to the average value for substitutions with alteration of definiteness. Five out of six examples in this group have a lengthened repaired segment. One example shows slightly rising pitch and another one low falling pitch, while the other instances show falling intonation patterns between 0.96 and 1.74 semitones.



What prosody reveals about the speaker's cognition

55

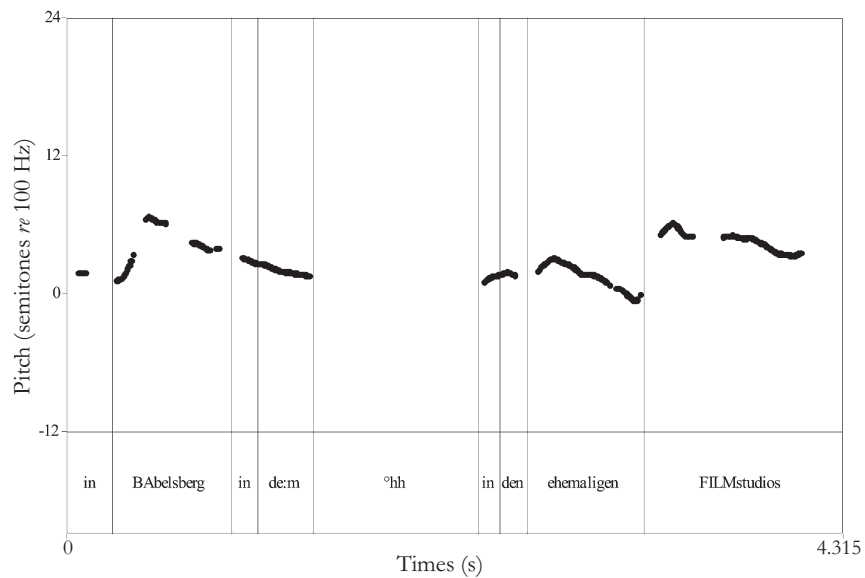


Figure 5: Intonation pattern in alteration of number

The next example represents substitutions with alteration of cliticization:

Example (9)

- 01 fr03: beim hErtie zum: zu dem CAfe do rüber;
next to hertie over **to.the-CL** to **the** cafe
- 02 i-fr03: ja
yes
- 03 fr03: sell schmale GÄSSle dort nuff-
this narrow street up there
- 04 i-fr03: hm_hm
hm hm
- 05 fr03: des war d_NUSSmannstro:ß.
that used to be nussmann street



56

Martin Pfeiffer

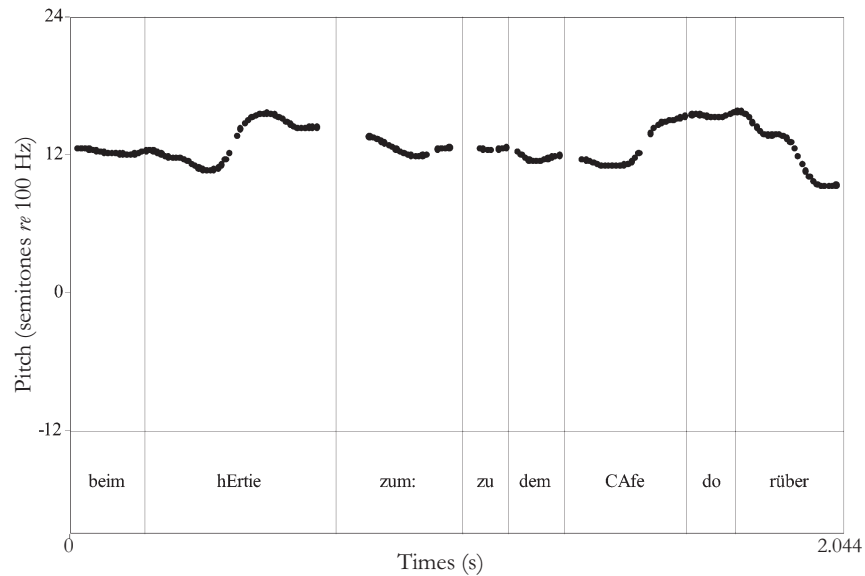


Figure 6: Intonation pattern in alteration of cliticization



In this example, fr03 describes the location of a street. She interrupts the emerging utterance after the cliticized form of preposition and determiner *zum:* (<to.the-CL>), retraces to the beginning of the prepositional phrase, and replaces the cliticized constituent with the non-cliticized form *zu dem* (<to the>). She does this presumably because the cliticized form cannot be used appropriately with the upcoming referent *CAfe* (<cafés>), which has not yet been introduced in the current conversation. As Figure 6 shows, the overall pitch pattern of the repaired segment is a slightly falling intonation of 0.89 semitones. The average value for intonation patterns in this group ($n = 8$) is 0.54, meaning there is slightly falling pitch on the first segment. Although the distribution of lengthened repaired segments and pauses between the segments is quite heterogeneous in this group (five lengthenings and three pauses, no coincidence of pause and lengthening), it is homogeneous with respect to intonation patterns (one slightly rising and seven slightly falling intonation patterns).





What prosody reveals about the speaker's cognition

57

Let us now turn to substitutions with alteration of gender. In this category, the pitch of the repaired segment falls considerably lower:

Example (10)

01 P57: un: danAch hat er (-) GESsen eigentlich bis: (--)
bis **zum:** (-) bis **zur** Tageschau;
and afterwards he kept on eating actually until
until.the-CL-M/N until.the-CL-F tv news

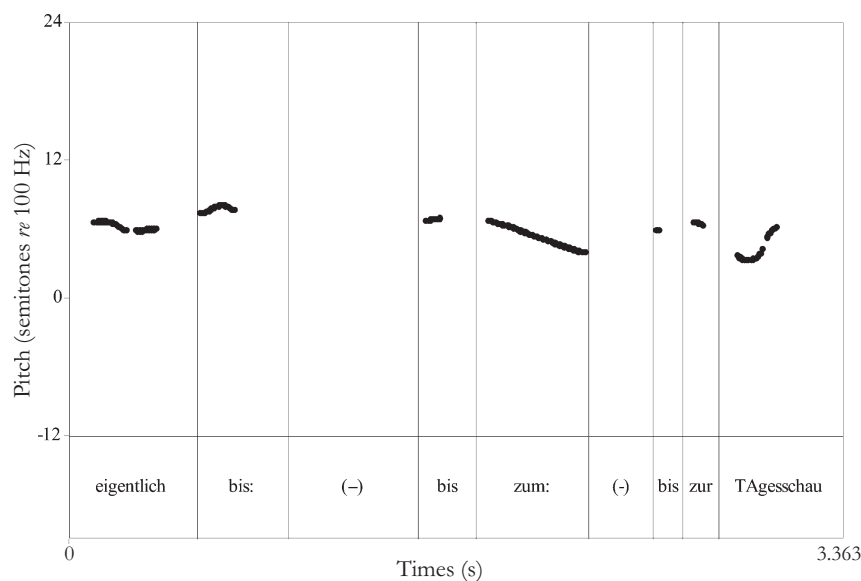


Figure 7: Intonation pattern in alteration of gender (repaired segment lengthened)

In this example, the speaker interrupts the emerging utterance after the lengthening of the cliticized form of the second preposition and the masculine/neuter determiner *zum:* (to.the-CL-M/N). Subsequently, he retraces to the beginning of the prepositional phrase in order to repeat the preposition *bis* and to substitute the cliticized determiner.¹² The altered cliticized form *zur*

¹² The prepositional phrase in which this substitution occurs is somewhat special because it contains two prepositions, *bis* (until) and *zu* (to). The combination of both prepositions, which also means until, is obligatory in this example, meaning that the preposition *bis* cannot stand alone in this context. In addition to the combination of two prepositions, there is more than one retraction involved in this instance of repair: Before carrying out self-repair with alteration of gender, the speaker breaks off after the production of the first preposition *bis:* (until) and repeats this constituent.



(to.the-CL-F) contains a feminine determiner that agrees with the subsequent feminine noun *Tagesschau* (tv news). As can be seen in Figure 7, intonation on the repaired segment *bis zum*: (to.the-CL-M/N) falls lower than in the examples analyzed above (Figs. 3–6). The value for this falling intonation pattern is 2.72 semitones, which is similar to the average of this group (2.2 semitones; $n = 33$).

The next example shows another instance of substitution with alteration of gender:

Example (11)

01 hh04: wir müssen ja mit dEn gesetzen arbeiten die DA sind.
we have PTCL to work with the laws that exist

02 i-hh04: hm_hm hm_hm
hm hm hm hm

03 hh04: wir ham ja keinen EInfluss auf **das** °hh auf **die**
entSTEHung der gesetze.
we have PTCL no influence on **the-N** on **the-F**
development of laws

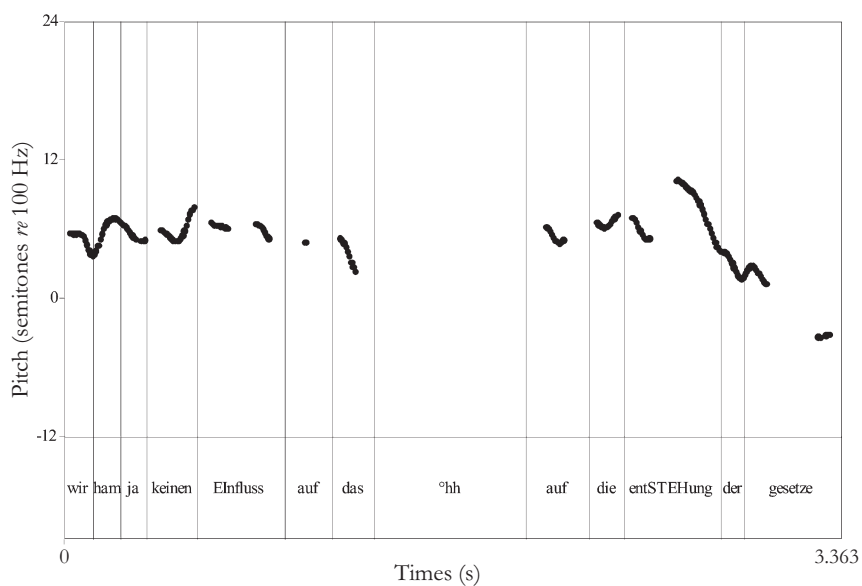


Figure 8: Intonation pattern in alteration of gender (repaired segment not lengthened)





What prosody reveals about the speaker's cognition

59

In accordance with the group of substitutions with alteration of gender, intonation falls quite low (2.39 semitones) on the repaired segment *auf das* (on the-N). Note that although the repaired segment in Figure 8 is not lengthened, the value of the falling intonation pattern is comparable to the lengthened segment in Figure 7. It is interesting that, on the one hand, there are no repaired segments with rising intonation in this group (though there are several examples with only slightly falling pitch); on the other hand, this group includes some of the lowest falling patterns (seven patterns between 3.25 and 5.24 semitones) within the entire corpus. There are 18 examples with a lengthened repaired segment and 23 examples with at least a micro-pause between the segments. However, similar to the other groups, there seems to be no correlation between low falling pitch and these two features.

This analysis demonstrates the existence of intonational difference between categories: The examples of repetition and alteration of definiteness, number, and cliticization show slightly falling pitch on the repaired segment, whereas the intonation patterns in the instances of alteration of gender fall much lower.

4.2. Similarities and differences between the intonation patterns

Let us now turn to the similarities and the main differences between the groups. An independent samples t-test was conducted to compare the average value of intonation patterns on the repaired segments (dependent variable) associated with the different types of repair/types of alteration (factors).

Table 2 gives an overview of the results for all comparisons:

Table 2: Statistical evaluation of similarities and differences between the groups.

Abbreviations: p = probability of error, df = degrees of freedom, ns = not significant, (*) = statistical tendency, * = significant difference, ** = highly significant difference.

1	Alteration of number – Alteration of cliticization	ns / t-value 1.11 / df 12
2	Alteration of definiteness – Alteration of cliticization	ns / t-value 1.05 / df 17
3	Alteration of definiteness – Alteration of number	ns / t-value -0.82 / df 15
4	Alteration of gender – Alteration of cliticization	p = 0.01* / t-value 3.46 / df 39
5	Alteration of gender – Alteration of definiteness	p = 0.013* / t-value 2.58 / df 42
6	Alteration of gender – Alteration of number	p = 0.061(*) / t-value 1.93 / df 37





- | | | |
|---|--|--|
| 7 | Alteration of definiteness, number, and cliticization – Repetitions | ns / t-value -0.66 / df 131 |
| 8 | Alteration of gender – Repetitions | $p = 0.000^{**}$ / t-value 6.04 / df 139 |
| 9 | Alteration of gender – Alteration of definiteness, number, and cliticization | $p = 0.000^{**}$ / t-value 4.09 / df 56 |

As can be seen, there is no significant difference between substitutions with alteration of number, cliticization, and definiteness, respectively (lines 1 to 3). This means that the intonation patterns in these three groups are statistically identical.

However, the intonation patterns associated with alteration of gender differ from these groups. The t-test shows significant differences between alteration of gender and alteration of cliticization (line 4), and between alteration of gender and alteration of definiteness (line 5). Given these significant differences, one might also expect to find a significant difference between alteration of gender and alteration of number (line 6), yet the test shows only a statistical tendency ($p = 0.061^{(*)}$). However, as the p-value for this comparison is quite close to significance, the absence of significance is most likely due to the small group size of substitutions with alteration of number ($n = 6$) and the heterogeneity of intonation patterns within this group ($M = 1.088$, Standard Deviation = 1.204).

Against this background, and given the similarity of substitutions with alteration of number, cliticization, and definiteness with respect to intonation patterns (see lines 1–3), it seems reasonable to compare these three subgroups, taken together, with the group of repetitions and the group of alteration of gender. Figure 9 shows the average falling intonation values for repetitions (left column), substitutions with alteration of definiteness, number, and cliticization (middle column) and substitutions with alteration of gender (right column).

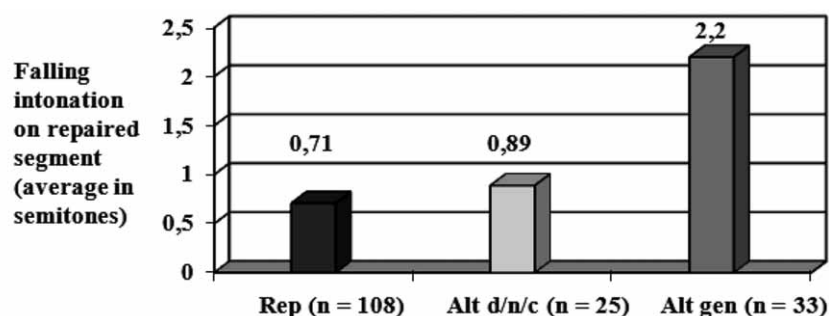


Figure 9: Differences in intonation between the groups (Rep = Repetition, Alt d/n/c = Alteration of definiteness/number/cliticization, Alt gen = Alteration of gender).





Repetitions and alterations of definiteness, number, and cliticization have slightly falling values (0.71 and 0.89 semitones), whereas the intonation value for alterations of gender falls considerably lower (2.2 semitones). As Figure 9 suggests, there is no significant difference between repetitions and substitutions with alteration of definiteness, number, and cliticization (Table 2, line 7), but there is a highly significant difference between substitutions with alteration of gender and repetitions (line 8) as well as between substitutions with alteration of gender and substitutions with alteration of definiteness, number, and cliticization (line 9).

Additionally, correlation analysis was conducted to investigate whether the intonation patterns on repaired segments correlate (i) with the duration of (possibly occurring) pauses between the repaired and the repairing segment and (ii) with the lengthening of the repaired segment. Both correlations proved to be extremely weak, meaning there is no significant influence of the duration of pauses or of the lengthening of the repaired segment on the intonation pattern of the repaired segment.

What conclusion can we draw from this statistical evaluation? The analyses indicate that the differences between the intonation patterns do not occur randomly, but are in fact systematically organized. This is an important observation with regard to the question as to whether the intonational characteristics of self-repair should be considered a part of grammar. From a usage-based perspective, one way to demonstrate that something belongs to grammar is to show that it is part of a more or less stable pattern of linguistic elements. Concerning the question of whether prosodic patterns are among the elements that can possibly become part of the grammatical system, the Cognitive Grammar approach gives a clear answer:

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)

According to this position, all the formal properties of self-repair, a recurrent commonality in any linguistic community, are capable of emerging as a linguistic unit. Indeed, numerous studies on structural aspects of self-repair have shown that language-specific syntactic regularities play a major role in the design of self-repair (Fox, Hayashi, and Jasperson 1996; Fox, Maschler, and Uhmman 2009; Uhmman 2001, 2006; Birkner et al. 2010, 2012; Pfeiffer 2010, in press). Based on observations about recurrent syntactic patterns in self-repair, some researchers concluded that there is a “grammar of repair” (Fox and Jasperson 1995: 79) that is oriented to by the speakers of a lan-





guage. Besides this syntactic orderliness, the present research shows that there are also prosodic regularities in the organization of self-repair. As was demonstrated in the intonation analysis, falling intonation occurs regularly with alterations of gender, i.e., lemma substitutions. Given the speakers' ability to reproduce this intonation pattern, we can assume that the intonation format of this type of self-repair, a "recurrent commonality", is represented as a part of the speakers' knowledge about the design of self-repair. In this regard, a central concept is *entrenchment*: "With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit (...)" (Langacker 1987: 59; see also Bybee 2006, 2007, 2010). Apparently, the process of entrenchment, i.e., the abstraction from single usage events, has led to the cognitive representation of the intonational contrast between the different types of projective repair in the minds of the speakers. Thus, there are good reasons to include this linguistic aspect in a grammatical description of this type of self-repair.

As mentioned above (Section 4.1), there are also substitutions of the determiner whose intonation contours do not fit neatly into the overall pattern (e.g. alterations of definiteness with low falling pitch on the repaired segment). Thus, there are other factors that can have an impact on the intonation pattern of the repaired segment. However, these factors cannot be treated here and must be left for further research.¹³

Despite several counter-examples within each group of self-repair, it is difficult to exclude prosody from a grammatical description of projective self-repair. As the alteration of gender seems to be responsible for the differences in intonation, prosody and syntax cannot be viewed as isolated in these instances of self-repair.

Let us summarize. The main finding of this intonation analysis is that the pitch of the repaired segment in substitutions with alteration of gender falls considerably lower than in all the other types of alteration and in repetitions. In what follows, this finding will be subject to a detailed discussion.

¹³ One possibility might be that the intonation pattern of the preposition and the determiner is part of the originally intended prosodic contour involved in a certain activity-type (e.g. teasing, mimicry, etc.) that is interrupted by self-repair. In some cases, the repaired segment might exhibit an intonation pattern that was planned in a way to lead to a nucleus accent on the originally intended noun. In other cases, it might also be that the intonation contour on the preposition and the determiner are part of a fixed collocation that is aborted when self-repair is initiated.





5. Discussion

This section discusses the intonation pattern associated with alteration of gender from two different (but interrelated) perspectives on language. While the first part gives a cognitive explanation for the identified intonation pattern (5.1), the second part reflects on a possible consequence of the intonation pattern for the interaction between speaker and recipient (5.2).

5.1. Cognitive perspective

Let us try to explain the main finding of the analysis from a psycholinguistic perspective and ask what the observed intonation pattern can tell us about the cognitive processes involved in the production of determiners. In psycholinguistic research on speech production, the predominant view is that the selection of the determiner is dependent on the noun (e.g. Levelt 1989). This means that the noun, as the lexical head of the phrase, is selected first. In a second step, according to the syntactic properties of the noun, certain “syntactic building procedures” (Levelt 1989: 11) are activated to add the other required constituents to the head (e.g. the determiner).

However, based on their analysis of speech pauses in Dutch, Schilperoord and Verhagen (2006) suggest a divergent model for the production of function words. They show that pauses between determiners and nouns occur quite often in speech production – a finding they interpret as problematic for lexically driven models of speech production (e.g. Kempen and Hoenkamp 1987; Levelt 1989). These models, they argue, cannot accommodate speech pauses between determiners and nouns, because they predict that the formulation of the entire nominal phrase (based on the grammatical properties of the retrieved lexical head) is completed prior to articulation. If this is the case, how, then, can the occurrence of speech pauses between the determiner and the noun be explained, if not as a tip-of-the-tongue phenomenon? For Schilperoord and Verhagen, there is only one way out of this dilemma: They view determiner selection as a process independent from noun selection. The determiner, more precisely the construction [*de/het/een* + ... + N]¹⁴ in Dutch is seen as carrying a meaning of its own, which can be paraphrased as “grounded entity” (2006: 152). This, as Langacker (1990: 321) articulates it, means that a single noun names “a ›type‹ of thing”, whereas the combination of a determiner and a noun designates “an ›instance‹ of that type”. Therefore, as any form-meaning pair can be activated and re-

¹⁴ For an overview of the different forms of determiners in Dutch, see footnote 17.



trieved independently, determiners can be seen as part of an autonomous construction.

Based on the empirical results of the present study, which of these opposing points of view should we subscribe to? The intonation pattern of alterations of gender clearly gives reason to subscribe to the traditional psycholinguistic position (e.g. Levelt 1989). In contrast to languages with no gender distinction, like English, the grammatical form of the determiner in German depends on the gender of the noun. As the determiner precedes the noun in German, the speaker needs to access the gender information of the following noun quite early in order to be able to produce the correct form of the determiner. Of course, the speaker also needs information about case, number, and definiteness to be able to produce the correct grammatical form of the determiner. However, these grammatical categories are – unlike gender – not intrinsic to the noun but determined by “external” sources, namely by the context or referent during conceptual stages of speech production (definiteness and number) and the preposition (case). Thus, this information cannot be provided by the retrieval of lemma information alone. In English, for instance, speakers can use determiners as fillers while searching for the next item due, whereas the agreement according to gender in German nominal phrases forces the speaker to decide on the following noun before producing the corresponding form of the determiner (cf. Birkner et al. 2010, 2012). Therefore, as the speaker needs to have the noun in mind when uttering the determiner, the substitution of this determiner with a determiner of a different gender is in fact a covert substitution of the lemma on the level of speech planning. In order to capture the cognitive processes that underlie this type of self-repair, I will refer to it as *lemma substitution*.¹⁵ As my results suggest, the cognitive processes involved in this lemma substitution, which is not necessarily carried out in other types of alteration or repetitions of the determiner, have a correlate on the phonetic level, namely a low falling intonation pattern of the repaired segment. This intonation pattern, which predominantly occurs in alterations of gender, supports the hypothesis that the determiner selection is dependent on the noun in German.

However, while it seems to be quite evident that there is indeed a link between lemma substitutions and falling intonation, nothing definite can be said about the exact processes that cause the falling intonation. It is thus still undetermined whether it is the detection of the initially selected problematic

¹⁵ Lemma substitutions also exist within nominal phrases that are not embedded in a prepositional phrase, of course. However, these cases are not considered here.





What prosody reveals about the speaker's cognition

65

lemma based on monitoring processes, the subsequent abandonment of the original lemma, the selection of the new lemma, a combination of these processes, or another process involved in the cognitive processing or in the interactional management of lemma substitutions that is responsible for the falling intonation pattern.

However, it is not only intonation that gives reason to endorse the traditional position, but also the quantitative distribution of repair types. To play devil's advocate, let us consider: Which evidence might be most convincing in supporting the assumption that the determiner is indeed selected independently from the noun?

1. *The form of the determiner is selected randomly.*

This is obviously not the case, as the overall occurrence of substitutions of determiners in prepositional phrases is rare (only 58 instances in 45 hours of speech). This type of self-repair should occur much more often if determiners were selected randomly. Moreover, there are more repetitions of the determiner than substitutions (108 vs. 58 instances), which should not have occurred if the above statement were true.

2. *There is a "default article" in German: In substitutions with alteration of gender, frequent forms of the determiner are more often replaced by rarer forms than vice versa.*

If speakers of German had a "default article" at their disposal that could be retrieved independently and used as a filler when the noun to follow had not yet been decided, the homomorphous forms of the masculine and neuter determiner (with dative prepositions, *dem* (the-M/N), *einem* (a-M/N) etc.) would certainly be good candidates for this purpose. By using these high frequency forms, speakers would run a lower risk of having to substitute the determiner retrospectively.¹⁶ Thus, if this second statement were true, default articles should occur more frequently in the repaired segment than the feminine forms of the determiner (with dative prepositions: *der* (the-F), *einer* (a-F) etc.). However, this is not the case in this quantitatively limited corpus: The respective forms of the determiner are substituted with the same fre-

¹⁶ As a matter of fact, these considerations depend on the overall distribution of gender in German nouns. According to an estimation by Bauch (1971), 50% of the nouns in German are masculine, 30% are feminine and 20% are of neuter gender. Supposing that nouns of all genders are selected equally frequently, this would mean that the use of the homomorphous forms *dem* (the-M/N) or *einem* (a-M/N) within dative prepositional phrases would cover about 70% of the subsequently selected nouns – self-repair would become necessary in only 30% of all the cases.



quency (15 masculine/neuter and 15 feminine forms). In the remaining three cases, two neuter forms and one masculine form (all occurring with accusative prepositions) are altered.

We can see that these quantitative distributions, too, call into question the hypothesis of an independent determiner selection put forth by Schilperoord and Verhagen (2006), at least in the case of German, which has three different gender categories.¹⁷

In this section, I suggested that the falling intonation pattern in substitutions with alteration of gender relies on the lemma substitution that, contrary to the other types of alteration, must always be carried out in these instances of repair. With respect to the current discussion about the cognitive processes involved in the production of determiners, this finding (as well as the quantitative distributions of repair types) gives reason to assume that the determiner selection in German, more precisely the selection of the gender form of the determiner, is dependent on the noun.

There are several other questions, regarding, for example, the stages of lemma access in German, that could be addressed on the basis of the collection presented in this chapter. Although these interesting issues cannot be treated here, this first part shows that an intonation analysis of naturally occurring talk can help shed light on questions of psycholinguistic research.

5.2. Interactional perspective

As was argued in the previous section, the process of lemma substitution on the level of speech planning is linked to the intonation pattern of the repaired segment. Let us now turn to the question of how this intonation pattern can be relevant for the interaction between speaker and recipient.

Every instance of self-repair must be designed in such a way as to fulfill two central interactional functions: It has to make clear to the hearer (1) what

¹⁷ Dutch determiners have three gender categories (masculine, feminine, and neuter) as well. But contrary to German, Dutch has relatively few determiner forms: only two different forms of the singular definite article (*de* for masculine/feminine nouns, *het* for neuter nouns). For all genders, there is only one form of the singular indefinite article (*een*) and only one form of the plural definite article, which is homomorphous to the form of the singular masculine/feminine definite article (*de*). One might conjecture, although Schilperoord and Verhagen (2006) do not address this issue, that *de* functions as a “default article” which, due to its high frequency of occurrence in Dutch, is used as a filler item that can be retrieved independently from the noun to follow – without running a high risk that a substitution of the determiner will turn out to be necessary retrospectively.



is being repaired and (2) how the repairing segment is linked to the repaired segment. Levelt and Cutler (1983: 212) describe the use of accents on the repairing segment in lexical repair in Dutch to meet these two requirements. Their study shows that the prosodic marking of the lexical item in the repairing segment refers back in time and retrospectively shows the hearer that the previous lexical item is rejected by establishing a prosodic contrast. However, the intonational characteristics of lemma substitutions as found in the present study cannot fulfill the same “marking” function as accents in lexical repair. In contrast to the prosodic marking in lexical substitutions, the prosodic characteristics associated with alterations of gender are to be found on the repaired segment and not on the repairing segment. Therefore, as the falling intonation is produced before the syntactic alteration is articulated, its function cannot be retrospective, but only prospective in character.

Therefore, it seems likely that the intonation pattern in lemma substitutions has a projective function. From the hearer's perspective, the gender of the determiner restricts the number of possible following nouns to a certain grammatical class of nouns. In other words, the determiner conveys information about the gender of the noun through syntactic projection. It is this strong projective potential that characterizes determiners in German (see Section 2.2 on the properties of projective repair). However, the intonation pattern in alterations of gender additionally conveys information about the type of repair to follow. From the hearer's point of view, falling intonation projects, or at least potentially projects, the substitution of the originally intended noun – a far-reaching change in the structure of the emerging syntactic gestalt. Auer (1996: 75) observes that intonation “brings in its local flexibility to revise and adjust [syntactic] gestalts while they are being put to speech” in projecting turn-continuations. This finding seems to hold for the interplay of prosody and syntax in certain types of self-repair as well: In alterations of gender, the prosodic projection modifies the syntactic projection. The intonation pattern tells the listener that the noun will not be the one syntactically projected by the determiner, but will be a noun of a different gender.

The intonation patterns associated with the different types of repair also manifest similarities with regard to turn-final intonation contours (e.g. Chafe 1980; Ford and Thompson 1996; Selting 1996; Szczepek Reed 2004; Couper-Kuhlen, to appear, for observations on turn-final intonation in English and German). Level intonation at the end of a turn-constructive unit (Sacks, Schegloff, and Jefferson 1974) can be used to project turn continuation to the listener. Similarly, level intonation on the repaired segment projects global continuation of the utterance, despite local production problems that need to



be repaired. Falling intonation, on the other hand, can be used to signal turn completeness – the speaker is ready to relinquish the floor. We might argue, then, that falling intonation on the repaired segment projects that the speaker is about to abandon the intended noun. The speaker does not show readiness for speaker transition in the latter case, but in both examples the function of falling intonation may have to do with signaling “abandonment” – abandoning the turn and abandoning a syntactic constituent, respectively.

6. Conclusion and outlook

The analysis presented in this chapter has shed some light on the relationship between prosody and syntax. The conversational action of self-repair provides yet one more piece of evidence for a link between syntax and intonation. This becomes obvious in the fine coordination of changes in gender and falling pitch on the repaired segment in lemma substitutions. Thus, we can conclude that self-repair is a highly organized phenomenon – not only, as previous studies have shown, on the sequential and the syntactic level – but also on the intonational level.

This finding strongly argues for an integration of this intonational aspect into an “Interactional Grammar” (Bergmann et al. this volume). Although there are certainly factors other than the cognitive processes involved in lemma substitutions that exert influence on the intonational design of the repaired segment, the evidence presented in this chapter suggests that speakers possess some kind of implicit knowledge about the intonation pattern necessary to produce and consistently reproduce substitutions with alteration of gender.

At this stage of research, we can only speculate about the possible relevance of the described intonation pattern for the co-participants involved in the activity of self-repair. In this respect, one interesting question to be pursued in future research could be whether the falling intonation pattern in lemma substitutions has a priming effect that can help the listener process the grammatical changes involved in this type of repair.

Acknowledgments

This chapter was written during a stay at UC Santa Barbara as a visiting scholar from January to August 2010. I would like to thank the DAAD for supporting this stay with a scholarship. I am grateful to Sandy Thompson for very helpful discussions about the syntax and the prosody of self-repair and to Jack Du Bois and the Dialogic Syntax Working Group for their valuable feedback dur-



What prosody reveals about the speaker's cognition

69

ing my stay. Furthermore, I would like to thank Bracha Nir for her statistical advice and Peter Auer, Pia Bergmann, Jana Brenning, Theodoros Papantoniou, Elisabeth Reber, and Sandy Thompson for various insightful comments on a previous draft of this chapter. All remaining errors are of course mine.

Appendix

Transcription conventions (GAT 2, Selting et al. 2009)

[]	overlap and simultaneous talk
[]	
°h / °hh	inbreaths of appr. 0.2–0.5/0.5–0.8 sec. duration
(.)	micro pause, estimated, up to 0.2 sec. duration appr.
(-) / (--)	estimated pauses of appr. 0.2–0.5/0.5–0.8 sec. duration
und_äh	cliticization within units
äh	hesitation marker
=	fast, immediate continuation with a new turn or segment (latching)
: / ::	lengthening, by about 0.2–0.5/0.5–0.8 sec.
'	cut-off by glottal closure
SYLlable	focus accent
sYllable	secondary accent
?	rising to high (final pitch movement of intonation phrase)
,	rising to mid (final pitch movement of intonation phrase)
-	level (final pitch movement of intonation phrase)
;	falling to mid (final pitch movement of intonation phrase)
.	falling to low (final pitch movement of intonation phrase)

Abbreviations

ACC	Accusative
CL	Cliticized
DAT	Dative
DEF	Definite
F	Feminine
INDEF	Indefinite
M	Masculine
N	Neuter
PL	Plural
PTCL	Particle
SG	Singular
TAG	Question tag



References

- Auer, P. 1996 On the prosody and syntax of turn-continuations. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in Conversation*, 57–100. Cambridge: Cambridge University Press.
- Auer, P. 2005a Delayed self-repairs as a structuring device for complex turns in conversation. In: A. Hakulinen and M. Selting (eds.), *Syntax and Lexis in Conversation*, 75–102. Amsterdam: John Benjamins.
- Auer, P. 2005b Projection in interaction and projection in grammar. *Text* 25 (1): 7–36.
- Barth-Weingarten, D., E. Reber and M. Selting (eds.) 2010 *Prosody in Interaction*. Amsterdam: John Benjamins.
- Bauch, H.-J. 1971 Zum Informationsgehalt der Kategorie Genus im Deutschen, Englischen und Polnischen. *Wissenschaftliche Zeitschrift der Universität Rostock, Gesellschaft- und Sprachwissenschaftliche Reihe*, XX. Jahrgang: 411–419. Rostock: Universität Rostock.
- Berg, T. 1986 The aftermath of error occurrence: Psycholinguistic evidence from cut-offs. *Language and Communication* 6: 195–213.
- Bergmann, P., J. Brenning, M. Pfeiffer and E. Reber this volume Towards an Interactional Grammar.
- Birkner, K. this volume Prosodic formats of relative clauses in spoken German.
- Birkner, K., S. Henricson, C. Lindholm and M. Pfeiffer 2010 Retraction patterns and 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.
- Boersma, P. and D. Weenink 2010 Praat: Doing phonetics by computer [Computer program]. Version 5.1.18, URL: <http://www.praat.org/> (accessed March 5, 2012).
- Bybee, J. 2006 From usage to grammar: The mind's response to repetition. *Language* 82: 529–551.
- Bybee, J. 2007 *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.
- Bybee, J. 2010 *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Chafe, W. L. 1980 The deployment of consciousness in the production of a narrative. In: W. L. Chafe (ed.), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, 9–50. Norwood, NJ: Ablex.
- Couper-Kuhlen, E. to appear Some truths and untruths about final intonation in conversational questions. In: J. P. de Ruiter (ed.), *Questions*. Cambridge: Cambridge University Press.
- Couper-Kuhlen, E. and M. Selting 1996 Towards an interactional perspective on prosody and a prosodic perspective on interaction. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in Conversation*, 11–56. Cambridge: Cambridge University Press.
- Couper-Kuhlen, E. and M. Selting 2001 Introducing Interactional Linguistics. In: M. Selting and E. Couper-Kuhlen (eds.), *Studies in Interactional Linguistics*, 1–22. Amsterdam: John Benjamins.





What prosody reveals about the speaker's cognition

71

- Couper-Kuhlen, E. and C. E. Ford (eds.) 2004 *Sound Patterns in Interaction. Cross-Linguistic Studies from Conversation*. Amsterdam: John Benjamins.
- Cutler, A. and M. Pearson 1986 On the analysis of prosodic turn-taking cues. In: C. Johns-Lewis (ed.), *Intonation and discourse*, 139–155. London: Croom Helm.
- Gallmann, P. 2006 Der Satz. In: M. Wermke, K. Kunzel-Razum and W. Scholze-Stubenrecht (eds.), *Duden. Die Grammatik*, 773–1066. Mannheim: Dudenverlag.
- Ford, C. E., B. A. Fox and S. A. Thompson 1996 Practices in the construction of turns: The “TCU” revisited. *Pragmatics* 6: 427–454.
- Ford, C. E. and S. A. Thompson 1996 Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 185–237. Cambridge: Cambridge University Press.
- Fox, B. A. and R. Jaspersion 1995 A syntactic exploration of repair in English conversation. In: P. W. Davis (ed.), *Alternative Linguistics. Descriptive and Theoretical Modes*, 77–134. Amsterdam: John Benjamins.
- Fox, B. A., M. Hayashi and R. Jaspersion 1996 Resources and repair: A cross-linguistic study of syntax and 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. Uhmman 2009 Morpho-syntactic resources for the organization of same-turn self-repair: Cross-Linguistic variation in English, German and Hebrew. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 245–291.
- Fox, B. A., F. Wouk, M. Hayashi, S. Fincke, L. Tao, M.-L. Sorjonen, M. Laakso and W. Flores Hernandez 2009 A cross-linguistic investigation of the site of initiation in same-turn self-repair. In: J. Sidnell (ed.), *Comparative Studies in Conversation Analysis*, 60–103. Cambridge: Cambridge University Press.
- Jaspersion, R. 2002 Some linguistic aspects of closure cut-off. In: C. E. Ford, B. A. Fox and S. A. Thompson (eds.), *The Language of Turn and Sequence*, 257–286. Oxford: Oxford University Press.
- Kempen, G. and E. Hoenkamp 1987 An incremental procedural grammar for sentence formulation. *Cognitive Science* 11: 201–257.
- Langacker, R. W. 1987 *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. 1990 *Concept, Image and Symbol. The Cognitive Basis of Grammar*. Berlin: de Gruyter.
- Langacker, R. W. 2001 Discourse in Cognitive Grammar. *Cognitive Linguistics* 12: 143–188.
- Levelt, W. J. M. 1983 Monitoring and self-repair in speech. *Cognition* 14: 41–104.
- Levelt, W. J. M. 1989 *Speaking: From Intention to Articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J. M. and A. Cutler 1983 Prosodic marking in speech repair. *Journal of Semantics* 2: 205–217.
- Papantoniou, T. 2010 Zur zweitsprachlichen Spezifik von Signalisierungsmitteln bei Sprachproduktionsproblemen: Die Verwendung des Heckenausdrucks “irgendwie” in der mündlichen Kommunikation. In: W.-D. Krause (ed.), *Das Fremde und der Text. Fremdsprachige Kommunikation und ihre Ergebnisse*, 119–152. Potsdam: Universitätsverlag.





- Pfeiffer, M. 2008 Reparaturen vor syntaktischem Abschluss im gesprochenen Deutsch. Unpublished Master's Thesis, University of Freiburg.
- 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.
- Plug, L. 2006 Speed and reduction in postpositioned self-initiated self-repair. *York Papers in Linguistics*, Series 2, 6: 143–162.
- Rieger, C. L. 2003 Repetitions as self-repair strategies in English and German conversations. *Journal of Pragmatics* 35: 47–69.
- 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. 1979 The relevance of repair to syntax-for-conversation. In: T. Givón (ed.), *Syntax and Semantics 12: Discourse and Syntax*, 261–286. New York: Academic Press.
- Schegloff, E. A., G. Jefferson and H. Sacks 1977 The preference for self-correction in the organization of repair in conversation. *Language* 53: 361–382.
- Schilperoord, J. and A. Verhagen 2006 Grammar and language production: Where do function words come from? In: J. Luchjenbroers (ed.), *Cognitive Linguistics Investigations. Across languages, fields and philosophical boundaries*, 139–168. Amsterdam: John Benjamins.
- Selting, M. 1995a *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Max Niemeyer.
- Selting, M. 1995b Der »mögliche Satz« als interaktiv relevante syntaktische Kategorie. *Linguistische Berichte* 158: 298–325.
- Selting, M. 1996 On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6: 357–388.
- Selting, M. 2000 The construction of units in conversational talk. *Language in Society* 29: 477–517.
- Selting, M. 2001 Fragments of units as deviant cases of unit production in conversational talk. In: M. Selting and E. Couper-Kuhlen (eds.), *Studies in Interactional Linguistics*, 229–258. Amsterdam: John Benjamins.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzlufft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock and S. Uhmann 2009 Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.
- Selting, M. and E. Couper-Kuhlen 2001 Forschungsprogramm »Interaktionale Linguistik«. *Linguistische Berichte* 187: 257–287.
- Szczepek Reed, B. 2004 Turn-final intonation in English. In: E. Couper-Kuhlen and C. E. Ford (eds.), *Sound Patterns in Interaction*, 97–118. Amsterdam: John Benjamins.
- Uhmann, S. 1992 Contextualizing relevance: On some forms and functions of speech rate changes in everyday conversation. In: P. Auer and A. Di Luzio (eds.), *The Contextualization of Language*, 297–335. Amsterdam: John Benjamins.
- Uhmann, S. 2001 Some arguments for the relevance of syntax to same-sentence





What prosody reveals about the speaker's cognition

73

- self-repair in everyday German conversation. In: E. Couper-Kuhlen and M. Selting (eds.), *Studies in Interactional Linguistics*, 373–404. Amsterdam: John Benjamins.
- Uhmann, S. 2006 Grammatik und Interaktion: Form follows function? Function follows form? In: A. Deppermann, R. Fiehler and T. Spranz-Fogasy (eds.), *Grammatik und Interaktion*, 179–201. Radolfzell: Verlag für Gesprächsforschung.
- Wells, B. and S. Peppé 1996 Ending up in Ulster: Prosody and turn-taking in English dialects. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in Conversation*, 101–130. Cambridge: Cambridge University Press
- Wouk, F. 2005 The syntax of repair in Indonesian. *Discourse Studies* 7: 237–258.
- Zhang, W. and K.-K. Luke 2010 Sentence planning and execution in conversation: Evidence from item replacement in Chinese. Unpublished manuscript.



Jana Brenning
University of Freiburg/ENS Lyon

Speakers' orientation to the nucleus accent in syntactic co-constructions

1. Introduction

The syntactic co-construction of one single *syntactic gestalt* (Auer 1996) by two (or more) speakers is an everyday phenomenon in spoken interaction. It has been the subject of various studies on different languages in Conversation Analysis and Interactional Linguistics (cf. Bolden 2003; Hayashi 2003; Helasvuoto 2004; Iwasaki 2009; Lerner 1991, 1996; Mondada 1999; Müller and Kläger 2010; Sacks 1992; Szczepek 2000a, b), but as of yet no systematic research has been done on collaborative turn sequences in spoken German.¹ In the present study, the term *syntactic co-construction* will be used to refer to the completion of a preliminary speaker's syntactic gestalt by another speaker. Sacks was the first to discuss this phenomenon in his lectures (cf. 1992: 56f.; 320f.) as collaborative utterances, collaboratively built sentences, or joint production. It has also been referred to as anticipatory completion or collaborative turn sequence (Lerner 1996, 2004), coénonciation (Jeanneret 1999), and co-construction (Ono and Thompson 1995; Helasvuoto 2004). The present study adopts the term co-construction, as this approach avoids any functional presumption and is based on the structural characteristics of the instances in our corpus. In contrast to this, the term "collaborative" that is frequently used in the literature suggests that the completion by another speaker cannot be turn competitive.²

This chapter will focus on *terminal item completion* as one possibility to preempt another speaker's syntactic project. The term terminal item completion, which is adapted from Lerner (1996), refers to syntactic co-constructions in which the completion by another speaker starts at the beginning of the turn transition space. Our analysis revealed the importance of integra-

¹ However, instances of syntactic co-constructions in German are mentioned in several studies to illustrate, for example, the interactional reality of projections (Auer 2007: 105) or to provide evidence for participants' orientation to special constructions such as pseudo-cleft constructions in German (Günthner 2006: 82).

² Even if the sequential and functional analysis of my collection showed this assumption to be true for many sequences, there are also cases in which the other speaker's completion has a competitive character.



ting the prosodic design of the co-construction into the study in order to describe the linguistic categories which are relevant for the participants within this intra-turn speaker change. Thus far, only a few studies on English have taken into account the prosody of syntactic co-constructions (Local 2005; Szczepek 2000a, 2006). However, as we will show in this chapter, it was discovered that speakers orient to a projected possible position of the nucleus accent syllable to start their completion of a previous speaker's emergent syntactic project.

We first give a brief overview on projection in spoken language, as it is a key concept in the analysis of co-constructions, thereby explaining how recipients can anticipate the completion of another speaker's syntactic gestalt (section 1.1). We will then go on to review Lerner's previous findings and the notion of the turn transition space giving a first instance of terminal item completion in German in order to illustrate the phenomenon (section 1.2.). After a short description of the data and the method of analysis (section 1.3.), we will raise the question of how the position of the intra-turn speaker change within terminal item completion in German can be defined (section 2). In section 3, we will present several instances of terminal item completion in which the second speaker comes in articulating the nucleus accent syllable. These analyses suggest that incoming speakers orient towards a projected possible position of the nucleus accent syllable to begin the completion of a preliminary speaker's emerging syntactic gestalt. Based on Uhmann's rules on focus projection (1991), we will show how syntax can provide an explanation for the projection of a possible position of the nucleus accent syllable (section 4). Finally, instances in which the second speaker comes in after the nucleus accent syllable already has been articulated by the first speaker will provide further insight. Even if, at first glance, they seem to contradict the hypothesis that the completion is oriented to a projected possible position of the nucleus accent syllable, these instances will also suggest that a possible position of the nucleus accent syllable is relevant for the participants. To summarize, we will first argue for a description of terminal item completion in German, including the prosodic design of the co-constructions. Second, these results will be shown to point towards a description of the grammar of spoken language that has to equally take into account prosody and syntax.

1.1. Projection in spoken language

As has been widely shown in Interactional Linguistics (cf. Auer 2005; Günthner 2009; Hausendorf 2007; Günthner and Hopper 2010; Mondada 2006), we cannot analyze spoken language without seriously taking into ac-



count its temporal organization. The crucial difference between spoken language and written language is its moment-by-moment emergence in real-time. For this reason, we have to examine emerging syntactic gestalts from an on-line perspective, always keeping in mind the irreversible, transitory and dialogical character of spoken language (Auer 2007, 2009).

The syntactic completion of a preliminary speaker's utterance provides strong evidence that speakers orient to the ongoing projections built up by an emerging syntactic gestalt, while monitoring closely what the current speaker is going to say. The co-construction of a syntactic unit by two (or more) speakers is therefore a conversational practice that proves that participants closely coordinate their actions. Co-constructions are only possible if the "emerging syntactic constructions [...] are processed with only a short delay by the recipient" (Auer 2009: 3).

It is projection on a syntactic (cf. Helasvuo 2004, Lerner 1996), prosodic, semanto-pragmatic, and visual level (cf. Bolden 2003; Iwasaki 2009) that enables speakers to anticipate a possible completion of the ongoing utterance. In this chapter, we are going to focus on syntactic and prosodic projections that are built up by the first speaker's emerging syntactic gestalt.

1.2. Syntactic co-construction through terminal item completion

Lerner, who was the first to make an extensive study of the joint production of one (syntactic) unit by two speakers (cf. Lerner 1991: 441), describes terminal item completion as one possibility of anticipatory completion by a second speaker.

The final word or two – the terminal item of a TCU – can be co-produced by a recipient [...] or actually co-opted by an interposing speaker. [...] It is the turn-taking mandated orientation to the imminent possible closure of the turn and with it the opening up at the point of pre-completion of the relevance of transition (and not any particular turn-constructional format) that furnishes an opportunity for completion. Terminal item completion begins in the same place in current turn (its pre-completion) as the pre-beginning of a next turn could begin, and therefore, in one sense, might be thought of as an alternative to it. (Lerner 1996: 256)

The emergent turn of the current speaker projects a possible point of completion and thus creates an opportunity for speaker change before an actual transition relevance place (TRP; Sacks, Schegloff, and Jefferson 1974) has been reached. Thus, according to Lerner, the terminal item completion occurs in the turn transition space that he describes referring to Schegloff's discussion of the pre-completion position in a turn construction unit (TCU) (1996: 84). Accordingly, terminal item completion might be an alternative to



the early start of a next turn by another speaker, as a second speaker comes in immediately prior to the first speaker reaching a possible TRP. This phenomenon has also been found to be true in instances of German, as we will see in the example given at the end of this section. It is the orientation to the projected possible closure of the turn that provides an occasion for a co-constructed completion by a second speaker.

However, we wondered if this point of intra-turn speaker change could be described in more linguistic detail. One clue might be the occurrence of a pitch peak, which, according to Schegloff (1988: 144; 1996: 84), can characterize the pre-completion position in a TCU:

A pitch peak thus can project intended turn completion at the next grammatically possible completion point. In doing so it can also open the "transition relevance space", the stretch of time in which transition from current to next speaker is properly done. (Schegloff 1988: 144)

Schegloff points out that speakers orient to the occurrence of a pitch peak by showing that orderly phenomena like early started next turns, rush throughs, or continuers often occur after a pitch peak has been articulated. However, he makes clear that a pitch peak can only be one possible resource to identify "designed possible completion at next grammatically possible completion" and claims that "there are various resources, and that we know relatively few of them" (1996: 84). Nevertheless, confirming the important role of a last pitch peak, Selting found in her research on the prosody of German (1995) that an accent near the end of a syntactic unit can serve as an additional local contextualization cue for a projected turn ending (1995: 201). Moreover, in their study of overlap in English, Wells and Macfarlane (1998) investigated the resources that allow participants to anticipate possible TRPs, and they also identified the crucial role of a pitch peak: they claim that "[...] the last major accent syllable is the earliest point at which turn exchange mechanisms can unproblematically come into play" (1998: 281).³

Thus, taking into account Lerner's analysis and the interactional relevance of the pitch peak found in previous research, it will be of interest in the present study to examine the position of the last pitch peak within co-constructed completions. Since, according to Lerner, terminal item completions occur in the turn transition space, we may expect that the second speaker comes in after a possible last pitch peak has occurred. Before dealing with

³ Wells and Macfarlane show that the TRP is projected by the specific phonetic design of the accent and claim that speakers cannot rely on the syntax and the information structure of an utterance to identify an accent as the final accent (1998: 290).



78

Jana Brenning

this question in detail in sections 2 and 3 of this chapter, we will first discuss a prototypical instance for terminal item completion in German in order to illustrate this phenomenon and its characteristics in German. In this sequence, speaker A tells the interviewer (I) and speaker C a story about his friend Holger, who does not drink alcohol during Lent. Speaker A thinks he should try to follow his example.

(1) Leistung (mu06)

- 1 A: =ich sollte an un für sich AUCH, (0.6)
in principle I should also
 2 meinem freund hOlger (.) FOLgen-
follow (the example of) my friend Holger
 3 des is ja: ein ganz H:ARter bÜrsche,
he is PTCL quite a tough guy
 4 der macht jetzt seit JAHren schon-
for years he hasn't had
 5 ab ascherMITtwoch (.) bis karFREItag, (.)
from Ash Wednesday to Good Friday
 6 kein tropfen ALkohol;
a drop of alcohol
 7 C: [ja (.)ich weiß-
yes i know
 8 A: [(schnalzt))
((clicks his tongue))
 -> 9 u:nd des is schOn also=äh:
and that is quite well uh
 ->10 I: a LEISTung;=a,
an achievement
 ->11 A: des is scho a LEISTung;=a
that is quite an achievement
 12 ma konn also auf alle fälle SAgen,
it's obvious to say that
 13 er ist be[stimmt k(h) kein SÄUfer;
he is most certainly n not an alcoholic

On the syntactic level, the emerging syntactic gestalt in line 9 *und des is schon also äh:* (and that is quite well uh) is interrupted before a possible point of syntactic completion has been reached. There is an open syntactic projection of the copula *is* which could be fulfilled by a predicative in the form of a noun phrase or an adjective when speaker A starts to hesitate. Moreover, line 9 does not constitute a complete intonation phrase, but is broken off without the nucleus accent syllable, i.e. the obligatory element of a complete intonation phrase⁴, having occurred. Thus, the speaker does not start his com-

⁴ The British School of intonation as well as autosegmental approaches agree that an intonation phrase (IP) has one obligatory accent (cf. Cruttenden 1997; Grice



pletion in the turn transition space as described by Lerner. Accordingly, the emergent unit is neither complete on the syntactic level nor on the prosodic level. After the lengthened filled pause at the end of line 9 (*also=äb:*), the interviewer comes in and adds a noun phrase *a LEISTung*, that constitutes one possibility to fulfill the projection. She completes the syntactic gestalt and the intonation phrase with falling intonation, and the noun phrase she suggests as a candidate completion carries the nucleus accent (which is transcribed in capital letters). In line 11, the first speaker A ratifies this completion, repeating the whole syntactic unit with the same prosodic design. As Lerner observes, this repetition is one possibility to ratify a candidate completion offered by another speaker (2004: 231).⁵

1.3. Method and Data

This study is based on a data set of 154 co-constructed syntactic completions that are taken from 16 hours of everyday conversation from two different corpora:

- Ten hours are audio recordings of informal dialect interviews with 3–4 participants that were recorded for the DFG-funded research project *Dialektintonation* (Universities of Freiburg and Potsdam, 1998–2005) under the direction of Peter Auer and Margret Selting.⁶ The interviews chosen for this study were conducted in rather informal settings and are comparable to everyday conversation. The participants know each other (i.e., they are family and/or friends) and speak Munich or Cologne dialect.
- Six hours are audio and video recordings of informal conversations among 2–5 friends or family members recorded and transcribed by the author. The speakers in these recordings speak mostly standard German.

Based on repeated listening to all recordings, we included in the analysis every completion by another speaker that occurred before a possible point of syntactic completion.⁷ We then analyzed and coded these sequences ac-

and Baumann 2000: 3; Ladd 1996: 209; Pheby 1975: 63; Uhmann 1991: 123). Following Pierrehumbert (1980), the nucleus accent is defined as the last accent of an intonation phrase independently of its prominence.

⁵ Oloff's (2011) recent multimodal study also found this to be true in French data.

⁶ I would like to thank the members of this project – the two principal investigators as well as Peter Gilles and Jörg Peters – for kindly providing me with the data.

⁷ The scope of this study is restricted to co-constructed completions. Co-constructed syntactic expansions are also a possibility of syntactic co-construction (cf. Szczeppek 2000a).



according to formal and sequential characteristics, using the tools provided by the online data base [moca] (Multimodal Oral Corpus Administration)⁸. As a result, this chapter focuses on 95 terminal item completions that were found to constitute the major type of co-constructed syntactic completions in the collection.⁹

The transcription of the examples follows GAT 2-conventions (cf. Seltin et al. 2009). Whenever useful, a translation will be offered to allow for an understanding of the syntactic structure, followed by an idiomatic translation in English. In some cases, less idiomatic translations have been chosen to give an impression of the German syntactic order. The context of the collaborative turn sequences is only provided in idiomatic translations.

We transcribed the prosody of all instances based on auditive criteria. Whenever possible, the auditive impression was verified using the phonetic software *praat*.¹⁰ However, this proved difficult within syntactic co-constructions, as they often occur in overlap (cf. Examples 2, 3, 6, 7, 8, 10).

2. Terminal item completion in German

In our collection of 154 co-constructed syntactic completions, 95 collaborative turn sequences correspond at first glance to Lerner's terminal item completion. As was shown in the sequence *Leistung* discussed above, the second speaker comes in immediately prior to a possible point of syntactic completion, i.e., there is only one more constituent projected that the second speaker has to articulate to complete the syntactic gestalt.

According to Lerner, the "terminal item" delivered by the second speaker corresponds to the final word or the final two words of a TCU (see citation above, Lerner 1996: 256); however, he does not specify this category further. Thus, he does not state whether or not these words correspond to the actual or to the projected completion of the TCU. In examining German data, we find this description only to be true if we consider the projections that are built up by the first speaker's utterance to define the "terminal item" of the syntactic gestalt and not the actual co-constructed completion. At the point in time where the second speaker comes in, we can always anticipate a syn-

⁸ <http://moca.phil2.uni-freiburg.de/web/index.html> (accessed April 18, 2011).

⁹ Other formats that occurred in the German data are, for example, co-constructed completions after coordinating conjunctions (e.g., *und* [and]) or within hypotactic sentence structures, even if co-constructions within conditional sentences that are reported to be frequent in English (cf. Szczepek Reed this volume) are only rather rare in my data.

¹⁰ [<http://www.fon.hum.uva.nl/praat/>], (accessed March 30, 2011).





tactic constituent consisting of one or two words which could fulfill the projection, but within the actually articulated completions, the length of this constituent varies considerably. In all of the cases analyzed here, a completion by adding one or two words is possible, as we have seen in Example 1 *Leistung*. Nevertheless, in our data, the second speaker might add more than two words to complete the previous utterance, as can be seen in the following analysis of Example 2. Speaker A tells about the sudden death of her father when two recipients (I and B) complete the previous utterance in overlap.

(2) Arzt (mu04)

- 1 B: aha
ah
(0.4)
- > 3 A: **und mei pApa war keiner der (.)**
and my dad was no one who
and my dad was not somebody who
- > 4 I: **zum [ARZT geh]t=[oder]-**
to-CLIT doctor goes TAG
goes to the doctor right
- > 5 B: **[JAMmert]-**
whine-3SG
whines
- 6 A: [der ir][gendwie geJAMmert hätt;=
who anyhow would have whined
- 7 =<<p>oder S]ONST was.>
or anything else
- 8 I: [hm_hm hm_hm]

This sequence is particularly interesting because two recipients suggest a completion in overlap after a short pause by A. Speaker B and the interviewer (I), who are mother and daughter, choose completions that are syntactically and semantically different, but both respond to the projection in play (l. 3). The relative pronoun in line 3 (*der*) opens up a new projection that is very wide and allows for a variable amount of constituents to be placed in the middlefield of the emerging syntactic structure, but the gestalt can only be closed by articulation of the right sentence bracket. Both speakers fulfill the projection of the relative pronoun, but in different ways. While the interviewer proposes a directional adverbial in the form of a prepositional phrase followed by a finite verb (*zum ARZT geht* l. 4), B, who starts a bit later with the completion, brings the syntactic gestalt directly to a possible point of completion by articulating only a finite verb (*JAMmert* l. 5).

Even though both speakers provide syntactically correct and perfectly fitting candidate completions, the number of words that are articulated differ.





Hence, it is unclear whether the number of words provides further insights on the formal structure of the co-constructions that are investigated in this chapter, or if there are instead other perspectives that allow us to describe the syntactic and prosodic design of terminal item completion in German. If we look again at Example 2, we can see that both speakers choose almost the same point in time to begin their completion. How could this locus of speaker change within the emergent syntactic and prosodic unit occurring before a possible TRP be described?

One suggestion could be to analyze this point of intra-turn speaker change referring to the notion of the turn transition space that begins, according to Schegloff (1996: 84), with the occurrence of a pitch peak (amongst other resources that might be exploited). However, neither in Example 1 nor in Example 2 is an accent assigned to the constituent that is articulated before the completion occurs. By contrast, the nucleus accent of the co-constructed intonation phrase is articulated by the second speaker, i.e. before the turn transition space (if it is defined by the occurrence of a pitch peak) has been opened. Therefore, the position of the intra-turn speaker change within terminal item completions in spoken German remains to be specified. Nevertheless, the finding that second speakers articulate the nucleus accent syllable suggests that its position might provide further insight to the structure of co-constructions. This question will be discussed in the following section.

3. Speakers' orientation to the nucleus accent

As we have seen in the examples presented thus far, the incoming speakers articulate the nucleus accent syllable when they complete the syntactic gestalt and the intonation phrase of a previous speaker. In the following analysis, it will be shown that the nucleus accent syllable is indeed an important cue for the organization of the incoming speaker's completion, which speakers orient to in specific ways. We can illustrate the special characteristics of this practice, looking again at examples 1 and 2.

(1)' Leistung (mu06)

9 u:nd des is schOn also=äh:
 and that is quite well uh
10 I: a LEISTung;=a,
 an achievement





(2)' Arzt (mu04)

- 3 A: und mei pApa war keiner der (.)
and my dad was no one who
and my dad was not somebody who
- 4 I: zum [ARZT geh] t=[oder]-
to-CLIT doctor goes TAG
goes to the doctor right
- 5 B: [JAMmert]-
whine-3SG
whines

As has already been mentioned above, it can be seen that the nucleus accent syllable often lies within the completion by the second speaker (*a LEISTung=a* (Ex. 1' l. 10); *zum ARZT geht* / *JAMmert* (Ex. 2' l. 3 and l. 4)). But what exactly is the relationship between the placement of the nucleus accent syllable and the beginning of the co-constructed completion? The following empirical observations, which we will illustrate by reanalyzing examples 1 and 2, can give some insight into how the description of this point of speaker change can be refined. As for our claim that the incoming speaker orients to the nucleus accent syllable, we can observe that it is not crucial that the immediately following word carries the accent syllable. Rather, the nucleus accent syllable in the examples is assigned to the next projected syntactic constituent that is articulated by the incoming speaker. For example, the predicative *a LEISTung* in Example 1 is in the form of an indefinite noun phrase. This constituent corresponds to the "terminal item" of the emerging syntactic gestalt, which is then brought to a possible point of completion.

Example 2, in which speakers B and I choose different completions, allows us to develop our claim further. As B only adds a finite verb, the nucleus accent syllable in B's suggestion does follow directly. In contrast, in I's completion in line 3, the accent is assigned to the noun that is part of the prepositional object of the verb. For this reason, it neither directly succeeds the hesitation in line 2, nor is it assigned to the finite verb, as in B's completion. Nevertheless, Example 2 demonstrates what all these instances have in common: first, the nucleus accent has not been articulated yet, and second, the completing speaker comes in at a point in which the nucleus accent syllable – depending on the lexical choice for the completion – could have followed immediately.

These specific characteristics of the point at which the second speaker comes in can also be observed in the third example. This exchange is taken from a conversation between two friends, Elena and Markus, in which they discuss different CD covers of the band Bad Religion.





84

Jana Brenning

(3) Gezeichnet (Italienurlaub)

8 El: ja
yes
(1.4)
10 und die against [the grAIn,
and the against the grain
11 Ma: [the grAIn=,
the grain
12 =da sind so:(--)MAISkolben-
there are like corncobs
13 [und so_ne paTROne dann [irgendwie;
and like a cartridge then somehow
14 El: [ja,=
yes yes
->15 =aber so_n bIssche:n ähm (--)
but like-CLIT little bit uh
but a little bit like uh
->16 Ma: ja so ge[ZEICHnet;
PTCL like drawn
like drawn
->17 El: [stiliSIERT jaJa (.) auch hm- (---)
stylized yeah=yeah(.) also hm
18 Ma: und die no control isch so BUNT,
and the no control is like colored



The co-constructed completion by Marcus in line 16 completes Elena's description by closing the right sentence bracket with a past participle (*ge[ZEICHnet*) that carries the nucleus accent. He retracts to the particle *so* (l. 15), anchoring the completion in the previous emerging gestalt. It is again important to see that the nucleus accent syllable is assigned to the next constituent in the sentence which can possibly fulfill the projections that are in play and thus complete the syntactic and prosodic unit. The two particles *ja* and *so* preceding the past participle could only be emphasized in a very different semantic and pragmatic context (l. 16). For this reason, the accent has to be assigned to the participle. Attention should be paid to the fact that the past participle could easily follow the hesitation directly without the expansion by the particle *ja* (creating a complete syntactic gestalt like *aber so_n bisschen geZEICHnet* (but a little bit like drawn)). Thus, the second speaker again starts his completion in a position in which a constituent carrying the nucleus accent syllable could have been placed.

From the analysis of the first three instances we can summarize the following observations concerning terminal item completions in German: speakers orient to a projected possible position of the nucleus accent syllable, i.e. the position for the constituent which the nucleus accent syllable is assigned to, when they complete a previous speaker's syntactic gestalt. It is this projected





position that is relevant for the beginning of the co-constructed completion and not the position of the actual constituent that carries the accent, because the completion itself can be preceded by a retraction or/and followed by an expansion beyond a possible point of syntactic completion.¹¹

The following extract confirms these observations, and also shows that a completion can be expanded quite far-reaching. Here the interviewer I tells her interlocutor about an illness she had lately and her medical examination in the hospital (l. 1–6).

(4) Halsentzündung (mu04)

- 1 I: und des war aber nUr (.) postinfekTIÖse;
and that was only after the infection
- 2 ich war dann a woche in großHAdern, (.)
then I spent one week in großhadern
- 3 A: mhm (-)
- 4 I: weil alles ABge[checkte wurde=]
because everything was checked
- 5 A: [hm
hm
- 6 I: =und unterSUCHT wurde,=
and examined
- >7 **=aber das war tatsächlich nU:r (.) [ähm**
but that was actually only uhm
but actually it was only
- >8 A: **[e verschlEppte sache**
a spread thing
a spread infection
- vom HALS her;**
from CLIT throat
coming from the throat
- 9 I: des wa:r
that was
- 10 also mein ärztin hats SO umschrieben-
well my doctor described it like that

A starts her anticipatory completion in line 8 after a lengthened vowel (*nu:r* l. 7) and a micro-pause in the interviewer's turn. The demonstrative followed by the copula in line 7 project a predicative to close the emergent syntactic unit (cf. Example 1). What is unique in this example is that A does not choose a short completion, as is the case for most of the completions. (In this example, you could imagine without any problems a syntactic unit like *aber das war tatsächlich nur ne HALSentzündung* (but that was actually only a throat infection)). Instead, she produces the complex noun phrase *e versch-*

¹¹ see Auer (2009) for a discussion of the syntactic basic operations of projection, retraction, and expansion.





leppete sache (l. 8) which is further expanded by a postmodifying prepositional phrase *vom HALS her*. In this way, the candidate completion is expanded after a possible point of syntactic completion has already been reached.

This again demonstrates that we must start from projections in analyzing terminal item completions. Even if this completion is, at first glance, quite long for a “terminal item” (especially if we consider it as the last one or two words), the second speaker comes in at the same point in time as the co-constructing speakers in Example 1, 2 and 3. In other words, second speakers enter at the point in which the syntactic projection allows for the anticipation of the constituent that can bring the emerging syntactic gestalt to completion. Moreover, as we have already seen in section 2 of this chapter, it is not crucial how many words the incoming speaker uses to complete the emerging gestalt, but rather at what point in time he starts the completion. In this example, this point in time is again identical to a projected possible position of the nucleus accent syllable in the first speaker’s ongoing turn, meaning that speaker A could have started her completion with a constituent carrying the nucleus accent syllable (if we imagine a completion by an indefinite noun phrase like *ne HALS entzündung* ‘a throat infection’). Thus, we can conclude from the sequences which have been analyzed so far, that the constituent that is articulated after the speaker change has to carry the nucleus accent. However its syntactic complexity can vary and additionally retractions, particles and hesitation markers can be included in the co-completion (cf. Example 3 *Gezeichnet*).

Until now, we have only analyzed instances in which the second speaker comes in after a hesitation in the talk of the first speaker, and one might suggest that this hesitation provides an alternative explanation for the start of the second speaker. 70 of the 95 terminal item completions in our corpus occur after different types of hesitation phenomena such as pauses, filled pauses, lengthening, or self-repair. Certainly, participants in interaction orient to these interruptions when doing such actions as, for example, helping out another speaker during a word search (cf. Lerner 1996: 261–263).

However, as we will see in the next two instances, co-constructed completions are not only found after the progressivity in the talk of the ongoing speaker has been disrupted (cf. Local 2005: 269). The analysis of these instances without hesitation provides strong evidence that speakers orient to a projected possible position of the nucleus accent syllable to time their completion.

In Example 5 *Schmerzhaft*, the participants discuss a common friend with a pelvic fracture. C and A told I the story about the different stages of the healing process.





(5) Schmerzhaft (mu02)

- 1 C: a:ber wenn natürlich blöde beWegungen oder so machst,
but of course if you make stupid moves or something
->2 **das_s**
that's
->3 I: **SCHMERZhaft** [wahrscheinlich- [=ge,
painful probably PTCL
probably painful right
4 C: [ja
yes
5 A: [naja (0.7) °h
well
6 also am omfang wars SEHR schmerzhaft-
well in the beginning it was very painful
7 ZEIT[weise;
sometimes

In this case, before the co-constructing speaker (I) starts in line 3, two projections are activated. First, the if-sentence in line 1 (*aber wenn natürlich blöde beWegungen oder so machst*) projects a matrix sentence. Second, speaker C continues in line 2 with the pronoun, which – together with the cliticized copula verb¹² (*das_s* (that's)) – projects a predicative to follow (cf. Example 1), which in the end is articulated by incoming speaker I.

This co-constructed completion follows without any preceding hesitation. I starts the completion precisely timed without overlap and the completion is confirmed on the part of the first speaker by a simple acknowledgment token (l. 4 *ja*) when the first point of possible completion has been reached (after the predicative *SCHMERZhaft*). Here, it is again a possible (and in this case actual) position of a constituent that can carry the nucleus accent that is exploited for the beginning of the completion and the intra-turn speaker change. We can observe a similar pattern in the next instance. Speaker A and B talk about a regional meal called *Sauerbraten*, a marinated beef roast that was originally prepared with horse meat.

(6) Pferdefleisch (k09)

- 4 A: dass das in den normAlen gÄngigen brAUhäusern
FERdef_fleisch is;
that they serve horse meat in the normal current breweries
5 B: ne_ne [ich wollt grad sagen das is RINDfleisch;=
no_no i was just going to say that it is beef

¹² One could argue that in line 2 we have to deal with the lengthening of the article's final consonant. It is difficult to exclude this possibility, but in taking the participant's perspective, one can argue that the interviewer treats line 2 as a copula construction that projects a predicative.



88

Jana Brenning

- 6 A: [dat is RINDfleisch;
that is *beef*
- 7 B: =dat is klar;
that is *obvious*
- 8 A: ja
yes
- 9 B: h° also dat denk ich AUCH; (--)
well i also *think that*
- 10 weil(-)ja sehr vIEle äh: auch et ferdefleisch
nit ESSen;
because *PTCL a lot of people uh: also don't eat horse meat*
- 11 A: [ja
yes
- 12 B: [grade AUSwärtige [leute;=ne,
above all *non local people right*
- 13 A: [ja aber et hat [natürlich en Image,
yes but it *certainly has an image*
- 14 B: [un wenn se dat so
Anbieten WÜ:Rden,
and if they *offered it like that*
- 15 A: **der Eschte rhEInische sAUerbraten is [FERdefleisch];**
the original *rhenisch marinated beef roast is horse meat*
- 16 B: **[is vom PFE:R]D-**
=ja_ja;
is from-*CLIT horse*
- yeah_yeah
- 17 A: [aber dat is der der nit jeJESSen wird;
but that is *the one that isn't eaten*



Again, due to the lack of hesitation preceding the syntactic completion (l. 15), it is even more evident that the speaker orients to a possible position of the nucleus accent syllable. In what follows, speaker B retracts to the finite verb before he articulates the missing constituent (l. 16). Both speakers terminate the turn in simultaneous overlap. However, they choose two alternative options for the predicative (*FERdefleisch* vs. *vom PFE:RD*).

This instance especially provides strong evidence for the orientation of the incoming speaker to a projected position of the nucleus accent syllable, as we can examine the actual completion of the first speaker A and the completion that the incoming speaker B suggests at the same time. In A's own completion, the nucleus accent syllable follows immediately after the copula verb (l. 15 *is FERdefleisch*), co-occurring with the point in time in which speaker B starts. However, B designs the completion differently, as he first repeats the copula verb *is* and then fulfills the projection, delivering the predicative in the form of a prepositional phrase (l. 16) with the noun carrying the accent (l. 16 *is vom FE:RD*). It is A's own completion of the syntactic gestalt which demonstrates that the point at which B starts is a possible position of the nucleus accent syllable. B's completion co-occurs with the articu-





lation of the nucleus accent syllable by the first speaker A. Consequently, in this extract as well as in the instances discussed above, the beginning of the completion corresponds precisely to a projected possible position of the nucleus accent syllable. However, this starting point is independent of the amount of words chosen for the completion and can only be described by taking into account the ongoing projections speakers orient to. This empirical observation is true for 74 out of the 95 terminal item completions in my collection of co-constructions.

How can we explain that a possible position of the nucleus accent syllable is projected and that speakers can anticipate this position? In order to exploit this position systematically to complete another speaker's turn, one would think that the speaker needs intuitive linguistic knowledge about which element the accent has to be assigned to. Thus in the following section, we will raise the question of which resources allow the recipient to anticipate a possible completion for the utterance, with particular consideration given to a possible position of the nucleus accent syllable. We will consider the emerging syntax of the syntactic gestalt of the first speaker as a strong projection device (cf. Auer 2009, Günthner and Hopper 2010) that is exploited in interaction by the participants.



4. The syntactic projection of the nucleus accent position



As we have seen, co-constructed completions demonstrate that speakers can predict a possible position of the nucleus accent syllable. Since the nucleus accent in sentences with broad focus corresponds to the focus accent (Grice and Baumann 2000:1; Uhmann 1992: 242; Välimaa-Blum 2005: 9), it seemed useful to consider research that has been done on linguistic rules for the assignment of the focus accent to different syntactic constituents. In her dissertation on focus phonology, Uhmann (1991) raises the question of how the focus structure of a sentence interacts with its intonational realization (1991: 3) and thus, how the assignment of focus accents to syntactic constituents in German can be described. Even if Uhmann's study is based on controlled experimental data in which focus was controlled by the use of question/answer sequences, her results inform our research in that she aims to discover grammatical principles that underlie the intonation of a sentence. Therefore, we will briefly present Uhmann's main results and then turn to a discussion of our terminal item completions applying these insights.





Uhmann (1991) discusses the phenomenon of focus projection.¹³ The term focus projection means that in sentences with broad focus, a pitch accent on one word can indicate a larger focus domain, e.g. the whole sentence, while in sentences with narrow focus only the word that carries the accent is focussed. Thus, there are some constituents that, when accented, can mark both focus on themselves and focus on the larger constituent that encompasses them (cf. Peters 2005: 98). In general, an accent can be assigned to every constituent of a sentence, but this essentially changes its interpretation (cf. Selting 1995: 117). Focus projection rules, therefore, deal with the question of which constituent of a German sentence has to be accented and thus become the focus exponent¹⁴ of a sentence, i.e. in order to project the feature focus on the whole sentence and reach the most ambiguities in focus interpretation.

According to Uhmann (1991: 209–215), the following rhematic hierarchy describes which syntactic constituent is potentially accentuated to project focus on the whole sentence:¹⁵

adv.III	<	subj.	<	verb	<	verb-addition	<	adv.II	<	adv.I	<	'objects'	<	predicative
<i>temp</i>		<i>agent.</i>				<i>instr.</i>		<i>dir.</i>						
<i>caus.</i>								<i>loc.</i>						

This model (Uhmann 1991: 213) leads to the interpretation of broad and narrow focus. Speakers interpret broad focus when the strongest element in the hierarchy is accentuated. In contrast, the focus cannot be on the whole sentence or constituent (broad focus) when one of the weaker constituents is prominent.¹⁷ It is shown, for example, that temporal adverbs (adverbs III) are not likely to be accented as long as there are other constituents, while in German, the predicative preferentially carries the accent.

¹³ We want to point out that the term *projection* in this case does not correspond to the notion of projection as it was introduced in section 1.2. of this chapter and used in the analysis of co-constructions thus far. Focus projection refers to the projection of the abstract feature focus onto the rest of the sentence.

¹⁴ The focus exponent is defined as the syntactic unit within a focused constituent that can project the abstract feature focus on the whole constituent. (Peters 2005: 98)

¹⁵ The focus accent is always assigned to the syllable that carries the word stress.

¹⁶ The category object comprises all arguments with thematic roles that are specific for objects, that are, for example also non-agentive subjects like in the sentence taken from Uhmann: Ich glaube, dass einem Kind ein FÜLLer gestohlen wurde. ›I think that a pen was stolen from the child.‹ (translation, J.B.), (Uhmann 1991: 212).

¹⁷ See Selting (1995: 119–122), who gives a clear illustration of these principles by discussing German conversation data.





Furthermore, Uhmman identifies two focus projection rules that make predictions about the focus exponent in sentences with broad focus, one for predicate-argument and one for modifier-head structures (Uhmman 1991: 207–215).

According to the first rule, in a predicate-argument structure, the argument placed closest to the right will be the focus exponent, as illustrated in the following example taken from Uhmman (1991: 221).¹⁸

Karl	hat	dem Kind	ein BUCH
SUBJECT	FINITE VERB	OBJECT_DAT	OBJECT_ACC

geschenkt.¹⁹
PAST PARTICIPLE

However, if the argument is realized as a pronoun, the accent is assigned to the verb (Uhmman 1991: 218), which is quite common in spoken language (cf. Baumann 1991: 211).

Uhmman's second rule describes the accent assignment for modifier-head structures. It is always the latter constituent that carries the accent, i.e. either the head or the modifier, depending on its position. She gives the following examples (Uhmman 1991: 208):

a) die	Jugend	von	HEute	b) die	heutige	JUgend ²⁰
DET	NOUN	PREP	ADVERB	DET	ADJECTIVE	NOUN

As Uhmman's study is based on controlled experimental data, her results cannot be applied to spoken language. Nevertheless, these rules and hierarchy were confirmed by Baumann (1999) in his analysis of a spontaneously told story in German. Even though this study is based on a small corpus, it indicates that the link between syntax and the assignment of accent as described by Uhmman can also be found in spontaneous language data.

Turning again to our study, the present analysis of co-constructions demonstrates that speakers have an intuitive knowledge on which syntactic con-

¹⁸ "In einer Struktur [F ... a ... b ...] oder [F ... b ... a ...], in der a Argument von b ist, fungiert a als Fokusexponent, wenn es das am weitesten rechts stehende interne Argument von b ist." (In a structure [F ... a ... b ...] or [F ... b ... a ...], where a is the argument of b, a serves as the focus exponent, when it is the rightmost internal argument of b.) (translation, J.B.) (Uhmman 1991: 215).

¹⁹ »Karl offered a BOOK to the child.« (translation, J.B.).

²⁰ a) »the young people of toDAY« b) »today's young PEOple« (translation, J.B.).





stituent the nucleus accent should be placed. The syntactic projection that is built up in the emerging utterance allows the recipient to predict not only what comes next, but also where possible positions for the accented item may be. Therefore, it seems worthwhile to investigate whether Uhmman's results can help to explain how speakers foresee a possible position of the nucleus accent syllable within terminal item completions.

As was illustrated by the examples above, in our corpus (cf. Table 1 below), predominantly predicatives and verbs occur as accented items (which can be explained by general German sentence structure), which made it difficult to test Uhmman's rules. Nevertheless, if the terminal item completion occurred within a modifier-head or predicate-argument structure, the rules made the correct predictions about the constituent the accent was assigned to in sentences with broad focus. The other instances do correspond in general to the rhematic hierarchy.

Tab. 1: Syntactic functions of the accented constituents

Example	Syntactic function of the accented constituent
1) 9 A: und des is schOn also=äh: and that is quite well uh 10 I: a LEISTung ;=a, an achievement	- predicative
2a) 2 A: und mei pApa war keiner der (.) and my dad was no one who and my dad was not somebody who 3 I: zum [ARZT geh]t=[oder]- to-CLIT doctor goes TAG goes to the doctor right	- adverbial (directional) ²¹
2b) 2 A: und mei pApa war keiner der (.) and my dad was no one who and my dad was not somebody who 3 I: zum [ARZT geh]t=[oder] to-CLIT doctor goes TAG goes to the doctor right 4 B: [JAMmert]- whine-3SG whines	- finite verb
3) 15 El: =aber so_n bIssche:n ähm(--) but like-CLIT little bit uh but a little bit like uh 16 Ma: ja so ge[ZEichnet; PTCL like drawn like drawn	- infinite verb (past participle)
4) 6 I: =aber das war tatsächlich nu:r (.) [ähm but that was actually only uhm but actually it was only uhm 7 A: verschlEppte sache vom HALS her; [e spread thing from CLIT throat a a spread infection coming from the throat	- predicative ²²





5) 2 C: das_s <i>that's</i> 3 I: SCHMERZhaft [wahrscheinlich -[=ge , <i>painful probably PTCL</i> <i>probably painful right</i>	- predicative
6) 15 A: der Eschte rhEInische sAUerbraten is [FERdefleisch]; <i>the original rhenisch</i> <i>marinated beef roast is horse meat</i> 16 B: [is vom PFE:R]D= ja_ja; <i>is from-CLIT horse</i> <i>yeah yeah</i>	- predicative

In every sequence we have seen thus far, the accented item is the strongest with regard to the rhematic hierarchy shown above. In four instances, the predicative carries the nucleus accent; that is, the strongest item in the model and thus the accent has to be assigned to it (in broad focus sentences).

The following example provides even more evidence that the position of the accent is defined by the emerging syntax of the unit. Analyzing the examples discussed, one could think that in most cases, it is simply the last word delivered by the second speaker that is accented. Therefore, it is interesting to examine instances with predicate-argument structures in which the finite verb is placed after the argument (e.g., in subordinate clauses in which the finite verb is obligatorily placed at the end). Following Uhmann's first rule for argument-verb structures, in this type of syntactic structure, the accent is not assigned to the last word of the syntactic gestalt, as the argument is articulated before the right sentence bracket is closed. Example 7 illustrates this point well:

(7) Bier (Caipirinha)

- 1 Ank: naja auf JEden fall,=
well in any case
->2 **=ham wir dann imm(hin)**
have-3PL we then (at least)
we have at least
3 (---)
->4 Sar: **BIER getrunken;**
beer drunk
drunk some beer
5 ((alle lachen))
((everybody laughs))
6 Ana: schnaps hält WÄRmer-
schnaps keeps warmer





Anke and three other friends are having drinks in a bar, and she tells them about the bad weather at a music festival she went to. She is about to resume her story (l. 1f. *naja auf jeden fall ham wir dann*) when Sarah completes her utterance after a long pause (l. 3)²¹, fulfilling the projection of the finite verb (*ham* l. 2) that is in play at this point and delivering the past participle preceded by a direct object. Here again the second speaker, Sarah, comes in, exploiting the projected position of the accent syllable to begin the completion.

This instance is interesting, as it provides evidence that speakers know which syntactic constituent has to carry the accent to articulate a sentence with broad focus. In this case, it is not the past participle that is accented (as, considering the rhematic hierarchy, the past participle would have been accented if a fictive completion like *»danced«* had been proposed), but, according to Uhmman's first rule, the direct object.

Syntactic and semantic projections allow the recipient to anticipate a candidate for the terminal item, i.e. a constituent that could fulfill the projections and close the emergent syntactic gestalt. As we have shown, the combination of the fact that the nucleus accent has not yet occurred and the possibility of predicting the syntactic structure which could follow allow the incoming speaker to anticipate the position of the nucleus accent syllable and orient the beginning of the co-constructed completion to it. Comparing accent placement within syntactic co-constructions to Uhmman's description, we can provide further evidence – from natural interaction – that there are grammatical regularities that can help to explain accent placement within an utterance.

Having described the syntactic projection of the nucleus accent syllable in instances where the second speaker orients its co-completion to a position for the accent, we will now turn to a discussion of the remaining sequences in the corpus. In the 21 instances left, the second speaker begins the completion after the nucleus accent syllable has already been articulated in the first speaker's ongoing turn. In the following analysis we will present some of these instances and show how they also provide evidence for our claim that speakers orient to a possible projected position of a constituent that carries the nucleus accent syllable.

First, 18 of these 21 instances show some common characteristic that can be illustrated by Example 8 *Eisbein*, taken from the same discussion as

²¹ In this case, we only have the audio recording, but the sounds of objects (probably the cocktail glasses) can be heard being moved on the table during the pause. This involvement in another activity might provide an explanation for the length of the pause.





Example 6. Still talking about food, the participants now turn to the question of what kind of regional terms exist to name different German meals, for example the meal *Eisbein* (knuckle of pork).

(8) Eisbein (k09)

- 16 A: =in der kAntine auf der SPEIsekarte stEht,
in the cafeteria on the menu is written
17 wenn et dat JIBT-
if they serve it
->18 EI[Sbein] ;
knuckle of pork
->19 B: [EISbei]n;=ja_ja,
knuckle of pork yeah yeah
20 I: un dAt is doch so KLASsisch-
and that is so classic
21 was_was glaub ich woanders als richtig typisch DEUTSCH
gesehen wird;
what what I think somewhere else is regarded as really typically German

In this case, the completion of the second speaker occurs quite late, but B still begins before a possible point of completion has been reached during the articulation of a possibly last item by A. As speaker A has already almost articulated the first syllable (*EIS*l. 18) that carries the nucleus accent syllable, B can be relatively certain of the word A is about to articulate, especially as it has already been established as the topic of conversation. The completing speaker B starts immediately after the occurrence of the nucleus syllable, anchoring the completion via a retraction to the beginning of the word in the previous speaker's emergent syntactic gestalt.

This instance illustrates some formal characteristics of the 18 instances in which the completing speaker comes in after the articulation of the nucleus accent syllable by the first speaker.

- 1) Overlap is always found in the first and the second speaker's completion.
- 2) The retraction goes at least back to the nucleus accent syllable, even if that means that the speaker has to retrace further back than to the next word boundary.

One can see the same structure in the following short extract (Example 9). B's completion also occurs in overlap, with a retraction to the nucleus accent syllable and not to the closest word boundary (*zumachen* l. 2).²²

²² The emerging syntactic gestalt of the first speaker has a narrow focus on the adverb *ganz* (l. 1). As this accent position is not projected, the second speaker cannot



96

Jana Brenning

(9) Schalke (mu02)

->1 A: **aber schalke kannst ja GANZ zu[mache]**.
but schalke you can PTCL completely close
but you can close the schalke stadium completely
(Schalke=German soccer team)

->2 B: **[GANZ zum]ache.=ja-**
completely close yes

3 (1.8)

Thus, we can see that the nucleus accent syllable seems again an important device speakers orient to. These cases correspond to Schegloff's (1996) and Selting's (1995) findings concerning the last pitch accent as a contextualization cue of the turn transition space. After the occurrence of a possible last accent, there may be a final opportunity for a second speaker to co-construct a completion before a possible point of syntactic, prosodic, and pragmatic completion has been reached. In short, the incoming after the articulation of the nucleus accent syllable and the retraction back to this position in Example 8 and 9 argue for the interactional relevance of the nucleus accent syllable. Moreover, all the other examples discussed thus far demonstrate that the nucleus accent is a linguistic category speakers within terminal item completion orient to. In a majority of instances (74 of 95), the second speaker starts its completion at a projected possible position of a constituent the nucleus accent syllable can be assigned to (Examples 1, 2, 3, 4, 5, 6).

There are 3 exceptions of the patterns described thus far that can be illustrated by the following instance. In the sequence *Eingetauscht*, the nucleus accent syllable has already been articulated and it is placed on the direct object (1. 3 *BUTter*), which means that the syntactic constituent to which the focus accent is assigned in order to project broad focus has already occurred. Hence, it is not possible for the incoming speaker to exploit a possible position of the nucleus accent syllable.

(10) Eingetauscht (mu06)

1 A: und und dann sam_mer übers LAND gefahrn,
and and then we drove across the countryside
 2 (0.3)
 ->3 **und ham dafür BUTter-**
and have-1PL for it butter
 4 (0.6)
 ->5 I: **EItauscht;**
exchanged

anticipate it in advance. This might be one possible explanation for the late incoming of the co-constructing speaker.





Speakers' orientation to the nucleus accent in syntactic co-constructions

97

6 A: MEHL (--) EIER; (-)
flour (--) eggs
7 bei die BAUern mit dem SCHNAPS ();
with the farmers with the liquor ()
8 (0.6)

If we look at the suggested completion *EItauscht* in line 5, we can see that the incoming speaker accentuates its completion and produces the last accentuated syllable in the syntactic gestalt. This happens very rarely in our data (i.e., in 2 other sequences).²³ One could argue that in this case, the interviewer tries to perform the pattern that is normally used (i.e., integrating in the completion an accented syllable as a candidate for the nucleus accent syllable). Moreover, it is also questionable whether it is at all possible for the second speaker to perform his completion without accentuating an item. It can also be assumed that the accent increases the relevance of the co-completion.²⁴

Table 2 below illustrates the possibilities discussed in this chapter, noting how many instances of each category could be found in the whole data set.

Table 2 Distribution of the examples in the corpus

Terminal item completion n=95

1) Nucleus accent within the completion of the second speaker (74)	2) Possible candidate for the nucleus accent syllable has been articulated by the first speaker	
	Possibility A	Possibility B
	Retraction to (at least) the position of the nucleus accent syllable (18)	Co-completion with an additional accent and without retraction (3)

In most cases, the recipient is not late but exploits the projected position of the constituent to which the nucleus accent syllable will be assigned (number 1 in the table). Alternatively, a pattern as in Example 8 and 9 is performed (number 2A in the table), in which the speaker includes the accented constituent in its completion by retracing at least to the nucleus accent syllable. Category 2B comprises rare deviant cases, which correspond to the last instance, *Eingetauscht*, discussed above (Example 10). In this case, the second speaker comes in after the strongest element in the rhematic hierarchy al-

²³ Mostly speakers retrace to the constituent that carries the nucleus accent syllable as in Example 8 and 9.

²⁴ Auer shows that syntactic expansions are more relevant on the level of information processing when they are prosodically exposed and have an own accent (1996: 94).





ready has been accentuated and articulates an additional accent syllable. In all other instances, we find clear evidence that first, speakers orient to a projected possible position of the nucleus accent, and second, that there are syntactic rules that can provide explanations for how incoming speakers anticipate this position.

5. Conclusion

The analysis in this chapter demonstrated that one frequent format exploited for syntactic co-construction in German is what Lerner (1996) describes as terminal item completion. We discussed how speakers exploit the strong syntactic projections towards the end of a syntactic gestalt to complete a previous speaker's turn. Again, this provides strong evidence for the interactional reality of projections on different levels. However, in contrast to the studies about co-constructions in English, it is shown that the intra-turn speaker change in German can systematically be described by examining the prosodic design of the two components the co-constructed syntactic gestalt consists of. Lerner's (1996) previous study in English suggests that the starting point of the second speaker within terminal item completions can be described by the notion of the turn transition space that might begin with the occurrence of a pitch peak. Our analysis of German data shows that completing speakers come in even before a possible last pitch peak has been articulated. They orient to the projection of a possible position of the nucleus accent syllable to begin their completion. This result holds for the majority of the examples examined in this study. Thus, posing the question of how speakers can anticipate a possible position of the nucleus accent syllable, we argued for the important role of syntax.

However, even in cases in which the second speaker comes in late, i.e. after a candidate for the nucleus accent syllable has already been articulated, further patterns can be detected that point towards the interactional relevance of this position. In the majority of these sequences, the speaker re-traces to the nucleus syllable to add his candidate completion.

Having shown the interactional relevance of the position of the nucleus accent syllable within terminal item completions, our analysis of co-constructions confirms that we have to consider syntax and prosody when analyzing the grammar of spoken language. As an utterance emerges over time, the more its syntactic structure is developed, and the more a possible point of completion can be anticipated. This projection is frequently exploited by the participants in interaction to complete a previous speaker's syntactic gestalt. However, the point of intra-turn speaker change within terminal item



completions cannot be described based only on syntactic criteria. Instead, it is crucial to integrate into the analysis the position of the nucleus accent syllable within the co-construction. As we have shown, the second speaker comes in at a possible (or actual (cf. Example 6)) position of a constituent that carries the nucleus accent syllable, and we claimed that this position is projected by the syntax. In the analysis of the grammar of spoken language, we have to take a holistic view on language and combine prosodic and syntactic analysis²⁵ in order to show how these resources are exploited by participants in their emergence in real-time.

Acknowledgments

I would like to thank Peter Auer, Pia Bergmann, Ina Hörmeyer, Martin Pfeiffer and Elisabeth Reber for their helpful comments on earlier versions of this chapter. All remaining errors are mine.

Appendix

Transcription conventions (GAT 2, Selting et al. 2009)

[]	overlap and simultaneous talk
[]	
°h/ °hh	inbreath of 0.2–0.5 seconds / 0.5–0.8 seconds
(.)	micro pause, estimated, up to 0.2 sec. duration
(-) / (--) / (---)	estimated pause of 0.2–0.5 / 0.5–0.8 / 0.8–1.0 seconds
(2.8)	measured pause
und_äh	cliticization within units
:	lengthening
=	fast, immediate continuation with a new turn or segment
SYLlable	strong primary accent
sYllable	secondary accent
<<p> so>	piano, soft
((coughs))	non-verbal vocal actions and events
?	rising to high (final pitch movements of intonation phrases)

²⁵ We are aware that visual resources like gaze, gesture, pointing, and body posture might be important, but the audio data employed in this study did not allow us to take them into account. Furthermore, there are pragmatic constraints that influence the grammar of co-constructions, which will be the subject of our future research.



100

Jana Brenning

- , rising to mid (final pitch movements of intonation phrases)
- level (final pitch movements of intonation phrases)
- ; falling to mid (final pitch movements of intonation phrases)
- . falling to low (final pitch movements of intonation phrases)

Abbreviations

DET	Determiner
PTCL	Particle
CLIT	Cliticized
PREP	Preposition

References

- Auer, P. 1996 On the prosody and syntax of Turn-Continuations. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in conversation*, 57–98. Cambridge: Cambridge University Press.
- Auer, P. 2005 Projection in interaction and projection in grammar. *Text – Interdisciplinary Journal for the study of discourse* 25: 7–36.
- Auer, P. 2007 Syntax als Prozess. In: H. Hausendorf (ed.), *Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion*, 95–124. Tübingen: Gunter Narr Verlag.
- Auer, P. 2009 Online-Syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31: 1–13.
- Baumann, S. 1999 Zum Verhältnis von Akzentform und kognitivem Status von Diskurseinheiten. *Convivium, Germanistisches Jahrbuch Polen*: 201–224.
- Bolden, G. 2003 Multiple modalities in collaborative turn sequences. *Gesture* 3: 187–212.
- Cruttenden, A. 1997 *Intonation*. Cambridge: Cambridge University Press.
- Grice, M. and S. Baumann 2000 Deutsche Intonation und GToBI. *Linguistische Berichte* 181: 1–33.
- Günthner, S. 2006 Was ihn trieb war vor allem Wanderlust. Pseudo-Cleft Konstruktionen im Deutschen. In: S. Günthner and W. Imo (eds.), *Konstruktionen in der Interaktion*, 59–90 Berlin: de Gruyter.
- Günthner, S. 2009 Between emergence and sedimentation: Projecting constructions in German interactions. *Gidi Arbeitspapierreihe* 22.
- Günthner, S. and P. J. Hopper 2010 Zeitlichkeit und sprachliche Strukturen: Pseudoclefts im Englischen und Deutschen. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 11: 1–28.
- Hausendorf, H. 2007 Die Prozessualität des Gesprächs als Dreh- und Angelpunkt der linguistischen Gesprächsforschung. In: H. Hausendorf (ed.), *Gespräch als Prozess*, 11–31. Tübingen: Gunter Narr Verlag.
- Hayashi, M. 2003 *Joint utterance construction in Japanese Conversation*. Amsterdam: John Benjamins.
- Helasvuo, M.-L. 2004 Shared syntax: the grammar of co-constructions. *Journal of Pragmatics* 36: 1315–1336.





Speakers' orientation to the nucleus accent in syntactic co-constructions 101

- Iwasaki, S. 2009 Initiating Interactive Turn Spaces in Japanese Conversation: Local Projection and Collaborative Action. *Discourse Processes* 46: 226–246.
- Ladd, R. D. 1996 *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lerner, G. 1991 On the syntax of sentences-in-progress. *Language in Society* 20: 441–458.
- Lerner, G. 1996 On the “semi-permeable character” of grammatical units in conversation: Conditional entry into the turn space of another speaker. In: E. Ochs, E.A. Schegloff and S.A. Thompson (eds.), *Interaction and Grammar*, 238–276. Cambridge: Cambridge University Press.
- Lerner, G. 2004 Collaborative turn sequences. In: G. Lerner (ed.), *Conversation Analysis. Studies from the first generation*, 225–256. Amsterdam: John Benjamins.
- Local, J. 2005 On the Interactional and Phonetic Design of Collaborative Completions. In: W. Hardcastle and J. Beck (eds.), *A Figure of Speech: a Festschrift for John Laver*, 263–282. New Jersey: Lawrence Erlbaum.
- Mondada, L. 1999 L’organisation séquentielle des ressources linguistiques dans l’élaboration collective des descriptions. *Langage et société* 89: 9–36.
- Mondada, L. 2006 Participants’ online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse Studies* 8: 117–129.
- Müller, F. E. and S. Kläger 2010 Collaborations syntaxiques – Formes et fonctions de leur usages dans un groupe subculturel lyonnais. *Pratiques* 147/148: 223–243.
- Oloff, F. 2011 L’hétéro-répétition suite aux complétions collaboratives: une étude multimodale de tours produits conjointement. Paper presented at the Conference (Dés-) organization de l’oral? De la segmentation à l’interprétation, Rennes, 24.–25. March 2011.
- Ono, T. and S.A. Thompson 1995 What can conversation tell us about syntax? In: P.W. Davis (ed.), *Alternative Linguistics. Descriptive and theoretical modes*, 213–271. Amsterdam: John Benjamins.
- Peters, J. 2005 Intonation. In: M. Wermke, K. Kunzel-Razum and W. Scholze-Stubenrecht (eds.), *Duden. Die Grammatik*, 95–128. Mannheim: Dudenverlag.
- Pierrehumbert, J. 1980 *The Phonology and Phonetics of English Intonation*. Bloomington: Indiana University Linguistics Club.
- Pheby, J. 1975 *Intonation und Grammatik im Deutschen*. Berlin: Akademischer Verlag.
- Sacks, H. 1992 [1964–1972] *Lectures on conversation*. (2 Vols.) Oxford: Blackwell.
- 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. 1988 Discourse as an interactional achievement II: An exercise in Conversation Analysis. In: D. Tannen (ed.), *Linguistics in Context: Connecting Observation and Understanding*. Lectures from the 1985 LSA/TESOL and NEH Institutes, 135–158., Norwood, New Jersey: Ablex.
- Schegloff E. A. 1996 Turn organization. One intersection of grammar and interaction. In: E. Ochs, E.A. Schegloff and S.A. Thompson (eds.), *Interaction and Grammar*, 53–133. Cambridge: Cambridge University Press.
- Selting, M. 1995 *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W.





- Schütte, A. Stukenbrock and S. Uhmann 2009 Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.
- Szczepek, B. 2000a Formal Aspects of Collaborative Productions in English Conversation. *InLiSt – Interaction and Linguistic Structures* 17.
- Szczepek, B. 2000b Functional Aspects of Collaborative Productions in English Conversation. *InLiSt – Interaction and Linguistic Structures* 21.
- Szczepek Reed, B. 2006 *Prosodic Orientation in English Conversation*. Basingstoke: Palgrave Macmillan.
- Uhmann, S. 1991 *Fokusphonologie*. Tübingen: Niemeyer.
- Välismaa-Blum, R. 2005 *Cognitive Phonology in Construction Grammar: Analytic Tools for Students of English*. Berlin: de Gruyter.
- Wells, B. and Macfarlane, S. 1998 Prosody as an Interactional Resource: Turn-projection and Overlap. *Language and Speech* 41: 265–294.





Pia Bergmann
University of Freiburg

The prosodic design of parentheses in spontaneous speech

1. Introduction

This chapter is concerned with the prosodic design of parentheses in spontaneous dialogue. Currently, few studies have examined prosody in naturally occurring, unscripted parentheticals in German (cf. Section 2.2. for a brief overview of research to date). This chapter addresses this gap in the literature by investigating the prosodic and phonetic features that coincide with entry into the parenthetical, exit from the parenthetical, and the parenthetical itself. By way of illustration, consider example (1): The speaker begins a syntactic project *also ich geh morgen aben?* (line 1), disrupts it with a syntactically unrelated insertion *geburtstagsgeschenk meiner frau* (line 3), and finally closes the project by *zu glenn miller* (line 4). (The transcription of all examples follows GAT 2, cf. Selting et al. 2009).

(1)

01 hh03: also ich geh MORgen aben?
 PTCL I go tomorrow evening
tomorrow evening I'm going
 02 (0.9)
 03 <<p,l>gebUrtstagsgeschenk meiner <<creak>FRAU>>;
 birthday present from my wife
a birthday present from my wife
 04 zu glenn MILLer,
 to glenn miller
to see Glenn Miller

At the beginning, the speaker breaks off the first part of the host structure with a segmental deletion and a glottal stop:

(1a)

01 hh03: also ich geh MORgen aben?
 PTCL I go tomorrow evening
tomorrow evening I'm going





104

Pia Bergmann

He then produces a pause and inserts the parenthetical at a lower pitch register and with less volume:

(1b)

02 (0.9)
03 <<p,l>gebUrtstagsgeschenk meiner <<creak>FRAU>>>;
birthday present from my wife
a birthday present from my wife

The end of the parenthetical is marked by creaky voice, and in resuming the original syntactic project, the speaker increases the pitch register and volume of the first part of the host:

(1c)

04 zu glenn MILLer,
to glenn miller
to see Glenn Miller

From an interactional point of view, the speaker can be said to contextualize both the transition from one project to the other and the status of the parenthetical insertion by means of prosodic devices that co-occur with the points in time where the transitions take place.

Instances such as these have attracted the scientific interest of scholars from different realms of research. Formal syntacticians examine which types of constituents are inserted and how the insertion can be explained within a hierarchical model of syntax (cf. e.g. Fortmann 2005; Kaltenböck 2005, 2007). Formal phonologists investigate how the syntactic elements are phrased prosodically and how the phrasing can be derived from syntactic structure (cf. Dehé 2009). Scholars from an interactional background, finally, explore how the disrupted structure is produced by the participants in interaction, and how the participants solve the problem of suspending and re-entering a structure as an interactional achievement (cf. Local 1992; Selting 1995; Mazeland 2007).

At the risk of simplifying matters, two main perspectives on this phenomenon can be identified: the first focuses on a grammatical description and general explanation of the phenomenon, and the second focuses on detailed descriptions of concrete and authentic pieces of interaction. The former is located in the framework of Prosodic Phonology (referred to henceforth as PP) (cf. Nespor and Vogel 2007), the application of Autosegmental-metrical Phonology (referred to henceforth as AM-Phonology) to intonation (cf. Ladd 2008), and derivational approaches to syntax, whereas the latter be-



longs to the framework of Interactional Linguistics, incorporating conversation analytic approaches and the *phonology for conversation* approach (Local, Kelly and Wells 1986; Local, Wells and Sebba 1985), among others (cf. Couper-Kuhlen and Selting 1996; Selting and Couper-Kuhlen 2000).

The question of whether insights gained from both approaches could be combined in order to achieve a comprehensive account of parentheticals will be discussed in Section 2.1. The PP/AM-Phonology approach is built on a complex body of theorizing, but often suffers from a lack of attention to language as it occurs in spontaneous interaction, whereas interactional studies achieve a high level of descriptive accuracy of naturally occurring talk, but often lack the theoretical underpinning that permit the formulation of general grammatical explanations. As will be shown, the difference between these approaches lies not only in the data and methods chosen for analysis, which would make a simple combination feasible, but also in their conceptualization of language. While Interactional Linguistics focuses on how conversation unfolds over time, PP/AM-Phonology views language as consisting of a production of pre-conceived, ready-made units.

Section 2.2 turns to the phenomenon of parentheticals itself and gives an overview of research examining the prosody of parentheses in German. Section 3 includes a description of the materials and methods used in the current analysis, and Section 4 includes an analysis of 63 parentheticals demonstrating which prosodic cues co-occur at the boundaries under discussion and how they serve the task of setting apart the different parts of the interactional sequences (Section 4). Finally, Section 5 summarizes the discussion and draws conclusions for the theoretical conceptualization of prosody in interaction.

2. Background

2.1. Interactional Linguistics and Prosodic Phonology/AM-Phonology?

The term *Interactional Linguistics* (IL) refers to the approach delineated by Selting and Couper-Kuhlen (2000), whereas *Prosodic Phonology* (PP) and *Autosegmental-metrical Phonology* (AM-Phonology), in their application to intonation, refer mainly to the approaches theorized by Nespor and Vogel (2007) and Ladd (2008). In all three approaches prosody is acknowledged as playing a central role in grammar. One seemingly insurmountable difference between the approaches lies in the data and methodology chosen. However, if it were only for data and methodology, could it not be possible to either widen the PP/AM-approach to include analysis of naturally occurring, inter-



actional data, or to simply apply the theory of the PP/AM-approach to Interactional Linguistics?

One similarity between IL and PP, though termed differently, is the notion of *interfaces* (PP) or *co-occurrence* (IL). The concept of interfaces lies at the very heart of PP-theory, which originated from the observation that the speech stream is not simply a linear sequence of sounds, but that these sounds are hierarchically organized to prosodic constituents, such as the syllable, the prosodic word, or the intonation phrase. Therefore, the question becomes, for example, how the syllable or the prosodic word relates to the morpheme and how they serve to proliferate the linguistic structure in speech. One important insight is that units of the different levels of linguistic organization need not be isomorphic. In other words, unit boundaries of different levels do not always coincide, which clearly necessitates conceptualizing prosodic constituents as autonomous (cf. Nespor and Vogel 2007: 2ff.). Another important insight is that boundaries vary in strength. For example, Fougeron and Keating (1997) show in their experimental study that boundaries on different levels of the prosodic hierarchy are marked by differences in articulatory effort: the higher in the hierarchy the stronger the boundary strength. One major research interest, therefore, addresses the question of how the edges or boundaries of units on different linguistic levels coincide and how they are marked by phonetic means (thus the notion of »phonetic encoding of prosodic structure«, Keating et al. 2003). Gradience in boundary strength and phonetic detail, therefore, are central concepts in this area of research.

Comparing the idea of interfaces with the interactional concept of co-occurrence reveals interesting similarities but crucial differences as well. In the IL approach, co-occurrence belongs to the concept of contextualization and expresses the idea that often resources of different linguistic levels occur together, i.e. redundantly, in order to serve certain contextualizing functions (cf. Couper-Kuhlen and Selting 1996). For instance, it could be shown that in turn-taking, the probability of a speaker change increases when devices on the lexico-semantic, syntactic, pragmatic, and prosodic level co-occur. The more boundaries on different linguistic levels coincide, the more often a speaker change occurs (cf. Ford and Thompson 1996; Bergmann 2008).

Although the co-occurrence of boundaries and the ways in which they are marked seems to be of interest to both approaches, the conceptualization of how boundaries are reached is crucially different. In IL, the term *projection* plays a major role, springing from the idea that structures emerge in time and project their (possible) completion (cf. Auer 2005, 2009). In the PP approach, on the contrary, boundaries or edges of constituents are not ar-





rived at in the process of language production, but are conceived of as belonging to pre-existent units that simply have to be mapped onto each other. Therefore, in the case of parentheses, it is asked how the three syntactic units are mapped onto a set of possible prosodic constituents. First, this completely neglects the phenomenon of *online-syntax* that becomes obvious in many cases of parentheses in which we find retrospective changes in the (linear) syntactic structure, when, for example, a parenthetical insertion is retrospectively integrated into the front field of the sentence (cf. Stoltenburg 2003). Second, the PP assumption also neglects the possibility that prosody may not be amenable to a clear segmentation into constituents. This idea has been formulated by Auer (2010) and Barth-Weingarten (2011), and both opt for a solution that omits the necessity of segmenting talk into units and replaces it with an analytical focus on boundaries, or “cesuras” of varying strength (cf. Barth-Weingarten). Although at first sight the PP framework seems appealing in this respect, because it introduces gradience in boundary strength, we must consider that it does not introduce gradience in phrasing per se, as all phonetically phrased stretches of talk must be mapped categorically on certain prosodic constituents.

Comparably to Auer (2010) and Barth-Weingarten (2011), Couper-Kuhlen (2007) points out that a reasonable theoretical model for prosody in interaction must be based on the assumption that prosodic structure unfolds in time. For example, the descriptive framework of AM-Phonology superficially seems to lend itself quite well to this objective by relating intonation contours to a linear structure of phonetic targets that are associated with specific elements on the text level (tune-text-association). Despite this linearity in structure, the problem with the model lies in its orientation towards language as a “product”: the linear structure is built upon a pre-existing, complete syntactic structure (cf. Couper-Kuhlen 2007: 70). On the contrary, Couper-Kuhlen (2007) argues for a “process-oriented” approach to prosody and to language in general. Prosody emerges in time in interaction; it is “janus-headed” by looking forward and backwards at the same time. Therefore, prosodic cues are situated in a specific context and always relate to what came before and what comes next. These two directions are called the *prospective* and the *retrospective* nature of prosody.

From a theoretical perspective, therefore, the idea of combining the theory of AM-Phonology with the data and methodology of IL, although at first sight appealing, has to be rejected. The neglect of the situatedness and emergence of linguistic structures in time makes the discussed theories unsuitable for the explanation of language in interaction.





2.2. Parentheticals and how they are designed prosodically

As introduced in Section 1, a threefold structure is constitutive of parentheticals, in which a host structure is interrupted by an insertion – the parenthesis – and taken up again at a later point. In the present chapter, parenthetical insertions are defined on syntactic grounds solely according to Stoltenburg (2003). As Stoltenburg states, the definition of parentheticals often suffers from a non-sufficient separation of linguistic levels on which the disruption is to be located: Pragmatic, phonological (intonational), and syntactic criteria are in many cases not separated strictly enough, the consequence being that many different phenomena fall under the heading of “parenthetical”. Therefore, the results of different studies become difficult to compare (cf. Stoltenburg 2003: 3ff.). Indeed, Stoltenburg proceeds, the “typical” parenthesis is one in which a bundle of features on the pragmatic, phonological and syntactic level come together in order to set apart the parenthetical insertion from its host. That is, the typical parenthesis would be non-integrated into the host structure on both the syntactic level and the intonational level, and it would at the same time fulfil a function that is external to the semantic content of the utterance, e.g. a meta comment. However, these different levels need not go together. Moreover, if we want to describe the prosodic design of parentheticals, it is of vital importance to separate the different linguistic levels in order to avoid circularity.

Therefore, parentheticals are defined on the basis of syntactic criteria alone in the present chapter. Neither phonological nor pragmatic criteria enter into either the definition or the selection process of the parentheticals. One basic characteristic of syntactic parentheses is that they do not exhibit any “internal” criteria that would consistently identify them as a parenthesis. Instead, “[t]hey derive their existence, as it were, from their interaction with a host clause” (Kaltenböck 2005: 27). This interaction is defined as a “non-relation”, in which the parenthetical insertion stands in a linear but not a dependency relationship to the host clause (cf. Kaltenböck 2005: 27). It must not fulfill a syntactic function in the host clause, such as object or adverbial. This gives us one possibility to distinguish parentheses from instances of same-turn self-repair where often a (syntactically integrated) element is modified or replaced (cf. Pfeiffer this volume). Furthermore, Stoltenburg (2003) differentiates between interruptions (*Unterbrechung*) and expansions (*Ausbau*) of a syntactic structure through, for example, relative clauses and attributes. Only the interruptions count as parentheticals, whereas the expansions – due to syntactic agreement with the host clause – should be viewed as integrated into the host structure (cf. Stoltenburg 2003: 11). A par-



enthetical is thus defined as a syntactically non-integrated interruption of an emergent syntactic structure that is resumed and completed after the interruption. The emergent syntactic structure can be a fragment in such instances as when the right sentence bracket is still missing. It can as well be a subordinate clause (e.g., prepositioned if-clauses) that needs a second element to be completed or a main clause where the complement has not yet been produced. The syntactic type of the parenthetical insertion is not constrained and can include main clauses as well as short meta comments like *ich sag mal* (let's say). An overview of the occurring types is given in Section 3.

From an interactional perspective, Mazeland (2007) identifies three relevant interactional tasks that must be resolved by current speakers in order to make the parenthetical structure transparent to their partners in interaction: the initiation task, the maintenance task, and the return task. The following schema exemplifies this structure:

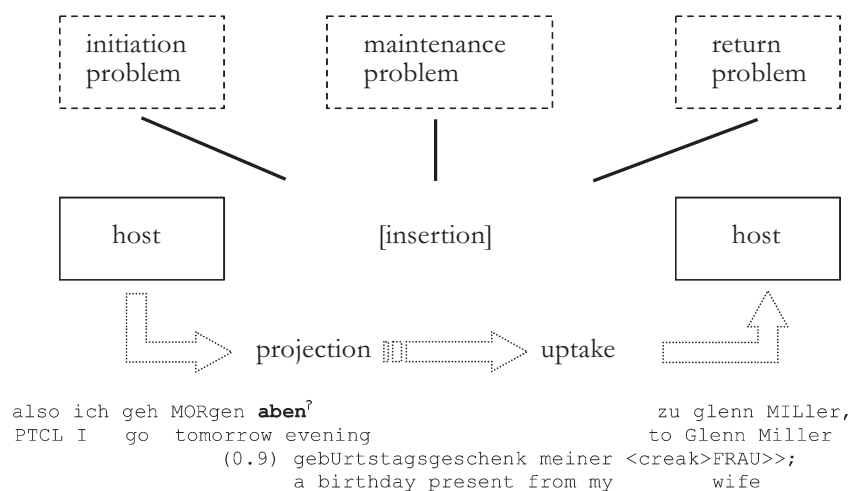


Figure 1: Parentheticals from an interactional point of view

The host structure projects its completion and thereby bridges the gap until it is taken up and completed in the second part of the host. The first task of the speaker is to initiate that a break in the on-going structure is to follow without the host being completely abandoned. The second task is that the speaker contextualizes the parenthesis as a unit, and the third task finally yields the uptake or continuation of the host, so that the unity of the complete host structure is guaranteed and becomes transparent to the current hearer. Completing these tasks, of course, is a complex undertaking in which



not only prosodic means, but also syntactic, lexico-semantic, and pragmatic means, play a role (cf. Mazeland 2007; Duvallon and Routarinne 2005). In the present chapter, however, we will exclusively focus on the prosody of parentheticals, particularly on syntactic parenthesis, whereas Mazeland's definition of parentheses refers to the domain of the turn constructional unit (TCU).

The research literature on the prosody of parentheses in German spontaneous speech is astonishingly scarce. To our knowledge, only Schönherr (1993), Peters (2006), and Döring (2007) have explicitly dealt with this issue. All studies are based on a syntactical definition of parentheses. Schönherr investigates Austrian German political talk shows, while Peters' data consist of Hamburg German interviews, and Döring's data are taken from political debate speeches in the German parliament. Typical prosodic characteristics of German parentheses include changes in speech rate, loudness, and pitch height. The "prototypical" parenthesis usually induces higher speech rate, lower volume, and lower pitch height, as described by Schönherr (1993). She states, however, that about two thirds of the analyzed parentheses are not marked at all in these aspects. Schönherr additionally observes changes in accent type and diminished strength of accentuation in the parenthesis. Döring (2007) tests three hypotheses developed through the analysis of research on English parentheticals: (1) that "parenthetical constructions are quieter, faster, and lower than their surrounding anchor clauses", (2) that "parenthetical constructions are clearly set off by pauses", and (3) that "parenthetical constructions have a clear intonation contour of their own" (Döring 2007: 290). Her analysis of the five examples in her data set show that neither changes in intensity, nor diminished pitch range, pitch jumps, or pauses are obligatory features of parentheticals. However, changes in speech rate and a lower pitch register seem to regularly set apart the parenthetical from its preceding host. Moreover, she states that each parenthetical construction constitutes an intonational domain of its own; the end of the preceding host carries an edge tone and "is often lengthened" (Döring 2007: 306). Her overall conclusion is that, if anything, it is the entry into the parenthesis that is marked.

Peters (2006) takes a somewhat different angle to the prosody of parentheses. He bases his research in a more phonologically-oriented approach in which the prevailing research question concerning parentheses is not *how* they are phrased, but *if* they are phrased at all. This relates to an assumption made in the theoretical framework of prosodic phonology (cf. Nespor and Vogel 2007), which predicts that a break in the syntactic structure is mirrored by a break in the prosodic structure (cf. Dehé 2009; cf. Dehé 2007 for a com-





prehensive account of the theoretical relationship of syntax and prosody in parentheticals). Dehé (2007) casts some doubt on the viability of this assumption for naturally occurring speech data (i.e., the British component of the International Corpus of English, ICE-GB, cf. Dehé 2007: 263), showing, among other things, that not all parentheticals constitute an intonation domain of their own. Peters (2006), too, demonstrates that the syntax-prosody interface is more complex than assumed. He identifies four different types of prosodic integration and therefore deviates from the prediction that each parenthesis should form a distinct intonation phrase and break up the host sentence into two other distinct intonation phrases. Additionally, he finds a significant link between the size of the parenthetical (i.e., the number of syllables) and the prosodic phrasing as well as between prosodic phrasing and the syntactic type of parenthesis.

Despite his departure from Prosodic Phonology in this respect, Peters' work is still located within this framework and within the AM approach to intonation in general. Assuming a phonological point of view in which phrasing is considered to be categorical, he states that

[t]o identify prosodic integration types we did not make use of prosodic cues like pauses or discontinuities in pitch scaling and speech rate, as these cues may be optional. (Peters 2006: 2)

Prosodic cues are thus explicitly disregarded and deemed unimportant for phrasing speech into intonation phrases. While this may be true from a phonological point of view, the question remains whether there is nothing to be gained from taking into account prosodic cues and examining the extent to which these cues play a role in phrasing.

To reach this aim, one focus of the present analysis is on cases of parentheticals that would be described as prosodically integrated from a phonological point of view. It will be shown that despite their integration with the preceding host, all of them display prosodic or phonetic discontinuities of some sort (cf. Section 4.1.1.1.). It will therefore be argued that all parentheses coincide with minimal prosodic phrasing, and that prosodic cues should not be viewed as purely optional when considering the initiation of the syntactic change as an interactional task. To view prosodic cues as such means to assume a "data-driven" approach to prosody in interaction rather than a "theory-driven" approach. Therefore, no prosodic features will be neglected beforehand due to theoretical assumptions.



3. Materials and methods

The data set of the current study consists of 63 parentheticals that were taken from two hours of informal interviews with elderly male speakers of German collected in Hamburg. The interviews were gathered for a research project entitled *Regional variation of intonation* funded by the German Research Foundation at the University of Freiburg and the University of Potsdam under the guidance of Peter Auer and Margret Selting. The data coincide in part with data used by Peters (2006).

The parentheticals, defined as outlined in Section 2, were selected by using the transcripts of the recordings only, in order not to be influenced by phonetic and prosodic features. The parenthetical insertions include main clauses, subordinate clauses, syntactic phrases, and stereotypical idioms. The idioms were treated as a separate group, even if they sometimes take the form of a main clause or a subordinate clause. The following four examples illustrate the four main types of parenthetical insertions. The parenthetical sequences are indicated by an arrow, the parenthetical insertion by bold face.

(2) main clause, hh04_28

-> 01 hh04: theorEtisch kAnn ein Arzt ja:
theoretically can a doctor PTCL
in theory a doctor can
-> 02 **und⁷ (.) in der prAxis ist es ja auch VORgekommen-**
and in the practice is it PTCL PTCL happened
and (.) it indeed has happened
-> 03 lEistung abrechnen die er gar nicht erBRACHT hat.
service bring to account that he not provided
bill services he has not provided

The parenthetical insertion (line 2) is a coordinated main clause in the middle field of the host utterance. It introduces additional information.

(3) subordinate clause, hh03_9

-> 01 hh03: =von der orthograPHIE,=
from the orthography
with respect to the orthography
-> 02 =**we_man das mal so AUSdrücken soll.=**
if one that PTCL like that express shall
if you want to express it like that
-> 03 =is das nich nich nich SAÜber genug;ne?
is that not not not clean enough; right
it is not not not good enough is it





The prosodic design of parentheses in spontaneous speech

113

In this example the parenthetical (line 2) is inserted between the front field and the left sentence bracket of the host clause. The first part of the host clause is thereby positioned into the pre-front field of the clause. The parenthesis is an if-clause that introduces a meta comment concerning the formulation of the host clause.

(4) syntactic phrase, hh03_25

```
-> 01  hh03:  also ich geh MORgen aben?
          PTCL I   go  tomorrow evening
          thus tomorrow evening I'm going
-> 02          (0.9)
-> 03          gebÜrtstagsgeschenk meiner FRAU;
          birthday present   from my wife
          a birthday present from my wife
-> 04          zu glenn MILLer,
          to glenn miller
          to see Glenn Miller
```

The parenthesis of this example is a noun phrase (line 3). It is positioned in the middle field of the host clause and provides additional background information concerning the host clause.



(5) stereotypical idiom, hh04_5

```
-> 01  hh04:  da wurde also a?
          there was PTCL [a]
          there was e
-> 02          (-) ich würde sagen AUSnahmslos HOCHdeutsch
          I   would say   exceptionless standard German
          gesprochen.
          spoken
          I would say spoken standard German without exception
```



The parenthesis (line 2) in this extract was classified as a stereotypical idiom. It has a fixed form, serves as a meta comment concerning the formulation of the second part of the host clause, and is inserted into the middle field of the host clause. Other stereotypical idioms found in this data set are *sag ich mal* (let's say) and *wie gesagt* (like I said).

The quantitative overview (see Table 1) shows that the majority of insertions are stereotypical idioms (n = 28), followed by subordinate clauses (n = 14), main clauses (n = 14), and syntactic phrases (n = 7).





Table 1: Syntactic types of the parenthetical insertion

syntactic type	number
main clause	14
subordinate clause	14
phrase	7
stereotypical idiom	28
total	63

After the syntactic classification, all parenthetical sequences were extracted as wav-files and analyzed as to their phonetic and prosodic features using *praat* (cf. Boersma and Weenink 2009). In accordance with the threefold syntactic definition, the following schema exemplifies the points in time that were considered relevant for prosodic analysis.

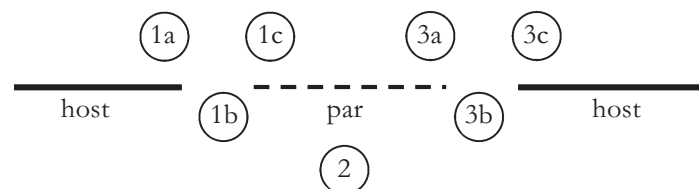


Figure 2: Relevant points in time for prosodic analysis

In this schema, the solid lines refer to the beginning and ending of the host, whereas the dotted line represents the parenthesis. The numbers 1–3 signify the initiation of the parenthetical, the maintenance of the parenthetical, and the return to the host, respectively. 1 and 3 are divided into three points in time each, namely the ending of the host/parenthetical, the time between the two stretches of talk, and the beginning of the parenthetical/host. Each such point was analyzed separately for its phonetic and prosodic design. The following list includes all features encountered at positions 1 and 3.

- 1a) or 3a) end of host / parenthesis
 - break-offs indicated by glottal stops / glottalization (bo)
 - (final) lengthening (le)
 - creaky voice (cv)
 - boundary tones (bt)
- 1b) or 3b) stretch between host and parenthesis / parenthesis and host
 - pauses (p)
 - breathing (b)
 - hesitation markers (hm)





1c) or 3c) beginning of parenthesis / host

- pitch jumps (pj)
- creaky voice (cv)
- anacrustic syllables (as)
- rush-throughs (rt)

These features equal the boundary cues of intonation phrases (cf. Grabe 1998; Bergmann and Mertzluft 2009) with the exception of the break-off signal “glottal stop” and “hesitation markers”. The former is typical of broken-off structures in which the turn is not at stake (i.e. they are used as a turn-holding device) (cf. Local and Kelly 1986).

For position 2, the parenthetical insertion itself, the investigated parameters are loudness, speech rate, strength of articulation on the segmental level, voice quality, pitch range, and pitch register. With “strength of articulation on the segmental level”, we refer to either segmental deletions or lenitions of, for example, voiceless fricatives to voiced fricatives as well as to centralization of peripheral vowels. The analysis of all prosodic-phonetic features was carried out auditorily and by visual inspection of the spectrograph.

In addition to the prosodic and phonetic design of the points in time, all instances of parentheticals as well as their hosts were analyzed for their intonational completeness. Stretches of talk were considered intonationally complete if they consisted of at least one (nuclear) accent and had a boundary tone.

4. Results

The data set was subdivided into two groups consisting of single parentheses ($n = 50$) on the one hand and multiple parentheses ($n = 13$) on the other. For the single parentheses, a simple threefold structure of host – parenthetical – host is constitutive, which coincides with the interactional tasks initiation, maintenance and return elucidated by Mazeland (2007). The multiple parentheses are more complicated in this respect since they involve a complex structure in which, for instance, the parenthetical itself is broken up by another parenthetical, or two or more parentheticals follow one after another. The first Section of this chapter concentrates on the single parentheticals by relating the prosodic and phonetic signals to the interactional tasks in the unfolding of the utterance in time (Section 4.1.1.) and by summarizing all occurring prosodic and phonetic signals at certain points in time (Section 4.1.2.). In sum, while 4.1.1. focuses on specific instances of parentheti-



cals and examines if and how the signals are used to solve the task of indicating syntactic structure in the initiation phase of the parenthetical, 4.1.2. gives general insight into the observed combinations of prosodic-phonetic signals of discontinuity. Section 4.2 will investigate multiple parentheticals.

4.1. Functional embedding of phonetic-prosodic signals in single parentheticals

The investigated points in time follow one after another. To do justice to the online production of speech by participants, the following description of the initiation task (positions 1–2) will take into careful consideration the sequential production of the prosodic parameters under discussion. According to Couper-Kuhlen (2007), a distinction between prospective and retrospective prosodic devices is crucial, and will therefore be taken up. The signals occurring at the end of the first part of the host (position 1a) can be viewed as prospective by foreshadowing an interruption of the prosodic and syntactic structure under way. The signals at the beginning of the parenthetical (position 1c) or throughout the parenthetical (position 2) are retrospective, meaning they relate back to the prosodic design of the host. The signals that occur between the host and the parenthetical (position 1b) are somewhat ambiguous with respect to the direction of their contextualizing function and will be subsumed under the retrospective signals.

4.1.1. Prosodic devices for initiating the parenthetical

As has been outlined in Section 2, an important point of interest is whether prosodic-phonetic cues bring extra value to the phrasing of parentheticals, or if they can be dismissed as “optional”, as suggested by Peters (2006). The analysis of the initiation phase, therefore, contrasts a description based on phonological intonation phrases in a strict sense with phonetic-prosodic boundary marking. It will be argued that it may be worthwhile to consider phonetic-prosodic cues as more than optional for phrasing.

The discussion of the initiation of the parenthetical will begin from two different types of parenthesis. *Type 1* entails all cases in which the first part of the host consists of a phonologically *incomplete* intonation phrase. *Type 2*, on the other hand, includes all cases where the first part of the host consists of a phonologically *complete* intonation phrase. The question now becomes if and how phonetic-prosodic signals of discontinuity co-occur with the syntactic disruption. Given the case of a phonologically incomplete host (*Type 1*), the parenthetical would be considered completely integrated, i.e. not separated



from the first part of the host, if no additional signals of discontinuity occurred. The hearer would, in these cases, have no indication of a change in the syntactic structure under way. If, however, we encounter systematic usage of phonetic-prosodic discontinuity markers at the syntactic break, we might interpret these as contextualizing the syntactic disruption. Thus, despite the fact that the host and the following parenthetical belong to the same intonation phrase from a phonological point of view, the speaker contextualizes the transition by phonetic-prosodic means and thereby initiates the parenthetical. After a complete intonation phrase (Type 2), the parenthetical is necessarily disintegrated, since the first part of the host constitutes an intonation phrase of its own. Additional phonetic-prosodic cues may, however, serve to contextualize the parenthetical as either an “unexpected” succession of the syntactic project under way (e.g., by introducing a hesitation marker in position 1b), or simply as different from the surrounding host (e.g., by higher speech rate in position 2).

In our data, the majority of cases involves incomplete intonation phrases for the host ($n = 38$) as compared to complete intonation phrases ($n = 12$). Table 2 gives an overview of the prosodic types as well as their distribution over different syntactic structures of the host. The category “fragment” refers to all instances in which the syntactic structure is incomplete (at least) due to a missing element in the right sentence bracket. “Subordinate clause” refers to those cases where a pre-positioned clause (e.g., an if-clause) projects a second part in order to be complete. The category “main clause” is somewhat problematic because it refers partly to structures which, from a syntactic point of view, could be complete, but only if they are viewed in isolation. For example, the host structure *also ich geh morgen abend* (cf. example 1) would be a complete sentence in isolation, but is clearly hearable as incomplete in its context because the projected local adverbial is still missing. Such cases were not omitted from the data set but are subsumed in the category “main clause” below.

Table 2: Distribution of Type 1 and Type 2 parentheticals / syntax of the first part of the host

Syntax (first part of the host)	Prosody	
	Type 1 (incomplete)	Type 2 (complete)
fragment	36	6
subordinate clause	0	4
main clause	2	2
total	38	12





As can be seen from Table 2, only two cases of incomplete intonation phrases do not coincide with a fragment, but with a “main clause”. However, it is important to acknowledge that these two cases are probably not perceived as syntactically complete in their context of occurrence, as has been mentioned above. Complete intonation phrases occur with syntactic fragments in six cases; the other six complete intonation phrases are produced with subordinate or main clauses. Although the data set is too small to draw a conclusion, the distribution seems to corroborate the findings of Peters (2006). He found that in cases of parentheticals that are inserted between two clauses, the first part of the host clause constitutes an intonation phrase of its own, whereas in cases where the parenthetical is inserted within a clause, the host does not coincide with a complete intonation phrase of its own but is either interrupted by the parenthetical, or incorporates it into its own prosodic structure.

Table 3 illustrates the distribution of the phonologically complete and incomplete intonation phrases with respect to the syntax of the following parenthetical insertion.

Table 3: Distribution of Type 1 and Type 2 parentheticals / syntax of the parenthetical

Syntax (parenthetical)	Prosody	
	Type 1 (incomplete)	Type 2 (complete)
main clause	7 (18.4%)	2 (16.7%)
subordinate clause	8 (21.1%)	6 (50%)
phrase	3 (7.9%)	2 (16.7%)
stereotypical idiom	20 (52.6%)	2 (16.7%)
total	38	12

It is striking that the majority of Type 1 parentheticals are stereotypical idioms (52.6%) as compared to subordinate clauses, which constitute the majority of cases in Type 2 parentheticals (50%). The distribution is not statistically significant, however ($\chi^2_{(3)} = 6.13; = 0.106$).

Given this background on the distribution of phonologically complete and incomplete intonation phrases, the next section turns to the phonetic-prosodic design of the initiation phase. Type 1 parentheticals (= phonologically incomplete host) are presented first.





4.1.1.1. Type 1 parentheticals: Phonologically incomplete host

In the data set of 50 single parentheticals, 38 hosts are phonologically incomplete (cf. Table 2). In all of these cases, the speaker is found to use phonetic-prosodic discontinuity markers. First and foremost, 23 instances are characterized by a break-off signal (= glottal stop) at the end of the host. In the remaining 15 instances, the speaker does not produce a break-off signal, but produces instead phonetic-prosodic discontinuity markers in at least one of the later positions. If no break-off signal is produced, the speaker can produce other signals of discontinuity at the end of the host (position 1a, $n = 6$, ex. 6–9), which can be referred to as prospective marking of the upcoming break. Other examples show signs of discontinuity between the host and the parenthetical (position 1b, $n = 1$, ex. 10) and at the beginning of the parenthetical (position 1c, $n = 6$, ex. 11, 12), the latter retrospectively marking the break in the syntactic structure. The latest position for signals of discontinuity to occur is position 2, which is the parenthetical itself. Like position 1c, the initiation task would in these cases be solved by retrospectively marking the parenthetical as separated from the host ($n = 2$, ex. 13). The following extracts exemplify these different types of parentheticals, beginning with cases of prospective marking and proceeding on to those with retrospective marking.

Prospective marking

The first example (ex. 6) illustrates one of the prevailing cases with a glottal stop as a break-off signal.

(6) hh04_5

```
-> 01   hh04:   da wurde also a'
               there was PTCL [a]
               there was e
-> 02       (-) ich würde sagen Ausnahmslos HOCHdeutsch
               I   would say   exceptionless standard German
               gesprochen.
               spoken
               I would say spoken standard German without exception
```





120

Pia Bergmann

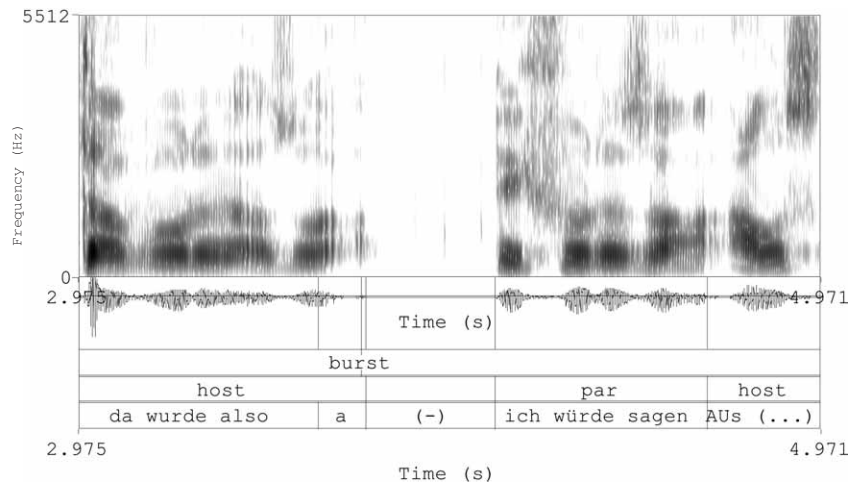


Figure 3: Sound extract with a glottal stop as break-off signal (hh04_5)

The speaker begins to utter the sound segment [a], which is already glottalized, and interrupts it with a glottal stop. (The location of the burst is indicated in the sound extract.) A pause follows before the speaker inserts the parenthetical *ich würde sagen* (I would say, line 2). The speaker then resumes the projected structure by taking up the interrupted sound segment, which is then produced as the first part of the diphthong [au]. This extract exemplifies the majority of Type 1 parentheticals, in which the first part of the host is clearly interrupted by a break-off signal. The speaker solves the initiation task by prospectively marking that something different from the expected structure will follow. At the same time, due to the “janus-headed” nature of prosodic devices, the break-off signal indicates retrospectively that the preceding structure has come to a halt.

The following examples illustrate cases in which the first part of the host is not interrupted by a glottal stop, but by other phonetic-prosodic signals at the end of the host, such as lengthenings and/or creaky voice. Examples (7) to (9) illustrate such cases.



(7) hh04_21

```

01   hh04:   und zwar hängt das DAMit zusammen dass meine
              and PTCL is linked to the fact   that my
              MUTter?,
              mother
              this is linked to the fact that my mother
-> 02        (-) die mUss: ↓ <l> SO genau WEISS_ich_s
              she must          so exact know I it
              <<creak>nich>>.
              not
              she must I don't exactly know
-> 03        die muss also als KLEINKind mit ihren ELtern (-)
              she must PTCL as small child with her parents
              nach hamburg gekommen sein
              to   hamburg have come
              she must have come to Hamburg as a small child with her parents

```

Example (7) is characterized by a lengthening of the lexical item *muss* (must, line 2), which occurs immediately prior to the parenthetical main clause. Later on, i.e. in position 1c and 2, the parenthetical is additionally set off from the host by an initial slight step down in pitch and a lowered pitch register. The speaker thus combines prospective and retrospective signals in order to set the parenthetical apart from its host. The host structure, however, neither constitutes an intonation phrase of its own nor is it clearly disrupted by a break-off signal. Prospective prosodic cues such as lengthening and retrospective cues such as pitch jumps and changes in pitch register nevertheless serve to initiate the change in syntactic structure.

In addition to lengthening, the next example (8) is characterized by a stretch of creaky voice before the insertion of the parenthetical.

(8) hh04_50

```

-> 01   hh04:   sind_die lEUTE denn heut<<creak>e: <<p,l>al>so
              are the people PTCL nowadays          PTCL
              sach_ich_ma>> (.) FLAcher.
              say I   once      more shallow
              are people nowadays well let's say more shallow

```



122

Pia Bergmann

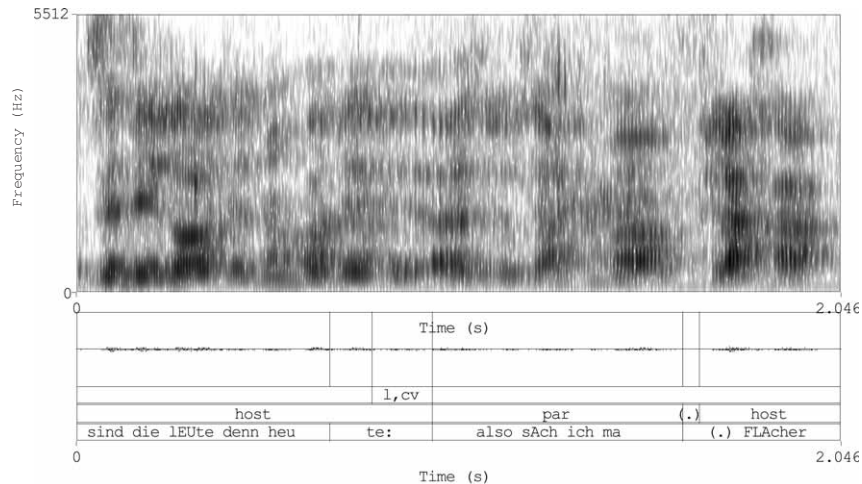


Figure 4: Sound extract with lengthening and creaky voice in position 1a (hh04_50)

The speaker lengthens the central vowel of the last syllable of the host, [tə], which is produced simultaneously with creaky voice, as can be seen in the spectrograph. The relevant stretch is marked by “l, cv” in the first tier of the text grid. Creaky voice is retained through the beginning of the first item of the parenthetical *also* (line 1). It is unclear, however, if this glottalization can be interpreted as a boundary signal (on a higher level than the word level), since the word begins with a vowel, which in German is often preceded by a glottal stop. In addition to lengthening and creaky voice, the parenthetical is produced with diminished intensity, pitch range, lower pitch level, and more articulatory reduction on the segmental level. Again, prospective and retrospective devices combine to set apart the parenthetical from the preceding host.

The last example (9) of this group serves to indicate the problematic status of some of the investigated phonetic and prosodic parameters, especially pauses and lengthenings. The speaker inserts a meta comment *sa_ch_ma* (let’s say, line 3) into the middle field of the utterance.

(9) hh04_45

```
01  i:  nee aber es geht ja jetzt Zunehmend daZU-
      no but it is happening more and more
02      dass äh alle sachen zu privatiSIeren,
      that uh to pass everything into the hands of private companies
-> 03  °hh un:d m (-) UNabhängige: <<p>sa_ch_ma> (.) FIRmen
      and m independent say I once companies
      so (--) da: (--) zu BAUen,
```



In addition to breathing between the host and the parenthetical insertion (position 1b), the inserted material is set off from the first part of the host by a step up in pitch (position 1c), as well as an increase in speech rate combined with articulatory reduction (position 2). A word of caution is in order here: The transcript shows that the first part of the host is produced with creaky voice except for the last lexical item *ja* (line 1). The categorization process of all parentheticals required that this instance be subsumed under the cases



with no signals in position 1a because the last item of the host does not bear any of the investigated features. It is doubtful, however, if this categorization is appropriate, or if the creaky voice stretch should be considered as initiating the upcoming break. What can be stated with certainty is that the occurrence of pauses or other markers of discontinuity in position 1b alone (i.e. without any preceding signals at the end of the host), is extremely rare. It does not seem sufficient to prospectively initiate the syntactic break by a pause or a hesitation marker alone. This aligns with the critical remark given for example (9): Since pauses, breaths, and hesitation markers are ambiguous with respect to their boundary marking function, they are rarely used as indications of a break without being accompanied by other signals of discontinuity.

Retrospective marking

This section turns to the examples in which neither a break-off signal nor any other signal of discontinuity occurs either at the end of the host or between the host and the parenthetical. The initiation task is thus solved by retrospectively marking the parenthetical at its beginning (position 1c) or throughout the duration of the parenthetical (position 2).

Extract (11) parenthetically inserts the meta comment *sa_ch_ma* (let's say, line 1) into the middle field of the ongoing syntactic structure.

(11) hh04_53

```
-> 01  i:  un:d DESwegen: sInd ja auch jetzt: (-) LEUTE
        and therefore are PTCL also now      people
        ↓ <<p>sa_ch_ma> in IHrem Alter (.) AUSgesucht worden-
          say I once in your age              chosen
        and that is why they chose people let's say of your age
02      die: dann wahrscheinlich früher (-) äh noch MEHR
        kontakt zu HAMBurgischen (-) äh_ch oder
        plAttdutschen LEUTen hatten,
        who probably had more contact with people from Hamburg or those who speak Low German
```

From a phonological point of view, there is no indication of a phrase boundary before the inserted material. Phonetically, however, the parenthetical is marked by a (slight) step down in pitch at its beginning (position 1c), as well as diminished loudness, pitch range, and articulatory reduction on the segmental level (position 2).

Extract (12) exemplifies another instance of a seemingly integrated parenthetical.



(12) hh04_31

-> 01 hh04: solAnge ich für dich KINder_**das_sacht_er_immer**;
as long as I for you child that said he always
as long as I get child that is what he always said
-> 02 =solAnge ich für dich KINdergeld kriege;
as long as I for you child benefits get
as long as I get child benefits for you
-> 03 (.) kannst_du: (.) äh: (---) kannst_du zur SCHULE
can you uh can you to school
gehen.
go
you can uh you can go visit the school

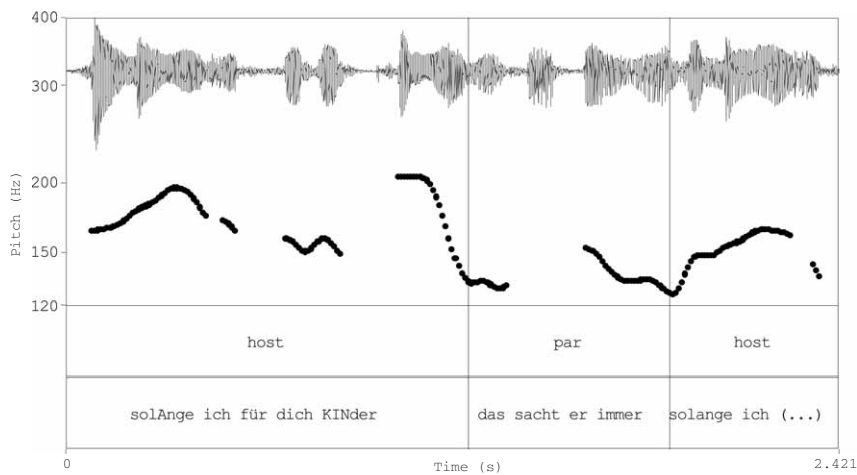


Figure 5: Pitch extract with phonetic-prosodic changes in position 1c and 2 (hh04_31)

A step down in pitch in position 1c, diminished loudness and pitch range, as well as a lower pitch register serve as cues to the separated status of the parenthetical. Although introduced after the first constituent of a noun compound (*kinde*, line 1), no signals for a break-off occur.

The last example (13) is the only case in which the parenthesis is marked as late as in position 2.



126

Pia Bergmann

(13) hh04_44

```
01 i: die EInen ham sich das problem vom HALS geschafft,
    some of them got rid of the problem
-> 02 und: auf die allgeMEINheit <p>sa_ch_ma> über?
    and on the public say I once trans
    [(.) TRAgen; ne,
    ferred; right
    and let's say transferred it to the community
03 hh04: [mh
    mb
```

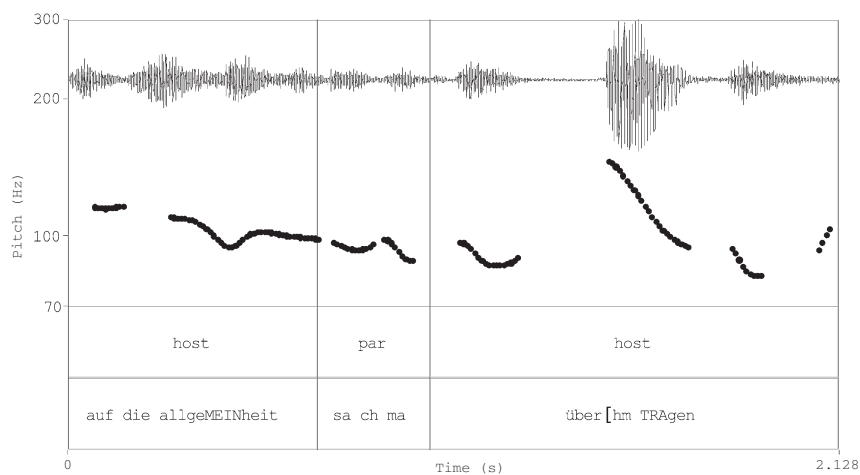


Figure 6: Pitch extract with phonetic-prosodic changes in position 2 (hh04_44)

Three parameters are combined in the prosodic design of this example: It is of lower volume, higher speech rate, and a higher amount of articulatory reduction on the segmental level. Compared with example 12, it is notable that there is no pitch jump at the beginning of the parenthetical. Rather, the parenthesis seems to be integrated into the preceding intonation contour. If it were not for loudness, speech rate, and articulation strength, this example could be viewed as completely integrated into the first part of host. We interpret this as an indication of the relevance of these phonetic-prosodic cues, since they coincide with the duration of the parenthetical and therefore can be viewed as retrospectively contextualizing its external status to the host structure.

To conclude this subsection, it can be argued that in the investigated parentheticals there is no instance of complete integration when phonetic and prosodic devices are taken into account. Despite the fact that they are not set





off by a clear intonation boundary in the phonological sense, and in some cases are not broken off by a glottal stop, the break in syntactic structure coincides with a phonetic-prosodic discontinuity in all cases. Therefore, there seems to be a tendency to minimally mark the syntactic break and to contextualize the parenthetical as something “different” from the ongoing syntactic structure. Furthermore, the analysis showed that in most cases, prospective and retrospective signals combine in the prosodic design of the parentheticals. We encountered no case of prospective marking without retrospective marking. Five cases of the investigated data are marked only retrospectively.

Within the group of Type 1 parentheticals there do not seem to be any strong tendencies concerning the relation between the prosodic design of the initiation phase and the syntactic type of the parenthetical. This is true for the distribution of break-off signals (see Table 4):

Table 4: Distribution of break-off signals

	– break-off signal	+ break-off signal
main clause	3 (20 %)	5 (21.8 %)
subordinate clause	3 (20 %)	4 (17.4 %)
phrase	0	3 (13 %)
stereotypical idiom	9 (60 %)	11 (47.8 %)
total	15	23

Both groups consist of about 20% each for main clauses and subordinate clauses. Phrases occur after a break-off signal only. Parentheticals without a preceding break-off signal consist of stereotypical idioms like *sag ich mal* in 60% of cases, whereas parentheticals with a preceding break-off signal are stereotypical idioms in only 47.8% of the cases. It would be interesting to see if this distribution is confirmed in a larger data base. In addition, it could be interesting to compare the parentheticals that are marked prospectively to those that are marked only retrospectively. However, as has been mentioned above, prospective marking alone does not exist. The five cases marked by retrospective devices only include three stereotypical idioms and two subordinate clauses. Using a data base as small as the present one obviously does not allow for the discovery of systematic trends.

On the basis of the given data, it can be concluded that prospective devices do not foreshadow the syntactic type of the parenthetical but simply serve to signal a break in the on-going syntactic structure. Thus, for the time being, the prosodic design of the initiation phase does not yield an internal differentiation of the parentheticals to come. It would be interesting, how-



ever, to compare the prosodic design of parentheticals to those of other interruptions of an on-going syntactic structure, such as repairs. If not serving an internal differentiation, it is possible that they serve to differentiate between different kinds of interruptions such as repairs or expansions. We assume that it is especially the prosodic design of the parenthetical itself that can serve to distinguish parentheticals from other kinds of interruptions. Moreover, it is possible that on the basis of a larger data set, internal differentiation of parentheticals through prosodic design may be illuminated.

4.1.1.2. Type 2 parentheticals: Phonologically complete host

Type 2 parentheticals are defined by a complete intonation phrase for the first part of the host. As shown in Table 2, twelve out of 50 parentheticals fall into this group. No cases coincide with a break-off signal. This means that at the end of the first part of the host the listener does not receive any prosodic cues as to the syntactic status of the parenthetical. It is clearly separated from its host, but at that point in time it is indistinguishable from an “ordinary” succession of the syntactic project where no disruption by an unrelated element is to follow. Moreover, with the exception of hesitation markers, cues in position 1 such as pauses and breathing may appear, because they are typical for “normal” intonation phrase boundaries (cf. Bergmann and Mertzlufft 2009). As a consequence, if the speaker wishes to contextualize the parenthetical as different from the expected succession of the syntactic project underway, he/she has to do this retrospectively by designing the parenthetical insertion differently from the first part of the host. This is indeed what we find in the data: All instances are characterized by changes in the prosodic design in position 2.

Example (14) is a case in point where changes in the amount of articulatory reduction on the segmental level combined with pitch register are apparent.

(14) hh04_40

```
-> 01  hh04:  und da kann man aber dann: ja jetzt da als
              and there can one but then PTCL now here as
              STEUerprüfer, =
              tax auditor
              but as a well let's say tax auditor
-> 02  <<1>ja_sa_ich_ma:> (.) gAr nichts gegen MACHen;
              PTCL say I once  nothing  against do
              ne,
              right
              you can't do anything against that, can you
```



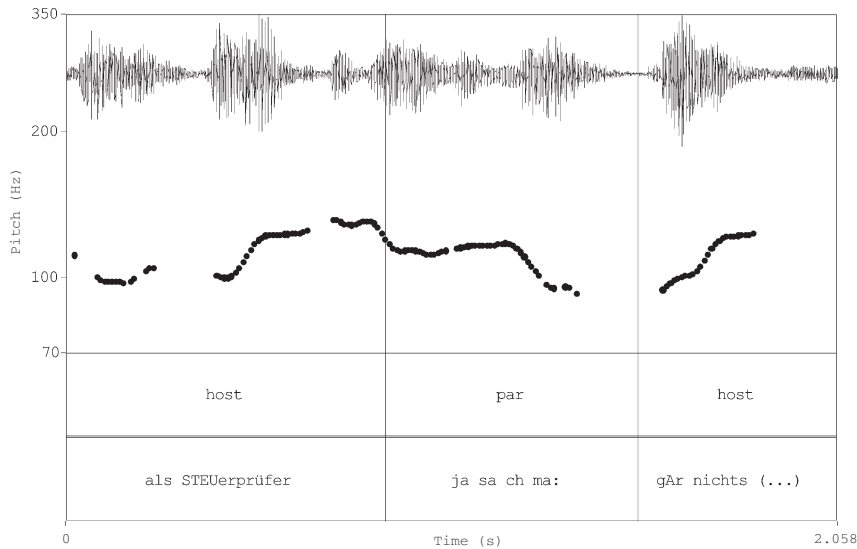


Figure 7: Pitch extract with phonetic-prosodic changes in position 2 (hh04_40)

Example (15) is characterized by a lowering in pitch register. No other cues are used to set off the parenthetical from the preceding host.

(15) hh04_14

- 01 hh04: und mir ist das mAnchmal AUFgefallen;=
and I sometimes noticed
- > 02 =wenn ich so in der mÖnckebergstrasse meine
when I so in the mönckebergstrasse my
MITtagspause mache,=
lunchbreak make
when I am having my lunch break in the Mönckebergstraße
- > 03 =<l>ich arbeite also in der INnenstadt>,
I work PTCL in the inner city
I work in the city centre
- > 04 °hhh äh (.) dass ich dann auf einmal (.) hInter
uh that I then all of a sudden behind
mir (.) jemanden HÖ<<creak>re>,
me someone hear
uh that all of a sudden I hear someone behind me
- > 05 (-) der PLATTdeutsch <<creak>spricht>.
who Low German speak
who speaks Low German dialect

Finally, the last example illustrates a case in which the parenthetical itself is produced with a change in five (all) of the investigated parameters.





130

Pia Bergmann

(16) hh03_9

```
01  hh03:  also sagen_ma vom °hh vom DEUTschen-=
           so say with respect to German
-> 02      =von der orthograPHIE,=
           from the orthography
           to the orthography
-> 03      =<<p,l,acc,creak>we_man das mal so AUSdrücken
           if one that PTCL like that express

           soll.=
           shall
           if you want to express it like that
-> 04      =is das nich nich nich> SAÜber genug;ne?>
           is that not not not clean enough; right
           it's not good enough, is it?
```

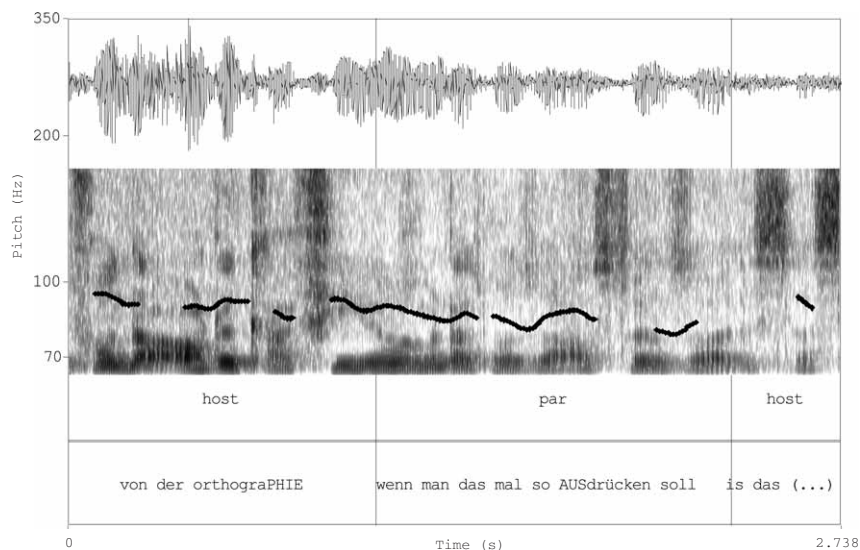


Figure 8: Sound extract with phonetic-prosodic changes in position 2 (hh03_9)

The parenthetical insertion is less loud, diminished in pitch range, and lower in register. Moreover, it is produced with higher speech rate and creaky voice, the latter increasing towards the end of the parenthetical.

To summarize this section, it was shown that all Type 2 parentheticals ($n = 12$) coincide with at least one of the investigated phonetic or prosodic features. A systematic relationship between the prosodic design and specific syntactic types of the parenthetical is not obvious on the basis of the given data. Thus, no conclusion can be made as of yet about whether the prosodic cues are associated with specific contextualizing functions that go beyond





the indication of “otherness”. Again, the question of whether the specific prosodic design serves an internal differentiation of the different (syntactic and functional) types of parenthetical insertions remains a question for future research.

4.1.2. Summary of phonetic-prosodic signals in parentheticals

This section gives a descriptive account of all phonetic and prosodic signals that occur at the transition phases in the host – parenthetical – host structure. For the sake of readability, the list of all features encountered at each specific point in time is listed from Section 3. Following this recapitulation, the quantitative distribution of the features, including their combined occurrence, will be demonstrated.

1a) or 3a) end of host / parenthesis

- break-offs indicated by glottal stops / glottalization (bo)
- (final) lengthening (le)
- creaky voice (cv)
- boundary tones (bt)

1b) or 3b) stretch between host and parenthesis / parenthesis and host

- pauses (p)
- breathing (b)
- hesitation markers (hm)

1c) or 3c) beginning of parenthesis / host

- pitch jumps (pj)
- creaky voice (cv)
- anacrustic syllables (as)
- rush-throughs (rt)

The following table is based on the 50 instances of parentheticals in our data set. The columns refer to points in time, and the rows to phonetic-prosodic cues. The rows are separated by the number of combined cues ranging from zero to four. For instance, the box in the first column in the first row refers to the fact that in position 1a, i.e. at the end of the first part of the host, there are 5 cases in the data set where no phonetic-prosodic signals occur.



Table 5: Prosodic design at the syntactic breaks

	host – parenthesis (position 1)			parenthesis – host (position 3)		
	a	b	c	a	b	c
0	5	28	11	12	27	11
1	19 bo	12 p	31 pj	16 bt	12 p	29 pj
	8 bt	2 b	1 rt	4 le	6 b	2 as
	5 le	5 hm		2 rt	1 hm	1 cv
	1 cv			1 cv		
2	4 bt, le	3 p, hm	3 pj, cv	8 bt, cv	1 p, b	7 pj, cv
	5 bo, le		1 pj, as	3 bt, le	5 b, hm	1 pj, vq
			1 pj, rt	2 rt, cv		
3	–	2 p, b, hm	–	2 bt, le, cv	–	–
4	–	–	1 pj, rt, cv, as	–	–	–

This distribution reveals considerable variation in the way the breaks in syntactic structure are signaled. Most commonly, a boundary tone accompanies the endings of either the host or the parenthetical (position a). It occurs alone or together with lengthening and/or creaky voice. In some cases, however, we find typical phonetic indicators of a boundary that do not coincide with a boundary tone: 9 instances of lengthening, 2 instances of creaky voice, 2 rush-throughs, and 2 rush-throughs combined with creaky voice. These instances exemplify cases with discontinuities that, from a phonological perspective, would not be categorized as a phrase boundary but may still be of some relevance for the listener to recognize that something “different” is to follow. In addition to those hosts that are marked by a boundary tone and/or other boundary signals, the data set includes 23 instances of broken-off hosts. There are five hosts (1a) and 12 parentheticals (3a) in which none of the mentioned parameters occurs.

Position b) is characterized by no occurrence of pauses, breaths, or hesitation markers in roughly one half of all utterances. In the majority of the other half, only one of the investigated parameters arises.

The analysis of position c) reveals that the majority of beginnings are indicated by a pitch jump. In only four cases do we encounter other signals without the occurrence of a pitch jump; these are rush-throughs, creaky voice, and anacrustic syllables. In spite of the fact that the beginnings of parentheticals (1c), as opposed to the beginnings of hosts (3c), are marked by one combination of these four signals, crucial differences in the design of the breaks are not obvious.





After this account of the local discontinuity markers occurring at well-defined points in time, Table 6 presents the prosody of the parenthetical itself. The investigated parameters are loudness, speech rate, strength of articulation, voice quality, pitch range, and pitch register. “+” refers to an increase of the given parameter (e.g., an increase in loudness), “-” to a decrease of the parameter, and “0” to no audible change in the parameter under investigation, as compared to the first part of the host.

Table 6: Prosodic design of the parenthetical

	loudness	speech rate	articulation	voice quality	pitch range	pitch register
+	4	17	1	–	2	2
–	24	2	22	3 creaky	19	27
0	22	31	27	47	29	21

According to these results, pitch register is the parameter most affected by parenthetical structure, whereas change in voice quality rarely occurs. Many parentheticals follow the expectations that they should be less loud, have a higher speech rate, be diminished in pitch range, and be lower in pitch height. Of interest here is that 4 parentheticals are produced at a louder volume than the first part of the host; 2 cases show slower speech rate, 1 case more precise articulation, 2 cases a wider pitch range, and 2 other cases higher pitch register. Cases such as these are opposite to what was expected based on previous findings concerning the prosody of parentheticals. A functional analysis of the parentheticals in their context might lead to interesting explanations for this unexpected design.

4.2. The prosodic design of multiple parentheses

This section analyzes the multiple parentheses that were found in the data. In addition to 50 single parentheses, the data set consists of 13 parentheticals with a structure more complex than the threefold host – parenthesis – host structure. Two examples will be discussed.

The first example (17) demonstrates how the prosodic design of two consecutive parentheses serves to contextualize their status of being subordinate to a shared host and of the second parenthesis as being simultaneously subordinate to the first one. It is argued that this is achieved by an up- and downgrading of prosodic boundary strength, and by a context-sensitive design of the parentheses as a whole. The extract begins with the first part of the host, which is syntactically incomplete and whose completion is uttered in line 4. Between both parts of the host, two parentheses are inserted (line 2 and line



134

Pia Bergmann

3). The first gives background information, while the second produces some kind of ironic self-assessment:

(17) hh03_26–27

```
-> 01   hh03:   im nÄ:chsten MO:nat-
               in the next month
               next month
-> 02           °hh g hat meine frau mir AUCH zum
               g has my wife me also to the
               gebÜrts<<creak>tag geschenkt-=
               birthday given
               my wife has given me that as a birthday present too
-> 03           =also (.) ich war wahrscheinlich> gAnz ARTig
               PTCL I was probably very well-behaved
               <<p,breathy>dieses jahr>;
               this year
               so I probably was very well-behaved this year
-> 04           °hhh äh gehen wir zu <<creak>hElmut LOTti>;
               uh go we to helmut lotti
               we are going to see Helmut Lotti
05   i:         ah ja
               I see
```



With respect to phrasing, each parenthesis constitutes a complete intonation phrase. Comparing the boundaries between the parentheses with the ones between the whole parenthetical insertion and its host, however, reveals interesting differences in prosodic design. It seems that the boundary between the parentheses is blurred by timing and voice quality. The phrases are latched onto each other, and creaky voice serves not only to indicate the end of the first parenthetical but also continues until the first accent in the second parenthetical (line 2 – line 3). Compared to this, the boundaries that separate both parentheticals from their hosts are marked more strongly (line 1 – line 2; line 3 – line 4). In the initiation phase, we find breathing, a broken-off segment, and a step down in pitch before the parenthetical begins. In the return phase, we have breathy voice and a low boundary tone at the end of the parenthetical, followed by breathing, a hesitation marker, and a pitch reset as well as normal voice quality at the beginning of the host. The second part of the host thereby continues not only syntactically, but prosodically too the suspended first part of the host. Through fine calibration of the phrase boundaries, the parentheticals are contextualized as belonging together more than belonging to their host. Moreover, both parentheticals are downgraded with respect to the host by means of lower pitch register. Still, despite the fact that both parentheticals are thus set apart from the host, they them-





selves are contextualized as different by subordinating parenthesis 2 under parenthesis 1. This is achieved by changing loudness and voice quality. In parenthesis 2, only the accented syllables are produced in modal voice. The syllables preceding the accented syllables are produced with creaky voice, and the syllables after the accented stretch are produced with breathy voice and diminished loudness. In conclusion, it is argued that the speaker uses prosodic devices to display the complex relationship of the parentheticals to each other as well as to their shared host.

The next example (18) is comparable to the preceding one in introducing two parentheticals, one after the other (line 5, line 6). Parenthesis 1 introduces a meta comment concerning the epistemic stance of the speaker, parenthesis 2 formulates an assessment of modern young people's interest in politics. Both parentheticals serve to search for a topic that, according to the speaker, resembles in relevance the topic "death penalty" for young people nowadays. Before resuming the host by a retraction to its beginning, the speaker utters another parenthetical (*sagen_wir_ma*, 'let's say', line 7). The host is then completed (line 8).

(18) hh04_25–27

```
-> 01  hh04:  als Ich zur SCHUle gegangen <<creak>bin>;
           when I went to school
-> 02      war das thEma TODEsstrafe;; (.)
           the topic death penalty was
-> 03      °hhh (1.1) nja unter uns SCHÜlern so;;
           well for us students
-> 04      (-) so ein THEma wie;;
           a topic like
-> 05      (--) <<pp,l>ja ich WEISS nich->
           PTCL I know not
           I don't know
-> 06      (--) <<p,l>hEUte' (--) sind die meisten °hh
           today are the most
           jUGendlichen ja kAUm noch polItisch interessIERT.>
           adolescents PTCL hardly anymore politically
           interested
           nowadays most young people are not interested in politics
-> 07      °hh sagen_wir_mal so_n thEma,=
           say we once such a topic
           let's say a topic
-> 08      =wie der wie das schIcksal des ha es FAU.
           like the fortune of the HSV [name of a soccer team]
09  i:      ah JA.
           I see
```





Prosodically, both parentheticals are set apart from the host by diminished loudness and lower pitch register. The initiation is achieved by lengthening the end of the first part of the host, falling pitch, and a pause. Considering the return phase, it is interesting that the break between parenthesis 2 and parenthesis 3 is more strongly marked than the break between parenthesis 3 and the host. Parenthesis 1 and 2 constitute a complete intonation phrase each. Relative to each other, parenthesis 1 is downgraded with respect to parenthesis 2 by means of lower volume. Comparable to example (17), the speaker achieves a double orientation of two parentheses towards each other and towards their host by the use of prosodic devices.

These examples highlight the fact that prosodic cues are used as a subtle device to calibrate categorical phrasing by weakening or enhancing boundary strength. Moreover, both examples demonstrate that prosodic cues are interpretable only relative to their context of occurrence. The concrete loudness or pitch level of an utterance as well as, for instance, the occurrence of breathy voice in example (17), are only interpretable when being perceived as changes from attributes such as higher pitch level or creaky voice in the previous utterance. Thus, these prosodic cues can only be attributed to their function when taking into consideration their articulation in the process of speech production.



5. Summary and conclusions

This chapter aimed to develop a formal description of the prosodic design of parentheticals in spontaneous German. It focussed on the syntactic breaks in parenthetical sequences and the way they coincided with prosodic and phonetic cues. A phonological view of phrasing that disregards prosodic boundary cues was rejected in favor of a view including phonetic variance as a potentially important cue to syntactic structure.

The results of the empirical investigation of 50 *single parentheticals* demonstrate that different phonetic and prosodic cues and combinations thereof coincide with the syntactic break. The most common cues are related to pitch, meaning we find boundary tones and pitch jumps as well as changes in pitch register in the parenthesis itself. The analysis of the initiation phase of the parenthesis revealed that there is no case in our data set of a completely integrated parenthesis. If the host structure and the parenthetical are not set apart by an intonation phrase boundary, there are indications of a prosodic discontinuity in all cases. These can occur at the end of the host structure, thereby serving as a contextualization cue for the current hearer to anticipate the upcoming syntactic break even in the absence of either an intonation





phrase boundary or a break-off signal. Phonetic-prosodic cues at the end of the host combine predominantly with cues during the parenthetical, i.e. in most cases, prospective and retrospective signals combine in order to set apart the parenthetical insertion from the preceding host. Retrospective signaling is more stable in this respect, as only five cases of only retrospective signaling, but no cases of only prospective signaling, were found. This means that the external syntactic status of the parenthetical always correlates with at least one signal of prosodic discontinuity or change. The same is true for those parentheticals that follow a complete intonation phrase: If the parenthesis is disintegrated in a phonological sense, the parenthetical is always marked by additional phonetic-prosodic cues. We interpret this stable correlation of the syntactically external element with prospective and/or retrospective phonetic-prosodic signals to be an indication of the relevance of these signals for the contextualization of the interruption. Many questions, however, remain for future research. First, no systematic relation between the phonetic-prosodic design of the initiation phase and the syntactic type of the parenthetical insertion was found. The prosodic design does not seem to yield an internal differentiation of the parenthetical insertions. An analysis of a larger data set may allow for deeper insight into this matter. Second, prosodic design may be influenced by the functional type of the parenthetical, a point of interest which has not been analyzed in the present article. A third promising direction may be to compare the prosodic design of parenthetical insertions with the design of other insertions/changes of a syntactic structure already underway. Although the systematic correlation of phonetic-prosodic detail with a syntactic structure like parentheticals is already a case in point to argue for the relevance of these signals in Interactional Grammar, the argument could be further strengthened by showing that the prosodic design serves to differentiate between different kinds of insertions. This, however, is an empirical question for future research.

In a detailed analysis of two cases of *multiple parentheticals*, it was shown how prosodic cues served to package both parentheticals as belonging together when compared to their shared host structure, but as being clearly distinct with respect to each other. It was argued that the up- and down-grading of phonological boundaries plays a role, as well as the way the parentheticals themselves are prosodically designed. The importance of recognizing that prosodic signals are only interpretable in relation to their (prosodic) context was also demonstrated. General conclusions drawn from the empirical investigation are (1) that because prosodic cues are systematically used in phrasing, the phonetic details of how (syntactic) breaks are articulated should therefore not be dismissed beforehand; and (2) that it is crucial to



view prosodic cues in their relation to the immediate context and therefore as embedded into the sequence as it unfolds in time.

In addition to the empirical investigation, it was discussed whether there was a way to combine the theory of PP and AM-Phonology in its application to intonation with the IL approach. Two concepts which superficially expose some similarity in both approaches were chosen for the discussion: the concept of interfaces and the concept of co-occurrence. The discussion concerning the compatibility of the approaches yielded a negative result. It was argued that despite the appealing nature of theorizing coinciding boundaries on different levels of linguistic structure, more basic aspects of language production are not conducive to a combination of these approaches. Due to two diverse conceptualizations of language – as either pre-conceived sentences that have to be mapped on a certain inventory of pre-existing prosodic constituents or as linguistic structure as emergent over time – these approaches are not amenable. The idea of the online-production of language has been theorized for prosody by Couper-Kuhlen (2007). The “cesura”-approach suggested by Barth-Weingarten (2011) points in a similar direction by assuming that it may be more fruitful to take into account boundaries of varying strength without the need for them to segment the speech stream in clearly definable and separable units. From the present study, no conclusion can be drawn with respect to whether units or boundaries should be the primary focus of investigation. Relating back to the general conclusions above, it can be said, however, that for a theory that is based on naturally occurring talk, it would be beneficial to incorporate the ideas of context-dependence and process-orientation, as well as those of phonetic detail at interfaces between different levels of linguistic structure.

Acknowledgments

I would like to thank Jana Brenning, Martin Pfeiffer, Elisabeth Reber and Peter Auer for many helpful comments on an earlier version of the chapter. All remaining errors are of course mine.

Appendix

Transcription conventions (according to GAT 2, cf. Selting et al. 2009)

[] overlap and simultaneous talk
[]

°hh / °hhh inbreath of appr. 0.5–0.8sec./0.8–1.0 sec. duration





The prosodic design of parentheses in spontaneous speech

139

(.)	micro pause, estimated, up to 0.2 sec. duration
(-)	short estimated pause of appr. 0.2–0.5 sec. duration
(--)	short estimated pause of appr. 0.2–0.5 sec. duration
(---)	short estimated pause of appr. 0.2–0.5 sec. duration
(0.9)	measured pause
we_man	cliticizations within units
äh	hesitation markers, so-called “filled pauses”
=	fast, immediate continuation with a new turn or segment
: / :: / :::	lengthening, by about 0.2–0.5/0.5–0.8/0.8–1.0 sec.
?	cut-off by glottal closure
MORgen	focus-accent
mOrgen	secondary accent
?	rising to high (final pitch movement)
,	rising to mid (final pitch movement)
-	level (final pitch movement)
;	falling to mid (final pitch movement)
.	falling to low (final pitch movement)
↑	smaller pitch upstep
↓	smaller pitch downstep
<<1>	> lower pitch register
<<p>	> piano, soft
<<pp>	> pianissimo, very soft
<<creaky>	> glottalized
<<breathy>	> change in voice quality stated
Abbreviations	
PTCL	Particle



References

- Auer, P. 2005 Projection in interaction and projection in grammar. *Text – Interdisciplinary Journal for the study of discourse* 25: 7–36.
- Auer, P. 2009 Online-Syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31: 1–13.
- Auer, P. 2010 Zum Segmentierungsproblem in der Gesprochenen Sprache. *InLiSt – Interaction and Linguistic Structures*, No. 49.
- Barth-Weingarten, D. 2011 The fuzziness of intonation units: Some theoretical considerations and a practical solution. *InLiSt – Interaction and Linguistic Structures*, No. 51.
- Bergmann, P. 2008 *Regionalspezifische Intonationsverläufe im Kölnischen. Formale und funktionale Analysen steigend-fallender Konturen*. Tübingen: Max Niemeyer Verlag.
- Bergmann, P. and C. Mertzlufft 2009 Die Segmentierung spontansprachlicher Daten in Intonationsphrasen. Ein Leitfaden für die Transkription. In: Karin Birkner and Anja Stukenbrock (eds.), *Die Arbeit mit Transkripten*, 83–95. Verlag für Gesprächsforschung.
- Boersma, P. and D. Weenink 2009 *Praat: doing phonetics by computer* [Computer program]. <http://www.praat.org/>.
- Couper-Kuhlen, E. 2007 Prosodische Prospektion und Retrospektion im Gespräch. In: H. Hausendorf (ed.), *Gespräch als Prozess*, 69–94. Tübingen: Gunter Narr Verlag.
- Couper-Kuhlen, E. and M. Selting 1996 Towards an interactional perspective on prosody and a prosodic perspective on interaction. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in conversation*, 11–56. Cambridge: Cambridge University Press.
- Dehé, N. 2007 The relation between syntactic and prosodic parenthesis. In: N. Dehé and Y. Kavalova (eds.), *Parentheticals*, 261–284. Amsterdam: John Benjamins Publishing Company.
- Dehé, N. 2009 Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics* 45: 569–615.
- Döring, S. 2007 Quieter, faster, lower, and set off by pauses? Reflections on prosodic aspects of parenthetical constructions in modern German. In: N. Dehé and Y. Kavalova (eds.), *Parentheticals*, 285–307. Amsterdam: John Benjamins Publishing Company.
- Duvallon, O. and S. Routarinne 2005 Parenthesis as a resource in the grammar of conversation. In: A. Hakulinen and M. Selting (eds.), *Syntax and lexis in conversation*, 45–74. Amsterdam: John Benjamins Publishing Company.
- Ford, C. E. and S. A. Thompson 1996 Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and grammar*, 134–184. Cambridge: Cambridge University Press.
- Fortmann, C. 2005 On parentheticals (in German). In: M. Butt and T. Holloway King (eds.), *Proceedings of the LFG05 Conference*, 166–185. University of Bergen.
- Fougeron, C. and P. A. Keating 1997 Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101 (6): 3728–3740.
- Grabe, E. 1998 *Comparative intonational phonology: English and German*. Wageningen: Ponsen and Looijen.





- Keating, P. A., T. Cho, C. Fougeron and C. Hsu 2003 Domain-initial strengthening in four languages. In: J. Local, R. Ogden and R. Temple (eds.), *Papers in Laboratory Phonology 6: Phonetic interpretations*, 145–163. Cambridge: Cambridge University Press.
- Kaltenböck, G. 2005 Charting the boundaries of syntax: a taxonomy of spoken parenthetical clauses. *VIEWS* 14: 21–53.
- Kaltenböck, G. 2007 Position, prosody, and scope: the case of English comment clauses. *VIEWS* 16: 3–38.
- Ladd, R. 2008 *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.
- Local, J. 1992 Continuing and restarting. In: P. Auer and A. di Luzio (eds.), *The contextualization of language*, 173–286. Amsterdam: John Benjamins Publishing Company.
- Local, J. and J. Kelly 1986 Projection and ›silences‹: Notes on phonetic and conversational structure. *Human Studies* 9: 185–204.
- Local, J., J. Kelly and W. Wells 1986 Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics* 22: 411–437.
- Local, J., W. Wells and M. Sebba 1985 Phonology for conversation. Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9: 309–330.
- Mazeland, H. 2007 Parenthetical sequences. *Journal of Pragmatics* 39: 1816–1869.
- Nespor, M. and I. Vogel 2007 *Prosodic phonology*. 2nd edition. Berlin: de Gruyter.
- Peters, J. 2006 Syntactic and prosodic parenthesis. *Speech prosody* 2006, Dresden, Germany. ISCA Archive, <http://www.isca-speech.org/archive>.
- Pfeiffer, M. this volume What prosody reveals about the speaker's cognition: Self-repair in German prepositional phrases.
- Schönherr, B. 1993 Prosodische und nonverbale Signale für Parenthesen. *Deutsche Sprache* 21: 223–243.
- Selting, M. 1995 *Prosodie im Gespräch*. Tübingen: Niemeyer.
- Selting, M. and E. Couper-Kuhlen 2000 Argumente für die Entwicklung einer ›interaktionalen Linguistik‹. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*: 76–95.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock and S. Uhmann 2009 Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.
- Stoltenburg, Benjamin 2003 Parenthesen im gesprochenen Deutsch. *InLiSt – Interaction and Linguistic Structures*, No. 34.





Beatrice Szczeppek Reed
University of York

Prosody, syntax and action formation: Intonation phrases as ›action components‹

1. Introduction

This chapter asks whether stretches of talk, similar to *intonation phrases* in size and boundary cues, are oriented to by conversational participants as relevant for spontaneous spoken interaction. The focus is on native speakers of British and American varieties of English. Both the existence and the relevance of intonation phrases are taken for granted by many interactional, and other, linguists, but they play hardly any role in the work by conversation analysts with a background in sociology. This is most probably the case because linguists are trained in a tradition that places high emphasis on units (of language), whereas sociologically minded analysts of language-in-conversation are predominantly interested in the system of social interaction, specifically the order and organization of sequences (Schegloff 2007), and the resulting sequential locations, or “slots” (Raymond 2010). It therefore seems sensible to ask whether intonation phrases are indeed a unit of interaction, that is, whether participants orient to them as separate, meaningful entities. If we cannot find convincing evidence that intonation phrases are indeed relevant for participants themselves, it does not seem appropriate to continue to use them, neither as an analytical category, nor as a unit for transcription. If, on the other hand, we do find proof that participants observably employ a phrasing mechanism for the accomplishment of actions which is not already captured in established concepts such as the turn and the Turn Constructional Unit (TCU), then it may be desirable to introduce this unit into a broader conversation analytic framework.¹ This chapter does not claim to answer this question definitively. It simply repre-

¹ Research on Japanese (Tanaka, 1999; 2000) and Korean (Kim, 1999; Young and Lee, 2004) has shown that speakers of these languages place response tokens at intra-turn locations, which is initial evidence for an orientation to sequentially relevant slots earlier than the Transition Relevance Place (TRP). However, research on English so far has claimed that response tokens are placed at TRPs and their expansions, arguing that response tokens are practices for ›passing up an opportunity to propose a full turn at talk‹ (Schegloff 1982: 81).





Kolumnentitel gekürzt, da zu lang

Prosody, syntax and action formation

143

sents a first step in one of several directions that could be taken in its exploration.²

Before we go on to explore these questions further, extract (1) provides several typical examples of what previous linguistic literature has referred to as intonation phrases. The extract comes from the Santa Barbara Corpus of Spoken American English, and occurs during an extended turn in which Doris expresses her surprise at a local politician's behavior.³

(1) SBC011 This Retirement Bit

```
01   Dor:   and you KNOW-
02         he MUST 'a been THINKing,
03         (0.74)
04         <<all> when he sAId he wAsn't a polItical appointEE;>
05         (0.37)
06         he MUST have been a- a- asSOciating,
07         (0.51)
08         his reLAtionship;
09         (0.25)
10         WI:TH (0.14) stEtItler
```

Throughout this extract the speaker produces her talk as a series of short chunks, often leaving considerable pauses in between them. In listening to the recording, one can hear that each chunk is delivered as one overarching intonation contour, and that each contour contains at least one pitch accent. All chunks are either complete syntactic clauses (lines 1, 2, 4), or phrases (lines 6, 8, 10).

Showing that speakers divide their talk into shorter chunks is relatively easy, given that no one is physically able to produce infinite stretches of speech. However, intonation phrases are linguistically conceptualized as more than simply “talk between breaths”. Many students of language have described chunks such as those above as linguistically meaningful units.⁴ Providing evidence for the interactional relevance of such chunks as a defined unit of talk – in addition to, and distinct from, units already described in the lit-

² Another step, in a different direction, but exploring the same issues, is Dagmar Barth-Weingarten's continuing work on intonation units, see, for example, Barth-Weingarten (2007) and (forthcoming).

³ See section 2 for more information on the data. See the Appendix for transcription notations.

⁴ Cf. Palmer (1922), Armstrong and Ward (1926), Schubiger (1935, 1958), Kingdon (1958), O'Connor and Arnold (1961/1973), Halliday (1967), Crystal (1969), Couper-Kuhlen (1986), Du Bois (1991), Chafe (1993), Brazil (1997), Cruttenden (1997), Selting (1995), (Wells 2006), Barth-Weingarten (2007).



erature, such as TCUs (cf. Sacks, Schegloff, and Jefferson 1974; Selting 1996, 2000), increments (cf. Ford, Fox, and Thompson 2002; Walker 2004; Couper-Kuhlen and Ono 2007), and parentheses (cf. Thompson and Mulac 1991; Mazeland 2007; Bergmann 2010, this volume) – is a much greater challenge.

In this chapter an attempt is made to approach chunks of talk such as those above from an actional, rather than a linguistic perspective. If the data show convincing evidence that conversationalists treat chunks as relevant units for the accomplishment of social actions, in addition to other action units such as the turn and the TCU, then their linguistic definition can perhaps follow – although we can already hypothesize that such a definition will have to be maximally flexible, given the high tolerance by conversational participants regarding the flexible syntactic and phonological structure of other units of talk. Turns and TCUs come in a variety of linguistic shapes and sizes, and this smaller action unit may be equally hard to pin down precisely. However, in contrast to purely linguistic treatments of the intonation phrase, in which a syntactic and phonological definition is the final analytical goal, this chapter attempts to define chunks interactionally, without the constraints of prescriptive definitions, and with the description of a repertoire of formal practices for action as the ultimate aim. This chapter pursues the notion that the contingent and emergent nature of spontaneous talk requires turn- and action-building resources that are able to expand or contract as required. After a short introduction of the data (section 2) and a brief overview of previous literature on intonation phrases (section 3), section 4 presents an exploration into whether short stretches of talk may be such a resource. The final section offers some concluding observations.

2. Data

The data presented below come from a corpus of 52 hours of audio recordings of spontaneous interactions, some face-to-face, some over the telephone. Extracts (1) and (4) – (7) are available via the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000, 2003, 2004, 2005), which is available at <http://www.talkbank.org/data/Conversation>, (MacWhinney 2007), or at the Linguistic Data Consortium, <http://www.ldc.upenn.edu>. Extracts (2) and (3) come from a private corpus of spoken English collected by Elizabeth Couper-Kuhlen, and currently held at the University of Helsinki. Extract (3) is taken from a broadcast phone-in programme; all other extracts are instances of private talk.

The examples of collaborative turn sequences come from a specified collection of 200 such sequences. The extracts for list construction and se-





quence initiation are not parts of specifically collated corpora of these phenomena, but instead are provided for showing local interactional order by single case analyses (Schegloff 1987). All data extracts presented in this chapter have been transcribed according to an adapted version of the original GAT transcription conventions (Selting et al. 1998). The most relevant transcription symbols can be found in the Appendix.

3. Previous work on intonation phrases

Various existing approaches to intonation have conceptualized the intonation phrase in different ways. One important distinction exists between the notion of a single, holistic entity in the so-called British school of intonation research, and the concept of groups of individual tones in autosegmental-metrical phonology. This chapter focuses on the British approach, as it has been the primary influence on current research on discourse prosody. For summaries of the concept of the intonation phrase in autosegmental-metrical phonology see, for example, Ladd (1996: 235–251) and Grice (2006).

The intonation phrase is typically defined as a linguistic unit consisting of a coherent pitch movement and accentual pattern, and typically associated with clearly defined syntactic entities. This unit has been referred to as intonation phrase (Wells 2006), but also as intonation unit (Du Bois 1991; Chafe 1993; Barth-Weingarten 2007), intonation-group (Cruttenden 1997), tone unit (Crystal 1969; Brazil 1997), tone group (Halliday 1967; Brown, Currie, and Kenworthy 1980), rhythm unit (Pike 1945), breath group (Liebermann 1967), and speech bar (“Sprechtakt”, Klinghardt and Klemm 1920; Klinghardt 1923). In very broad terms an intonation phrase contains a spate of talk delivered as one recognizable overall intonation movement. In an unproblematic case, this would contain a pitch accent near the beginning, and another, more prominent one on the final stressed syllable. The contour would start with a comparatively high pitch onset on the first stressed syllable, which would be followed by gradual declination in overall pitch register and loudness. The final syllable would be lengthened, irrespective of its degree of stress, and the whole phrase would be preceded and followed by short pauses (Crystal 1969; Couper-Kuhlen 1986; Cruttenden 1997; Wells 2006).

Many phonological approaches consider intonation in close relation to other linguistic systems, typically syntax, focus, and information structure (cf. Halliday 1970; von Heusinger 1999; Gussenhoven 1984). For example, Wells (2006) defines the function of individual pitch accents as related to specific syntactic constructions, such as falling pitch for WH-questions. He also defines the location of the syllable with primary emphasis (the *nucleus*)



within the intonation phrase as indicative of whether the speaker is expressing broad, or narrow focus, and old or new information. In terms of the structure of intonation phrases, Cruttenden (1997) defines “internal” and “external criteria” for “intonation-groups”. His internal criteria include at least one stressed syllable, and pitch movement on, to, and/or from that stressed syllable. External intonation-group criteria define their potential boundaries. Criteria include pausing; anacrusis, that is, fast delivery of unstressed syllables before the first pitch accent; lengthening of the final syllable; and a change in pitch from one intonation-group to the next.⁵

Some discourse analysts and interactional linguists have also been interested in intonation phrases, mainly with a focus on the transcription of naturally occurring talk. Du Bois (1991) and Du Bois et al. (1993) put forward the transcript notation known as Discourse Transcription (DT), with the “intonation unit” as one of its central categories. Du Bois et al. (1993: 47) define it as “a stretch of speech uttered under a single coherent intonation contour”, with potential initial cues of pausing and an upward shift in overall pitch, and a potential final cue of syllable lengthening. Similar to the phonological literature, where the main final accent – the nucleus – is an important focal point, analysts of naturally occurring interaction have also focused primarily on the pitch movement at the end of an intonation phrase. However, this has been with a principle interest in discourse function, rather than phonological form. Put very simply, the final pitch movement of an intonation phrase is frequently interpreted in terms of whether it projects *completion* or *continuation* (of a sentence, an idea, or a turn-at-talk; see also Chafe 1993; Gumperz 1993; Selting 1995). This interest is due to the role ascribed to prosody for turn-taking and turn construction (cf. Local, Wells, and Sebba 1985; Local, Kelly, and Wells 1986; Selting 1996; Wells and Peppé 1996; Schegloff 1998; Wells and Macfarlane 1998; Fox 2001; Szczeppek Reed 2004) and narrative structure (cf. Chafe 1980, 1987, 1988, 1993). Investigations of these relationships routinely link the prosodic form of intonation phrase endings to turn continuation or closure. It is therefore possible to argue that for most, if not all discourse-related approaches, an interest in intonation phrases as a holistic category, while relevant for transcription, takes second place behind a primary interest in the prosodic shape of the final syllable(s), and thus in intonation phrase *boundaries*.

⁵ More in-depth explorations of intonation phrase structure can be found in publications by some of the most prominent contributors to the British school of intonation analysis, notably Palmer (1922), Armstrong and Ward (1926), Schubiger (1935, 1958), Kingdon (1958), O'Connor and Arnold (1961/1973) and Halliday (1963/1973, 1967, 1970). Reviews of the British school of intonation can be found in Gibbon (1976), Crystal (1969) and Couper-Kuhlen (1986).



This priority – i.e. an interest in the nature of intonation phrase endings – is also present in one of the most recent investigations into intonation phrases in English from an interactional linguistic perspective. Barth-Weingarten (2007) asks whether “intonation units” actually exist in everyday talk-in-interaction. The author starts from the assumption that if they do, then the way in which participants design phrase endings can be expected to show similarities to the design of turn endings. Barth-Weingarten finds that prosodic strategies that mark turn endings, such as pitch peaks, syllable lengthening and diminuendo, are indeed also present at the end of potential turn-internal intonation units, albeit in a reduced form. Her finding provides initial evidence that participants do indeed structure their talk by orienting to a speech unit of intonation phrase-like length and form.

In a number of recent publications (Szczepek Reed 2010a, 2010b, 2010c) I have addressed the difficulties of applying the model of intonation phrases as described in phonological literature to naturally occurring talk. I have found that, while participants do indeed produce talk in chunk-sized components, the prosodic and syntactic features of these chunks are not easily defined along the lines of previous literature on intonation phrases. For example, while phonological approaches in the British tradition stipulate that there can be only one major accent (nucleus) per phrase, ordinary talk shows an abundance of instances in which participants place equally strong emphasis on more than one syllable per chunk.⁶ Moreover, the frequent claim that intonation phrases always overlap with syntactically defined entities, such as clauses or phrases, does not hold up in an analysis of spontaneous, everyday speech (cf. Cruttenden 1997). In fact, participants often design chunks onto which it would be difficult to impose an interpretation of some form of traditional syntactic entity. Furthermore, as all transcribers of natural talk who use an intonation phrase-based transcription system know, major problems arise over the question where the boundaries between individual phrases lie in each individual case.⁷ As with many other interactional practices, chunks seem to come in many shapes and sizes, and in attempting

⁶ Peters (2006) describes as *prosodic parenthesis* instances in which parentheticals are prosodically integrated into a matrix sentence, but maintain their nuclear accent. The result are intonation phrases containing more than one nuclear accent.

⁷ This is also mentioned by Cruttenden (1997), who, however, ascribes this issue to the imperfections of spontaneous talk: ›When we consider spontaneous speech (particularly conversation) any clear and obvious division into intonation-groups is not so apparent because of the broken nature of much spontaneous speech, including as it does hesitation, repetitions, false starts, incomplete sentences, and sentences involving a grammatical caesura in their middle.‹ (p. 29)



to define their structure, one is soon reminded of Hopper's (1992) famous finding concerning the diverse nature of telephone openings, where the previously established model "shows close detailed fit with only a minority" of cases (Hopper 1992: 90).

It has been pointed out in this section that two very different fields of language study draw on the notion of a prosodically defined unit of language, situated roughly above the word and below the sentence (although, of course, most would agree that intonation phrases can contain single words, or whole sentences). While phonological definitions go into great detail regarding the prosodic and syntactic structure of this unit, interactional approaches are more interested in the prosodic forms and interactional functions of its boundaries.

4. Units of talk and the formation of actions

Recent research in CA and interactional linguistics (Selting and Couper-Kuhlen 2001) has detailed the way in which resources for everyday interaction, such as lexis, syntax, prosody, gesture and body posture, have to be flexible enough to deal with the emerging nature of interaction, and the contingencies that continuously accompany it. Unsurprisingly, the turn, one of the most central units of interaction, is also among the most flexible ones. In terms of its component parts, it can consist of anything from a single monosyllabic word or token, up to an entire conversational narrative. Turns are defined retrospectively, by observations of co-participants treating them as finished. In order to deal with those conversational units that could be turns, but for specific interactional reasons are not treated as such, a second type of interactional unit has been introduced, the TCU, defined by most as a potential turn, that is, a stretch of talk after which "transition to a next speaker becomes *relevant* (although not necessarily accomplished)" (Schegloff 1996: 55, emphasis in the original). Adding to this definition, Selting (2000) argues that conversational participants also orient to TCUs that do not make turn transition relevant, and thus do not end in Transition Relevance Places (TRPs). With this line of argumentation, Selting maintains the category of the TCU as the "smallest interactionally relevant complete linguistic (unit) in their given context. They end in TRPs, unless particular linguistic and interactional resources are used in order to project and postpone TRPs to the end of larger turns" (Selting 2000: 512). Where conversation analytic literature has in the past discussed *multi-unit turns*, what is usually referred to by this term are turns that contain more than one TCU, as in the case of story tellings (Selting 2000), or turns that are designed to accomplish more than one action (Local and Walker 2004).





Ford, Fox, and Thompson (1996) and Ford (2004) sound a note of caution concerning any attempt to define TCUs, and interactional units in general. Firstly they argue that “the ultimate ›indefinability‹ of TCUs is essential to their functionality” (Ford, Fox, and Thompson 1996: 428). Secondly, Ford expresses concern that “the drive to define units may cause us to miss systematic practices that make conversation work for participants in real contexts of use” (Ford 2004: 38). In this chapter, this problem is readily acknowledged. The more rigorous definition of TCUs suggested by Schegloff (1996) is adopted here, which assumes their capability to act as potential turns if and when participants decide to treat them as such, thus always ending in an opportunity for speaker transition. Stretches of talk that are clearly not potentially complete turns, but that are nevertheless packaged by participants as entities, are described below as *action components*. However, the data show that a definition of the internal structure of potential action components-as-units is highly problematic. For example, a syntactic definition of them is not always possible.⁸ Instead, it is suggested that future analysts focus on the way participants demarcate units, i.e. on unit boundaries, rather than on essential features of the units themselves (Auer 2010; Barth-Weingarten forthcoming).

Unlike linguistic units, such as the clause or the sentence, both the turn and the TCU are fundamentally units of interaction (although the TCU is often defined in linguistic terms, see Sacks, Schegloff, and Jefferson 1974; Selting 2000). Turns provide the main formatting resources for the accomplishment of actions, and complete turns are typically defined, among other things, as being complete in terms of the conversational action accomplished through them. Ford and Thompson (1996) define turn endings as points of syntactic, prosodic and pragmatic completion. However, as we look more closely at the formats of certain conversational actions, the data show that many turns and TCUs are made up of two or more shorter stretches of talk, which do not in themselves accomplish complete actions, but which contribute specific action components. Lerner (1991, 1994), in his investigation of compound TCUs, describes how TCUs can be made up of more than one syntactic clause. One example analyzed in more detail by Lerner is the *if X, then Y* format, consisting of a preliminary component *if X* and a final component *then Y*.

⁸ This distinguishes our *action component* from Selting’s (2000) TCUs which do not end in TRPs, as Selting defines those as “complete linguistic units” (2000: 512). In addition, for Selting “the TCU is not identical with an ›intonation unit‹ or ›prosodic unit‹” (Selting 2000: 490).



In this chapter it is argued that in order to represent the flexible structure of TCUs as designed for dealing with interactional contingencies, it may be helpful to have an additional concept of a unit of action formation below that of the TCU. The extracts below present instances of action formats that require more than one component. It will be argued that individual components are designed as coherent entities, and as separate from other such entities, and that they can at times take the form of units below the level of the TCU. In this they are similar to the compound TCU components described by Lerner (1991, 1994); however, in contrast to Lerner's components, the chunks introduced here do not rely on a syntactic definition. The stretches of talk under investigation here are not potential turns – although turns and TCUs can, of course, take the size of single chunks; and as they are not necessarily syntactically complete they are not TCUs that do not end in a TRP (Selting 2000).

As has been mentioned above, most approaches that adopt the intonation phrase as a relevant analytical concept treat it as a structuring mechanism for information flow, narrative development and/or turn-taking. So far, the intonation phrase has not been linked to the domain of conversational actions. This may explain the lack of interest in such a unit in sociologically oriented CA approaches, where action and action formation take centre stage, and linguistic structuring mechanisms are seen exclusively as resources for social order. However, below a number of sample practices are introduced, which seem to suggest that intonation phrase-like chunks of talk play a role in the formation of some actions. In the following data analyses, these stretches of talk will initially be simply referred to as *chunks*, firstly, because such neutral terminology allows the inclusion of social, as well as linguistic features; and secondly, because chunks of spontaneous talk that are in theory candidates for “intonation phrases” often defy traditional phonological definitions of IPs. The term action component will only be used once the data seem to suggest the interactional relevance of these chunks.

In the remaining part of this section, three practices are singled out in which the use of individual chunks, and participants' orientation to chunks as chunks, is particularly salient and frequent. The actions described below are collaborative turn production, list construction and *well*-prefaced sequence initiation. While aspects of all of them have already been studied extensively in their own right⁹, what will hopefully be shown below is partici-

⁹ Regarding collaborative turn production, see, for example, Falk (1980), Lerner (1991, 1994, 1996, 2002), Ferrara (1992), Ono and Thompson (1995), Szczeppek (2000a, 2000b); Local (2005), Szczeppek Reed (2006), and Krause (2010).



pants' employment of, and orientation to, chunks as individual discourse entities, which do not constitute (potential) stand-alone turns, and as such are not TCUs in their own right.

4.1. Collaborative turn production

Collaboratively constructed turns, in which an incoming participant continues a previously unfinished turn by a previous speaker, are by definition prime locations for potential unit fragmentation. Concerning our question regarding the interactional relevancy of chunks, it is precisely those places at which prior speakers break off, and, even more importantly, the kinds of chunks with which next speakers come in to continue, that display participants' interpretation of an interactionally acceptable division of turns into smaller units. In building a turn collaboratively, participants show which parts of the turn-in-progress they consider to be potentially separate components of the kind that can be split off from prior talk.

In the extract below, an *if X, then Y* construction is begun by a first speaker, and completed by an incoming participant. In this extract from a multi-party dinner table conversation, Peter talks to his daughter-in-law Sally about the profit she has made over time by the increased value of her house.



(2) Wally

01 Pet: if you ↑HAD to sell it nO:w.
 02 ↑yOU've got FIFteen thOUsand pOUNds in ↑↑CASH.
 03 (1.65)
 04 but if you'd've ↑bEEn RENTing,
 05 (0.41)
 06 Sal: <<all> you wOULDn't have HAD [that;>
 07 Pet: [STILL,
 08 Rob: <<p> yeah - >
 09 Pet: you're ↑nOt likely to have save[d up fIfteen [thOUsand
 10 Sal: [NO;
 11 [NO;
 13 Pet: pOU[nds in thrEE or fOUr ↑YEARS;
 14 Sal: [no -
 15 Pet: ↑WOULD you.

Regarding list construction, see, for example, Jefferson (1990), Lerner (1994), Selting (2007).

Regarding sequence initiation, see, for example, Schegloff (1968, 1986, 2007), Clayman (1991), Couper-Kuhlen (2004), Szczepek Reed (2009).





If-clauses are frequently used as opportunities for collaborative turn production, at least in English (Lerner 1991, 1994, 1996; Szczepek Reed 2006).¹⁰ By collaboratively constructing an *if X, then Y* format, participants show that they treat the second part of the format as a component that can be split off from the first, and can even be delivered by another participant. Lerner (1991, 1994) refers to such multi-component TCUs as “compound turn-constructional units”:

On occasion, speakers produce turn-constructional units that project in their course that the current unit is in some way a preliminary component and that a second component will be produced to bring the turn-constructional unit to completion. For example, when a speaker begins a turn-constructional unit with an *if* component, a second *then* component can be foreshown. The turn unit is only properly complete on the completion of the *then* component. (Lerner 1994: 26)

In the extract above, neither of the two turn chunks are TCUs in their own right, given their overall actional design. While both the prosodic and the syntactic structure of the second speaker’s component (line 6) could also occur as a stand-alone turn, its sequential location following the *if*-clause clearly identifies it as a continuation of prior, incomplete talk. This sequential perspective is vital to our analysis: while in theory many utterances could occur as complete turns in the right context, even linguistically incomplete ones (Selting 2001), this specific utterance (*you wouldn’t have had that*) in this specific sequential slot following the *if*-clause *if you’d have been renting*, cannot. In addition to the sequential location, the second speaker’s chunk shows another particularity which clearly identifies it as part of the prior turn. Sally, who is the recipient of Peter’s current turn, nevertheless uses the second person pronoun *you* to refer to herself. In doing so, she shows that her completion is not designed as her own talk, but instead attributed to Peter, designed as a completion of his turn.

The above extract is an example of a single TCU being delivered by two – separately delivered – chunks of talk, or, as Lerner calls them, preliminary and final components. In addition to being a single turn, we can also tentatively argue that the social action accomplished by the two speakers is a single one. The extract shows Peter’s closing of a longer multi-unit turn in which he has praised Sally’s decision to buy, rather than rent a house. The *if*-clauses are designed to close this argument, and Sally’s contribution is arguably not an independent action in its own right, but a collaborative participation in the sequence closing-in-progress.

¹⁰ Krause (2010) finds hardly any instances of collaborative constructions of *if* clauses in her corpus of German conversations.



The same extract also shows a single speaker instantiation of an *if X, then Y* format, *if you had to sell it now you've got fifteen thousand pounds in cash* (lines 1–2). This TCU, too, is divided by the participant into two chunks. The boundary between them is constituted by lengthening of the final accent of the first chunk (*now*, line 1), and a pitch step-up to the first accent of the new chunk (*you've*, line 2). Both of these prosodic features (final lengthening and initial step up) are frequently mentioned boundary cues in the literature on intonation phrases. However, regarding the internal prosodic nature of the two chunks, we find some variation with traditional phonological theory in that the second chunk, *you've got fifteen thousand pounds in cash*, contains two equally strong nucleus accents (*fif*- and *cash*), of which previous definitions of intonation phrases allow only one.

A similar division of a single turn into chunks by two speakers occurs in extract (3) from a US American radio phone-in program. Caller Helen complains about a Senator, and runs into trouble during turn production. Host Barbara Carlson helps out by providing a collaborative completion.

(3) Carlson KSTP

01 Hel: we ↑jUst HEAR:D the other day about ↑WELLstone by the
 02 way bArbra,
 03 <<all> hE never mentioned this to YOU;=
 04 but i don't s'pose he WOULD;>
 05 .hh
 06 that hE is one of the ↑THREE ↑LARGest;
 07 .h
 08 uhm (0.24) b- uh -
 09 .h
 10 intro↑DUcer (0.08) of ↑BILLS,
 11 (0.17)
 12 ui
 13 .h
 14 (0.27)
 15 uh that uh -
 16 (0.15)
 17 <<all> that um - >
 18 (0.44)
 19 Host: [<<f> that SPEND MOneY.>
 20 Hel: [well Anyway.
 21 (0.32)
 22 MOneY spen[ding bills.
 23 Host: [Okay;>

Once again, an incoming participant splits a component of the current turn-in-progress off from its host turn and produces it separately. In this case, the



second component is a projected relative clause (*that spend money*, line 19). Between this component and the beginning of the collaborative turn sequence (lines 1–10) the caller displays difficulties with the continuation of her turn, which most probably prompts the collaborative production of the final component by the recipient. As in the above case, this may not seem particularly spectacular at first sight, because the components that are being split off are clearly defined syntactic entities in both instances. Nevertheless, both extracts highlight an observable participant practice of separating ongoing turns and actions into smaller chunks of talk, which could not stand alone – in this case neither as complete syntactic units, nor as complete actions. The distinction between linguistic completion on the one hand, and action completion on the other is an important one. Incomplete syntactic units frequently accomplish complete conversational actions (Selting 2001). Similarly, “incomplete” prosodic units may be treated by participants as turn-final, depending on the interactional context (Szczepiek Reed 2004). Instead, it is the interactional status of an utterance as complete or not which decides its potential to stand alone as a full turn.

Collaborative turn productions are a good starting point for highlighting the chunking of speech, because the separation of turns into chunks is accomplished across speaker transitions. This gives analysts firm, participant-based evidence for the practice of chunking, without having to draw on other – linguistic, gestural or cognitive – turn-internal mechanisms. The chunks in the above extracts seem to be employed and conceptualized by participants as entities that are sufficiently unit-like for them to be treated as possible candidates for separate, cross-speaker delivery. The way in which they are employed suggests a treatment by participants as components of actions, or *action components*.

4.2. List construction

Turning to single-participant turns, another clear case of TCU division into smaller chunks is list construction. Conversation analytic investigations of lists by Jefferson (1990), Lerner (1991) and Selting (2007) agree that most lists are designed and treated as whole turns; at the same time, both Selting and Jefferson introduce the possibility of lists as sequences of TCUs. For example, Jefferson defines lists “partially (...) as a serial recycling of a given turn constructional unit (word, phrase, sentence, etc.)” (1990: 89). Selting (2007) describes closed lists, which are designed as single TCUs; and open lists, which are frequently delivered as strings of TCUs. Given the wide variety of utterances that can be defined as lists, particularly according to Jef-



ferson (1990), it is clear that some are indeed designed as strings of separate TCUs. However, lists are also frequently delivered as single TCUs, constructed of individual components that are not potential turns in their own right. All of the above mentioned studies use terms such as “component”, “item” and “part” to refer to the chunks that conversational lists are made of. The following extract from the Santa Barbara Corpus is an example of list construction. Jill summarizes, and begins to close a sequence that has centred on her fear of pregnancy with a list of narrative components of a prior story, “suspense, relief and ecstasy”, which she originally introduced at the very beginning of the sequence.

(4) SBC028 Hey Cutie Pie

```
01  Jeff:  so was thAt all the DRama?
02        (0.22)
03  Jill:  tch
04        <<h> thAt was the ↑DRama;>
05        and thAt was the susPENSE,
06        .hh
07        and thAt was the reLIEF,
08        and thAt was the ECStasy.
09        (0.93)
10  Jeff:  <<h> REALLY?>
11  Jill:  uHU?
```

Following the Second Pair Part *that was the drama* (line 4), which is both an answer to Jeff’s question (and thus a TCU and potential turn in its own right), and a preface for the upcoming list, Jill begins a new TCU in which she lists the story elements of her previous narrative (lines 5–8). All of the three list components follow the same structure, involving verbal and syntactic repetition (*and that was the NOUN*), a feature described by Jefferson (1990: 89) as “serial-unit-replication”. Prosodically, the chunks satisfy all criteria for phonological intonation phrases: each is delivered as one overarching intonation contour, each has one initial secondary, and one final primary accent, and each ends in a clearly identifiable unit-final pitch movement. Although each individual list item could in theory stand alone syntactically as a main clause, the chunks are not prosodically delivered as ending in TRPs, but instead with (repeated) list intonation (Selting 2007). However, more important than the linguistic structure of single phrases is the conversational action accomplished by the turn in question. The action, i.e. opening up the sequence closing, is not complete before the end of the final list item; therefore, the turn locations following the first two items are not places for potential speaker change. Once again, it is the sequential slot





at which the chunks occur that makes them components of an ongoing action, rather than separate actions of their own.

The action accomplished here, i.e. the opening up of a sequence closing by way of a listed summary of previously mentioned items, does not lend itself easily to be delivered as a single speech chunk. Although this closing of a pregnancy-related sequence is a complex collaborative activity that continues for several exchanges beyond the transcribed extract, the specific action achieved here – opening up the sequence closing – is designed as an individual action, and thus as an individual TCU. The multi-component structure of this action is employed by Jill as a resource for offering an interpretative summary of her narrative.

Another example of a list, this time of individual syntactic phrases, rather than clauses, occurs in the extract below. In this extract from the Santa Barbara Corpus, Rosemary is having lunch at a restaurant with her two daughters; they are all studying the menu together.

(5) SBC031 Tastes very special

01 Ros: what in the WORLD is a philadElphia-
02 (0.16)
03 She: it's r- ROAST BEEF with;
04 (0.4)
05 mELted CHEESE and;
06 SAUtéed Onions.
07 Ros: m;
08 that [DOES sound GOO:D;
09 She: [on a HOAgY RO:LL.

Sherry's list-like description of the dish is delivered chunk-by-chunk, however without clear-cut syntactic boundaries (in the Prepositional Phrase *with melted cheese* the preposition is grouped with the prior Noun Phrase, i.e. *roast beef with*, line 3). The chunking here is accomplished in part through pausing, in part through prosodic phrasing. All items together constitute the Second Pair Part delivered in response to Rosemary's question (line 1). This time, the action could have been accomplished in a single chunk: one could imagine an uninterrupted TCU *it's roast beef with melted cheese and sautéed onions*. However, the multi-component facility of list construction provides the flexibility to deliver each item as prominently separate.

Although the practice by which the chunks are separated from each other is prosodic in nature, the principle interactional feature of the list components is their use as a resource for the formation of Sherry's response. In the same way in which turns are often delivered in the form of sentences and



global intonation contours, action components may take certain syntactic or prosodic forms. This does not mean, however, that in an analysis of interaction they are best described as prosodic or syntactic units, just as describing turns as sentences would not do them justice analytically.

As in section 4.1 on collaborative turns, the main point concerning conversational lists is that they show participants' practice of delivering actions in the form of individual action components, which are themselves incomplete TCUs, and actions. Components are therefore not potential turns, and thus, at least in the above instances, not TCUs. The very nature of lists as repetitions of individual tokens-of-a-type means that participants display their interpretation of every single list item as separate from others: by doing the same again, they show, first, that the prior item is finished, which constitutes an orientation to an existing boundary; and, second, that the prior item was indeed an item, by reproducing another one of its kind. Or, in the words of Jefferson (1990: 89): "each unit (consists) of or (contains) an item which is adequately representative of and adequately represented by each other unit's item."

So far, syntax and prosody have been shown to be relevant for the delivery (by participants) and the delimitation (by analysts) of speech chunks. This will come as no surprise to those readers familiar with the intonation phrase concept, as these two linguistic aspects are precisely those that are traditionally called upon as the defining features of intonation phrases. The following section attempts to show that it is not primarily syntax and prosody that are accountable for participants' separation of talk into chunks, but the flexible structure of actions, and thus the nature of interaction itself.

4.3. *well*-prefaced sequence initiation

The above extracts have shown that participants design chunks of talk as separate from others if the format of a specific action makes such a structure relevant. The examples have also shown that at times, those chunks can be of a unit-type that is smaller than the TCU. Finally, the above extracts have presented instances in which the structure and accomplishment of the activities in question (collaborative turns, conversational lists) has been going hand in hand with recognizable syntactic patterns (phrases or clauses). The following extracts show a further type of action, *well*-prefaced sequence initiation, that can involve the division of emerging TCUs into smaller chunks. In these cases, however, a clear syntactic definition, while sometimes possible, does not sufficiently describe the core action that is being accomplished. A first example is extract (6), from a recording of a family birthday party (Santa Bar-



158

Beatrice Szczeppek Reed

bara Corpus). Immediately prior to the transcribed extract, the participants have pretended to have forgotten Kendra's birthday. At the beginning of the transcribed section, her brother Kevin initiates a mock-sequence in which he pretends to suddenly remember it (cf. Ehmer 2011). Following this, his wife Wendy performs the sequence initiation in question.

(6) SBC013 Appease the monster

```

01  Kev:  oh that's ↑RIGHT;
02        there ↑IS more to tonight.
03        (0.42)
04  Wen:  there ↑IS;
05        <<h> ↑↑OH::;
06        we forGO::T.>
07        (1.62)
08  Ken:  ((sniff))
09        (0.45)
10  Wen:  well I: have;
11        (0.39)
12        <<p> FUN present.>
13        <<all+p> i'm exCITed about it.>

```

Following the mock-surprise sequence (lines 1–6) and a lengthy silence (lines 7–9), Kevin's wife Wendy initiates a new, non-mock sequence, starting with *well I have* (0.39) *fun present* (lines 10–12). The transcription as two separate lines represents Wendy's delivery of this TCU as two separate chunks of talk. Her first chunk *well I have* is produced with default loudness, and lengthening on the only pitch accent *I*. This is followed by a silence of 0.39 seconds, after which the second chunk *fun present* is delivered with noticeably reduced loudness, and a new overall pitch contour. This prosodic delivery shows that Wendy is designing the two parts of her TCU as separate. And indeed, they can be seen to be accomplishing two separate action components, which together form the overall of sequence initiation.

The first chunk, *well I have*, accomplishes the move away from the mock-surprise sequence. This is achieved primarily through the use of *well*. Sacks, Schegloff, and Jefferson (1974) describe *well* as belonging to a group of resources for doing turn beginnings:

Appositional beginnings, e.g. well, but, and, so etc., are extraordinarily common, and do satisfy the constraints of beginning. But they do that without revealing much about the constructional features of the sentence thus begun, i.e. without requiring that the speaker have a plan in hand as a condition for starting. (...) Appositionals, then, are turn-entry devices or PRE-STARTS (...). (Sacks, Schegloff, and Jefferson 1974: 719, emphasis in the original)





This description of *well* as a “turn-entry device” refers to an aspect of turn organization that is not linked to a specific next action. According to Sacks, Schegloff, and Jefferson, all that is being accomplished by *well* and other “pre-starts” is the beginning itself. Although the above quote is taken from a section on the “first starter goes” principle, and thus the comments on *well* were originally made with the potential for turn competition in mind, extract (6) also shows a primarily organizational use of *well*. The chunk *well I have* does little more than signal a move away from prior talk, and towards a new participant framework (*I*), without indicating which specific next action its speaker is moving towards.

In terms of the precise interactional nature of this sequential beginning, Schifffrin’s analysis of *well* as an indicator of potential non-coherence is helpful here. According to Schifffrin, “*well* anchors the speaker into a conversation precisely at those points where upcoming coherence is not guaranteed” (1987: 126). In spite of Schifffrin’s overarching analysis of *well* as a response marker, which does not fit the extract above, the more general use of *well* as an indicator of something new fits both our example, and also Sacks, Schegloff, and Jefferson’s (1974) interpretation. The action initiated by Wendy is not coherent with the sarcastic stance of previous talk. The *well*-prefaced chunk therefore accomplishes the move away from this talk, and signals the initiation of a new beginning.

It is important to note, however, that Wendy does not produce a stand-alone *well*, which may function as a TCU in its own right, but instead uses the interactionally incomplete chunk *well I have*. In doing so she secures the immediately following turn space – she does not end her chunk with a TRP – without yet committing to her main topic proffer, i.e. birthday presents. Thus, at the end of the first chunk, the turn is still missing a component that specifies what type of sequence/ action is being initiated. This second component is delivered as a separate entity, and as a new chunk (*fun present*). By introducing the word *present*, this second chunk completes the move to a new sequence involving the exchange of presents, and therefore to a whole new interactional activity. Thus, while the first component achieves the moving away from the previous action and maintains a claim to the floor, the second one accomplishes moving toward something new. Together, the chunks make up the full turn and accomplish sequence initiation.

It is noticeable that in the above extract the first chunk *well I have* is not a syntactically complete unit of the type we have encountered before: it is neither a clause, nor a full (verb) phrase. This contributed to our analytical noticing of this chunk as accomplishing primarily interactional work. Extract (6) draws our attention to the interactional role of *chunks*, which seems to be



related to the packaging of action components, and thus to the formation of actions. Unsurprisingly, chunks that are defined exclusively in the domain of prosody and action, without any clear syntactic boundaries are rare, given the very nature of syntax as a resource for structuring talk, and actions.

Another example of a sequence initiating action component, once again prefaced by *well*, can be seen in extract (7) from the Santa Barbara Corpus. Bernard and his friends are having dinner, and have been discussing the challenges of cooking in an apartment on the 12th floor. In initiating a new sequence (line 8), Sean returns to a topic discussed earlier in the conversation, which concerns living in places other than New York.

(7) SBC051 New Yorkers Anonymous

01 Ber: the ALtitude;
02 you have to be CAREful right,
03 Fra: [RIGHT -
04 Ber: [espEcially if you're BAKing.
05 hehe
06 Ali: <<h+p> e[hehe
07 Fra: [ehehehe
08 Sea: **well fran <<h> I: 'VE been looking: - >**
09 **(0.29)**
10 **REALly for a year or TWO at different plAcEs to ↓LIVE**
11 **i went bAck to EUrope -**
12 Fra: mHM -

Sean's sequence initiation is spread across two chunks of talk: *well Fran I've been looking* (line 8), which accomplishes the move away from the previous sequence, and draws attention to the beginning of something new; and *really for a year or two at different places to live I went back to Europe* (lines 10–11), which is delivered as one uninterrupted stretch of talk. Once again, the sequence initiation is accomplished in two stages: the first chunk, prefaced by *well*, introduces a move away from the immediately prior sequence, and a return to an earlier conversational topic of "looking for the best place to live". It does so by selecting one recipient in particular, and introducing the person (*I*) and part of the activity (*looking*) of an upcoming telling. All of these aspects are a return to a prior sequence and topic, which shows that *well* in this case is not used as a "pre-start" (Sacks, Schegloff, and Jefferson 1974), but instead as a *pre-restart*, i.e. re-starting a previous sequence.

However, the TCU is not yet complete at this stage, even though the speaker designs the chunk as separate: after the initial syllables it is delivered with a noticeably higher pitch register than subsequent talk, and it is separated from what follows by a pause (line 9). In the second part the remaining



action component is performed, i.e. the specification of the reported activity. Interestingly, this component defies all previous descriptions of intonation phrases, with its four primary pitch accents (*real-*, *two*, *live*, *Eu-*), inclusion of various syntactic phrases and their extensions, and a sentence boundary (*places to live* | *I went back*). Once again we find that what is relevant for an analysis of interaction are the units participants themselves orient to as separate. In this case, the actional phrasing does not coincide with predictable syntactic and prosodic phrasing; in many other cases, it does (see above). The most significant practice, however, is conversationalists' accomplishment of actions as separate chunks of talk. The linguistic form of those chunks seems to be as flexible as interaction itself.

The two extracts presented in this section show *well*-prefaced sequence initiating chunks, which accomplish the interactional work of moving away from prior talk; followed by chunks which continue the sequence initiating action, and through which the turn moves towards projected upcoming talk. What is interesting in both instances is that these sequence initiations are not prefaced by "preliminaries", such as "*Can I ask you a question*" (Schegloff 1980), but instead by incomplete turn components. For example, neither of the sequences above are initiated by a full TCU announcing the beginning of a new sequence, and asking co-participants' permission for such a move. One possible interpretation of the multi-component delivery of the *well*-prefaced sequence initiating TCUs above is that they replicate the typical preliminary structure:

preliminary (i.e. announcement of new sequence + invitation of agreement/
objection)
pause/potential for co-participant response
sequence initiation

However, the extracts presented here skip the opportunity for speaker change following the preliminary turn. While the *well*-prefaced sequence initiations also provide an initial projecting component, followed by a pause (*well i have* + pause; *well fran i've been looking* + pause), talk designed in this way does not invite co-participant uptake. The initial chunk, while announcing the beginning of a new sequence, is not a potential First Pair Part as in the case of a preliminary; instead, it is a syntactically, prosodically and pragmatically incomplete component (Ford and Thompson 1996), which does clearly not end in a TRP. These features make the multi-component sequence initiation a strategy that gives currently speaking participants greater control over the immediately upcoming sequential development, while allowing co-participants less influence.





5. Concluding observations

This chapter has tried to explore whether participants orient to intonation phrase-sized chunks of talk as relevant for the accomplishment of conversational actions. The sample extracts above show that for certain actions, participants seem to employ chunks of talk that could not stand alone as turns, but that are instead used as building blocks for turns and actions. These chunks take on a wide variety of forms, and some instantiations of them are difficult to define according to their semantic, syntactic or prosodic characteristics (Szczeppek Reed 2010b). For example, neither the postulation that they always contain only one primary accent, or nucleus, holds in naturally occurring data; nor can they always be defined according to syntactic phrasing rules. However, their boundaries are established by the participants themselves: by producing turns collaboratively, and by prosodically packaging chunks as separate entities, participants show that they are orienting to them as separate turn components.

An interactional perspective on language tries to unearth what is relevant for language use(rs) in an emergent conversational context. Descriptions of formal linguistic properties are not significant in their own right, but instead must be shown to be made significant by participants themselves. Regarding the construction of units of talk, it is not enough that traditional linguistics has been able to show the possibility of certain formal units, such as the intonation phrase. The question is whether and how such a unit is used in natural discourse. With regard to the chunks of talk discussed in this chapter, there are two reasons why their purely linguistic description is not appropriate for an interactional analysis.

Firstly, it does not satisfactorily describe their role for the accomplishment of social actions, just as the description as “sentence” would not do justice to the interactional role and relevance of a turn. Many instances of chunks may well be prosodic phrases, as well as traditionally definable syntactic constructions. However, in the study of language-in-interaction syntax and prosody are conceptualized as domains of formal practices, rather than social actions in themselves (Schegloff 1997). Therefore, the above-described chunks of talk may be intonation phrases, phonologically; and clauses or phrases, syntactically. Interactionally, they are components of actions.

A second, comparatively minor reason why a traditional phonological description of chunks as intonation phrases is not (always) appropriate is that in natural speech data many instances of chunks which could in theory be candidates for intonation phrases do not actually meet the traditionally prescribed phonological and syntactic criteria. Admittedly, this problem could



be solved by adapting the linguistic description of intonation phrases to findings from natural talk, and by adopting a maximally flexible definition, both in terms of the prosodic and the syntactic characteristics. However, the main argument of this chapter concerns the interactional role of chunks as action components: by limiting the reference to this interactional phenomenon to the domain of phonology (*intonation* phrases), the phenomenon itself is not adequately described as a unit of social action.

As the examples above have shown, participants seem to employ chunks of talk, smaller than the TCU, for the accomplishment of specific conversational actions. The above instances of collaborative turn productions, conversational lists and *well*-prefaced sequence initiations were used to show sample instances of actions which are accomplished – at least in the above cases – by two or more action components, which could not stand alone as individual TCUs. In the light of the data presented here, according to which chunks may indeed be relevant as action components, the term *intonation phrase* (and its variants, such as “intonation unit” or “tone group”) does not seem adequate. In previous work on this topic I have therefore suggested the term *Turn Constructional Phrase (TCP)*, to signify the nature of chunks as building blocks of turns at a level below the TCU (Szczepek Reed 2010b, 2010c). However, it seems to me that this may not have been such a wise choice. By referring to chunks by a term so closely related to the term *TCU*, a relation between the two units is being implied that future research on interaction may not corroborate: chunks may at times be building blocks for TCUs, at other times TCUs may come in the form of single chunks. Furthermore, the term *phrase* is closely linked to syntactic and phonological concepts. Therefore, if one were to decide that chunks are indeed relevant for the formation of actions, as the data presented here seem to suggest, a term that reflects their relevance for interaction would be preferable over one that reflects mainly their relevance for language. Besides the appealingly descriptive term *chunk*, the alternative term *action component* is suggested here which takes action formation as its primary reference point.

However, it is vital to mention at the end of this chapter that considerably more work needs to be done to show that (some) actions are indeed accomplished through orientation to a component structure of individual building blocks. A point of departure for future research may be a collection of observable actions accomplished through action components. After all, it is only if we can find evidence for the relevance of units in interaction that we can justify their use as categories in our analyses and transcripts.





Acknowledgments

I would like to thank Dagmar Barth-Weingarten and the editors of this volume for extensive comments on earlier drafts of this chapter.

Appendix

Transcription Conventions (adapted from Selting et al. 1998)

Pauses and lengthening

(.)	micro-pause
(2.85)	measured pause
:::	lengthening

Accents

ACcent	primary pitch accent
Accent	secondary pitch accent

Component-final pitch movements

?	rise-to-high
,	rise-to-mid
—	level
;	fall-to-mid
.	fall-to-low

Pitch step-up/step down

↑	pitch step-up
↓	pitch step-down

Change of pitch register

<<l> >	low pitch register
<<h> >	high pitch register

Volume and tempo changes

<<f> >	forte
<<ff> >	fortissimo
<<p> >	piano
<<pp> >	pianissimo
<<all> >	allegro
<<len> >	lento



Prosody, syntax and action formation

165

Breathing

.h, .hh, .hhh in-breath

h, hh, hhh out-breath

Other conventions

[overlapping talk

[

References

- Armstrong, L. E. and I. C. Ward 1926 *Handbook of English Intonation*. Leipzig and Berlin: Teubner. Second edition (1931). Cambridge: Heffer.
- Auer, P. 2010 Zum Segmentierungsproblem in der gesprochenen Sprache. *InLiSt* 49.
- Barth-Weingarten, D. 2007 Intonation units and actions – evidence from everyday interaction. Paper presented at the *Conference of the International Pragmatics Association*, Göteborg, 8–13 July, 2007.
- Barth-Weingarten, D. forthcoming From “intonation units” to cesuring – an alternative approach to the prosodic-phonetic structuring of talk-in-interaction. In: B. Szczepek Reed and G. Raymond (eds.), *Units of Talk – Units of Action*. Amsterdam: John Benjamins.
- Bergmann, P. 2010 Convergence or divergence of syntactic, prosodic, and visual unit boundaries – The case of parentheses. Paper presented at the *International Conference on Conversation Analysis*, Mannheim, 4–8 July 2010.
- Brazil, D. 1997 *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.
- Brown, G., K. L. Currie and J. Kenworthy 1980 *Questions of Intonation*. London: Croom Helm.
- Chafe, W. L. 1980 The deployment of consciousness in the production of a narrative. In: W. L. Chafe (ed.), *The Pear Stories. Cognitive, Cultural and Linguistic Aspects of Narrative Production*, 9–50. Norwood, New Jersey: Ablex.
- Chafe, W. L. 1987 Cognitive constraints on information flow. In: R. S. Tomlin (ed.), *Coherence and Grounding in Discourse*, 21–55. Amsterdam: John Benjamins.
- Chafe, W. L. 1988 Linking intonation units in spoken English. In: J. Haiman and S. Thompson (eds.), *Clause Combining in Grammar and Discourse*, 1–27. Amsterdam: John Benjamins.
- Chafe, W. L. 1993 Prosodic and functional units of language. In: J. A. Edwards and M. D. Lampert (eds.), *Talking Data. Transcription and Coding in Discourse Research*, 33–43. Hillsdale: Lawrence Erlbaum.
- Clayman, S. E. 1991 News interview openings: aspects of sequential organization. In: P. Scannell (ed.), *Broadcast Talk: A Reader*, 48–75. Beverly Hills: Sage.
- Couper-Kuhlen, E. 1986 *An Introduction to English Prosody*. London: Edward Arnold.
- Couper-Kuhlen, E. 2004 Prosody and sequence organization in English conversation: The case of new beginnings. In: E. Couper-Kuhlen and C. E. Ford (eds.), *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*, 335–376. Amsterdam: John Benjamins.





- Couper-Kuhlen, E. and T. Ono 2007 "Incrementing" in conversation. a comparison of practices in English, German and Japanese. *Pragmatics* 17: 513–552.
- Cruttenden, A. 1997 *Intonation*. Cambridge: Cambridge University Press.
- Crystal, D. 1969 *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press
- Du Bois, J. W. 1991 Transcription design principles for spoken discourse research. *Pragmatics* 1: 71–106.
- Du Bois, J. W., S. Schuetze-Coburn, S. Cumming and D. Paolino 1993 Outline of discourse transcription. In: J. A. Edwards and M. D. Lampert (eds.), *Talking Data. Transcription and Coding in Discourse Research*, 45–89. Hillsdale: Lawrence Erlbaum.
- Du Bois, J. W., W. L. Chafe, C. Meyer and S. A. Thompson 2000 *Santa Barbara Corpus of Spoken American English, Part 1*. Linguistic Data Consortium, Philadelphia.
- Du Bois, J. W., W. L. Chafe, C. Meyer, S. A. Thompson, and N. Martey 2003 *Santa Barbara Corpus of Spoken American English, Part 2*. Linguistic Data Consortium, Philadelphia.
- Du Bois, J. W. and R. Englebretson 2004 *Santa Barbara Corpus of Spoken American English, Part 3*. Linguistic Data Consortium, Philadelphia.
- Du Bois, J. W. and R. Englebretson 2005 *Santa Barbara Corpus of Spoken American English, Part 4*. Linguistic Data Consortium, Philadelphia.
- Ehmer, O. 2011 *Imagination und Animation: Die Herstellung mentaler Räume durch animierte Rede*. Berlin: de Gruyter.
- Falk, J. 1980 *The Duet as a Conversational Process*. Ann Arbor, Michigan: University Microfilms International.
- Ferrara, K. 1992 The interactive achievement of a sentence: Joint productions in therapeutic discourse. *Discourse Processes* 15: 207–28.
- Ford, C. E. 2004 Contingency and units in interaction. *Discourse Studies* 6: 27–52.
- Ford, C. E. and S. A. Thompson 1996 Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 134–184. Cambridge: Cambridge University Press.
- Ford, C., B. A. Fox and S. A. Thompson 1996 Practices in the construction of turns: The "TCU" revisited. *Pragmatics* 6: 427–454.
- Ford, C., B. A. Fox and S. A. Thompson 2002 Constituency and the grammar of turn increments. In: C. Ford, B. A. Fox and S. A. Thompson (eds.), *The Language of Turn and Sequence*, 14–38. Oxford: Oxford University Press.
- Fox, B. A. 2001 An exploration of prosody and turn projection in English Conversation. In: M. Selting and E. Couper-Kuhlen (eds.), *Studies in Interactional Linguistics*, 287–315. Amsterdam: John Benjamins.
- Gibbon, D. 1976 *Perspectives of Intonation Analysis*. Frankfurt am Main: Peter Lang.
- Grice, M. 2006 Intonation. In: Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 778–788. Second Edition, Vol 5. Oxford: Elsevier.
- Gumperz, J. 1993 Transcribing conversational exchanges. In: J. A. Edwards and M. D. Lampert (eds.), *Talking Data. Transcription and Coding in Discourse Research*, 91–121. Hillsdale: Lawrence Erlbaum.
- Gussenhoven, C. 1984 *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.





- Halliday, M. A. K. 1963 The tones of English. *Archivum Linguisticum* 15, 1–28. Reprinted in: W.E. Jones and J. Laver (eds.) (1973), *Phonetics in Linguistics. A book of readings*, 103–126. London: Longman.
- Halliday, M. A. K. 1967 *Intonation and Grammar in British English*. The Hague: Mouton.
- Halliday, M. A. K. 1970 *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.
- Heusinger, K. von 1999 *Intonation and Information Structure*. University of Konstanz: Habilitationsschrift. <http://elib.uni-stuttgart.de/opus/volltexte/2003/1396/pdf/heusinger.pdf>
- Hopper, R. 1992 *Telephone Conversation*. Bloomington: Indiana University Press
- Jefferson, G. 1990 List construction as a task and resource. In: G. Psathas (ed.), *Interaction Competence*, 63–92. Washington D.C.: University Press of America.
- Kim, K. 1999 Phrasal unit boundaries and organization of turns and sequences in Korean conversation. *Human Studies* 22: 425–446.
- Kingdon, R. 1958 *The Groundwork of English Intonation*. London: Longman.
- Klinghardt, H. 1923 *Sprechmelodie und Sprechtakt*. Marburg: Elwert.
- Klinghardt, H. and G. Klemm 1920 *Übungen im englischen Tonfall für Lehrer und Studierende*. Cöthen: Otto Schulze Verlag.
- Krause, S. 2010 Displays of understanding in collaborative turn sequences. Paper presented at the *International Conference on Conversation Analysis*, Mannheim, 4–8 July 2010.
- Ladd, D. R. 1996 *Intonational phonology*. Cambridge: Cambridge University Press.
- Lerner, G. H. 1991 On the syntax of sentences in progress. *Language in Society* 20: 441–458.
- Lerner, G. H. 1994 Responsive list construction. *Journal of Language and Social Psychology* 13: 20–33.
- Lerner, G. H. 1996 On the “semi-permeable” character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 238–276. Cambridge: Cambridge University Press.
- Lerner, G. H. 2002 Turn-sharing: The choral co-production of talk-in-interaction. In: C. E. Ford, B. A. Fox and S. A. Thompson (eds.), *The Language of Turn and Sequence*, 225–56. Oxford: Oxford University Press.
- Lieberman, P. 1967 *Intonation, Perception and Language*. Cambridge, Mass: MIT Press.
- Local, J. 2005 On the interactional and phonetic design of collaborative Completions. In: W. J. Hardcastle and J. M. Beck (eds.), *A Figure of Speech: a Festschrift for John Laver*, 263–82. New Jersey: Lawrence Erlbaum.
- Local, J., B. Wells and M. Sebba 1985 Phonology for conversation. Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9: 309–330.
- Local, J., J. Kelly and B. Wells 1986 Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics* 22: 411–437.
- Local, J. and G. Walker 2004 Abrupt-joins as a resource for the production of multi-unit, multi-action turns. *Journal of Pragmatics* 36: 1375–1403.
- MacWhinney, B. 2007 The TalkBank Project. In: J. C. Beal, K. P. Corrigan and H. L. Moisl (eds.), *Creating and Digitizing Language Corpora: Synchronic Databases, Vol.1*, Houndmills: Palgrave Macmillan.
- Mazeland, H. 2007 Parenthetical sequences. *Journal of Pragmatics* 39: 1816–1869.





- O'Connor, J. D. and G. F. Arnold 1961/1973 *Intonation of Colloquial English*. London: Longman.
- Ono, T. and S. A. Thompson 1995 What can conversation tell us about syntax? In: P. W. Davies (ed.), *Descriptive and Theoretical Modes in the Alternative Linguistics*, 213–71. Amsterdam: John Benjamins.
- Palmer, H. E. 1922 *English Intonation, with Systematic Exercises*. Cambridge: Heffer.
- Peters, J. 2006 Syntactic and prosodic parenthesis. *Conference Proceedings Speech Prosody 2006*, Dresden, Germany, May 2–5, 2006. http://www.isca-speech.org/archive/sp2006/papers/sp06_245.pdf (last accessed 19/04/2011).
- Pike, K. L. 1945 *Intonation of American English*. Ann Arbor: University of Michigan Press.
- Raymond, G. 2010 Positioning Responses: On the relevance of “slots” for the organization of responses to polar interrogatives in English. Paper presented at the *International Conference on Conversation Analysis*, Mannheim, 4–8 July 2010.
- 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. 1968 Sequencing in conversational openings. *American Anthropologist* 70: 1075–95.
- Schegloff, E. A. 1980 Preliminaries to preliminaries: “Can I ask you a question?”. *Sociological Inquiry* 50: 104–152.
- Schegloff, E. A. 1982 Discourse as an interactional achievement: Some uses of “uh huh” and other things that come between sentences. In: D. Tannen (ed.), *Analyzing Discourse: Text and Talk*, 71–93. Washington, D.C.: Georgetown University Press.
- Schegloff, E. A. 1986 The routine as achievement. *Human Studies* 9: 111–52.
- Schegloff, E. A. 1987 Analyzing single episodes of interaction: An exercise in Conversation Analysis. *Social Psychology Quarterly* 50: 101–114.
- Schegloff, E. A. 1996 Turn organization: One intersection of grammar and interaction. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 52–133. Cambridge: Cambridge University Press.
- Schegloff, E. A. 1997 Practices and actions: Boundary cases of other-initiated repair. *Discourse Processes* 23: 499–545.
- Schegloff, E. A. 1998 Reflections on studying prosody in talk-in-interaction. *Language and Speech* 41: 235–263.
- Schegloff, E. A. 2007 *Sequence Organization in Interaction. A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Schiffrin, D. 1987 *Discourse Markers*. Cambridge: Cambridge University Press.
- Schubiger, M. 1935 *The Role of Intonation in Spoken English*. Cambridge: Heffer.
- Schubiger, M. 1958 *English Intonation: Its Form and Function*. Tübingen: Niemeyer.
- Selting, M. 1995 *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.
- Selting, M. 1996 On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6: 371–388.
- Selting, M. 2000 The construction of units in conversational talk. *Language in Society* 29: 477–517.
- Selting, M. 2001 Fragments of units as deviant cases of unit production in conversational talk. In: M. Selting and E. Couper-Kuhlen (eds.), *Studies in interactional linguistics*, 229–258. Amsterdam: John Benjamins.





- Selting, M. 2007 Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics* 39: 483–526.
- Selting, M. and E. Couper-Kuhlen (eds.) 2001 *Studies in Interactional Linguistics*. Amsterdam: John Benjamins.
- Selting, M., P. Auer, B. Barden, J. R. Bergmann, E. Couper-Kuhlen, S. Günthner, C. Meier, U. Quasthoff, P. Schlobinski and S. Uhmann 1998 Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173: 91–122.
- Szczepek, B. B. 2000a Formal aspects of collaborative productions in English conversation.
- Szczepek, B. B. 2000b Functional aspects of collaborative productions in English conversation. *InLiSt* 21.
- Szczepek, B. B. 2004 Turn-final intonation in English. In: E. Couper-Kuhlen and C. E. Ford (eds.), *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*, 97–118. Amsterdam: John Benjamins.
- Szczepek, B. B. 2006 *Prosodic Orientation in English Conversation*. Basingstoke: Palgrave Macmillan.
- Szczepek, B. B. 2009 Prosodic orientation: A practice for sequence organization in broadcast telephone openings. *Journal of Pragmatics* 41: 1223–1247.
- Szczepek, B. B. 2010a *Analysing Conversation: An Introduction to Prosody*. Basingstoke: Palgrave.
- Szczepek, B. B. 2010b Intonation phrases in natural conversation: A participants' category? In: D. Barth-Weingarten, E. Reber, and M. Selting (eds.), *Prosody in Interaction*, 191–212. Amsterdam: John Benjamins.
- Szczepek, B. B. 2010c Units of interaction: Tone units or Turn Constructional Phrases? In: E. Delais-Roussarie (ed.), *Conference Proceedings: Interface Discourse and Prosody*, 351–363. University of Chicago. Paris, 9–11 September 2009.
- Tanaka, H. 1999 *Turn-taking in Japanese Conversation: A Study in Grammar and Interaction*. Amsterdam: John Benjamins.
- Tanaka, H. 2000 Turn projection in Japanese talk-in-interaction. *Research on Language and Social Interaction* 33: 1–38.
- Thompson, S. A. and A. Mulac 1991 A Quantitative Perspective on the Grammaticization of Epistemic Parentheticals in English. In: E. Traugott and B. Heine (eds.), *Grammaticalization II*, 313–339. Amsterdam: John Benjamins.
- Walker, G. 2004 On some interactional and phonetic properties of increments to turns in talk-in-interaction. In: E. Couper-Kuhlen and C. E. Ford (eds.), *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*, 147–169. Amsterdam: John Benjamins.
- Wells, B. and S. Peppé 1996 Ending up in Ulster: Prosody and turn-taking in English Dialects. In: E. Couper-Kuhlen and M. Selting (eds.), *Prosody in Conversation*, 101–130. Cambridge: Cambridge University Press.
- Wells, B. and S. Macfarlane 1998 Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech* 41: 265–294.
- Wells, J. C. 2006 *English Intonation. An Introduction*. Cambridge: Cambridge University Press.
- Young, R. F. and J. Lee 2004 Identifying units in interaction: Reactive tokens in Korean and English conversations. *Journal of Sociolinguistics* 8: 380–407.







II Embodiment







Lorenza Mondada
Freiburg Institute for Advanced Studies
ICAR Research Lab (CNRS, University of Lyon)

Deixis: an integrated interactional multimodal analysis

1. Introduction

This chapter focuses on the highly coordinated adjustments occurring between speaker and recipient in the production of deictic descriptions. By carefully transcribing and analyzing video-recorded instances in which a speaker points to a co-present object in order to show and explain it to a recipient, the chapter deals with the finely tuned formatting of turns at talk, as they progressively, emergently, incrementally unfold in time. This temporal dimension depends in a crucial manner on the sequential organization of action and interaction. Moreover, in co-present interactions, this temporality does not only concern talk but also other multimodal resources – such as gesture, gaze, facial expressions, body postures, etc. The chapter aims at showing that the temporally finely tuned mobilization of these resources in interaction is a systematic achievement of the participants. This systematic and methodic order is achieved for talk as well as for embodied conducts, within complex multimodal Gestalts. This invites to a grammar constituted by interactional descriptions taking into consideration complex layers of time.

The chapter first sketches some issues related to the emergent construction of talk-and-other conducts in interaction; then it proposes a systematic analysis of a collection of data, based on the organization of the deictic *ici* (‘here’ in French) in turn-beginning position occurring with a concurrent or delayed pointing gesture. The chapter shows that depending on the temporal distribution of verbal and embodied resources within the formatting of the turn and of the action, *ici* is organized in two patterns, in which it achieves different interactive tasks – as a device for introducing and referring to an object and as a device for getting the attention of the recipient.

1.1. An integrated approach of grammar and multimodality in interaction

By focusing on an apparently simple instance of deictic reference – a deictic word and a deictic gesture –, the chapter aims at showing the complexity of such Gestalts (a term frequently used by Auer, 2009 or Selting, 1996 for designating situated linguistic constructions), their multi-layered order, their



embeddedness in sequential organization, the flexibility of their temporal arrangement and the multiple issues the participants are orienting to. It suggests that the form of a grammatical-interactional description should integrate these multiple and dynamic aspects, in order to develop a grammatical model based on the actual usages of speakers in context.

Multimodality – considered as the integrated study of all the relevant linguistic, embodied, material resources participants mobilize for organizing social interaction – is a focus which deepens and enlarges the tasks of interactional linguistics and conversation analysis, and which invites to complexify their conception of temporality and sequentiality.

Interactional linguistics inspired by conversation analysis has shown the necessity to “retemporalize” language (Auer, Couper-Kuhlen, and Müller 1999) and to consider language as being used, changed, structured by the fact that linguistic structures in talk are incremental (Auer 2009) and emergent (Hopper 1988). Turn format is interactively built moment by moment in an interactive way (Goodwin 1979), the speaker orienting to and taking into consideration the conduct of the coparticipants, reflexively adjusting the emergent turn to what they are (or are not) doing. Thus, turns are formatted as an orderly series of sequential positions, from pre-beginnings, to turn-initial positions, and so on until pre-completion position and completion (Schegloff 1996). These positions are interactively and sequentially achieved, in a flexible and contingent way (Ford 2004), constructing units of talk which are a dynamic, negotiated, collective achievement.

Within the linguistic description of the features of turn-constructual units, the interplay of multiple dimensions have been taken into consideration, from syntax to prosody (Ochs, Schegloff, and Thompson 1996; Hakulinen and Selting, 2005; Couper-Kuhlen and Selting, 1996). The integration of multimodality enlarges this multiplicity, by considering, with talk, gesture, gaze, facial expressions, body postures, and movements too. Each of these dimensions unfolds in time too, concurrently with talk, constituting the sequential embodied organization of action. These temporalities are strongly coordinated – for example, gesture co-occurring or even preceding their lexical affiliates (Kendon 1980; McNeill 1979; Schegloff 1984). Their synchronization is the result of interactive work, by which talk and gesture, or talk and posture, are organized in a way that aligns them temporally, for example either by delaying talk to adjust to gesture or the reverse (Condon 1971; Kendon 2004: 135). Co-occurrence of gesture and speech has been largely demonstrated within gesture studies – namely within approaches considering that talk and gesture originate in the same conceptual process (McNeill 1985, 1992), a vision of talk and gesture as “composite signals”



(Clark 1996: 156), as constituting an “integrated message model”, showing that gesture and facial displays are used simultaneously with words, mobilized together to produce “visible acts of meaning” (Bavelas and Chovil 2000).

Despite these researches however, the way in which multimodal resources as a whole are mobilized within multiple temporal and sequential relationships in natural interaction remains to be systematically studied, not only focusing on the speaker but also on the addressees, taking into consideration the entire participation framework (Goodwin 1981, 2000). This chapter makes the consequences of this approach explicit for the conception of a grammar based on complex multimodal configurations in action.

1.2. The complexity of pointing

Pointing is a primordial site for the organization of human action and has been studied in a variety of disciplines (Kita 2003a). Deictic expressions have been largely described as depending on context for the interpretation of their reference (Fillmore 1975; Lyons 1991) and as being “pointers” to that context. But the way in which this context is defined and the reference precisely produced by speakers and recognized by recipients remains obscure within purely linguistic and monological models. Bühler’s (1934) notion of “origo”, which serves to define the coordinate system of subjective orientation in which personal, temporal, and spatial deictic reference is made, is defined in relation to the “I”, the speaker, which – even if Bühler himself admits that shifts to other than the speaker are possible (such as switches to the addressee or switches in an imaginary domain, for the deixis *am phantasma*) – have favoured a speaker-centred interpretation of deictic reference (see Fricke 2007 for discussions and developments). Bühler recognized that the boundaries of the origo which defines the “here” space are variable: *here* can refer to the country, the town, the building, the office where I am, as well as more microscopic portions of space depending on the kind of reference I am doing. Within classical models, even if *here* seems to be the prototypical case of deixis, and as such the most transparent and primitive way to refer, the question of how it refers precisely to a point remains unanswered.

Even within linguistics, the co-occurrence between the word *here* and the pointing gesture has been recognized as part of the prototypical way in which *here* works. Lyons (1991: 150) writes: “Identification by pointing, if I may use the term ›pointing‹ in a very general sense, is deixis at its purest”. Gesture studies have largely investigated this relation, insisting on the synchronicity of the word and the pointing gesture and proposing detailed de-



scriptions of the types (e.g. Fricke 2007) and the shapes (Kendon 2004) of deictic gestures. Other studies, as Kita's (2003b) have expanded the focus from the hand gesture to the entire body, showing that the speaker's pointing gesture, body (torso) and gaze do orient together while pointing and using a deictic word, within an integrated vision of the speaker's conduct. Nevertheless, Kita's analysis does not take into account the adjustment to the addressee: even if he admits that reciprocity and interaction may play a role, he sees the alignment between gesture, gaze, body and the surrounding space as a demonstration that gesture facilitate conceptual planning for speaking.

An interactionist view of deixis (Hanks 1990; Haviland 1993) integrates the recipient-designed dimension of deixis, as well as the fact that the context in which deixis is achieved is not determining it from outside, but is reflexively constituted in the very act of referring to it. In this socio-centred perspective, which takes into account both the speaker and the other co-participants, the linguistic and the embodied resources, deixis becomes a highly complex communicative phenomenon, where various dimensions are simultaneously brought together. As Goodwin (2003) shows, pointing and talking are formatted together by taking into consideration the surrounding space, the activity in which participants are engaged, and their mutual orientation. Using the example of archeologists excavating dirt, Goodwin shows how participants actively constitute a visual field which has to be scrutinized, parsed and understood convergently by the co-participants in order to find out where the speaker is pointing. The archeologists juxtapose language, gesture, tools – like trowels –, graphic fields – like maps – on a domain of scrutiny, which is surrounding them, but is also being delimited by the very act of referring to it. In this sense, gesture is environmentally coupled (Goodwin 2007) and not used as a separated resource coming from the exterior world into a pre-existing context: the domain of scrutiny is transformed and re-organized by the very action of pointing, done within the current task.

As Hindmarsh and Heath (2000) show, pointing gesture and body movements amplifying them are realized in a way that is recipient-designed, i.e. indicate and even display the referent for the co-participants, at the relevant moment, when the referent is visible for them. Pointing gestures are “produced and timed with respect of the activities of the co-participants, such that they are in a position to be able to see the pointing gesture in the course of its production” (2000: 1868). Thus the organization of the gesture and the body of the speaker are adjusted to the recipient in order to guide him in the material surround and towards the referent. Since recipient display their understanding and grasping of the action going on, speakers adjust to the production of these expressions, to their absence or delay (Mondada in press).





This mutual orientation involves not only talk and gesture but the entire body, gazing and bending on the object (Hindmarsh and Heath 2000) and, more radically, actively rearranging the surrounding environment. Mondada (2005, 2007) shows how speakers, prior to the utterance of the deictic, dispose their bodies within space, reposition objects within space, and even restructure the environment. The deictic and the pointing gesture are produced only after participants have relevantly organized the disposition of the spatial context. Thus, deictic words and gesture are not merely adapting to a pre-existing and immutable context; they are part of an action which actively renews and changes context, rearranging the interactional space in the most adequate way for the pointing to take place. In another study, Mondada (2009a) focuses on the way in which bodies are organized during the emergent opening of a focused encounter in which a passerby asks a direction to an unknown person in the street: the configuration of the interactional space is the prerequisite for the itinerary description to be launched and for pointing gesture to occur. These studies show that there is not a set of expressive, linguistic and gestural, resources on the one hand and a context on the other hand, but that the very mobilization of multimodal resources within a sequential trajectory of action reshapes the context.

Complex arrays of juxtaposed multimodal resources, recipient-design and active (re)construction of the interactional space are the main aspects on which the activity of showing, demonstrating and pointing will be described below – and which a grammar based on the situated participants' usages in interaction has to take into account.

2. The phenomenon and the data

The chapter focuses on the coordination of talk, gesture and gaze between two participants in a naturalistic setting in which a new referent is introduced by the deictic *ici* (here) and then described and explained. The data set is composed by videorecordings of a recurrent activity – explaining the technical feature of a car to a customer who just bought it – in which new items being topicalized are routinely introduced by a pointing gesture and the deictic *ici*. 7 videorecordings of the same car model explained to 7 different customers (for a total of 4 hours) coming to the garage to pick up their car have been made, in which the same dealer, who I call Jan, explains the specificities of the car before its release. Data have been recorded within extended ethnographic fieldwork in a garage in a big town in France. The activity being documented constitutes a routine practice for the dealer and has not been altered by the video recording. Informed consent has been obtained from all of the participants.



Within this activity, explanations and instructions (in which the customer is invited to manipulate the car) are often introduced with *ici*, the mention of the object explained and a pointing gesture to the object. Both participants are sitting in an arrangement side-by-side, within the car, in front of the dashboard. The dealer progressively points and manipulates all of the technical objects, going from the left hand side to the right hand side, as well as turning to the roof and to the space between the driver and the passenger seats. These recurrent actions offer an environment in which turn-initial and even sequence-initial *ici* is recurrently used to introduce the next item to be talked about and to be manipulated by the dealer and the customer.

My analysis focuses on the multimodal resources mobilized to introduce the new item within a turn beginning with the deictic *ici*. The focus is not only on the gesture done by the speaker while uttering the deictic, but also on his gaze on the recipient and the recipient's response. By taking into consideration talk, gesture, head movements, nods, gaze directions of both participants as they are coordinated in a finely tuned way within their social interaction, I aim at highlighting the way in which these resources are interactively and timely organized within the course of the activity.

This situation allows us to observe the complexity of deictic practices, the temporal and sequential organization of multimodal resources in interaction, and the finely tuned organization of shared focuses of attention. This casts light on referential practices, joint attention, visual perception and mutual understanding not as mere mental processes but as embodied actions which are interactively managed.

More generally, this approach to deictic practices casts a light on some specificities of an interactional grammar taking seriously the situated uses by participants of a variety of resources, which are both locally defined, relatively to the activities and the context at hand, and orderly mobilized, respecting the fundamental principles of turn-taking and sequence organization. In this framework, the grammar of *ici* takes the form of a complex multimodal and praxeological Gestalt, comprising not only the form *ici* but all other multimodal resources involved, their finely tuned temporalities (both simultaneous and successive) and their sequential positions.

3. A first sketch of the analysis: two forms of *ici*

A careful analysis of the videodata reveals that instructions beginning with *ici* can be organized in two different ways.



3.1. First pattern

The first pattern consists in *ici* co-occurring with the speaker's pointing gesture towards the object which is then described to the recipient. This pattern is characterized by an expected co-occurrence of the deictic form and the pointing gesture. It is also characterized by the fact that the recipient's attention is focused on the gesture and therefore on the object pointed at.

The pattern is observable in the following excerpt, in which the dealer shows to a customer, Marie, how to activate the windshield wipers by pushing a button on the steering wheel:

(1) (M5.33 essuie-glaces e6)

```
1 Jan:  *i#c*i#  *vos essuie-gla*ces,* (.) en +bA:s,#  
        here your windshield wipers, (.) below,  
        *...*points to the lever*pushes down*  
        >>looks b•looks at windshield----->  
mar    >>looks at button-----+looks up at  
                                   windshield wiper->  
fig    #1  #2                                #3
```



1



2



3



180

Lorenza Mondada

```

2          (0.3)
3 Jan:     ça s'ra une +saleté s+ur l'pare-bri.:#se,=
           this will be a dirtiness on the windscreen,
           jan                                -->•looks at M-->
           mar                                -->+looks down+looks in front-->
           fig                                #4
4 Mar:     =m•m
           jan     ->•,,
    
```



As Jan utters his turn-initial *ici*, he begins to point (Figure 1) in such a way that his pointing gesture reaches its maximal extension and encounters the lever on the steering wheel at the end of the deictic (Figure 2). From the very beginning of the turn, Marie's gaze is oriented on the "gesture space" (McNeill 1992) in front of them; she immediately focuses on the object indicated by the gesture.

The deictic *ici* is followed, without any pause, by the noun-phrase *vos essuie-glaces* (your windshield wipers). At the end of the NP, on the last syllable of *essuie-glaces*, Jan pushes the lever and activates the wiper. Just after a pause, he adds *en bas* which describes the direction in which he has just pushed the lever. At that moment, Marie looks up to the windshield, following the wiper. Her change of gaze direction displays her understanding of the effect of the lever activation. On his side, Jan, who was looking at the button while uttering the deictic, looks up earlier, as he produces the NP (Figure 3). After a pause (line 2) he gives a description of the circumstances in which this wiper has to be activated, in the future tense (line 3): at the end of this sentence, he looks at Marie (Figure 4).

This gaze to Marie projects Jan's turn completion and orients towards the relevance of a response from her, displaying her understanding and confirming the completion of the explanation: as soon as she produces a *mm*, Jan looks away and initiates the introduction of the next item.

In this fragment, Jan's and Marie's actions are coordinated in a finely tuned way: Jan's progression within his turn and his demonstrating action is adjusted to Marie's gaze and final acknowledgment. The production of the



instruction is interactively managed, his completion being defined by the co-orientation, the relevant bodily displays, the final gaze and recipient's confirmation.

3.2. Second pattern

But this first pattern does not constitute the unique possibility of introducing a new referent within the corpus. A second pattern is observable, characterized by another temporality and sequential organization – with a delayed gesture and the recipient gazing away at the beginning of the sequence. The deictic *ici* is produced without any deictic gesture, it is followed by a noun-phrase naming the item, which co-occurs with a pointing gesture once the recipient has turned his attention on the speaker's instructing action.

In the following fragment, after having explained the neutralization of the airbag activated between the seats, Jan shows to Guy the button which allows the driver to close the doors of the car:

(2) (G5.00 fermeture des portes)

```
1 Jan: I#+C*I,# à l'a#vant, +la *fermeture +des# *por•tes,#
      HERE, in the front, the closing of the doors
      *.....approaches hand*hand on button-*touches on
                                           button-->
                                           •looks at G----->
guy      +moves head-----+.....+looks -->>
fig      #5      #6      #7      #8      #9
2      (0.9)
3 Jan: pour rouler en sécuri+té. +•
      in order to drive safely.
      ---->•
guy      +small nod+
4      (0.*7)
guy      ->*,,,
```




182

Lorenza Mondada



5



6



7



8



9

Jan's turn, line 1, is perfectly coordinated with Guy's body realignment, from a position in which he looks away from the target object to a position in which he is focused on it.

When Jan begins to utter *ici*, he has not yet moved his arm (Figure 5). He begins to approach his hand on the end of *ici* (Figure 6) and expands progressively his pointing gesture (Figure 7), approaching the target, staying on the target and finally gently touching it, as he expands his turn and as Guy is turning towards the front.

Jan's turn inserts, after the deictic, another location (*à l'avant*), before producing the NP describing the newly introduced item (*la fermeture des portes*). The extra spatial indication orients to Guy's position, still looking behind. During the NP, Guy begins to look at the object and towards the middle of the NP his head is perfectly aligned with the target. At that moment, Jan touches gently the button that activates the closure of the car's doors (Figure 8).



On the last syllable of the NP, he also looks at Guy (Figure 9). At that point, Guy is gazing towards the indicated object. After a pause (line 2) where Guy does not respond at all, Jan adds a functional explanation (line 3): at his pre-completion, Guy produces a nod and Jan stops to stare at him and withdraws his gaze (line 4). Again, a response token done by the recipient closes the sequence.

In this case, Jan's turn is formatted in a way that takes into consideration the fact that at the beginning of the turn Guy is not in a position in which he can see his gesture. Not only the gesture trajectory adjusts to this, but also the turn's format – integrating an extra spatial indication. Jan also gazes at Guy as soon as the NP is uttered, as to check his understanding, and maintains his gaze until the second part of the explanation has been produced and Guy has marked its completion by a slight nod.

3.3. Introducing the referent vs. getting the attention of the recipient

Thus, turns beginning with *ici* and introducing a new referent are formatted in two different ways, depending on the attention of the co-participant. Their organization is temporally finely shaped by a constant adjustment to the recipient's state of attention. This reveals the detailed and timely interactional organization of referential practices and arrangement of multimodal resources for achieving it.

In the first case, *ici* is a referent introducer; in the second case, *ici* works as an attention getting device. In the latter case, the introduction of the referent is achieved by the NP and by the gesture co-occurrent with it.

Therefore, the corpus reveals that participants use two types of *ici*, characterized by a different temporal distribution of resources, in particular of pointing gesture, and achieving two different kinds of actions.

In the following analyses, I will show that these two formats are consistent through the corpus, by giving more occurrences of both of them. Finally I will focus on the systematic organization of recipients' responses.

4. Pattern 1: *ici* + pointing gesture as introducing a new referent

Pattern 1 is characterized by a concomitance of speaker's deictic and pointing gesture coordinated with recipient's gaze on the pointed object. Below are three occurrences of this pattern:



184

Lorenza Mondada

(3) (L12.43 air-bag e1)
1 + (0.*3)
jan *.....->
luc +looks-->
2 Jan: ici *<la# +coupure *air-bag* passag*er.>•
here the deconnection (of the) passenger's air-bag.
->*points-----*points-*holds----*
>>looks at the button-----•looks at L-->>
luc -->+leans over----->>
fig #10
3 Luc: d'ac[cord.
oka[y.



10

(4) (D6_42 essuie-glaces e9)
1 Jan: *+ici* vos* e#ssuie-+glaces, en ba*s.
here your windshield wipers, below.
*pts*pushes*finger stays near the button-->>
•looks at D-->>
dia +looks at button---+looks at wwiper---->
fig #11A=11B
2 (0.+3)+
->+nods+looks at button-->>
3 Jan: ((next item))



11A



11B



(5) (M10.30 changer 645/e2)

```

1      *(0.2)
      *.....->
2 Jan: ic+i, (.) la# flèche que vous avez en* haut,*+
      here, (.) the arrow you have above,
      ->*points-----*pushes*
      mar      +looks down at the lever-----+
      fig      #12
      *+ (.) et* en ba*s.*
      (.) and below.
      *,,.....*shows--*,,*
      +looks dashboard-->>
3      (0.3)
4 Jan: c'est pour changer, ces deux là.
      it's to change, these two there

```



12

In these three excerpts, *ici* is produced by the speaker while pointing to the object, while the recipient looks at the target too.

What characterizes this pattern is the fact that *ici* is followed – often without any pause (Extracts 1, 3, 4) – by the NP which describes the target object (either with a determined article or with a possessive adjective, like in Extract 4). In Extract 5, the NP is composed by a noun expanded by a relative phrase, referring to two directions in which the lever can be activated.

Excerpts 1, 4 and 5 also concern a double action: the action of showing by pointing towards a lever and the action of demonstrating how it works by pulling up the lever. This supposes on the side of the receiver a double focus of attention: on the button or the lever on the one hand, on the activated mechanism (the windshield wipers, or an indication appearing on the dashboard) on the other hand. Again, the recipient's attention switching from one target to the other displays her understanding (cf. Mondada 2011) of the technical detail which is being explained as well as of its consequences.





The disposition of the objects within the space of the car and the sequential ordering of the explanations contribute to the fact that the recipient's attention might already be focused on the region targeted by the next item – thus by the deictic and the pointing gesture. The dealer explains the various technical features by following an order of proximity from left to right; in this way the completion of an explanation projects the next item which might be situated in the same area.

This feature is visible in the following excerpt, which contains a series of indications:

(6) (G22.38 série e11)

```

1 Jan: → ic•i, (.) lunette arriè•re chauffa•nte,+
      here, (.) heating rear window,
      ..•looks at G-----•down--->
      >>points--->
      guy >>looks down---> +nods+
2      (0.4)
3 Jan: → ici, le recyclage d'air,• (0.6) +dans l'tu+nnel,
      here, air recycling, (0.6) in a tunnel,
      -->•looks at G--->
      guy +nods-----+
4      tout +c'qu'[est+ tunnel (.) poi•ds lou+rds,+
      everything which is tunnel (.) heavy goods vehicles,
5 Guy:      [ou+ais,
      [yeah,
      +nods-----+ +small nod+
      --->•down--->
6      (0.3)
7 Jan: → et ici la position, (0.4) a cé, (.) air climatisé
      qu'j'enlève,
      and here the position, (0.4) AC, (.) air conditioned
      which I take off,
8      (1.•2)
      ->•looks at G--->>
9 Jan: ((continues))
    
```

In this fragment, Jan is showing a series of functions regulating the air conditioning: every function is introduced by *ici* co-occurrent with his continuous pointing, followed by a NP. Likewise, Guy's gaze is constantly focused on the target of his gesture.

In this type of environment, the target can acquire a straightforward dimension, given by its visibility and accessibility in recipient's focus of attention. This has an impact on turn construction, which can be even more economical, as in the following excerpt, where the turn is reduced to the deictic:



(7) (D7_48 ici e10)

```
1 Jan:  *+alors* ici,*
        so here,
        *.....*points*
        dia +looks--->>
2      *(0.2)* (0.2) |
        jan  *pushes*
3 Mus:  | (0.2)•(0.2)[ ((music continues))
        jan  •looks at D-----•
4 Dia:  [oké
        [okay
```

As Diane is looking at the radio, Jan points towards one of the buttons and pushes it: after a slight gap (line 2), music is hearable (line 3), constituting a self-explanation of the device. As Jan turns to Diane, she immediately acknowledges the demonstration (line 4).

In this case, there is an extreme economy of the verbal and gestural deictic resources, which introduce the referent and predicate its features: no more words are necessary.

In contrast, when there is a delay within the attention of the recipient, the turn is expanded thanks to the addition of extra materials, such as an insertion in the following excerpt. Jan is explaining to Marie the disconnection of the airbag, a device that is specially meant for babies travelling on the passenger's seat. This explanation occurs when Marie is holding her baby in her arms and is momentarily impeded by him to look at the target:

(8) (M16.38/1031 airbag e3)

```
1      *(0+.3)
        jan  *.....->
        mar  +turns head-->
2 Jan:  .h i+c#*+i,# (0.2)* +comme vous avez*# l'bé*bé, (.)+*#
        .h here, (0.2) since you have the baby, (.)
        -->*points---*,,,,,,,,,,,,,,,,,,,,,*points baby*...*
        mar  Æ ->+l--+leans head+leans down on the target-----+
        fig  #13 #14 #15 #16
3      +*la >coupure airbag passager.<*(.) ça c'est *
        the >disconnection of the passenger's airbag.< (.) that's
        *3x pointings-----*finger on the button*
        mar  +looks--->>
4      •quand [vous laissez BIEN SUR• (.)
        when [you leave OF COURSE (.)
        •gazes at M-----•
5 Mar:  [ah ça c'est super
        [oh that's great
```



188

Lorenza Mondada

5 Jan: le bébé à l'avant (0.4) dans un couffin
the baby on the front (seat) (0.4) in a moses basket



When the speaker utters *ici*, the recipient has just turned her head (line 1) towards him and looks towards indicated direction (Figure 13). But her head movement is not complete and her body is not yet in the position of comfortably looking at the target. She continues to move – leaning down with not only the head but also the upper torso (Figures 14 and 15) – and her bodily arrangement is completed only at the end of Jan's insertion, *comme vous avez l'enfant* (since you have the baby) (line 2) (Figure 16). Thus, the insertion of the parenthesis allows Jan to wait that Marie's adequate body disposition is completed. Her movement and his insertion are perfectly coordinated, in such a way that when Jan utters the NP (*>la coupure airbag passager<*, lines 2–3), her attention is fully focused on the target.

The pattern “*ici* + NP concomitant with S's pointing and A's looking” is generally realized with a coordinated speaker's verbo-gestural introduction of the referent and addressee's focus of attention on the referent. In this case, *ici* works as an introducer of a new referent.

This pattern can be reduced to the deictics *ici* only, in cases in which the gesture is treated as straightforward enough for the reference (Excerpt 7). But the pattern can also be expanded by inserting some additional material between *ici* and the NP in cases in which the addressee's attention is not yet entirely focused on the target. These cases show that the simultaneity of the speaker's gesture and the addressee's gaze is an interactional process which has to be achieved within time by the participants; in case of delays, the speaker adjusts the format of his turn to the timing of the recipient's actions.



5. Pattern 2: *ici* as an attention getting device

The first pattern characterizing the use of turn-initial *ici* is most frequently adopted for the introduction of an object that is proximal to the previous one and which does not necessitate a reorientation of the addressee's attention. Alternatively, when the speaker introduces a referent located in another area than the one which is currently being looked at by the recipient, the same construction is used, but with a different distribution of gesture and gaze. In this second pattern, *ici* is not used to point to the new referent but much more as an instruction that there is something to look at, inviting the recipient to refocalize her attention and to search for the relevant object in the surroundings. Typically, this refocalization happens when the next object to be explained is not located next to the previous one, but elsewhere in the car interior.

Excerpt 2 already showed an instance of this pattern. The next excerpt is another instance: after having explained the regulation of the air conditioning below the radio, Jan goes on about how the glove box can be refrigerated:

(9) (G23.33 ouverture BteGants – e6)

1 Jan: #ic+i,* +vous avez, *(0.9)+ 1'ouverture, (.) c't-à-dire#
here, you have, (0.9) the opening, (.) that's to say
*.....*introduces hand in box----->>
guy +l up+lateral head's mvmt+leans progressively down-->
•looks at G----->>
fig #17 18#
2 qu'en gros si vous voulez, +réfri#gérer la boîte à gants
roughly if you want, to refrigerate the glove box
guy -->+stays in his position-->>
fig #19



In this case, we observe that Guy leans progressively on his right and that he stabilizes the adequate posture to look into the box only when Jan is already describing it. In this case, Jan's indicating gesture has reached his maximal





190

Lorenza Mondada

extension when he produces the NP (*l'ouverture*). This NP is introduced after a verb (*vous avez*) and a long pause, both delaying its production while Guy is still leaning down. After the NP, Jan further expands the explanation with a reformulation marker (*c'est-à-dire*). In this case, the speaker formats his turn in such a way that he can expand it, in order to coordinate with the delayed re-orientation of his addressee's body.

In the following excerpt a similar delay is observable. Jan has just entered the car, sitting beside Luc: at the beginning of a new set of instructions, he shows to Luc the automatic rear view mirror, situated in front of them, on the top of the windshield:

(10) (L6.00 retroviseur + ici final – e4)

```
1 Jan:  i#ci. si •on part par# le haut, +*on a le, #**+ (.) #*
        here. if we start from the top, we have the, (.)
        •looks up -----•l at L-->>
                                *.....*points 2x*
luc    >>looks away-----+.....+l up-->>
fig    #20                      #21                      #22      #23
2      *le rétrovi*seur automatique.
        the automatic rear view mirror.
        * , , , , , , , *
3 Luc:  m[M
```





In this fragment, when Jan utters *ici*, neither him nor Luc are looking at the target, which is located on the top of their heads. Luc is looking down, with a thinking face. His son, on the rear, is also looking down. Moreover, Jan's hand is still in its home position (Figure 20).

Jan's turn is composed, after the deictic, by a locative instruction (*si on part par le haut*, line 1), uttered while Jan looks up and Luc still gazes down (Figure 21). This gaze up can be considered as a kind of preliminary pointing. Luc begins to move his head up only at the end of this instruction, when Guy introduces the referent with the verb *on a*. Significantly, the NP, *le rétroviseur automatique* is introduced with a self-repair of the article (*le, (.) le*) which is related to the delayed participants' gesture and gaze (Figure 22). At this precise moment, Jan points twice at the target and Luc looks at it (Figure 23).

Thus, in this case, *ici* is clearly not used to point to the target. It is rather used as an attention getting device, projecting an imminent pointing to an object to be searched for.

The continuation of this fragment shows another instance of pointing towards the referent, which is done later on, while Jan expands his explanation and checks Luc's understanding. The sequence is closed by a turn-final *ici* (7):

(11) (continuation of Excerpt 10)

```

4 Jan: [qui lui,• (.) va s'as+sombrir •tout seul la nuit.
      [which (.) will get darker automatically at night.
      -->• •looks at L--->
luc -->+looks at J----->
5 vous +aurez plus de ++jour et nu*it•.***
  you will no longer have any day and night.
      *....*points 3x-->
      --->•
luc -->+looks up-->>
luc ++nods-----++
6 Luc: d'accord.=
      alright.=
7 Jan: =vous# rap•pelez, i*ci?•
      =do you remember, here?
      -->*,,,
      •looks at L---•
fig #24
8 Luc: >ouais ouais,<
      >yeah yeah,<

```





After Luc's first minimal acknowledgment (line 3), Jan continues his explanation, offering a delayed completion expanding his current turn with a relative phrase (line 4) and then with the addition of a new TCU (line 5). At the end of the latter, Jan points again (thrice) to the rear view mirror. But he does it in a different way than the first time: the first pointing touched laterally the target (Figures 22–23) whereas the second pointing touches it below (Figure 24). The latter gesture not only points to the target, but also does a gesture that was routinely done on the manual rear view mirror at night. With this gesture, Jan is re-enacting a familiar gesture that Luc is supposed to remember.

In this expansion of the explanation, Jan does not close the sequence after Luc's acknowledgment but pursues a more explicit confirmation of understanding. This is also the case of another sequence with a delayed gesture of pointing, in which Jan shows the closing of the doors to Marie – within a sequence beginning and ending with *ici*:

(12) (M5.33/359 fermeture des portes – e2)

```

1 Jan:  d*onc ici la ferme+ture des por+*tes.+
        so here the closing of the doors.
        *.....*touches w. medium->
mar    >>looks away-----+.....+looks at J's finger+
2      •+(0.7)+
      jan  •looks at M----->
      mar  +leans fwd+looks intensively-->
3 Jan:  °( * )°*
        *points*
4      (0.3)
5 Jan:  °(pour)° rouler en sécurité, hein vous
        (in order to) drive safely, right
6      vous [enfer]•mezl'*soi*r,
        you [lock yourself at night,
        ----->*pushes*,,,
        -->*
7 Mar:  [ouais]
        [yeah]
8      (0|.5|)
      eve  |door locks|

```



9 Jan: ça s'ra ici.
it will be here.

Like in the previous excerpts, when Jan utters *ici*, his recipient is looking away and himself does not point yet to the target. As Marie turns her gaze to the object, Jan points at it.

During the pause following the completion of his turn, Jan gazes at Marie who does not respond but further leans towards the target, looking intensively (line 2). He orients to her absence of response by adding something and pointing again (line 3). In absence of any response (line 4), he continues to talk. At the end of this new addition, as Marie finally produces an acknowledgment token (line 7), he pushes the button which was previously pointed at. The sequence is closed with a conclusive *ça s'ra ici*. (line 9).

These additions and this final insistence on *ici* seem to orient to the confirmation of a possible knowledge and expertise of the recipient. Jan does not close the sequence with his pointing, or with the recipient's acknowledgment, but with a repeat of target's location, after his co-participant's confirmation. For comparison, the same item is explained to Guy in a different way (see Excerpt 2): the sequence is shorter and Jan does not push the button at the end. Guy is here treated as an "expert", whereas Marie is treated as a "novice".

These two configurations of *ici* co-occurrent or not with speaker pointing gesture and recipient focus of attention on the referred to target show how deixis is a complex practice, not depending on the deictic word alone but on the temporally finely tuned and coordinated achievement of a complex multimodal Gestalt, which also constitutes a shared focus of attention.

This practice is sequentially articulated, in various phases:

- • a *designation phase* (pointing), responded to by recipient gaze on the target;
- • a *demonstration phase* (pushing), responded to by recipient gaze on the consecutive display or moving object (e.g. windshield or dashboard);
- Both gazes, as well as the switch between one and the other, constitute visible evidences of understanding of these two first phases.
- • an *explanatory phase*, responded to by recipient verbal acknowledgment or nodding – or solicited by a continuation of speaker explanation (cf. infra § 6.).

Thus, this organization in phases is mutually oriented to by both participants; their completion is interactively achieved, thanks to a constant monitoring of recipient's responses and understanding.



6. Orienting towards recipient's response

As shown by the previous analyses, speakers constantly adjust and adapt the format of their turn as well as of their gesture's trajectory to the conduct of their recipient. In this way, they achieve their timely coordination and secure a common understanding of the ongoing referential activity. More specifically, in this last section I focus on the way in which the speaker demonstrably orients towards the recipient at a particular sequential position: when the end of the explaining episode is projected. In this sequential environment, the speaker projects the completion of the episode and turns to the recipient, monitoring for his or her displays of understanding (cf. Mondada 2011) (6.1.). In absence of a response, or if the recipient does not produce an adequate response, the speaker pursues a response by expanding his current action (6.2.).

6.1. Speaker's gaze on recipient: different distributions in pattern 1 vs. pattern 2

Through pattern 1 and pattern 2, a regular distribution of gaze is observable: at the beginning of the sequence, the speaker generally looks at the target while pointing to it; then, his gaze switches from the object to the recipient. This switch tends to occur later in pattern 1 (towards the completion of the explanation) and earlier in pattern 2 (during or at the end of the NP, i.e. the first part of the explanation). The fact that the speaker gazes earlier at the recipient when the pointing gesture and the focus of attention are achieved later on (pattern 2) displays his orientation towards a difficult, delayed, i.e. not straightforward, establishment of a focus of attention on the target and the consequent check for an adequate understanding of the co-participant.

The role of speaker's gaze on the recipient has been variously commented on in the literature and depends on the sequential position in which the shift occurs as well as the kind of action that is going on and is expected on the side of the recipient.

Already Kendon (1967: 56) noticed that speaker's gaze on the listener work not only as checks on the latter's conduct but also as "signal" to obtain a confirmation of his understanding of what is being said.

Bavelas, Coates, and Johnson (2002) report that in narratives when the speaker looks at the recipient, the latter tends to respond with a *mbm* or a nod, conveying attentiveness and understanding, before the speaker with-



draws her gaze. Gaze seems to be the stronger cue for inviting to produce a response (stronger than tag questions, pauses, changes in intonation, or conversational gesture). Listener's response is a collaborative action: "speaker gaze creates the opportunity for a listener response, and the response then terminates that gaze" (2002: 572). In the data they used, face-to-face encounters in which A tells a story to B, speaker gaze opens a "gaze-window" (2002: 569) characterized by mutual gaze, and listener produces her response in the last part of the window. In our data, participants are sitting side-by-side and look together at an object; mutual gaze is relatively seldom in this configuration. But the interactional space, characterized by their proximity, makes speaker gaze visible for the recipient, who responds to it.

In a conversation analytic framework, exploring various techniques for "mobilizing response", Stivers and Rossano (2010) show that, beside interrogative prosody and lexico-morphosyntax, gaze is a resource for pursuing uptake, especially when a response is missing. In my data, this technique is systematically used for the achievement of the completeness of the instructions; nevertheless, gaze is not always enough to obtain a response (see 6.2).

Let's go back to some of the occurrences of pattern 1:

(13) (cf. Extract 1) (M5.33 essuie-glaces e6)

```
1 Jan:  ici vos essuie-glaces, (.) en +bA:s,
        here your windshield wipers, (.) below,
2      (0.3)
3 Jan:  ça s'ra une saleté sur l'pare-bri::se=
        this will be a dirtiness on the windscreen
        jan                                •looks at M-->
4 Mar:  =m•m/
        jan    ->•,,
5 Jan:  ((next item))
```

(14) (cf. Extract 4) (D6_42 essuie-glaces e9)

```
1 Jan:  ici vos essuie-glaces, en ba•s.
        here your windshield wipers, below.
        →                                •looks at D-->>
2      (0.+3)+
        →      +nods+
3 Jan:  ((next item))
```

(15) (cf. Extract 7) (D7_48 ici e10)

```
1 Jan:  alors ici,
        so here,
```





196

Lorenza Mondada

```

2      (0.4) |
3 Mus:      | (0.2) • (0.2) [((music continues))
      jan →      • looks at D-----•
4 Dia: →      [oké
              [okay
5 Jan:      ((next item))

(16) (cf. Extract 8) (M16.38/1031 airbag e3)
2 Jan:      .h ici, (0.2) comme vous avez l'bébé, (.)
              .h here, (0.2) since you have the baby, (.) the
3          la >coupure airbag passager.< (.) ça c'est
              >disconnection of the passenger's airbag.< (.) this is
4          •quand [vous laissez BIEN SUR• (.)
              when [you leave OF COURSE (.)
      → •gazes at Mar-----•
5 Mar: →      [ah ça c'est super
              [oh that's great

```

In all of these cases, the speaker looks at the recipient towards the projectable completion of his explanation: in Excerpt 13, on the last syllable of his turn (line 3), like in Excerpt 14 (line 1); in Excerpt 15, as soon as music is heard (line 3). In Excerpt 16, which is an instance of delayed gesture and gaze within pattern 1, speaker's gaze on the recipient occurs a bit earlier, after the production of the construction containing the NP (line 3).

In all of these cases, the recipient responds promptly to speaker's gaze, either with a vocal acknowledgment (Excerpts 13, 4; Excerpts 15, 4), with a positive assessment (Excerpt 16, 5 – cf. Mondada 2009b), or with a nod (Excerpt 14, 2).

Thus, both participants orient to the completion of the episode after the referent and its demonstration have been secured.

This mutual orientation to the completion is also observable for pattern 2, except that the gaze on the recipient is achieved earlier:

```

(17) (cf. Extract 2) (G5.00 fermeture des portes)
1 Jan:      ICI, à l'avant, la fermeture des por•tes,
              HERE, in the front, the closing of the doors,
      →      •looks at G----->
2      (0.9)
3 Jan:      pour rouler en sécuri+té. +•
              in order to drive safely.
              ----•
      guy →      +small nod+

```





(18) (cf. Extract 10) (L6.00 *retroviseur* – e4)
 1 Jan: *ici. si on part par# le haut, on a le,*
 here. if we start from the top, we have the,
 2 • (.) *le rétroviseur automatique.*
 (.) the automatic rear view mirror.
 → •looks at Luc-->>
 3 Luc: → m[M

In these cases, Jan looks at the recipient during the NP (Excerpt 17) or even before (Excerpt 18), although the response comes later, at his turn-completion. This response displays that the item's description has reached completion, for the practical purposes of their situated understanding.

These systematicities show that the speaker orients to the timing of the establishment of his recipient's focus of attention; he monitors (M.H. Goodwin 1980) his or her recognition of what he has said, the joint constitution of a common ground (Clark 1996), his displays of understanding (Mondada 2011), and even his perceived perception (Hausendorf 2003; Hindmarsh and Heath 2000: 1871) and reflexively adapts to them.

6.2. Expansions: Pursuing a response

A further evidence of speaker's adjustments to the recipient, generated by this other-monitoring (cf. M.H. Goodwin 1980; Clark and Krych 2004; Schmitt and Deppermann 2007), is constituted by instances in which the recipient's response is delayed. In these cases, the speaker produces an expansion of his ongoing explanation, orienting to a possible problem of understanding, and offers a renewed transition relevance point, where a response is again expected and projected.

Speakers confronted with a recipient not providing for an expected response can mobilize various techniques for pursuing it. Among the audible-verbal procedures, Jefferson (1981) points to the use (and abuse) of various response solicitations, such as *ne* in German. Pomerantz (1984) notes that when recipients do not produce a solicited response – for example not hearing/seeing it, not understanding it or not agreeing with it – speakers have various procedures through which they pursue a response: they can look for a possible trouble source and repair or revise their initial action; they can revise the assumptions concerning shared knowledge or they can orient towards a possible disagreement and modify their position. Other techniques exploit the grammatical and syntactic features of the incremental production



198

Lorenza Mondada

of talk in interaction: for example, speakers can produce an expansion of their turn in forms of increments (Schegloff 1996; Ford, Fox, Thompson 2002), which occasions a continuation of the current turn and offer new opportunities to respond in the next transition-relevance space. Other forms of expansions have been studied on the basis of this corpus, when the dealer is “fishing for assessments” (Mondada 2009b: 356).

In the following fragments, I focus both on gaze and turn expansions as a double technique mobilized by Jan for achieving the completion of explanations in both patterns 1 and 2.

The following fragment is an occurrence of pattern 1: Jan is explaining how to move from a radio station to another by activating a lever on the steering wheel:

(19) (cf. 5) (M10.30 changer 645/e2)

```

1      *(0.2)
      *.....->
2 Jan:  ic+*i, (.) la flèche que vous avez en*
      here, (.) the arrow you have
      ->*points-----*
      mar  +looks down at the lever-->
3      *haut,*+ (.) et* en ba*s.*
      above, (.) and below.
      *pushes*, , , , , *shows*, , *
      -->+looks at dashboard-->
4      (0.3)
5 Jan:  c'est pour changer,* ces deux là.
      it's to change, these two there
      →                                     •looks at Mar-->
6      → (0.+7)
      mar  -->+looks down-->
7 Jan: → on avance +et on re+cule.*
      we go on and we come back.
      -->•
      mar →      -->+nods----+
```

In line 2, Marie alternatively looks at the lever and at the dashboard, where the name of the radio station is displayed. After the introduction of the referred to object (line 2), Jan adds a second part, in which he explains its function (line 5). In pre-completion position, he turns to Marie (end of line 5) and looks at her. After completion, a long pause follows (line 6), in which Marie does not respond; instead, she looks again at the lever. In absence of her response, Jan adds a new TCU to his explanation (line 7), continuing to gaze at





her. Marie finally acknowledges Jan's action by a nod: just after, he turns his gaze away and completes his TCU.

Whereas in the previous extract Jan was orienting to an absence of response from his recipient, in the following extract he seems to pursue a more substantial response than just a minimal one:

(20) (L6.53 ordinateur de bord – e4)

```

1      *(0.+3)*
  jan  *.....*
  luc   +looks at button-->
2 Jan:  *ici on a le ordi•nateur de bo*+rd,* .hh qui nous donne
        here we have the car computer, .hh which gives us
        *points-----*pushes*
                •looks at dashboard----->
  luc   ----->+looks at dashboard --->
3      le nombre de kilomètres• qu'il vous reste à ++faire++
        the number of kilometers that we can drive
        → ----->•looks at Luc---->
  luc →                                     ++small nod++
4      [pour ne PAS tomber en panne +d'essen+ce. (.)
        [before to run out of gas. (.)
5 Luc → [°mh°
  luc →                                     +nods+
6      ave[c c'que vous avez dans l'réservoir, (.) aujourd'hui.
        wit[h what you have in the reservoir, (.) today
                ---->.
7 Luc: → [°d'accord°
        [°alright°
8 Jan:  (.) vous pouvez faire cent quatre-vingt kilomè[tres.
        (.) you can drive one hundred and eithy kilome[tters.
9 Luc: → [d'accord.
        [alright.

```

Once Jan has pushed the pointed at button, Luc immediately looks at the small display on the dashboard where the informations given by the computer are appearing. His gaze remains focused on this display during the entire explanation. Jan looks at him quite early, during the second part of the explanation, following the first in which the referent is introduced. As soon as Jan gazes at him, Luc responds with a small nod (line 3) and a continuer (line 5), then with a new nod, just prior to TCU completion. But Jan continues, adding a new TCU after a micro-pause (line 4), in overlap with an agreeing response by Luc (line 7). Jan withdraws his gaze on the last part of his compound TCU and Luc repeats his response, with louder voice, at its completion. This brings the episode to a close. In this case, Jan seems to pursue a more substantial response than a nod or a continuer, and continues





200

Lorenza Mondada

to expand his explanation, occasioning new transition relevance points in which Luc provides more explicit acknowledgments.

Within pattern 2, expansions are also used by Jan for pursuing a response from his or her recipient. This is the case of Extracts 10–11 and 12, in which his explanation is expanded and closed by the production of a new deictic *ici* (see above). This is also the case of the following fragment, in which Jan develops substantially his description of the uses of the glove box, as Luc continues to look at it but does not respond to him until line 14:

(21) (cf. Extract 9) (G23.33 ouverture BteGants – e6)

- 1 Jan: ic+1,* +vous avez, *(0.9)+ l'ouverture, (.) c't-à-dire
here, you have, (0.9) the opening, (.) that's to say
*.....*introduces hand in box----->
guy +l up+lateral head's mvmt+leans progressively down-->
•looks at G-->>
- 2 qu'en gros si vous voulez, +réfrigérer la boîte à gants
roughly if you want, to refrigerate the glove box
guy -->+stays in his position-->>
- 3 (0.5)
- 4 Jan: °et non les disques ge pé esse •hein,°
°and not the GPS disks right,°
-->•leans and looks down-->
- 5 (0.4)
- 6 Jan: donc en gros, ICI. (0.3)• dans ce sens• vous ouvrez,
so roughly, HERE. (0.3) in this direction you open,
-->••looks at L-->
- 7 (0.4)
- 8 Jan: y a l'air, (.) qui sort,
there is the air, (.) which goes out,
- 9 (0.3)
- 10 Jan: et dans c'sens vous fermez.
and in this direction you close.
- 11 (0.6)
- 12 Jan: en gros, si vous mettez des bouteilles d'eau fraîche,
roughly, if you put bottles of fresh water,
- 13 [vous mettez des sandwiches, °ou:• (.) quoi qu'ce soit.°
[you put sandwiches, °or: (.) anything else.°
-->•looks away-->
- 14 Guy: → [ouais
[yeah

In this fragment, Jan shows the glove box to Luc, who leans down in a visible way to see inside the box (cf. Extract 9). Jan's explanation is developed, after the introduction of the referent (line 1) by an if-clause beginning with *c't-à-dire* (line 1) and *en gros* (line 2), two conjunctions introducing a reformu-





lation. This if-clause is a first part projecting a second part (Lerner 1991), which is not produced here: after a contrastive negation (line 4) and a pause (line 5), Jan does not complete the projected structure but begins a new one, with *donc en gros* and a deictic (line 6). Again he constructs a two part structure, with a first part introduced by *dans ce sens* (line 6) and a second part introduced by another *dans ce sens* (line 10), acquiring its spatial direction indexically. He looks at Jan again on the first one, but Luc continues to look inside the glove box and does not display any acknowledgment. Finally, after a new longer pause (line 11), Jan introduces a new if-construction, prefaced again by *en gros* (line 12). Only the first part of this compound construction is realized, in the form of a list: this time, Luc responds (line 14) after the first item of the list, which prosodically projects more items to come, in overlap with the second item. Jan utters the third generic item of the list with a lower voice, orienting to the closure of the episode and looking away.

In this case, Jan adds multiple expansions before Luc responds; as soon the latter produces his *ouais* (‘yeah’), Jan treats the episode as completed.

The analyses of speaker’s gaze on the recipient show that in the context of the introduction of a referent and of a subsequent instruction and explanation, the speaker expects a response from his recipient, which interactively defines the completion of the episode and brings it to a close.

Gaze is distributed differently in turns initiated by *ici* as an introductory device or as an attention getting device: in the former, speaker gaze tends to occur towards the projected completion of the episode; in the latter it occurs earlier, as a confirmation of the successful achievement of the referents’ introduction. Thus, when *ici* is not working as a referent introducer, gaze seems to be more important for checking the ongoing introduction of the referent.

Gaze can be followed by the response it solicits, but it can also be followed by the absence of a response or by a response which is considered as not adequate or not strong enough in that context. In the latter case, other resources can be mobilized to pursue the response, namely various expansions of the initial explanation. As soon as a response is produced, the speaker brings the turn and the episode to a close and withdraws his gaze.

These responses to the multimodal introduction of a referent show that the various methods for achieving it – namely using *ici* within two different multimodal formats – are not only sensitive to the responsiveness of the recipient but also actively pursue it.





7. Conclusions

In the first part of this chapter I describe two configurations in which *ici* is used in turn-initial position. In the first configuration, *ici* is concomitant with a pointing gesture and with the recipient's gaze on the pointed object; this complex Gestalt works as a *device introducing a new referent* singled out within the material environment. In the second configuration, *ici* is uttered before speaker pointing and recipient gaze are achieved: in this case, *ici* works as an *attention getting device*, as a device announcing that something has to be searched for and seen, projecting imminent pointing. In this latter case, pointing occurs when the NP verbalizing the object is uttered, and is concomitant with recipient gaze. These two configurations show the importance of the temporal and interactional arrangement of a multimodal composite Gestalt and the need to describe such Gestalts in order to capture the praxeological and interactional meaning and function of deictic words, such as *ici*.

In the second part of the chapter, I show how these two configurations are attended to by the co-participants: speaker monitors the conduct of recipient, projecting and expecting a response displaying his perception, recognition and understanding of the deictic practice. More precisely, speaker gaze at recipient towards the completion of the referential action; thereby he solicits his response in order to treat the episode as complete and to bring it to a close. In absence of response, speaker mobilizes various techniques which provide for new occasions to respond. In this sense, the completion of the referential practice is interactively and reflexively defined by both participants.

The practice of referring to a new object with the deictic *ici* cannot be reduced to the verbal deictics but has to take into consideration in an integrated way a) the multimodal Gestalt, b) its temporal arrangement, c) its sequential unfolding, d) the way in which it is interactively treated by both participants. These different dimensions go together, forming a what could be called a *multi-modal construction* (for instance, Streeck 2009:96, speaks of "tri-modal construction" integrating speech, gesture and gaze).

These multi-modal constructions in interaction pave the way for an integrative model of grammar based on the situated use of visual, verbal and auditive resources. Grounding grammar on use and users means, within the framework developed here, a focus on interaction, time and context. A grammar based on usage entails the consideration of the interactive construction of turns at talk, as they are achieved by participants and within their local perspective. A focus on actual usages means to take into consideration



the temporal and emergent dimension of turn co-elaboration, as it unfolds moment by moment, as it can be constantly redesigned and reformatted when confronted to the contingencies of situated talk and action. This focus on actual usages also concerns what users in real time situatedly define as adequate and relevant resources for their communicative action: a range of multimodal resources, constituted by visual as well as auditive forms, by verbal as well as embodied patterns, are locally mobilized for the mutually intelligible organization of action and interaction. One of the consequences is the highly flexible, timely, and mutually adjustable character of these resources. As this study shows about deixis, their grammar relies both on systematic patterns, generated by their methodical use, and on indexical features, indispensable for their contextually situated use. Time, multimodality, and indexicality are the major landmarks of an interactional grammar.

Acknowledgments

Data analyzed in this chapter have been collected within the EMIC project (“Espace, Mobilité, Interaction, Corps”) funded by Peugeot-PSA in 2003–2004. Thanks to Jonathan Bergena and Caroline Cance for their collaboration during fieldwork and video recordings.

Appendix

Transcript conventions

Talk:

Talk has been transcribed according to conventions developed by Gail Jefferson (cf. Jefferson, 2004).

Embodied conducts:

An indicative translation is provided line per line.

Multimodal details have been transcribed according to the following conventions (see Mondada, http://icar.univ-lyon2.fr/projets/corinte/bandeau_droit/convention_icor.htm):

- * * delimitdescriptionofthedealer/Jan’sactions.
- • delimitdescriptionofdealer/Jan’sgaze.
- + + delimitdescriptionofthecustomer’sactions.
- *---> action described continues across subsequent lines.
- *--->> action described continues until and after excerpt’s end.
- >* action described continues until the same symbol is reached.



204

Lorenza Mondada

>>-- action described begins before the excerpt's beginning.
 action's preparation.
 , , , , , action's retraction.
 guy participant doing the action is identified in small characters
 when he is not the current speaker or when the gesture is done
 during a pause
 fig figure; screen shot
 # indicates the exact moment at which the screen shot has been
 recorded

Translation is indicative and aims at supporting the reading of the original

References

- Auer, P. 2009 Online syntax. Thoughts on the temporality of spoken language. *Language Sciences* 31: 1–13.
- Auer, P., E. Couper-Kuhlen and F. Müller 1999 *Language in Time. The rhythm and tempo of spoken interaction*. Oxford: Oxford University Press.
- Bavelas, J. B. and N. Chovil 2000 Visible acts of meaning. An integrated message model of language use in face-to-face dialogue. *Journal of Language and Social Psychology* 19: 163–194.
- Bavelas, J. B., L. Coates and T. Johnson 2002 Listener responses as a collaborative process: The role of gaze. *Journal of Communication* 52: 566–580.
- Bühler, K. 1965 [1934] *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Stuttgart: Fischer.
- Clark, H. H. 1996 *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H. and M. A. Krych 2004 Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50: 62–81.
- Condon, W. S. 1971 Speech and body motion synchrony of the speaker-hearer. In: D. L. Horton and J. J. Jenkins (eds.), *Perception of Language*, 150–173. Columbus: Merrill.
- Couper-Kuhlen, E. and M. Selting (eds.) 1996 *Prosody in Conversation: Interactional studies*. Cambridge: Cambridge University Press.
- Fillmore, C. 1975 *Santa Cruz Lectures on Deixis, 1971*. Bloomington: Indiana University Linguistics Club.
- Ford, C. E. 2004 Contingency and units in interaction. *Discourse Studies* 6: 27–52.
- Ford, C. E., B. A. Fox and S. A. Thompson 2002 Constituency and the grammar of turn increments. In: C. E. Ford, B. A. Fox and S. A. Thompson (eds.), *The Language of Turn and Sequence*, 14–38. Oxford: Oxford University Press.
- Fricke, E. 2007 *Origo, Geste und Raum: Lokaldeixis im Deutschen*. Berlin: De Gruyter.
- Goodwin, C. 1979 The interactive construction of a sentence in natural conversation. In: G. Psathas (ed.), *Everyday Language: Studies in ethnomethodology*, 97–121. New York: Irvington Publishers.
- Goodwin, C. 1981 *Conversational Organization: Interaction between speakers and hearers*. New York: Academic Press.





- Goodwin, C. 2000 Action and embodiment within situated human interaction. *Journal of Pragmatics* 32: 1489–1522.
- Goodwin, C. 2003 Pointing as situated practice. In: S. Kita (ed.), *Pointing: Where language, culture and cognition meet*, 217–241. Hillsdale: L. Erlbaum.
- Goodwin, C. 2007 Environmentally coupled gestures. In: S. Duncan, J. Cassell and E. Levy (eds.), *Gesture and the Dynamic Dimensions of Language*, 195–212. Amsterdam, Philadelphia: John Benjamins.
- Goodwin, M. H. 1980 Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry* 50: 303–317.
- Hakulinen, A., and M. Selting (eds.) 2005 *Syntax and Lexis in Conversation. Studies on the use of linguistic resources in talk-in-interaction*. Amsterdam: John Benjamins.
- Hanks, W. F. 1990 *Referential Practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.
- Hausendorf, H. 2003 Deixis and speech situation revisited: The mechanism of perceived perception. In: F. Lenz (ed.), *Deictic Conceptualization of Space, Time and Person*, 249–269. Amsterdam: John Benjamins.
- Haviland, J. B. 1993 Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3: 3–45.
- Hindmarsh, J. and C. Heath 2000 Embodied reference: A study of deixis in workplace interaction. *Journal of Pragmatics* 32: 1855–1878.
- Hopper, P. 1988 Emergent grammar and the a priori grammar postulate. In: D. Tannen (ed.), *Linguistics in Context: Connecting observation and understanding*, 103–120. Norwood: Ablex.
- Jefferson, G. 1981 The abominable «ne?»: A working paper exploring the phenomenon of post-response pursuit of response. *Occasional Paper No.6*, Department of Sociology, University of Manchester.
- Jefferson, G. 2004 Glossary of transcript symbols with an introduction. In: G. H. Lerner (ed.), *Conversation Analysis: Studies from the first generation*, 13–31. Amsterdam: John Benjamins.
- Kendon A. 1967 Some functions of gaze-direction in social interaction. *Acta Psychologica* 26: 22–63.
- Kendon, A. 1980 Gesture and speech: Two aspects of the process of utterance. In: M. R. Key (ed.), *Nonverbal Communication and Language*, 207–277. The Hague: Mouton.
- Kendon, A. 2004 *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kita, S. (ed.) 2003a *Pointing: Where language, culture and cognition meet*. Mahwah: Erlbaum.
- Kita, S. 2003b Interplay of gaze, hand, torso orientation, and language in pointing. In: S. Kita (ed.), *Pointing: Where language, culture and cognition meet*, 307–328. Mahwah: Erlbaum.
- Lerner, G. H. 1991 On the syntax of sentence-in-progress. *Language in Society* 20: 441–458.
- Lyons, J. 1991 *Natural Language and Universal Grammar*. Cambridge: Cambridge University Press.
- McNeill, D. 1979 *The Conceptual Basis of Language*. Hillsdale: Erlbaum.
- McNeill, D. 1985 So you think gestures are nonverbal? *Psychology Review* 92: 350–371.





- McNeill, D. 1992 *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- Mondada, L. 2005 La constitution de l'origo déictique comme travail interactionnel des participants: une approche praxéologique de la spatialité. *Intellectica* 2/3: 75–100.
- Mondada, L. 2007 Interaktionsraum und Koordinierung. In: R. Schmitt (ed). *Koordination. Analysen zur multimodalen Interaktion*, 55–94. Tübingen: Narr.
- Mondada, L. 2009a Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics* 41: 1977–1997.
- Mondada, L. 2009b The embodied and negotiated production of assessments in instructed actions. *Research on Language and Social Interaction* 42: 329–361.
- Mondada, L. 2011 Understanding as a embodied, situated and sequential achievement in interaction. *Journal of Pragmatics* 43: 542–552.
- Mondada, L. in press Organisation multimodale de la parole-en-interaction: Pratiques incarnées d'introduction des référents. *Langue Française*.
- Ochs, E., E. A. Schegloff and S.A. Thompson (eds.) 1996 *Grammar and Interaction*. Cambridge: Cambridge University Press.
- Pomerantz, A. 1984 Pursuing a response. In: J. M. Atkinson and J. Heritage (eds.), *Structures of Social Action*, 152–163. Cambridge University Press, Cambridge.
- Schegloff, E. A. 1984 On some gestures' relation to talk. In: J. M. Atkinson and J. Heritage (eds.), *Structures of Social Action*, 266–296. Cambridge: Cambridge University Press.
- Schegloff, E. A. 1996 Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Grammar and Interaction*, 52–133. Cambridge: Cambridge University Press.
- Schmitt, R. and A. Deppermann 2007 Monitoring und Koordination als Voraussetzungen der multimodalen Konstitution von Interaktionsräumen. In: R. Schmitt (ed.). *Koordination. Analysen zur multimodalen Interaktion*, 95–128. Tübingen: Narr.
- Selting, M. 1996 On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6: 357–388.
- Stivers, T. and F. Rossano 2010 Mobilizing response. *Research on Language and Social Interaction* 43: 3–31.
- Streeck, J. 2009 *Gesturecraft. The manufacture of meaning*. Amsterdam: John Benjamins.





Florence Oloff
ICAR Research Lab (CNRS, University of Lyon)

Withdrawal from turns in overlap and participation

1. Introduction

In this chapter, I will focus on the phenomenon of drop out, i.e., withdrawal from the turn due to overlapping talk, in order to reflect on the link between “unfinished” turns and participation framework. With the help of a sequential and multimodal analysis inspired by the conversation analytical approach, I will show that dropping out from a turn is strongly linked to the availability displayed by potential recipients of a turn-at-talk. Although conversation analysis has described in detail the systematics of overlapping talk, especially of its onset (Jefferson 1973, 1983, 1986) and its resolution (Schegloff 2000; Jefferson 2004), the phenomenon of withdrawal from a turn due to simultaneous talk has not been investigated in detail. While it seems to be difficult to describe this interactional practice by referring exclusively to syntactic features (incompleteness of the turn), I suggest looking at turn withdrawal from a multimodal perspective (e.g. Goodwin 1980, 1981; Mondada 2007a; Schmitt 2005), taking into account visible resources like gaze or gesture. The problem of continuing or stopping a turn-in-progress in overlapping talk can be closely linked to the participation framework (Goodwin and Goodwin 2004), as speakers do visibly take into account their recipient’s availability and coordinate their turn construction with the dynamic changes of the participation framework and the interactional space.

1.1. Turn-taking, overlapping talk and overlap resolution

The analysis of turn-taking as a systematic procedure in naturally occurring interaction has been a classical field of investigation of conversation analysis since the publication of the well-known article by Sacks, Schegloff, and Jefferson in 1974. As the authors point out, speakers alternate their turns by minimizing gaps and overlap (Sacks et al. 1974: 700–701), orienting to moments of possible completion in their interlocutors’ turns, the so-called *transition-relevance places* (TRP).¹ Overlapping talk is positioned with respect to the

¹ Though the omnipresence of simultaneous talk in conversation has led to a long lasting discussion on the validity of Sacks et al’s (1974) observations, the notion of “interruption”, or of “rules” for turn-taking, this chapter will not comment on



possible completion of the first speaker's turn, i.e., the next speaker starts to speak as soon as he is able to recognize a possible completion point in a *turn-constructive unit* (TCU) (Drew 2009; Jefferson 1983, 1986), which are interactively and praxeologically shaped organizational units of talk (Ford 2004; Ford, Fox, and Thompson 1996; Mondada 2007b). The length of overlap depends, on the one hand, on the possibility of anticipating a completion in the current speaker's turn, and, on the other hand, on the possibilities a current speaker has to continue his turn after a TRP by adding tags (Jefferson 1973), turn extensions, or even new TCUs. Long lasting overlaps in ordinary conversation are quite rare, as speakers generally orient to a quick resolution of the overlap and thus to a re-installment of the "one party at a time"-principle (Sacks, Schegloff, and Jefferson 1974; Schegloff 2000). This resolution can be either inherent to the format of the overlapping turns (e.g. in case of continuers), or, should the overlapping talk persist, be achieved by a set of systematic practices (Jefferson 2004; Schegloff 2000), like the increase of volume of the talk, the production of sound stretches, or the slowing down or acceleration of the pace. This upgrade to competition or *fight for the floor* (Schegloff 2000) aims at keeping the floor while making the co-participant drop out of the turn. Although Schegloff (2000: 293) suggests that it is generally the "receiver of the upgrade" who will drop out, sometimes both speakers persist in overlap until completion of their turn, sometimes both speakers drop out simultaneously well before turn completion, and a withdrawal from the turn can occur even before participants in overlap upgrade to competition. Indeed, Jefferson (2004) shows that withdrawal from the turn is not simply linked to matters of first and second position, or of speaking rights related to those positions.

The phenomenon of withdrawal raises at least two main research questions. First, to which kind of rule or interactional constraints do speakers in overlap orient in order to negotiate withdrawal from the turn? Second, do those turn withdrawals have a specific format, with systematic syntactic, prosodic, etc. patterns? Up to now, no work in conversation analysis has been dedicated to determining whether those "unfinished" turns possess a specific or recurrent format, i.e., if there is a systematic breaking point within the syntactic construction of the abandoned turn. This may be related to the fact that withdrawal from the turn is above all negotiated on interactional grounds. In fact, current speakers may drop out of their turn if an interlocutor is initiating repair (Oloff 2009: 423–450; Sacks, Schegloff, and Jeffer-

those methodological questions (but see Schegloff 1988/89, 1992, 2002 for a conversation analytical view on those issues).





son 1974: 720), which shows that speakers are handling overlapping talk not simply with regard to its position or to speaking rights (“first speaker goes”), but merely with respect to sequential trajectories and *recipient design* (Schegloff 2000: 45). It thus seems difficult to put forward a specific formal rule for withdrawal from the turn, as speakers are handling overlapping talk locally, orienting to it as being non-problematic or as rather problematic and misplaced, even if the onset may in both cases be close to a TRP. In this chapter, I will first reflect on some syntactic properties of withdrawal (1.2.) and then argue for a multimodal approach to this phenomenon, considering the participation framework and *recipient design* (1.3.). The notion of *recipient design* (Sacks, Schegloff, and Jefferson 1974) reflects the idea that every turn is actionally, grammatically, and epistemically formatted with respect to a potential addressee. Considering Goodwin’s work (1980, 1981) on phrasal breaks at turn beginnings (which are related to the absence of the recipient’s gaze), gaze and, more generally, the availability of the recipient, seem to play a role for turn abandonment, too. In my analysis (2.), I will, therefore, comment on the use of multimodal resources (like gaze) and the interplay between turn withdrawal and recipient design.

1.2. Syntactic features of withdrawals from the turn

When looking at some examples of turn withdrawal in the literature, it seems to be appropriate to describe withdrawal from the turn in overlap first and foremost as a syntactic phenomenon, i.e., as the withdrawal from an emerging turn before it reaches a possible syntactic completion. In the following excerpts, the hyphen (“-”) marks the fragmentary character of a word and/or the premature retraction from its production, “a cut-off or self-interruption, often done with a glottal or dental stop” (Schegloff 2000: 61):

Ex. 1 (Schegloff 2000: 23)

- 1 Vic: Be[cuz] I’m deh en I’m gon’ ...
2 Mike: [Did] ju-]

Ex. 2 (Jefferson 2004: 45)

- 1 Essie: I think Cookie [ta-
2 Janet: [I didn’ even know’e was i::ll.

Ex. 3 (Schegloff 2000: 23)

- 1 Ava: [‘B’t asi]de fr’m that it’s a’right.
2 Bee: [So what-]





Withdrawal from the turn seems to imply the withdrawal from a TCU in progress before reaching its possible syntactic completion. For this reason, we might wonder whether there is some kind of systematic breaking point within overlapped syntactic constructions. Put differently, do participants drop out from the turn when having reached a certain point within the emergent syntactic construction? And, therefore, is the withdrawal basically linked to syntactic constraints? In the three examples, the withdrawal occurs clearly before the speakers have reached a syntactic or semantic completion of their turn. Nevertheless, this small sample also shows that those breaking points occur at different syntactic positions: before, during or after having produced a (finite) verb. Moreover, the rest of the utterance is still underspecified at those points and could take various forms, “simple” continuations of the TCU (ex. 1 *Did you* -> “see him”), or also the addition of subordinated constructions (*Did you* -> “notice that I’ve bought some milk”). This vagueness inherent to turn- and TCU-beginnings (at least in languages such as English, Schegloff, Ochs, and Thompson 1996: 27–32) is especially visible in the following example. Here, Janet self-selects twice at a TRP in Polly’s turn, but withdraws immediately after having produced only a fragmentary item (l. 2, 4). Finally, she self-selects again after the end of Polly’s turn (l. 7) in order to produce a complete turn:

Ex. 4 (Jefferson 2004: 46)

- 1 Polly: I jus’ thought it was so kind of stupid=
- 2 Janet: =[Y-
- 3 Polly: [[I didn’ even say anything=
- 4 Janet: =[Eh-
- 5 Polly: [[when I came ho:me.
- 6 (0.3)
- 7 Janet: Well Essie jus’ called ‘n I- an’ I aftuh call ‘er back...

We may notice that the fragments in l. 2 and 4 do not allow her interlocutor Polly (nor the analyst) to anticipate the format Janet finally adopts when speaking in the clear at l. 7, where she produces one complete TCU (*Well Essie jus’ called*), one fragment of a TCU (*‘n I-*), and finally, recycling the preceding fragment, a second complete TCU (*an’ I aftuh call ‘er back*). This case shows that withdrawals do not only occur clearly during the production of a TCU (i.e., after a TCU-beginning), but also at *turn-beginnings*, where speakers do not always start immediately with the production of a TCU, but may use pre-placed appositionals or various discourse markers (Lindström 2006; Schegloff 1987, 1996). As overlap resolution is organized “beat by beat” (Schegloff



2000: 45), withdrawal from the turn may in fact occur at any point within the syntactic construction of an overlapping turn, depending also on the position of the overlap onset within the overlapped turn(s). The *emic* nature of overlap resolution makes it difficult to apply a formal rule (e.g., withdrawal from the turn occurs systematically after or before the production of certain syntactic constituents), as speakers handle the management of simultaneous talk in a local way, adapting reflexively to the conduct of the other party involved. Those first observations hint at the fact that withdrawal may be more interactionally grounded than based on purely syntactic or grammatical constraints.²

1.3. A multimodal approach to withdrawals from the turn

The concept of multimodality within interactional approaches is based on the idea that the communicative process is to be understood as holistic and inseparable from the participants' bodies (Schmitt 2005: 18–22). In face-to-face interaction, the speakers' bodies are a resource for performing interactional tasks and socially relevant actions. Those bodily resources are not restricted to the use of speech, but concern also gesture, gaze, body posture, facial expression, the constellation of bodies in space, or the manipulation of artifacts. Inspired by context analysis (e.g. Kendon 1990; Scheflen 1972) and conversation analysis, multimodal analysis considers both sequentiality and simultaneity as fundamental to interaction and claims that *coordination* between participants is a permanent interactional task (Deppermann and Schmitt 2007; Schmitt 2007), a notion that refers not only to recipient design, but primarily to the fact that participants mutually perceive and monitor each other's audible and visible conduct, and displays of availability (Goodwin 1980). If not only audible, but also visible resources are taken into account, turn shapes and their syntactic formats can be understood as being embedded in embodied interactional practices. Grammar is thus a resource for and at the same time shaped by interactional processes (e.g. Auer 2002, 2005; Hakulinen and Selting 2005; Ochs, Schegloff, and Thompson 1996, Selting and Couper-Kuhlen 2000), which triggers reflections on the appropriate definition and grammatical description of spoken, actional units, their delimitations, or their (in)completeness (e.g. Hayashi 2003; Laursen 2005; Lerner 2002; Mondada 2004, 2007a; Olsher 2004; Schegloff 1984).

² Indeed, withdrawal from the turn (without being in overlap) can be a recipient designed practice used to handle delicate topics or actions, as has been shown by the study of unfinished turns in French (Chevalier 2005, 2008; Chevalier and Clift 2008).



As early as 1979, Goodwin and Schegloff commented on the link between syntax and interaction. Goodwin (1979) shows how a speaker formats a syntactic construction in specific ways according to his designed recipients. Furthermore, he shows that phrasal breaks during utterance beginnings in face-to-face interactions are linked to the absence of the recipient's gaze toward the speaker (Goodwin 1980, 1981). As soon as a mutual gaze between both participants is established, the speaker continues his syntactic construction without further perturbation. Schegloff (1979) insists on the use of phrasal breaks for topic management in conversation: when speakers initiate a new topic or a topic shift, they frequently do so with a phrasal break. If the turn that introduces the new topic contains no perturbation, the recipient is likely to initiate repair in the next position (see also Drew 1997). Syntactic constructions are, therefore, sensitive to recipient design and, more specifically, to the availability displayed by co-participants within a given participation framework (Goodwin 1997; Goodwin and Goodwin 2004). Withdrawal from the turn could thus not only be linked to the emergent syntactic construction of the abandoned or of the continued concurrent turn, but also to actional features of the concurrent lines of action, and to the availability of the recipient(s) for the overlapping speaker(s). Consequently, I suggest adopting a multimodal approach to the phenomenon of withdrawal from the turn. Instead of investigating withdrawal as a primarily syntactic phenomenon, I will look at the speakers' audible and visible conduct during simultaneous talk and show the possible interplay between recipient availability and withdrawals from the turn.

1.4. Data

The data used in this chapter are naturally occurring interactions in French and German.³ Whereas the German corpus (3 hours) – an informal dinner conversation between German students in France with a number of participants varying from three to eight participants – has been collected by the author, the French data have been provided by the ICAR research lab (partly available on the CLAPI database of spoken French, <http://clapi.univ-lyon2.fr/>). The Saxe corpus (1,5 hours) is an informal work meeting at home

³ The decision to take French and German examples has not a comparative, but an illustrative objective. The data show that the practice of turn withdrawal exists in both languages, and that both French and German speakers are sensitive to issues of recipient availability. For a comparative study, the data set would have to be considerably enlarged (see section 3.).



between three colleagues working for a marketing start-up, and the MO-SAIC corpus (1,5 hours) shows three architects during a work meeting, where they exchange the last updates on a common project. All data have been videotaped using at least two cameras, the multimodal transcripts (see transcription conventions at the end of this chapter) being based on all available views of the recordings.

2. The contribution of video data to the analysis of withdrawals from the turn

When working with audio data, two different withdrawal positions with respect to the development of the turn can be identified. First, there can be a withdrawal from the turn before TCU-completion; in this case, the turn is already in progress and the TCU remains clearly syntactically incomplete, although its completion may be more or less projectable.⁴ Second, the withdrawal can be positioned at an even earlier point, during (ex. 4) or immediately after the turn-beginning. In this case, the withdrawal occurs after having audibly produced the beginning of an utterance. However, the production of a fragment or of an utterance framing discourse marker gives no precise hint about the syntactic construction that may follow. In her account of pointing gestures used by speakers in order to prepare a turn, Mondada (2004, 2007a) shows how the use of video data may add yet another observable sequential position when studying the incremental development of turns in interaction, the *pre-beginning*. The following excerpt illustrates that a third possible sequential position – during or immediately after the pre-beginning – exists also in the case of withdrawal from the turn. Here, in overlap with Jean-Baptiste's (JEB) turn (l. 1), Sophie (SOP) produces a tongue click at low volume, followed by an in-breath (l. 2). She then utters a minimal response (*ouais*, l. 4.), which seems to be a response to Jean-Baptiste's question in line 1 and 3. From the verbal transcript, it cannot be decided if the tongue click and the in-breath are indeed a preparation for Sophie's response to Jean-Baptiste. But by looking at the visible resources deployed by the speakers, we may interpret the audible pre-beginning in line 2 and the response in line 4 as two different turns, the first incipient turn having been visibly abandoned by Sophie:

⁴ Video data also show that syntactically incomplete turns may be completed by gestures (*embodied completions*, Olsher 2004), or that gestures can be used in order to link distant TCUs and may project increments of apparently complete turns (Laursen 2005).



Ex. 5 SAXE_ms_012327_mais t'en as qu'un

```

1 JEB [ah pour]quoi*tOI# [*t'achètes* la ]# qualité&
tra [oh why ] you*DO # [*buy qua*lity]#products&
2 SOP > [°°.mts°°] [* .h:::~:]
jeb >-gaze SOP----->>
sop >>gaze FAB-----*.....*gaze JEB----->>
                                     *...opens mouth----->
fig. #1 #2
3 JEB &pour ton gamin:/*=
tra &for your kid *=
sop >-open mouth-----*
4 SOP =ouais
tra =yeah

```

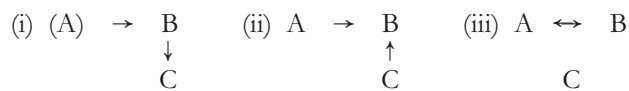


It can be noticed that before starting her turn, Sophie gazes at Fabien (fig. #1). After her tongue click, she visibly opens her mouth, projecting an incipient turn beginning also in visible ways (fig. #1, resembling the “a”-face described by Streeck and Hartge 1992), which is confirmed by the following deep in-breath. However, while breathing in, Sophie already turns her head in Jean-Baptiste’s direction, having engaged in mutual gaze just after having stopped her in-breath (fig. #2). She maintains the gaze in his direction beyond the end of the excerpt, which shows that her minimal response (l. 4) is addressed to Jean-Baptiste. It is thus mainly Sophie’s visible reorientation of gaze which shows that she indeed withdraws from an incipient turn addressed to Fabien in order to respond to Jean-Baptiste. This illustrates that withdrawal from the turn is not only an audible phenomenon (although it often implies the production of an “incomplete” utterance), but primarily embodied and implemented by visible resources as well. In this sequential position, it seems particularly difficult to grasp the phenomenon of withdrawal in syntactic terms – at least if syntax is considered as the production of at least one constituent of an emergent syntactic construction.

As excerpt 5 has shown, withdrawal from the turn in overlap seems to be closely linked to a modification of the participation framework. Although we may analyze a certain amount of withdrawals from the turn in overlap by fal-



ling exclusively back upon sequential features of the talk (e.g., action trajectories, speaking rights, repair initiation, etc.), the use of video data can account for supplementary features of this practice. On the one hand, the withdrawal in pre-beginning position becomes describable as an interactional phenomenon, which is not always possible or clear-cut if we rely only on audible resources for the analysis. On the other hand, video data help to shed new light on the relevance of the participation framework for negotiating turn-continuation or its abandonment. Notably, although Schegloff (2000) does not present video data, he already makes some useful remarks along these lines. He details the possible configurations of overlapping talk, which he schematizes in the following way (where each letter, A, B, C, corresponds to a speaker, whereas the arrows represent a turn addressed to the respective interlocutor):



Scheme 1: Basic overlap configurations (Schegloff 2000: 8)

Schegloff suggests that simultaneous talk between two speakers occurs exclusively in one of these three configurations, and he offers some basic observations on these configurations that include also some remarks on visible resources. In configuration (ii), where A and C are simultaneously addressing B, Schegloff (2000: 8) proposes that B's gaze direction will be decisive for deciding if either C or A will continue. The interesting idea that withdrawal from the turn may be related to displays of recipientship by the potential addressee is not further developed. On the contrary, Schegloff (2000: 8) underlines that

[a]lthough almost certainly the body can be deployed in a manner relevant to overlap in configuration (iii) [where A and B are addressing each other at the same time, FO] it does not appear to figure so centrally in that circumstance.

Schegloff does not comment on any particular role of visible resources in configuration (i) – where A is talking to B while B is simultaneously talking to C – although the parenthesis placed around speaker A hints at an underlying interpretation of A withdrawing systematically from the turn as B is unavailable as a recipient (interpretation that can be easily refuted if we think of repair sequences that are initiated in overlap, where a current speaker B can quickly suspend his line of action in order to respond to A). Schegloff's remarks point to the relevance of a multimodal approach to turn withdrawals.





However, he does not give any details on the way in which visible resources and turn withdrawals interact during simultaneous talk. In order to consolidate his preliminary observations, I will now show some cases where withdrawal from the turn is clearly related to displays of reciprocity.

2.1. Withdrawal from the turn when gazing at a non-gazing recipient

As Schegloff (2000) remarks, a speaker is likely to drop out from the turn when he notices that his recipient is not available at that moment. The following excerpts will confirm his observations, though it does not seem to depend on the overlap configurations suggested by Schegloff, but seems to be primarily a matter of displayed availability and its dynamic development during overlapping talk. While Schegloff's configurations may be useful for a first approach to overlapping talk, their schematized representation reflects only weakly the interactional dynamics of turn-taking and reciprocity. Instead of commenting further on those configurations, I will now underline the relevance of the recipient's availability for turn abandonment.

In excerpt 6, the three French colleagues are discussing a new marketing concept for biscuits, and how to extend the range of already existing biscuits for children. The excerpt starts at a point where Jean-Baptiste (JEB) has just stated that the brand *Marquise* is the market leader in biscuits. Along these lines, Fabien (FAB) is asking if high quality biscuits for kids do already exist (l. 1–2). This question receives two answers, the first is initiated by Jean-Baptiste, who, after a first negative answer (*non*, l. 4) initiates a new, complex turn in which he states once again that the brand *Marquise* is the highest quality product on the market and thus the most luxurious one (l. 4, 7, 9, 13). Sophie (SOP) simply confirms that *Marquise* is the best brand (l. 5–6), while Fabien, in a short overlapping turn, observes that high quality biscuits for children do not exist (l. 8). Although Sophie has aligned with her colleagues' reasoning, she then suggests a possible alternative, the biscuit brand *Crocta* (l. 12), which may also offer high quality biscuits for children. As we can see in the transcript, Sophie's second proposal is developed in overlap with Jean-Baptiste's complex turn, and even if she reinitiates the turn twice (l. 17), she does not complete it. Jean-Baptiste's overlapping competitive talk (l. 18) could account for her dropping out (Schegloff 2000). However, the multimodal analysis reveals that Sophie is withdrawing from each TCU shortly after having directed her gaze to a non-gazing recipient:



Withdrawal from turns in overlap and participation

217

Ex. 6 SAXE_ms_012739_crocta

```

1 FAB  donc\ pour EN+fants:/ est-ce que t`as+du haut
tra    so    for chIl+dren: do    you have+ high
sop    >>SOP flicks through a brochure, looks at it-->
jeb    +...gaze FAB-----+.broch. SOP>
2      d` gamme pour enfants/ °x(x)\°
tra    quality for children °x(x)°
3      (0.2)
4 JEB  non (j` veux dire)[*qu`t` [As\ ]
tra    no (i 'd say) [*that you [HAvE]
5 SOP  [*h::: [ben ]*c'est&
tra    [*h::: [well]*it's &
sop    >---gaze table-----*.....*gaze FAB->
6 SOP  &ma[rqui[se/]'
tra    &ma[rqui[se/]'
7 JEB  [<par co[ntre c'est ce]'rtAI]*N/ ((f))>&
tra    [<on the[other hand what's cer]' tAI]*N ((f))>&
8 FAB  [j:amais]' vu/]
tra    [never]' seen]
sop    >--gaze FAB-----*,,,
9 JEB  &[(.) c'est] que+ marquisef du+coup// (0.8) est:/
tra    &[(.) is ]that+ marquisefthere+fore (0.8) is:
10 jeb  [((biro))]
jeb    +...gaze FAB----+,,,
fab    >--gaze SOP-----f.....gaze twd JEB----->
11      (.)
12 SOP > *cro*c[f+ta/]
13 JEB  [f+le haut/] d`gamme\#
tra    [f+the best] product #
sop    *...*gaze FAB----->
fab    >-----f..gaze JEB----->
jeb    +..gaze FAB----->
fig.    #1
14      (0.2)
15 FAB  °ouais/°*
tra    °yeah° *
sop    >-----*,,,
16      (.)
17 SOP > croc*tA/+doit*ê t` (coin-)*cr[foc*ta/ il est- ]f
tra    croc*tA/+must*be (coin-)*cr[foc*ta it is- ]f
18 JEB  [f i*1 a TOUS LES]f&
tra    [f i*t has ALL THE]f&
sop    *.....*gaze FAB---*.....*gaze JEB----->
fab    >gaze JEB-----f..gaze SOP-----f
jeb    >--FAB---+,,,gaze table
fig.    #2
19 JEB  &#fCO*DES de celui# que: [+tu] (emmèneras)
tra    &#fCO*DES of the one# that [+you](will take)
20 FAB  [+hm:]
sop    >-JEB*..gaze table----->>

```





218

Florence Oloff

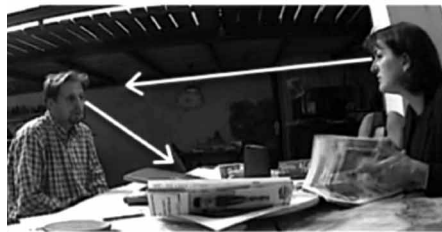
```
fab      £...gaze JEB----->>
jeb      #3                               #4      +...gaze FAB-->>
```

Sophie's gaze towards Fabien shows that her first answer (l. 5–6) as well as her second proposal *Crocta* (l. 12, fig. #1) are both addressed to Fabien. At the moment of her second suggestion, Fabien has already started to gaze in Jean-Baptiste's direction (l. 9, fig. #1). Thus, Fabien is not available as a recipient for Sophie at that moment. The fact that she reinitiates her turn (l. 17) after Fabien's minimal answer to Jean-Baptiste (*ouais* / °, l. 15) shows that she orients to a possible closing of the sequence initiated by Jean-Baptiste. This reinitialization (l. 17) shows that the name of the brand *Crocta* (l. 12) was more a turn-beginning than a complete turn. Sophie redirects her gaze from the brochure on the table in front of her to Fabien (fig. #2) shortly after having repeated the brand name *Crocta* (l. 17). But when her eyes are fully opened (*doit être*, l. 17), Fabien is still looking at Jean-Baptiste. Sophie does not complete the TCU (*crocta doit être*), but stops the emerging syntactic construction with an unidentifiable incomplete word (*coin*), probably an adjective).



Immediately afterwards, Sophie turns her head to the right and gazes at Jean-Baptiste, while restarting her turn for a second time (*crocta il est*, l. 17). But again, she encounters a non-gazing recipient, as Jean-Baptiste is looking at the table in front of him (fig. #3). He does not modify the orientation of his gaze, and three syllables later, Sophie withdraws from the turn, turning her gaze back to the brochure lying in front of her (fig. #4), and abandoning her turn in a definitive way (Oloff 2010). As Sophie has turned away from Fabien after having dropped out a first time, she cannot see that Fabien briefly orients to her (l. 18) when she reinitiates her turn for Jean-Baptiste. Sophie does not restart her turn after three trials without having been able to engage in mutual gaze with one of her interlocutors.





3



4

This example shows that in multi-party interactions (i.e., involving at least three participants), a speaker may try, after having encountered a first non-gazing recipient, to catch the attention of another recipient. Speakers orient to the displayed availability of their recipient in order to decide whether to continue a turn in progress or to suspend its production, possibly restarting it at a later point. The next example shows that speakers do not necessarily gaze at each other when they start to speak, but that it seems necessary to obtain the recipient's gaze at a certain point during a turn or a sequence (Goodwin 1981; Rossano, Brown, and Levinson 2009). Here, six German friends are having a raclette dinner at Isabelle's (ISA) place. While they are still eating, one of the participants sitting at the table, Christian (CHR), is gazing at the table in front of him. A few moments later, he self-selects in order to ask Dennis (DEN), sitting opposite him, if he is using the blue glass (hidden by JAN's head) in front of his plate (l. 7–8). In overlap with this turn, Isabelle and Dennis simultaneously self-select (l. 9–10). Both Christian and Isabelle drop out of their turn, even though they have initiated their turns significantly earlier than Dennis (cf. Isabelle's projective *mhm*, l. 5):

Ex. 7 RAC_po1_005656_trinksch nich aus dem glas

```

1      (0.4)$(0.6)
   chr      $...gaze table / glass-->1.10
2  DEN      °mhm/hm:\°
3      (1.1) [(0.2)
4          [((background music: beginning of new song))>
5  ISA      mhm:\
6          $(.)
   chr      $...starts pointing twd DEN's glass->
7  CHR > #ahm defnnis/&
   tra      #er defnnis/&
   den      f....>
   fig.      #1
8  CHR      &trin[fksch      nich $#      aus $      (de)m]
   tra      &you [fdon't      drink$#      from $      th(is)]
9  ISA >      [f>>(weiß nich      ob$#'s      $'n )<<\ ]
   tra      [f>>(don't know if$#there's$a-)<< ]

```





220

Florence Oloff

```

10 DEN      [fAH::$/      J]*A:: $f*°::\°*#&
tra      [fOH::$/      YE]*AH:: $f*°::\°*#&
isa      >gaze plate-----*.DEN*
                                     *.smiles-*
den      >.....f-----gaze ISA-----f.gaze plate>
chr      >gaze glass-----$......$gaze DEN----->
                                     >.....pppp DEN's glass-----$,,,retracts>
fig.      #2      #3
11 FAB      &*°ah\hm:/°=
isa      *...gaze plate->
12 ISA      = [<hörst $du:// (.) [f$ sch#ö$ n oder/ ((©))>]
tra      = [<do you$hear (.) [f$ n#i$ce isn't it(©))>]
13 CHR >    [° .mtk°      [f$ tr#inksch$nich aus (de)m]
tra      [f$you#don't $drink from th(is)]
chr      >.hold--$..extends fully.....$pppp DEN's glass->
                                     >gaze DEN-----$.gaze DEN's glass----->
den      >gaze down/plate-----f...gaze CHR----->
fig.      #4
14 CHR      &fglas/f
tra      &fglassf
den      >-CHR--f,,,
                                     f....>
15 DEN      NEINEfIN/fdas is$: \ kannst du neh[men\
tra      NO NfO fthat is$: you can ta[ke it
den      >....fpppf,,,,,
chr      >pppppppppppppppp$...RH moves twd glass->>

```

Christian initiates his turn (l. 7) without searching for Dennis' gaze (fig.#1). Instead, he seems to focus mainly on the blue glass that he starts pointing at during his turn (fig. #2). Only later in his turn (*aus (de)m*) does he direct his gaze to Dennis, looking at him on the last syllable of his emergent turn ((*de)m*), fig. #3). But they cannot engage in mutual gaze, as Dennis has already turned his head to the host Isabelle when he started his overlapping turn (his loud *AH::\$/ JA::°::\°*, l. 10, in fact recognizes and assesses the song which has started playing some seconds before, l. 4). We can observe that Christian withdraws from the turn only one syllable after having noticed the unavailability of his addressee, although only one lexical item seems to be missed in order to complete syntactically the turn (*aus (de)m* --> *glas*, cf. l. 13–14).

Christian also slightly retracts his fully extended pointing gesture (fig. #3, l. 10). Interestingly, he holds his gaze toward Dennis until they engage in a mutual, although minimal gaze (even if Christian anticipates the repeat of his turn well before, cf. his turn-beginning tongue click, l. 13, and the re-extension of his pointing gesture during Isabelle's *hörst du://*, l. 12). Precisely when Dennis is starting to look in Christian's direction, Christian is lowering his





gaze again to the glass, so that they exchange a brief mutual glance (fig. #4). He now completes his turn with the missing complement (*glas*) which has been projected by *aus (de)m* (l. 13–14). We may note that the overlapping talk by Isabelle does not disturb the repeat and completion of Christian's turn, which is of course related to the fact that Isabelle is not the recipient of this turn and that the recipient Christian addresses is now available.



Unlike Christian, Isabelle does not recycle her abandoned turn (l. 9). Both the unspecific format of her turn (*weiß nicht ob's 'n*), and the absence of a gaze to a participant (she is looking down at her plate while talking) make it impossible to know to whom in particular this turn is addressed. In her case, she does not withdraw from the turn after having noticed that no recipient was gazing at her. Nevertheless, it is worth noting that Isabelle's orientation is quickly changing precisely after the breaking off of her TCU: she quickly looks up to Dennis and starts smiling shortly afterwards (fig. #3). Her fol-





lowing assessment concerning Dennis' loud turn (l. 12) is already uttered with her gaze turned back to the plate. In this case, too, the mutual gaze between the two participants is restricted to a short glance in a mid-sequence position.

Both example 6 and 7 show that there is an interesting interaction between displays of reciprocity and withdrawals from the turn: when looking at a non-gazing recipient, speakers drop quickly out from the turn, the turn completion being linked to the establishment of at least a short lasting mutual gaze. Speakers who are addressed by a recipient they are not gazing at may change gaze direction immediately after the withdrawal from the turn, engaging in mutual gaze with the overlapping speaker. In both cases, withdrawal is closely timed with respect to gaze. In order to give more empirical evidence for this observation, we shall look at yet another instance of changing gaze orientation linked to a drop out, i.e., a case where the withdrawal occurs shortly after mutual gaze between the speaker and his addressee has ceased.

2.2. Withdrawal of reciprocity leading to withdrawal from the turn

If overlapping speakers already have established a mutual gaze, the display of availability may not be likely to intervene in the negotiation of turn continuation or abandonment. Nevertheless, one of the speakers may use the practice of turning to another participant, thus withdrawing reciprocity from his overlapping co-participant. It is this *recipient-shift* (Lerner 2003), in this case the re-orientation of one speaker toward an alternative addressee and, therefore, the loss of his recipient, that may lead an overlapping speaker to withdraw from the turn, as the following example will illustrate.

Preceding the excerpt, Manuela (MAN) has asked Isabelle (ISA) about her work, and how she generally handles video recorded data. Isabelle develops a complex turn in which she mentions first the digitization of the data (l. 1–2, 4–6). By the use of ›first‹ (*erstmal*, l. 2), she projects at least a second step to come and thus a continuation of her turn. Nevertheless, Isabelle stops after having given more details on the digitization process (i.e., import the data in order to obtain a mov file, l. 4–6), taking a long in-breath followed by a short pause (l. 7). She continues her turn as Dennis (DEN) self-selects simultaneously (l. 8–9). Dennis drops out before completing his TCU (*d(u) mUsst DANN DA::*, l. 9), leaving the floor to Isabelle. This fight for the floor is obviously due to a negotiation of the transition-relevance place (as an intra- or inter-turn TRP, cf. Lerner 1996) between Dennis and Isabelle, which is also indicated by the competitive, high volume of both overlapping turns.



The multimodal annotations of the transcript reveal that Isabelle changes her gaze direction during the simultaneous talk, turning her head from Dennis to Manuela. The withdrawal from the turn thus occurs shortly after this visible re-orientation of the withdrawer's recipient:

Ex. 8 RAC_po1_000642_movdatei

```

1 ISA .h: ABER/ ähm: (.) .h gut un:d wenn das jetzt so
tra .h: BUT er: (.) .h well an:d once this is PRT PRT
2 fertig is/dann werd ich die erstmal digitalisieren/=
tra finished then i will first digitize them=
3 MAN =mhm[:°:/°]
4 ISA [die] f[schliess ich dann halt an 'n computer f an/
tra [then]fi will connect it PRT to a computer f PVS
den f[nod-----f
5 ISA .h: un' da fhab ich nachherf so 'n: so 'n: m+ov:/
tra .h: an' thenflater i will havefa kind of: kind of m+ov:
den f°nod-----°f
man >>leaned forward, gaze ISA-----+,,,>
6 ISA (.)+öh 'ne mofvdatei/+f
tra (.)+er a mofv file +f
man ,,,,+,leans back,,,,+
den f[nod---f
7 ISA *+<.h:: (0.4)>(0.2)+
isa *...gaze DEN----->
man <+.....arm+((listening posture))>
den >>gaze ISA----->
8 ISA [# unf(d) DANN] [GU*CK fICH MIR DAS#AL]LES fA:N*:/&
tra [# anf(d) THEN] [I *WILL fWATCH #ALL] OF ITfSVP* &
9 DEN > [#d(u)f mÜsst] [DA*NN f DA::#:\]
tra [#y(ou)f thEn] [HA*VE TOf PRT::#:]
isa >gaze DEN-----*.....gaze MAN-----*..>
*rhythmical head movement-----*
den f.....f-square gesture-----f.....>
fig #1 #2
10 ISA &>>oder beziehungs(wei)<< hör mir das (.)*kurz an/ und&
tra & >>or rath(er)<< i will listen(.) *PRT to it and&
isa >...-gaze DEN-----*,,,,,
den >..scratches throat->>
    
```

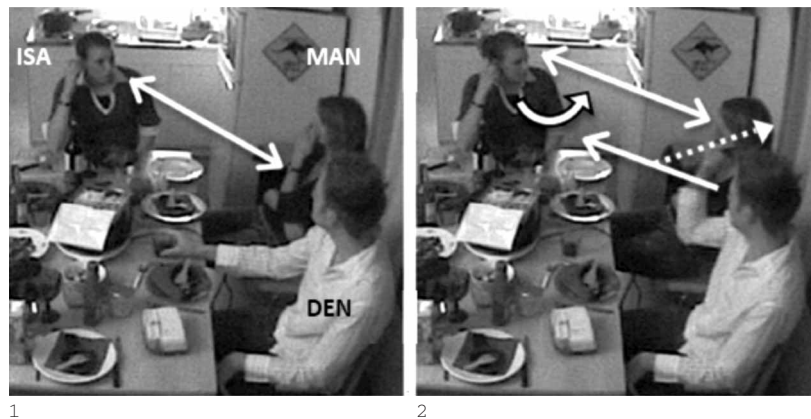
During the beginning of the excerpt, both Manuela and Dennis adopt a recipient position and intervene only minimally (Manuela in l. 3, Dennis nods several times, l. 4–6). During the pause (l. 7), Isabelle starts to gaze at Dennis who is already looking at her since the beginning of the excerpt. While the rising intonation of *movdatei* and the in-breath project a continuation of Isabelle's explanation, Dennis seems to interpret this moment as a possibility to self-select and to contribute in a more essential way to her complex turn. In-





deed, both speakers continue in a similar vein, as both use the adverb *dann* (>then<) within their turn and start the description of a possible second step of handling video recorded data.

While the beginnings of the overlapping turns are uttered at a normal volume, both upgrade quickly to competition. The mutual gaze at overlap onset (fig. #1) is dissolved by Isabelle's reorientation: during the third syllable of her turn (*GUCK*, l. 8), she turns her head with rhythmical nods to the left and thus looks at Manuela (fig. #2). Not even two syllables later, Dennis drops out of the turn, leaving his TCU syntactically incomplete (the complement of *d(u) mUsst* / >you have to<, l. 9 --> "do x" is left out). Shortly afterwards, Dennis also retracts his gesturing right hand – which has been tracing a square in the air – to his throat (l. 8–10). Just after overlap resolution, Isabelle reduces the volume of her voice to a normal level and redirects her gaze to Dennis, which shows that her display of unavailability has indeed been used as a resource for overlap resolution.



This case shows that speakers in overlap are sensitive toward the availability of their recipient. As long as a mutual gaze between the two speakers is maintained, the overlapping talk may persist. But if one of the speakers withdraws his gaze from his overlapping co-participant and turns to another possible addressee, the receiver of this recipient-shift ceases the production of his turn very quickly. The display of unavailability functions as a resource for overlap resolution, and the speaker who orients to a "new" recipient may, therefore, continue his turn in the clear. The timing of the withdrawal appears to be precisely positioned with respect to the interlocutor's re-orientation: one or two items after the beginning of the re-orientation toward another recipient, the former addressee drops out. In the examples examined





up to now, the abandonment of one of the overlapping turns does not seem to be grounded on syntactic criteria, but related to the dynamics of the participation framework. Nevertheless, the use of gaze as a resource for overlap management presupposes that the participants are actually looking at each other or seeking their recipient's gaze. This may not be the case for all occurrences of overlapping talk, as a last example will illustrate.

2.3. Withdrawal from the turn when no participant gazes

We might wonder if gaze is always relevant for overlap resolution and for interaction in general. As Rossano, Brown, and Levinson (2009) point out, gaze in interaction is not directly related to the systematics of turn-taking and should not exclusively be understood as a device for reciprocity. Speakers do not gaze continuously at each other (Goodwin 1980): sometimes, participants are not seeking another's gaze during a whole sequence or even for longer spans of talk: face-to-face interaction without gaze occurs, whether it is due to a specific cultural setting (Rossano, Brown, and Levinson 2009), or to a specific spatial arrangement of the bodies (Mondada 2008), or in case of telephone conversations or other types of mediated interaction (Mondada 2007c). Participants might also gaze less often at each other in settings with specific interactional ecologies (Goodwin 2000; Mondada *tbp*), where mutual gaze is less relevant, for example, because participants are handling specific artifacts that are the main visual focus of attention.

This is the case of the following excerpt, where three architects are discussing the consequences of the removal of a lift and a staircase for the light conditions of a large restaurant room on the ground floor of a castle that is being transformed into a hotel. At the beginning of the excerpt, Laurent (LAU) is arguing once more for the maintaining of the lift and the stairs in order to keep a visual contact with the outside (l. 1–4), whereas Cédric (CED) underlines that their customer wishes to have more rooms and more space, which is reason for removing those two elements (l. 6). During this turn, Marie (MAR) self-selects in overlap, but withdraws very quickly (l. 7). She recycles her turn-beginning twice, but withdraws each time from the turn (l. 9, l. 14). The participants are mostly looking at the plans spread out on the table. Therefore, the withdrawal or continuation of a turn in overlap cannot be linked to the absence or presence of mutual gaze between the interlocutors. First, Cédric's counterargument (the customer's need for space, l. 6) and the ensuing relevance of a response from Laurent could of course account for Marie dropping out, as she orients to Laurent's speaking rights at this moment and the ongoing sequence between him and Cédric (l. 9–14).





226

Florence Oloff

But a look at the participant's pointing gestures as they are deployed in the common interactional space of the table and the plans hints at an interesting connection with Marie's withdrawals:

Ex. 9 mosaic_122900_moi je

```

1 LAU  &.h:: et j` trouvais ça      bIEn: d'avoir ce::\ .h
tra   &.h:: and i thought it was gOOd to have this:: .h
2      cette perspective/ euh::\ `fin c:- de vOIr le:s
tra   this view          er:: well i:- to sEE the:
3      issues de s`cours/ >>et comment tu remontes<<
tra   emergency exits    >>and how you go back up<<
4      jusqu'à:: °l'extéFrieur°
tra   to:: °the outFside°
lau   F--RH on table->
5      (1.1)
6 CED  oui\ mafis(h) f(h)ace [à son*be[soin d`#F] plafce\
tra   yes buft(h) in view of[his *ne[ed for #F] spafce
7 MAR > [tsh [moi:/ #F]
tra   [tsh [i: #F]
lau   >-----f...RH moves twd plan.....f-plan-f.RH>
mar   >gaze plan----->
ced   >RH on table-----*lifts RH with biro...>
fig.  >gaze plan----->
8      (#)
9 MAR > [*m:oi/ F*#+je::°::° ]
tra   [*i: F*#+i::°::° ]
10 LAU [*et là/ F*#+on l'aFvait ic]i ave*c F#
tra   [*and thEref*#+we had fit he]re wi*thf#
lau   >.....f-pp plan---fbiro circle----f
mar   .*.....down*,,,,,,,,,,,,,,,,,,,,,*RH to mouth->
ced   >plan-----+...gaze MAR----->
fig.  >RH touches chin----->
11 LAU F l'ascen+seuFr/ on l'avait favec l'escal+ier/
tra   F the li+fft we had it fwiththe stair+s
lau   F.RH down..f--biro circle---f,,RH to the edge of plan->
ced   >gaze MAR+,,gaze plan----->
12 LAU o[n l'aura ][*F+plus ]=
tra   w[e won't have it][*F+anymore]=
13 CED > [moi j` suis p-][*F+j` suis][pas s:-#F]
tra   [i i'm n-][*F+i'm ][not s:-#F]
14 MAR > = [moi/ je#F]+*j+e\*
tra   = [i i #F]+*i+: *
lau   >,,RH to edge of planf,,retracts forearmf.RH to head>
mar   >RH in front of mouth*...RH twd plan.....*ppppp*,,,
ced   >..leans forward...+RH with biro twd plan+pp+,,,
    
```



```

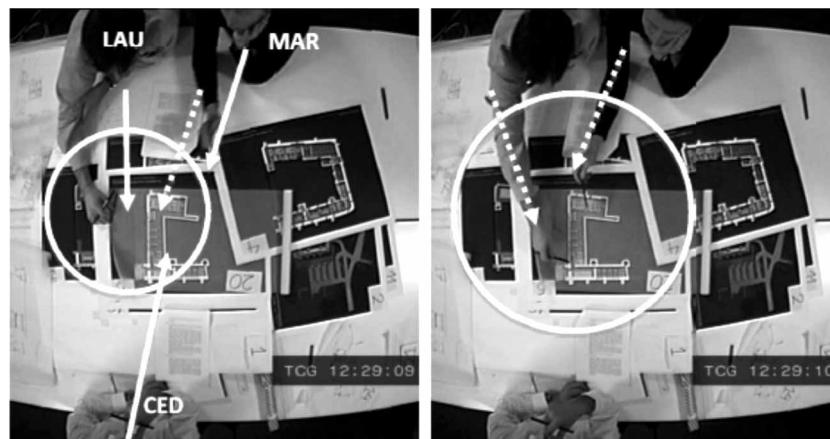
fig.                                +...opens biro                                #4
15      *(0.1)*(0.2)
mar      *...gaze CED->
          *..smiles->
16 MAR   °mhm+:+*:*° <h*::[: (laugh)>
17 CED   [vas-y+
tra      [go on+
mar      ,,,RH*mouth*,,positions RH to the right of her face
ced      >---+...gaze MAR-----+,,,gaze down-->
          ,,,,+RH on LH-----+
          >leaned fwd-----+straightens up-->
18      (0.1)      *(0.1)
mar      >gaze CED*,,gaze plan->
19 MAR > *h: moi j'aurais plu*+tôt envie+de:: quand*
tra      *h: i i would ra*+ther like+to:: when *
mar      *...lowers RH.....*...ppp to plan ppppppp*
ced      >...straightens up-----+ "listening posture"
          >gaze down-----+..gaze plans----->
20      *j'arrive:/ euh\ .h de voir/ que j'ai::
tra      *i'm going in er .h to see that i have:
mar      *traces lines----->
20      j'ai un rapport à:: l'ex*térieur:
tra      i have a connection to:: the out*side
mar      >--gaze plans-----*
22      (0.2)
23 CED   oui/
tra      yes

```

As Laurent is explaining his point of view (l. 1–4), his right hand is moving to different points on the plan in front of him. He retracts his gesturing hand at the end of his turn and puts down his right forearm on the table during the last word of his turn (°*extérieur* °, l. 4). He does not change position during the following long pause (l. 5), and his movement back to the plan (l. 6) is due to the fact that the transparent paper on the black plan starts rolling up as Cédric starts his turn (l. 6). His right hand is thus visibly directed to the outer edge of the plan in order to prevent it from rolling up again (fig. #1). Therefore, he projects no turn continuation at that moment. The timing of Marie's first self-selection (l. 7) is thus positioned at a TRP, as is also shown by the format she adopts for a turn-beginning (a tongue click and the French tonic pronoun *moi*), anticipating the end of Cédric's turn, which could be possibly complete after *son besoin* (filling the slot of the obligatory complement of the preposition *face à* -> x / in view of -> x). The post-overlap recycling of Marie's turn beginning (Schegloff 1987) at l. 9 aims at reinstalling the relevance of her turn in a first sequential position after the overlap resolution (Oloff 2009: 148–153).



However, Marie drops out after having repeated her turn-beginning (l. 9), although Laurent does not adopt a particularly competitive format (l. 10) and both are gazing at the plan, neither at each other nor toward Cédric. Instead, Laurent's new pointing gesture toward the plan at that moment could account for Marie's drop out, as he is investigating the space Marie prepares pointing at as well. As we can see in fig. #1 (l. 6–7), Marie starts to move her right hand – which has been in a resting position on the table until that moment – toward the plan during her turn-beginning. Her hand reaches the edge of the plan and is lowered in preparation for a possible pointing gesture with the biro when she starts to repeat her turn-beginning (l. 9). At that moment, Laurent has already made a quick movement with his right hand toward the plan (movement which began on the last syllable of Cédric's turn, *place*, l. 6). His arm reaches full extension after *et là*/ (l. 10, fig. #2). Precisely at that moment, Marie starts to withdraw her right arm from the plan and positions it in front of her mouth (fig. #3).



1

2

Marie drops out of the turn one syllable after Laurent's right arm has reached full extension. Thus, her withdrawal seems to be linked to the simultaneous investigation of the "pointing space" on the table by Laurent. In anticipation of a possible turn completion (l. 11), Cédric then starts a new turn (l. 13) that is overlapped by Laurent's conclusion (*on l'aura plus*, l. 12), preceded by a change in his body posture and a movement of his right hand toward the plan. The break in his turn (*j' suis p-*, l. 13) could be related both to the fact that he is opening his biro and to the anticipation of a next TRP in Laurent's turn. Precisely at the end of Laurent's utterance, Marie reinitiates her turn





(l. 14). During the pre-beginning, she has moved her right hand from the stand-by position in front of her mouth back toward the plan. Cédric and Marie are, therefore, moving their right arm simultaneously toward the middle of the table, pointing to relatively close points on the plan (fig. #4). As both gaze at the plan, we can assume a mutual visibility of the gestures, which reach their full extension at the same time. Although they do not point to the same space on the plan, Cédric retracts his arm nearly immediately (0.1 seconds) after overlap resolution. Interestingly, Marie also retracts her right arm just a fraction of a second later, after having recycled the pronoun *je* (l. 14). Due to limited space, I will not comment on the ensuing explicit negotiation of Marie's speaking rights and the participants' use of gaze (l. 15–23).



In this example, the three participants have to negotiate between the continuation of an ongoing sequence (between Laurent and Cédric) and the start of a new sequence, initiated by Marie. The multimodal transcript shows that both Marie's self-selections and withdrawals are done without gazing at one of her potential addressees. In this setting, the use of gaze and the visual check of the recipient's availability do not seem to have primordial relevance. Instead, the participants focus on the plan in the middle of the table to which they frequently point during their turns. The specific ecology of this setting results in a particular importance of the occurrence and visibility of pointing gestures or hand movements above and toward the plan. With regard to withdrawals from the turn, it can be said that the relevant feature here seems to be the availability of the pointing space. If, especially during overlapping talk, another speaker simultaneously enters this central space in the middle





of the table (typically by extending his arm, eventually holding a biro), the availability of the space and thus of the potential recipients is compromised. Consequently, even if withdrawal from the turn in overlapping talk cannot be systematically related to the absence or presence of (mutual) gaze, it is, at least in face-to-face interaction, clearly managed by the use not only of audible, but also of visible resources.

3. Conclusion

This chapter has been dedicated to the analysis of withdrawals from the turn during simultaneous talk, a phenomenon that has not yet been investigated in detail within conversation analysis and interactional linguistics. It does not seem possible to account for turn withdrawal exclusively in syntactic terms, as the breaking points of the emergent syntactic constructions appear to be at various distances with respect to a possible TCU completion and to the length of the overlapping talk. Those breaking points may be more usefully distinguished with regard to different sequential positions within an utterance: the pre-beginning of the turn, the turn-beginning, and the TCU, after the turn-beginning. As the examples in section 2 have shown, the concepts of recipientship and availability help to account for the phenomenon of turn abandonment during overlapping talk. A multimodal analysis of several cases of withdrawal from the turn led to some preliminary observations. First, if, in overlap, a speaker gazes at a non-gazing (and overlapping) addressee, he is likely to withdraw quickly from the turn as soon as he notices his interlocutor's absence of gaze. Second, if a speaker is addressed in overlap by one of his co-participants, his withdrawal from the turn is often closely timed with a re-orientation of his gaze toward the overlapping speaker. Third, if a mutual gaze between two participants is already established when overlap occurs, a re-orientation to an alternative, non-overlapping participant may lead to a quick abandonment of the turn by the speaker who lost his recipient.

However, we should not assume that gaze direction is the only or the most relevant feature that intervenes during overlap management. As the last example (ex. 9) has shown, withdrawal from the turn occurs also when speakers do not even seek to gaze at each other. This case, nevertheless, shows that participants in face-to-face interaction make use of visual elements in order to resolve overlap, and orient toward the availability of interactional space. In this example, the withdrawals from the turn appear to be in connection with visible pointing gestures that are carried out in the common space of attention (and that thus catch the attention of potential re-





cipients), the plans on the table, and that render this space (un)available for taking a turn.

These analyses confirm that withdrawal from the turn does not seem to follow a formal rule (which could for instance be described in terms of recurrent incomplete syntactic patterns), but that it is, just as turn-taking in general, the result of negotiations between speakers who make use of different audible and visible resources in dynamic, local, indexical, and praxeological ways (Mondada 2007b). Therefore, withdrawal from the turn is the result of an ensemble of basic interactional features like sequential pressure, the (dis)aligned character of turns or lines of action, problems of epistemic groundings (as in the case of repair sequences), the availability of the recipient and of the interactional space, the manipulation and implication of artifacts during the interaction, etc. My contribution aimed not at presenting an exhaustive inventory of “negotiable features” during overlapping talk, but at considering the visible aspects of turn withdrawal in particular. A multimodal approach may contribute in interesting ways to the understanding of turn abandonment and issues of (syntactic) turn completeness in interaction: if we consider visible elements that intervene during turn abandonment, the pre-beginning becomes a relevant sequential position where withdrawal can occur. As pre-beginnings may be visual (Mondada 2004, 2007a) or consist of (not always) audible non-lexical items and phenomena (in-breath, tongue clicks), the abandonment of the turn at this stage may not be understandable as such, being inaudible and literally invisible if participants’ visibly displayed orientations are not considered. The analyzed excerpts also emphasized the embodied nature of withdrawals from the turn. Consequently, if one wishes to investigate the specificities of syntax or other grammatical features within sequences of overlapping talk, those withdrawal formats should be studied within their specific multimodal settings. The analysis of visible resources like gaze, body posture, gesture, or the manipulation of artifacts, and their interplay with recurrent patterns of prosodic, syntactic, and actional features, allows to be reached a more complete understanding of recurrent phenomena like withdrawal from the turn in overlap and of the *gestalt*-like character of turns and fragments of turns-at-talk. As a consequence, a grammatical description of spoken language should consider that interaction is not primarily driven by a set of abstract rules, but that recurrent (syntactic, prosodic, etc.) patterns are intertwined with the management of interactional tasks, and thus with audible and visible actions.

This chapter should be understood as an invitation to investigate further the phenomenon of withdrawal from the turn. Two main points are worth emphasizing. First, if we wish to understand how reciprocity displays *within*



and availability *of* the interactional space intervene during overlapping talk, the preliminary observations I made in this chapter should be systematized by looking at a large variety of interactional settings and ecologies. More specifically, it would be interesting to consider in more systematic ways particular participative configurations by comparing withdrawals in interactions where two, three, or more speakers interact (Egbert 1997; Schegloff 1995). A second main research issue seems to be the timing of the withdrawals: the three different sequential positions for withdrawal (pre-beginning, turn-beginning, TCUs in the post-turn-beginning phase) could be studied in order to reveal recurrent features of turn abandonments at those positions, whether they are mainly visible or rather audible. At a later stage, those recurrent patterns could also be related to language specific grammatical or lexical resources, especially if larger amounts of data are available for analysis. Considering that in case of overlapping talk, two (or three, etc.) syntactic constructions emerge simultaneously, we might wonder how those emerging turns interact with each other, and especially how visible actions are positioned with respect to the withdrawal. Therefore, a fine-grained analysis of the moment where a visible re-orientation is displayed by one or more participant(s) and the moment where a speaker withdraws from the turn could be helpful in order to contribute to a full understanding of how issues of turn-taking and turn-completeness are handled by speakers in interaction.



Appendix

Transcription Conventions

1 ALE	original conversation
tra	approximate translation
[]	overlap (onset & end)
☺	smiley voice
(.)	micro-pause (< 0.2 seconds)
(2)	length of pauses in seconds
/	rising / falling intonation
&	continuation of current turn
.h/h	breath (in / out)
(h)	laughter particle, breathy voice
(il va)	uncertain transcription
((tv))	comments
^	liaison
]'	end of overlap (in case of multiple overlaps)
:	sound stretch





Withdrawal from turns in overlap and participation

233

xxx	incomprehensible segment
<u>extra</u>	prominence of talk
=	latching
par-	truncation
°bon°	low volume
BON	increased volume
< >	delimitation of phenomena noted within (())
>>bon<<	accelerated speech
*, +, £	delimitate a participant's gestures and actions
*-->	gesture or action described continues across subsequent lines
*-->>	gesture or action described continues until and after the end of the excerpt
>--	gesture or action described begins before the beginning of the line
>>--	gesture or action described begins before the beginning of the excerpt
....	gesture's preparation
--	gesture's apex is reached and maintained
, , , , ,	gesture's retraction
fab	pseudonym of participant doing the gesture
fig.	the exact point where a screen shot (figures) has been taken is indicated,
#	with a specific sign showing its position within turns-at-talk
ppp	pointing gesture
RH/LH	right hand / left hand
SVP	separable verbal particle (in German)
PRT	discourse particle (in German)

References

- Auer, P. 2002 Projection in Interaction and Projection in Grammar. *InLiSt-Interaction and Linguistic Structures* 33.
- Auer, P. 2005 Syntax als Prozess. *InLiSt – Interaction and Linguistic Structures* 41.
- Chevalier, F. H. G. 2005 To complete or not to complete: A conversation analytic investigation of unfinished turns in French. Ph.D. dissertation, University of Essex.
- Chevalier, F. H. G. 2008 Unfinished turns in French conversation: How context matters. *Research on Language & Social Interaction* 41: 1–30.



- Chevalier, F. H. G. and R. Clift 2008 Unfinished turns in French conversation: Projectability, syntax and action. *Journal of Pragmatics* 40: 1731–1752.
- Deppermann, A. and R. Schmitt 2007 Koordination. Zur Begründung eines neuen Forschungsgegenstandes. In: Reinhold Schmitt (ed.), *Koordination. Analysen zur multimodalen Interaktion*, 15–54. Tübingen: Gunter Narr.
- Drew, P. 1997 ›Open‹ class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics* 28: 69–101.
- Drew, P. 2009 ›Quit talking while I'm interrupting‹: A comparison between positions of overlap onset in conversation. In: M. Haakana, M. Laakso and J. Lindström (eds.), *Talk in Interaction: Comparative Dimensions*, 70–93. Helsinki: Finnish Literature Society.
- Egbert, M. M. 1997 Schisiming: The collaborative transformation from a single conversation to multiple conversations. *Research on Language and Social Interaction* 30: 1–51.
- Ford, C. E. 2004 Contingency and units in interaction. *Discourse Studies* 6: 27–52.
- Ford, C. E., B. A. Fox and S. A. Thompson 1996 Practices in the construction of turns: The “TCU” revisited. *Pragmatics* 6: 427–454.
- Goodwin, C. 1979 The interactive construction of a sentence in natural conversation. In: G. Psathas (ed.), *Everyday Language: Studies in Ethnomethodology*, 97–121. New York: Irvington Publishers.
- Goodwin, C. 1980 Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological Inquiry* 50: 272–302.
- Goodwin, C. 1981 *Conversational Organization. Interaction between Speakers and Hearers*. New York: Academic Press.
- Goodwin, C. 2000 Action and embodiment within situated human interaction. *Journal of Pragmatics* 32: 1489–1522.
- Goodwin, C. and M. H. Goodwin 2004 Participation. In: A. Duranti (ed.), *A Companion to Linguistic Anthropology*, 222–244. Oxford: Blackwell.
- Goodwin, M. H. 1997 Byplay: Negotiating evaluation in storytelling. In: G. R. Guy, C. Feagin, D. Schriffrin and J. Baugh (eds.), *Towards a Social Science of Language: Papers in Honor of William Labov*, 77–102. Amsterdam/Philadelphia: John Benjamins.
- Hakulinen, A. and M. Selting (eds.) 2005 *Syntax and Lexis in Conversation*. Amsterdam/Philadelphia: John Benjamins.
- Hayashi, M. 2003 *Joint Utterance Construction in Japanese Conversation*. Amsterdam/Philadelphia: John Benjamins.
- Jefferson, G. 1973 A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica* IX: 47–96.
- Jefferson, G. 1983 Notes on some orderlinesses of overlap onset. *Tilburg Papers in Language and Literature (Tilburg University)* 28: 1–28.
- Jefferson, G. 1986 Notes on ›latency‹ in overlap onset. *Human Studies* 9: 153–183.
- Jefferson, G. 2004 A sketch of some orderly aspects of overlap in natural conversation. In: G. H. Lerner (ed.), *Conversation Analysis. Studies from the First Generation*, 43–59. Amsterdam/Philadelphia: John Benjamins.
- Kendon, A. 1990 *Conducting Interaction. Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.
- Kendon, A. 1990 Some context for context analysis: A view of the origins of structural studies of face-to-face interaction. In: A. Kendon (ed.), *Conducting Interaction*, 15–49. Cambridge: Cambridge University Press.





- Laursen, L. 2005 Towards an embodied grammar. Gesture in tying practices. Constructing obvious cohesion, *Interacting Bodies*, Lyon. http://gesture-lyon2005.ens-lyon.fr/article.php?id_article=238 (accessed on August 2, 2012).
- Lerner, G. H. 1996 On the “semi-permeable” character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 238–276. Cambridge: Cambridge University Press.
- Lerner, G. H. 2002 Turn-sharing. The coral co-production of talk-in-interaction. In: C. E. Ford, B. A. Fox and S. A. Thompson (eds.), *The Language of Turn and Sequence*, 225–256. New York: Oxford University Press.
- Lerner, G. H. 2003 Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society* 32: 177–201.
- Lindström, J. 2006 Grammar in the service of interaction: Exploring turn organization in Swedish. *Research on Language and Social Interaction* 39: 81–117.
- Mondada, L. 2004 Temporalité, séquentialité et multimodalité au fondement de l’organisation de l’interaction: Le pointage comme pratique de prise de tour. *Cahiers de Linguistique française* 26: 269–292.
- Mondada, L. 2007a Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies* 9: 194–225.
- Mondada, L. 2007b 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. 2007c Imbrications de la technologie et de l’ordre interactionnel. L’organisation de vérifications et d’identifications de problèmes pendant la visioconférence. *Réseaux* 144: 141–182.
- Mondada, L. 2008 Exchanging glances while talking and driving: Issues in the analysis of multi-activity. *Sociolinguistics Symposium* 17. Amsterdam.
- Mondada, L. to be published An interactionist perspective on the ecology of linguistic practices: The situated and embodied production of talk. In: R. Ludwig, P. Mühlhäusler and S. Pagel (eds.), *Language Ecology and Language Contact*.
- Ochs, E., E. A. Schegloff and S. A. Thompson (eds.) 1996 *Interaction and Grammar*. Cambridge: Cambridge University Press.
- Oloff, F. 2009 Contribution à l’étude systématique de l’organisation des tours de parole: les chevauchements en français et en allemand. Ph.D. Dissertation, ENS LSH Lyon and Universität Mannheim. Mannheim, University of Mannheim: <https://ub-madoc.bib.uni-mannheim.de/29617/> (accessed on August 2, 2012).
- Oloff, F. 2010 Embodied claims to speakership following overlapping talk. *ICCA 2010*, Mannheim.
- Olsher, D. 2004 Talk and gesture: The embodied completion of sequential actions in spoken interaction. In: R. Gardner and J. Wagner (eds.), *Second Language Conversations*, 221–245. London: Continuum.
- Rossano, F., P. Brown and S. C. Levinson 2009 Gaze, questioning, and culture. In: J. Sidnell (ed.), *Conversation Analysis: Comparative Perspectives*, 187–249. Cambridge: Cambridge University Press.
- Sacks, H., E. A. Schegloff and G. Jefferson 1974 A Simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735.
- Schefflen, A. E. 1972 *Body Language and Social Order: Communication as Behavioral Control*. Englewood Cliff: Prentice Hall.





- Schegloff, E. A. 1979 The relevance of repair to syntax-for-conversation. *Syntax and Semantics* 12: 261–286.
- Schegloff, E. A. 1984 On some gestures' relation to talk. In: J. M. Atkinson and J. Heritage (eds.), *Structures of Social Action. Studies in Conversation Analysis*, 266–296. Cambridge: Cambridge University Press.
- Schegloff, E. A. 1987 Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. In: G. Button and J. R. E. Lee (eds.), *Talk and Social Organization*, 70–85. Clevedon: Multilingual Matters.
- Schegloff, E. A. 1988/1989 From interview to confrontation: Observations of the Bush/Rather Encounter. *Research on Language and Social Interaction* 22: 215–240.
- Schegloff, E. A. 1992 To Searle on conversation: A note in return. In: J. R. Searle, H. Parret and J. Verschueren (eds.), *(On) Searle on Conversation*, 113–128. Amsterdam/Philadelphia: John Benjamins.
- Schegloff, E. A. 1995 Parties and talking together: Two ways in which numbers are significant for talk-in-interaction. In: P. ten Have and G. Psathas (eds.), *Situated Order. Studies in the Social Organization of Talk and Embodied Activities*, 31–42. Washington: University Press of America.
- Schegloff, E. A. 1996 Turn organization: one intersection of grammar and interaction. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 52–133. Cambridge: Cambridge University Press.
- Schegloff, E. A. 2000 Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29: 1–63.
- Schegloff, E. A. 2002 Accounts of conduct in interaction. Interruption, overlap, and turn-taking. In: J. H. Turner (ed.), *Handbook of Sociological Theory*, 287–321. New York: Springer.
- Schegloff, E. A., E. Ochs and S. A. Thompson 1996 Introduction. In: E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), *Interaction and Grammar*, 1–51. Cambridge: Cambridge University Press.
- Schmitt, R. 2005 Zur multimodalen Struktur von *turn-taking*. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 6: 17–61.
- Schmitt, R. (ed.) 2007 *Koordination. Analysen zur multimodalen Interaktion*. Tübingen: Gunter Narr.
- Selting, M. and E. Couper-Kuhlen 2000 Argumente für die Entwicklung einer »interaktionalen Linguistik«. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 1: 76–95.
- 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/Philadelphia: John Benjamins.





Ina Hörmeyer
University of Freiburg

The importance of gaze in the constitution of units in Augmentative and Alternative Communication (AAC)

1. Introduction

The turn-constructive unit (TCU) is traditionally defined as the smallest interactionally relevant complete linguistic unit (Sacks, Schegloff, and Jefferson 1974), constructed “with syntactic and prosodic resources within their semantic, pragmatic, activity-type-specific, and sequential conversational context” (Selting 2000: 477). However, this definition is not sufficient in interactions characterized by a lack of linguistic resources. In these cases, visual signals become a substitute for missing modalities. They play an essential role in all kinds of interaction in which the participants are dependent on augmentative and alternative forms of communication due to severe language impairment.

The American Speech-Language-Hearing Association (ASHA) defines Augmentative and Alternative Communication (AAC) as “an area of research, clinical and educational practice. AAC involves attempts to study and, when necessary, compensate for temporary or permanent impairments, activity limitations, and participation restrictions of persons with severe disorders of speech-language production and/or comprehension, including spoken and written modes of communication” (ASHA 2004: 1). The term *aid* is used to refer to “a device, either electronic or non-electronic, that is used to transmit or receive messages” (ASHA 2004: 1). People who use AAC require adaptive assistance for speaking and/or writing because of different communication disorders including congenital illnesses such as autism or cerebral palsy and acquired impairments such as multiple sclerosis or traumatic brain injury.

Augmented communication is characterized by a number of distinctive properties. Much of the research concerning the interaction between aided and non-aided speakers deals with the so-called “asymmetry” of participants’ actions. Unlike their co-participants, the aided speakers predominantly produce responsive communicative actions and so demonstrate passive roles in interaction (Calculator and Dollaghan 1982; Clarke and Kirton 2003; Light, Collier, and Parnes 1985). In other words, they act as respondents to their co-participants. Typical actions are provisions of in-



formation, provisions of clarification, and confirmations/denials. These asymmetries are mainly seen in interactions between aided-speaking children and their parents. The parents use a high number of initiating actions, producing many questions, commands, and requests for clarification (Light, Collier, and Parnes 1985; Pennington and McConachie 1999; Von Tetzchner and Martinsen 1996). It is the parents who often determine the conversational topics (Ferm, Ahlsen, and Björck-Akesson 2005), ask questions to which the answer is already known (Pennington and McConachie 1999) and – despite the potential of many electronic communication aids – make use of established yes/no-interrogatives (von Tetzchner and Martinsen 1996). These strategies used by the co-participants require very short and predictable utterances by the aided speakers. By controlling the interaction in such a way, the co-participants can avoid a formidable communicative breakdown.

A further property of augmented communication is its collaborative aspect, which resembles conversation with aphasic and dysarthric speakers (Bauer and Auer 2009; Goodwin 1995, 2000, 2003). The co-participants of aphasic speakers do not act just as recipients; instead, they are often involved in producing and interpreting their aphasic partners' utterances. By doing so, the co-participants take over much of the communicative work and, in many cases, are solely responsible for controlling the conversation. This applies to conversations with dysarthric speakers as well (Bloch 2005; Bloch and Beeke 2008; Bloch and Wilkinson 2009; Clarke and Wilkinson 2009; Collins and Markova 1995). The responsibility for frequent communicative problems is shared by the conversationalists, so the non-aided speaker takes the part of a co-producer of the utterances (Collins and Markova 1995). Special collaborative techniques, such as conjectures or completions, become an everyday routine in such interactional scenarios and are no longer seen as uncommon or problematic (Bloch 2005; Bloch and Beeke 2008).

Furthermore, Clarke and Wilkinson (2007), in their analysis of the sequential structure of augmented conversations, discovered that aided speakers often produce their utterances as second parts of adjacency pairs – mostly giving answers to questions – and after meta-interactional prompts¹. The questions are primarily formulated in a way that requires very simple and short answers with little or no syntactic form, often consisting of just

¹ The term *meta-interactional* was used to “represent talk that was an explicit evocation of how the conversation should develop.” (Clarke and Wilkinson 2007: 341)



one word². As the aided speakers are able to produce elliptical and thus very clear utterances, the co-participants establish a structurally predictable form of interaction, providing a strictly pre-defined framework for what the aided speaker should do next. This framework of sequential implicativeness is useful not only on a semantic level to understand one-word utterances, but also on a more complex level of turn constructional units. The co-participants of aided speakers may orient to the sequentiality of conversational interaction as a resource for identifying the words which belong together and recognizing the defined units. The fact that most utterances produced by communication aids consist of only one word makes the defined units clear and unambiguous. In most cases, co-participants also have no problem identifying units consisting of two or more words, even if no syntactic or prosodic structure is present. This also applies within utterances that are not the second part of an adjacency pair or which follow after meta-interactional prompts when the sequential context is not particularly helpful in recognizing a unit. In these cases, the aided speaker has to indicate the boundaries of the defined unit. In the following investigation, I will focus on the role of gaze in constituting units in augmented communication.

The application of the principles of conversation analysis to aphasic communication has provided much research to date (Bauer and Auer 2009; Goodwin 1995, 2000, 2003). Recently, there have also been some investigations in the field of augmented communication based on conversation

² The following example shows the use of an electronic communication aid to produce a second part of an adjacency pair, N is the aided speaking person:

- 01 R: ich würde GERne mal hören,
I would like to know
- 02 wi:e ähm (---) wie bei ihnen im moment (--) das mit dem
WOHnen sich darstellt;
about your actual living situation
- 03 N: ((blickt auf Kommunikationshilfe))
((looks at electronic communication aid))
- 04 (0.53)
- 05 R: wie wie das für sie IS.
how it is for you
- 06 (3.68)
- 07 N: SUpEr;
brilliant
- 08 ((blickt zu R))
((looks at R))
- 09 R: ((lacht))
((laughs))





analytic principles (Clarke and Wilkinson 2007, 2008, 2009). Because of the examination of sequential relationships between participants' actions, conversation analysis seems an appropriate approach for the study of augmented communication, which is often realized over a sequence of turns by the aided speakers and their co-participants.

After a short overview of the data, I will discuss some examples of augmented communication where gaze seems to be an essential resource in constituting and recognizing units. I will then explore cases where a high projective context provides another resource on the level of pragmatics, showing that a conflict may exist between the gaze level and the pragmatic level. I will demonstrate that, between these two levels, it is the co-participant of the aided speaker who must decide which is more relevant to the context.

2. Data

The following extracts are from conversations between two young Germans with cerebral palsy combined with severe dysarthria and their conversation partners.

Cerebral palsy (CP) describes a group of chronic conditions affecting body movements and muscle coordination (Duffy 2005). This condition is caused by damage to one or more specific areas of the brain, usually occurring during fetal development or infancy. Depending on which areas of the brain have been damaged, people with cerebral palsy may experience muscle tightness or spasms, involuntary movement, or impairment of sight, hearing, or speech. The condition of dysarthria concerns motor speech impairment within any or all of the speech subsystems (Thiele 1999). Thus, the production of spoken language is impaired or impossible, depending on the severity of the impairment. However, there are no reliable indicators for cognitive or grammatical competence of people with CP.

The first participant, Nina, is a 19-year-old woman who has used her present electronic communication aid for about one year. Nina lives alone, aided round the clock by her personal assistants. The extracts are from different conversations with two of her personal assistants, Melanie and Lutz. Her conversation with the electronic communication aid is characterized by a high percentage of one-word utterances and the collaborative constitution of meaning with the help of her conversation partners.

The second participant, Max, is a 16-year-old student who has used his present communication aid for one and a half years. He lives at home with his parents, who are very committed to using the electronic communication aid. The extract is from a conversation at school with his teacher, Mrs H. His



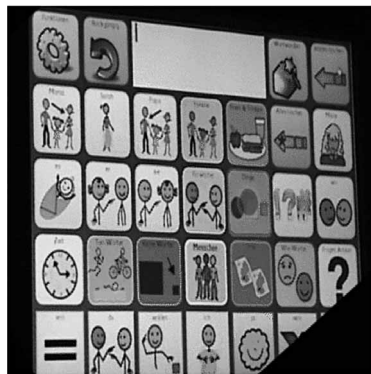
The importance of gaze in the constitution of units

241

utterances are highly elaborated, and he tries to communicate without much collaborative work from his conversation partners.

In both cases, the conversation partners have known the aided speakers for several years and are used to communicating with them using electronic communication aids.

The electronic communication aid used by both Nina and Max was invented by Tobii Technologies and is called “MyTobii”. It consists of grid-based, dynamic surfaces, with pictographs that can be selected directly via eye-tracking.



Gaze direction is indicated by a red light on the surface, and the user can decide whether he wants the word to be spoken by an electronic voice directly after its selection or if he wants to build a complete utterance and have the words be spoken after finishing. The communication aid also offers an integrated grammatical function, enabling automatic flexion of verbs in sentences with verb-second position. Other kinds of flexion can be operated by the user manually. Therefore, theoretically, the user has the ability to produce well-formed sentences, based on written German, with the help of his communication aid. However, in actuality, most of the utterances produced with the communication aid consist of only one word. One of the reasons for this is the long production time that is needed to find and activate the adequate word. If an aided speaker produces more than one word, in most cases he does so without using any morphological markers or “correct” syntax. He simply strings together several words in a telegraphic style.

Two different cameras were used in data collection, one to film the conversation partners and the other to film the surface of the electronic communication aid. In all extracts described here, the conversation partners sit in front of each other, so that the surface of the communication aid is only vis-





ible to the aided speaker. The co-participant can see only the back side and, thus, can hear only what the electronic voice says. The use of the second camera allows us not only to hear the final utterance produced by the communication aid but also to observe the production process, which can include word search or deleting of words.

The data was transcribed by the conventions of the *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)* (Selting et al. 2009). Some additional conventions, concerning the use of the electronic communication aid, are noted directly at the transcripts.

3. The role of gaze in constituting units in augmented communication

Previous work on gaze has shown that gaze not only serves as a substitute for missing resources in interaction under limited conditions, but also has important functions in ordinary every-day interaction in its coordinated use with gesture and talk. Research has shown the functions of mutual gaze in face-to-face conversations and the interactional needs it fulfills (Goodwin 1981; Kendon 1967). Different patterns of gaze direction are used to organize stance taking in assessments (Haddington 2006), right-side boundaries in reenactments (Sidnell 2006), and question-answer pairs (Rossano 2010).

Some research has been done on the role of gaze in aphasic conversations (Bauer and Auer 2009; Laakso and Klippi 1999). Gaze is often used by the aphasic speaker to invite the co-participant to collaboratively take part in the word-search. Here again, gaze functions as a signal to organize special issues of turn-taking.

Researchers have also examined the communicative function eye gaze has in interaction with augmented speakers (Clarke and Kirton 2003; Light, Collier, and Parnes 1985; Pennington and McConachie 1999). When the aided speakers in these studies used eye gaze alone, they usually did so to achieve very special communicative functions, namely to provide information in response to co-participants' requests, to provide yes/no responses, or to request objects within the environment. However, gaze seems to be perfectly suited to aspects of organizing turn-taking in augmented communication as well. Actually, there are arrangements between aided speakers and their conversation partners that organize turn-taking. Nina, for example, uses her gaze direction to select a next speaker. Otherwise this would not have become clear because the role of gaze in organizing turn-taking is directly connected to the issue of constituting units. This data reflects that the constitution of units is equivalent to turn-taking. Even if the utterances produced by the elec-





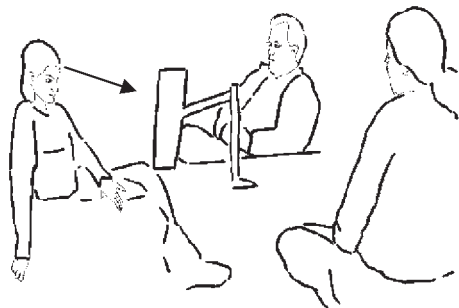
tronic communication aid consist of more than one word, it can generally be concluded that only one unit per turn is produced. This means that by shifting gaze, the aided speaker signals the boundaries of the intended units.

In what follows, I will show the role of gaze in constituting units in augmented interactions where syntactical, morphological, and prosodic resources are missing, as well as interactions where the aided speaker uses syntax and morphology to produce electronic utterances.

Extract one is taken from a conversation between Nina and her two assistants, Melanie and Lutz. The conversationalists discuss some of the other, part-time assistants, many of whom will soon finish their studies, requiring Nina to find new assistants. One of the part-time assistants is Anna. In the following section, Nina asks Melanie and Lutz for how long Anna will study.

(1)³

- 01 N: du
you
02 L: ich
I
03 mich um erstsemester kümmern oder so
I should look for new students
04 N: ((blickt auf Tobii))
((looks at Tobii))



- 05 L: die stuDIERN noch schön lang;
they are going to be at their studies for a long time
06 (0.6)
07 <<lachend>hm>
<<laughing>hm>
08 (1.9)

³ The electronic communication aid is called “Tobii” in the transcript. Utterances produced by the communication aid are written in italics.

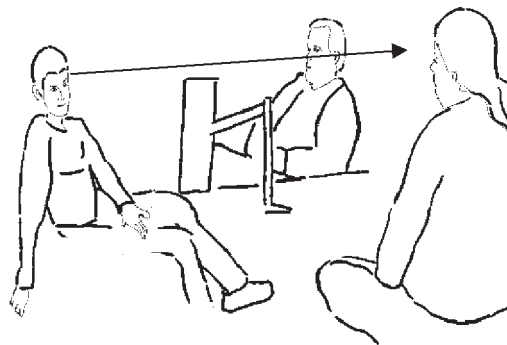




244

Ina Hörmeyer

09 hups
whoops
10 (6.0)
11 N: ANna
12 (1.1)
13 L: ANna,
14 (6.2)
15 N: eins
one
16 ((blickt zu Melanie, nickt ihr zu))
((looks at Melanie, nods at her))



17 M: [ein JAHR,
one year
18 L: [ANna is noch (-) <<p>ein jahr da ja->
Anna will stay for another year

The extract starts with the mentioning of new students, who are going to be at their studies for a long time and are thus potential assistants for Nina. Nina has just produced the word *du* (you) (line 01) with her communication aid, which Lutz interprets as a request for looking for new students (line 03). It seems that for Nina, Lutz's interpretation does not fit her intention, because she does not confirm or decline the proposed interpretation. Rather, she gives new information by using the communication aid to help her conversation partner understand her intention. This action of giving new information starts in line 04, where Nina's gaze shifts from her conversation partner Lutz to her communication aid. With this orientation to her communication aid, Nina signals her preparedness to use the machine and, therefore, to speak. Because of the long production process, including her word search and fixation of a symbol, the shifting of gaze represents the beginning of her turn.

In the following utterance, Lutz gives a funny explanation why it should be new students (line 05). This little joke and his following laughter (line 07) serve as a communicative strategy that can often be observed with the con-





versation partners of aided speakers to bridge the long production time. To select one new word can last several minutes. The interaction partners cannot continue the conversation because that would result in a sequential dislocation – the aided turn would no longer fit. By giving this funny explanation, Lutz finds a way to reduce the pause while ensuring the relevance of Nina's turn. After laughing, Lutz hits the table where the communication aid is fixated with his leg, and then comments on this accident with a *hups* (‘whoops’, line 09).

This is followed by another six-second pause until Nina produces the name *ANna* (line 11). Lutz repeats this name (line 13), which can also be seen as a communicative strategy for constituting meaning and is often observed in conversations with dysarthric speakers (Bloch 2005). It is not meant, therefore, as an initiation of repair, but as “part of an established routine of sequencing that enables the understanding of talk through a successive build-up of turns” (Bloch 2005: 52). This insert expansion assures the understanding of the particular turns and, therefore, the meaning of the whole utterance. The repetition does not have to be confirmed in this special kind of sequence.

Nina continues looking at her communication aid and produces the next word, *eins* (‘one’, line 15). Afterwards, she looks at Melanie and nods, and by doing so, signals the turn allocation and invites her to collaborate, what means that Melanie should guess what she means with her two-word utterance. Both of her conversation partners follow this prompt and give a proposal of interpretation, namely that her assistant Anna will stay (as an assistant) for another year. Lutz's interpretation (‘Anna will stay for another year’ line 18) shows that he identifies the two words produced by Nina as one unit, although there is no morphological indication. Thus, by shifting gaze from her conversation partners to the communication aid, Nina signals the beginning of her turn and, with that, the beginning of her turn-constitutive unit. The end of the turn and the end of the unit is shown in the same way, by shifting gaze from the communication aid back to her partners. This example not only shows how gaze is used by the augmented communicating person to constitute a unit, but also how gaze can help the conversation partners to identify the defined unit. All the words that are produced by the communication aid while looking at it are intended and interpreted as a unit; therefore, it can be said that gaze functions as a unit boundary.

Extract 2 is from the same conversation. In this extract, Nina, Melanie, and Lutz discuss Lutz's work schedule. Nina believes that Lutz works too much and that he does not have enough time for her.

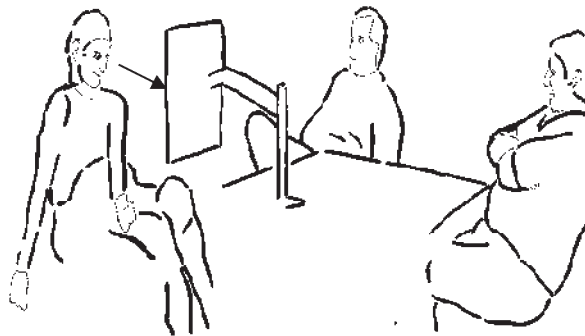


246

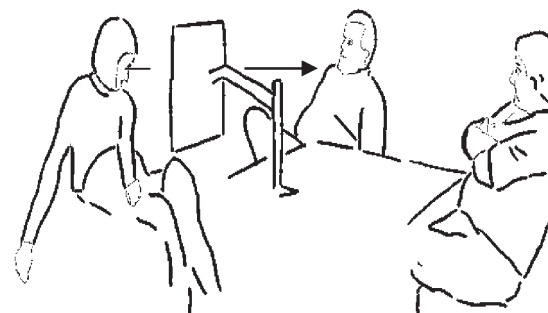
Ina Hörmeyer

(2)

02 L: du bis beim AR[beiten noch; hm?
you are still talking about work, aren't you?
 03 N: ((nickt leicht))
((nods slightly))
 04 ((blickt auf Tobii))
((looks at Tobii))



05 (1.0)
 06 L: (genau)
(right)
 07 (2.1)
 08 <<p>übernächste wo[che sieht ganz gut aus;>
the week after next looks quite good
 09 N: [geht
goes
 10 (3.1)
 11 arbeiten
to work
 12 ((blickt zu Lutz))
((looks at Lutz))





13 (0.4)
 14 ((nickt tief zu Lutz, blickt auf Tobii))
((big nods towards Lutz, looks at Tobii))
 15 hm_m
 16 (0.5)
 17 hm:
 18 (0.3)
 19 L: ARbeiten [gehen;
go to work
 20 N: [gn
 21 (4.0)
 22 L: ganz regulär in nächste woche MO:Ntag-
as always next week on Monday

Directly before the extract starts, Lutz tries to discover what Nina wants to say. He makes several guesses until he realizes that Nina is still on the “work”-theme. After expressing this guess in line 02, Nina nods slightly, signaling confirmation. Nina then shifts her gaze to the communication aid and signals the beginning of the production of an utterance (line 04). Lutz makes a short comment on his work schedule for the following week, but he does not change the topic, and when Nina produces the first word *geht* (‘goes’, line 09), Lutz stops talking and gives her the turn. Nina produces a second word, *arbeiten* (‘to work’, line 11). She then looks at Lutz, signaling turn allocation and the end of the unit.

In this extract, Nina’s utterance is not directly followed by a reaction from Lutz. This, of course, happens quite often because it can be difficult to ascertain what Nina wants to say. In many cases, extensive sequences constructed of different hints and guesses are necessary to discover the meaning of an utterance (Laakso and Klippi 1999). In this extract, Nina makes use of other modes of communication: she first nods at Lutz (line 14) and then produces sounds (line 15 and 17). Nina orients back to the communication aid when she recognizes Lutz’s difficulties, signaling that she may want to provide more semantic information to aid him. However, Lutz gives an interpretation (line 19) without another aided turn from Nina followed by an overview of his work schedule. Again, Lutz treats the two words Nina produces while looking at the communication aid as one unit; therefore, gaze again functions as a unit boundary. While Nina is oriented towards her aid, Lutz waits until he receives her gaze to take his turn.

This orientation to gaze can also be observed in utterances that are highly elaborated and offer a syntactic structure. Extract 3 is from a conversation between the second participant, Max, and his teacher Mrs H. Immediately before the extract begins, Max tells his teacher and his classmates what he has done the previous weekend; he went shopping with his mother



248

Ina Hörmeyer

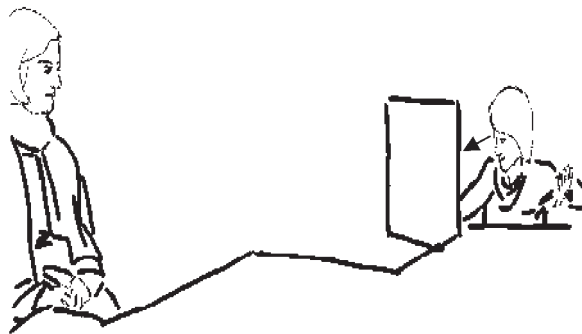
and had ice cream. The following conversation is about the kind of ice cream he had.

(3)

01 H: und mich würd interessieren was du für_n EIS
what I'm interested in is what sort of ice cream
gegessen hast?

you've eaten?

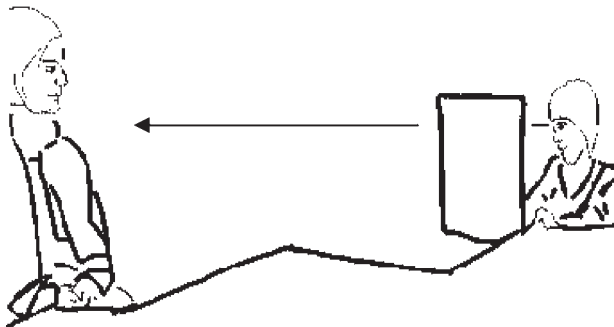
02 M: ((blickt auf Tobii))
((looks at Tobii))



03 (22.0)

04 ich haben zitr[one;
I have lemon

05 [((blickt zu H))
((looks at H))



06 (1.0)

07 H: [((macht kreisende Bewegung mit beiden Händen))
((makes rotary movements with both of her hands))

08 M: [((blickt auf Tobii, dann wieder zu H))
((looks at Tobii, then at H))





09 H: <<pp>weiter>
go on

After finishing his report, Mrs H. asks him what sort of ice cream he had (line 01). Max orients to his communication aid (line 02), and his utterance appears after 22 seconds. This is a very long period of time, even in augmented communication, and it is possible that in other communicative contexts Max would not have had so much time. In this case, where the teacher gives Max the time to communicate via electronic aid, nobody takes a turn until Max has finished producing his utterance (line 04). This utterance is an interesting answer. While the question is formulated in a way that requires a very simple and short answer consisting of only one word, *zitronen* (lemon), Max decides to produce more than an elliptical answer; he produces a syntactically complete utterance with the subject, the first part of the verb, and the object. Max does not choose the morphologically correct first person singular of the German verb *haben*; however, he does use the correct syntax except that he does not produce the second part of the verb, the participle⁴. Although Max gives more information than he would have given in an elliptical answer, the decision to produce a complete utterance makes the second verbal part relevant. After producing his answer, Max orients to his teacher by shifting gaze (line 05) and signals the end of his turn and the end of the unit. Mrs H. does not react directly; there is a short gap of one second. After the gap, Mrs H. first demonstrates that Max should go on with his utterance by gesturing (line 07) and then by verbalizing (line 09). This signifies that she does not agree with the end of his unit, as she asks him for the missing verbal element. Because Max started to utter a “complete sentence”, his answer is now interpreted as a break-off that needs to be repaired. By comparing these different data, it can be seen that in the school context a strong orientation towards “completeness” in interaction exists which also holds for non-aided speakers. This orientation, however, was not observed in interactions with friends, parents, or assistants.

In fact, after some time, Max repairs his unit to create a grammatically complete sentence. In his production time, the teacher orients and gestures towards a classmate of Max, Tim. Consider Extract 4, which resumes where Extract 3 left off.

⁴ The German brace construction is a basic principle of the German syntax and applies to the disconnection of the two parts of a predicate which “bracket” the other elements of the sentence (Wöllstein-Leisten et al. 2005).





250

Ina Hörmeyer

(4)

10 (4.4)
11 T: ((hustet))
((coughs))
12 (2.5)
13 so
so
14 H: ((öffnet Hände))
((opens her hands))
15 T: JA
yes
16 [ich weiß
I know
17 H: [((nickt leicht))
((nods slightly))
18 (0.53)
19 M: ich HABen zitrone_eis
I have lemon ice cream
20 ((blickt zu H))
((looks at H))
21 (1.0)
22 T: das ist
this is
23 (1.2)
24 H: ((klopft sich mit Händen auf die Brust))
((taps with her hands on her chest))
25 (2.0)
26 M: arrg::[_n
27 H: ((blickt zu Max, macht kreisende Bewegung mit beiden
Händen))
((looks at Max, makes rotary movements with both of her hands))
28 M: ((blickt auf Tobii))
((looks at Tobii))
29 (11.8)
30 ich HABen zitr[one_eis gegessen;
I have eaten lemon ice cream
31 [((blickt zu H))
((looks at H))
32 H: ((nickt))
((nods))

While Max is producing his first repair, Mrs H. is interacting with Tim. The repaired answer is articulated in line 19, followed by a shift of gaze by Max to his teacher. Because Mrs H. is still looking at Tim, she does not realize this gaze shift and does not react to Max's utterance directly. Max has to request attention before she re-orient to him (lines 26–27). She then repeats the rotary movements with her hands to show him that she still does not agree with his unit (line 27). In his first repair, Max added only another noun to





specify the object, so the participle is still missing. Max seems to realize this, because in his second repair, he adds the adequate verb. After producing the repair in line 30, Max orients to Mrs H. by shifting his gaze (line 31). Now Mrs H. agrees with his unit and nods (line 32). We can again observe an orientation to gaze direction when defining and mediating the boundaries of turn-constructive units. The words produced while looking at the communication aid are defined and recognized as a unit, serving as a basis for discussing the grammatical correctness of the unit.

The two interactional settings analyzed so far are quite different. The first involves a kind of everyday conversation between an augmented speaker and her personal assistant. The second involves a kind of institutional conversation between an augmented speaker and his teacher. In both settings, the shifting of gaze signals a completed unit on both syntactical and sequential levels. However, a difference can be observed in the reaction to this unit by the interaction partners. In the conversation between Nina and Lutz, Nina's utterance serves as semantic information, or a hint, that Lutz tries to interpret. In the conversation between Max and Mrs H., his teacher evaluates his utterance on the level of grammatical correctness; the correctness has to be mediated before the teacher reacts to the information Max provided with his utterance.

To summarize, we can say that gaze can function as a unit boundary in augmented communication. This applies not only to everyday interactions, but also to interactions in school contexts where the produced units are more negotiable in terms of grammatical completeness. Another question is whether other levels of unit-constitution exist which can be in conflict with gaze.

4. Other levels of unit constitution

When we discuss the constitution of units in augmented communication, we must always consider the lack of resources. Aided speakers cannot fully access resources, such as prosody or syntax, when building units and must use other resources, such as eye gaze. However, even without prosodic or syntactic markers, the units are produced within semantic, pragmatic, activity-type specific and sequential conversational contexts. In some cases, this leads to a conflict between the pragmatic level and the gaze level.

Extract 5 is taken from another conversation between Nina, Lutz, and Melanie. In this configuration, Melanie is sitting in the background. Nina and Lutz discuss a Christmas present for Nina's father that Nina would like to purchase in town the next day, together with Melanie and her mother.



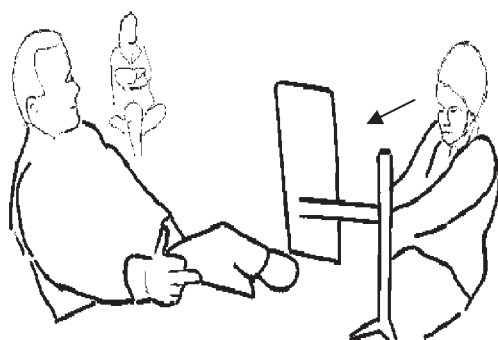


252

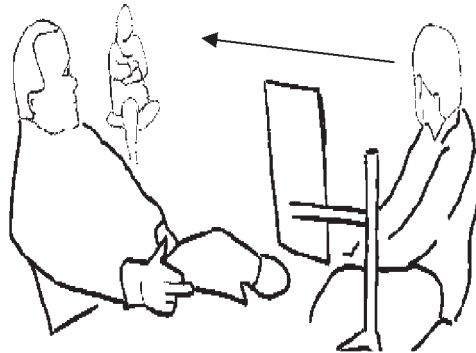
Ina Hörmeyer

(5)

03 L: HAST du das schon?
have you got that already?
 04 h: :_m
 05 (0.3)
 06 M: *hm*hm
 07 L: .hh
 08 [kuckst du dir das noch AN;
are you still going to have a look at it
 09 (0.4)
 10 N: hm
 11 ((nickt))
 ((nods))
 12 hm_m ((blickt auf Tobii))
 ((looks at Tobii))



13 (0.8)
 14 L: in der STADT?
in town?
 15 (2.5)
 16 N: melanie
 17 (0.2)
 18 [mama
mum
 19 L: [mi
ni
 20 mit [MElanie;
with melanie
 21 N: [((blickt zu Melanie))
 ((looks at Melanie))



- 22 (0.2)
- 23 L: ma MELanie und mama kuckst [du dir das an;
ma melanie and mum have a look at it
- 24 N: [((nickt tief zu Melanie))
((big nod towards Melanie))
- 25 L: deine MAMA kommt ja: kommt ja heut nachmittag;=
your mum is coming this afternoon
- 26 =und dann (.) kuckt ihr euch das dann=
and then you can have a look at it

Just before the beginning of the extract Lutz has discovered exactly what Nina wants to give her father, namely a small model of a motorbike. After discovering this, Lutz asks for more information about the present. At the beginning of the extract he asks Nina if she has already purchased it. Here, Melanie gives a negative answer for Nina (line 07), as she knows about the plans concerning the present. This “speaking for”-technique is a communicative strategy that is also common in conversations with aphasic speakers (Bauer and Auer, 2009). This strategy is often used in interactions with conversation partners unfamiliar with augmented communication. In such cases a familiar person, who has the same background as the aided speaker, acts as a translator of the utterances produced with a communication aid.

After the answer uttered by Melanie, Lutz continues putting questions by asking for the action that has to follow, the looking for the present (line 10). By looking at Nina, he orients towards her and shows that he wants an answer directly from her. Nina confirms the question by nodding (line 13), and, at the same time, articulating some sounds. She then looks at her communication aid (line 14) and signals her turn, showing that she wants to give more detailed information. Before Nina starts to speak, Lutz asks another question. By doing this, he follows a communicative strategy that is very successful in conversations with non-speaking persons who do not use a communication aid, the so-called “yes/no-interrogatives” (von Tetzchner and



Martinsen 1996). In using this strategy, the conversation partner tries to discover what the non-speaking co-participant wants to say by formulating guesses in form of yes/no-interrogatives. These kinds of questions can be answered with minimal effort, so that even people with severe language impairments can make use of this technique. This communicative strategy is, to some extent, an integral part of conversations with non-speaking partners and is often practised after adopting a communication aid.

However, Nina does not react to the question but concentrates on producing her intended utterance, which Lutz realizes, thus not insisting on an answer. Nina produces the words *melanie* and *mama* (lines 18 and 20). While producing the word *mama*, Lutz starts to collaborate by giving a proposal that overlaps with Nina. After producing her second word, Nina looks towards Melanie, and in doing so, signals turn allocation (line 23). Lutz makes a self-initiated repair of his proposed interpretation (line 25). Again, by shifting gaze, Nina signals the unit boundaries. What is interesting here is the fact that Lutz does not wait until he perceives a visual signal but reacts directly after the production of the first word, conflicting with the defined unit. What happens here is that by asking Nina a question (‘Are you going to have a look at it?’ line 10), Lutz produces the first part of an adjacency pair which makes the second part – the answer – conditionally relevant. The answer is given with the word *melanie*. This provides a pragmatically complete answer which Lutz is able to react to. Thus, his first interpretation of Nina’s utterance includes only the word *melanie* and only in the repaired interpretation does he react to the second word.

It is important to bear in mind that in most cases, Nina produces just one-word utterances, and therefore, Lutz’s reaction would usually not cause any problems. In this case, it is clear that the answer (and so the unit) is not complete and that here, as in the other examples, the gaze signals the end of the unit. This example shows that in cases of a highly projective context, it is the conversation partner who must decide whether to react immediately or to wait until a visual signal is given. Waiting for the visual signal makes the defined unit clear for the co-participant. However, this means, of course, another slowing down of the whole interaction. This could be avoided by reacting directly, which would construct a more fluent interaction. In this case, Lutz’s decision to pursue fluent interaction causes a self repair that would not have been necessary if he had decided to wait for the gaze. However, it should be noted that Lutz’s decision is influenced by the fact that more than 90 % of Nina’s utterances produced by the electronic aid consist of only one word. This is a particular pattern in the augmented communication with Nina. Therefore, it can be assumed that Lutz orients to a rule of frequency





The importance of gaze in the constitution of units

255

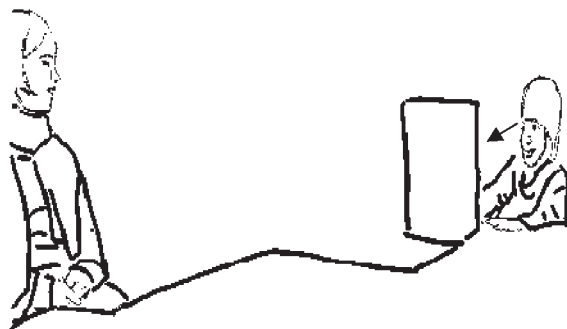
in his interaction with Nina; he adapts to the characteristics of the communicative situation and behaves accordingly.

This same decision also must be made by the conversation partner in situations where the aided speaker produces highly-elaborate utterances. As was demonstrated previously, in such situations, gaze plays an important role in the constitution of units.

The following extract is from the same teacher-student interaction in the school for physically handicapped children as Extracts 3 and 4. Max has just described his weekend, where he recounts his first experience sending an email via his electronic communication aid.

(6)

- 01 H: jetzt würd mich ganz kurz intressieren-
what I am interested in is
 02 (0.2)
 03 funktionNIERT denn jetzt das emailprogramm=
does the email program work
 04 =kannst du ver[SENDen endlich;
can you send at last
 05 M: [((nickt))
((nods))
 06 H: und wer hats [(-) HINbekommen?
and who's worked it all out?
 07 M: [((nickt))
((nods))
 08 ((blickt auf Tobii))
((looks at Tobii))



- 09 (18.9)
 10 MAma hat
mum has
 11 (25.7)
 12 A: <<pp>hach je:>

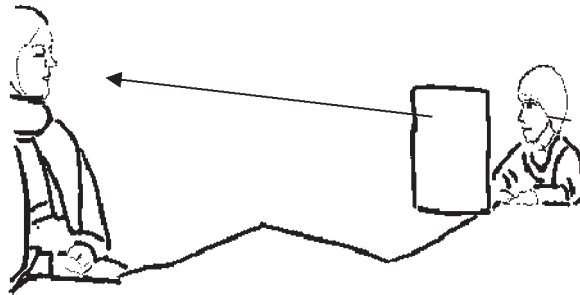




256

Ina Hörmeyer

13 ((sighs))
 13 (12.87)
 14 M: MAma hat das_Email
 mum has the email
 15 (31.6)
 16 MAma hat das_Email geöffne
 mum has ope the email
 17 (1.3)
 18 MAma [hat das_Email geöffnen;
 mum has open the email
 19 [(blickt auf Mitschüler, dann auf H)]
 ((looks at classmates, then to H))



20 H: geÖFFnet?
 opened?
 21 (0.5)
 22 M: ((nickt))
 ((nods))
 23 H: und wer hats verSCHICKT?
 and who has sent it?

The extract begins with the teacher asking whether the program finally works and if Max is able to send emails (lines 03–04). Max confirms by nodding. The teacher's next question (who has worked it all out, line 06) makes a different kind of answer relevant, asking for the name of a person. Therefore, Max gives the answer with the help of his communication aid, which he signals by shifting his gaze away from the teacher towards his aid. He then builds an incremental utterance starting with *MAma hat* (mum has, line 10) and ending with *MAma hat das_Email geöffnen;* (mum has open the email, line 18). After this, he shifts his gaze to his classmates and then to his teacher. Here, again, the direction of the gaze is used to signal turn allocation and the completeness of the intended turn-constructual unit, even though Max





uses the resources of syntax and morphology as well. Only the verb *öffnen* (to open) is not used with the correct flectional suffix of the participle. Mrs H. repeats the verb ›open‹ (line 20) using the correct form. By doing so, she makes a morphological repair on the one hand, and she makes sure that it was the verb *geÖFFnet?* (opened) which Max meant on the other hand. Max affirms her repair, and the teacher then asks who has sent the email.

In this extract, the conversation partner acts differently than in the previous example with Nina and Lutz. The question *und wer hats [(-) HINbekommen?* (and who has worked it all out) is the first part of an adjacency pair and makes an answer conditionally relevant, specifically an answer that contains a name or another reference to a person. This condition is fulfilled after Max's first utterance by mentioning *MAma*. The teacher does not react to this first word or the following increments, but waits until Max signals the end of his utterance by shifting gaze, even though this takes a long time. For Max, it seems necessary to produce a whole sentence (or rather it seems necessary for him in this institutional context). The German verb *hinbekommen* (to work it all out) is semantically not very specific. Thus, the request ›and who worked it all out‹ does not clarify what the teacher means. Instead of making a request himself that could have made the teacher's intention clear, Max gives a specification in his answer. By mentioning the verb ›open‹, he demonstrates he has understood the request by Mrs H. and clarifies that it was his mother who opened the email. Due to German syntax, the past participle ›opened‹ belongs in the last position, and so Max must produce the entire sentence to give his answer. He must state what his mother did exactly, so he can show his interpretation of his teacher's request. By waiting until Max has signaled the end of his unit, this interpreted meaning becomes clear to the teacher. By repeating the verb (and repairing it), Mrs H. makes certain that the selection of ›open‹ was not a mistake. Her next request (›who has sent it‹, line 23) shows that she is not fully satisfied with Max's interpretation; she wanted to know not only who opened the email, but also who sent it. However, she can do this initiation of repair only because she waited until Max had finished his unit and his interpretation became apparent. Therefore, she accepts a very long production time and a large discontinuation of interaction. In this case, then, the interaction partner of the augmented communicating person decided she wants the unit to be clear and not the interaction to be fluent. Certainly, this can again be dependent on the teaching context where a strong orientation to grammatical completeness is present.

These examples demonstrate that there are projectable structures in augmented communication on a pragmatic level to which the co-participants



orient themselves. However, this orientation to the pragmatic level can cause a conflict with gaze level and so with the intended unit. Thus, gaze is a necessary resource to identify the intended units, even in communicative situations with a highly projective context.

5. Discussion

This chapter has been concerned with exploring the role of gaze in the constitution of units in augmented communication. The analysis has focused on a relatively small number of examples that are illustrative of recurring features in conversations with augmented speakers. In these examples, we are confronted with a kind of conversation that is characterized by a lack of different levels of interaction. In many cases, an examination of the level of turn-constructive units reveals a lack of resources, such as prosody and syntax. These resources, normally essential for constituting and identifying units in conversational talk, are not relied upon here, as aided speakers and their co-participants access other resources. What was observed in the data is that eye gaze is the clearest and most unambiguous resource in the constitution and recognition of units in augmented communication. In addition to the role gaze plays in speaker selection, it is a strong instrument to signal unit boundaries, which both aided and non-aided speakers use. Even in utterances in which syntactical and morphological markers are used, gaze serves as an important observable interactional resource in unit-constitution. Thus, as a first result, it can be said that gaze does function as a unit boundary. It is the task of a grammar to describe the basic units of a language and to systemize their regularities (Fiehler 2006). However, to understand and describe the special regularities of Augmentative and Alternative communication, a concentration on the traditional levels of description is not very useful. The data has shown that within a grammar of AAC, the inclusion and systematization of explicit visual signals, such as eye gaze, is indispensable. These resources compensate for the failing potential of projectability due to the lack of prosody and syntax. The examples with Max show that even when syntax is used to constitute units, the conversational partner still orients to gaze. It is important to note that the syntactical resources the electronic communication aid possesses do not have the same function as in spoken language. Compared to spoken language, it takes an extremely long time to produce a multi-word utterance electronically. The pauses between the single words are several seconds long and the morphology of the words often has to be repaired manually (see Extract 6). Thus, even when syntactical features are used, it is quite difficult for the conversation partner to project the intended





end of a unit. By orienting to gaze, she can rely on a resource that is not projectable but clarifies the boundaries of the units in an explicit way.

It was also demonstrated that other levels of unit constitution can potentially come into conflict with the level of gaze. Cases with a highly projective context where the pragmatic function can supersede the gaze function were noted. In unambiguously located aided turns, the non-aided conversation partner may orient towards the sequentiality of conversational interaction; that is, the local context in which an aided unit is produced as a resource for understanding that unit, as is demonstrated in Extract 5. For this reason, a second result is that although there may be a lack of resources in this kind of interaction, conversation partners can resort to other resources that are the same as in “normal” conversational talk. These resources are the semantic, pragmatic, activity-type specific, and sequential conversational context a unit is always embedded in.

The investigation of augmented communication can also be useful in the discussion of a grammar of spoken language. The analysis of interactions with an electronic communication aid has shown that prosody is an essential resource in spoken language that must be substituted with explicit visual signals, even in cases where syntax and morphology are used. Therefore, prosody should also play an essential role in the grammar of spoken language.

Another conclusion is, once again, the importance of the special role conversation partners play in aided speaker interaction. The conflict between the level of pragmatics and the level of gaze is primarily a conflict for the non-aided speaking person. Finally it is the non-aided participant who must decide whether the unit itself should be clarified or whether the interaction should be fluent.

Acknowledgments

I would like to thank Peter Auer, Jana Brenning, Martin Pfeiffer, Elisabeth Reber and Jennifer Fielding and her mother for their helpful comments on earlier versions of this chapter.

Appendix

Transcription conventions (GAT 2, Selting et al. 2009)

[]	overlap and simultaneous talk
[]	
°h / °hh	inbreath of 0.2–0.5 seconds / 0.5–0.8 seconds
(.)	micro pause, estimated, up to 0.2 sec. duration





260

Ina Hörmeyer

(-) / (--) / (---)	estimated pause of 0.2–0.5 / 0.5–0.8 / 0.8–1.0 seconds
(2.8)	measured pause
und_äh	cliticization within units
:	lengthening
=	fast, immediate continuation with a new turn or segment
SYLlable	strong primary accent
sYlLlable	secondary accent
((coughs))	non-verbal vocal actions and events
?	rising to high (final pitch movements of intonation phrases)
,	rising to mid (final pitch movements of intonation phrases)
–	level (final pitch movements of intonation phrases)
;	falling to mid (final pitch movements of intonation phrases)
.	falling to low (final pitch movements of intonation phrases)

References

- American Speech-Language-Hearing Association 2004 Roles and responsibilities of speech-language pathologists with respect to augmentative and alternative communication: Technical report. *ASHA Supplement* 24: 1–17.
- Bauer, A. and P. Auer 2009 *Aphasie im Alltag*. Stuttgart: Thieme.
- Bloch, S. 2005 Co-construction meaning in dysarthria: word and letter repetition in the construction of turns. In: K. Richards and Paul Seedhouse (eds.): *Applying Conversation Analysis*, 38–55. Basingstoke: Palgrave Macmillan.
- Bloch, S. and S. Beeke 2008 Co-constructed talk in the conversations of people with dysarthria and aphasia. *Clinical Linguistics and Phonetics* 22: 974–990.
- Bloch, S. and R. Wilkinson 2009 Acquired dysarthria in conversation: Identifying sources of understandability problems. *International Journal of Language and Communication Disorders* 44: 769–783.
- Calculator, S. and C. Dollaghan 1982 The use of communication boards in a residential setting. An evaluation. *Journal of Speech and Hearing Disorders* 47: 281–287.
- Clarke, M. and A. Kirton 2003 Patterns of interaction between children with physical disabilities using augmentative and alternative communication systems and their peers. *Child Language and Therapy* 19: 135–151.
- Clarke, M. and R. Wilkinson 2007 Interaction between children with cerebral palsy and their peers 1. *Augmentative and Alternative Communication* 23: 336–348.
- Clarke, M. and R. Wilkinson 2008 Interaction between Children with Cerebral Palsy and their Peers 2. *Augmentative and Alternative Communication* 24: 3–15.



- Clarke, M. and R. Wilkinson 2009 The collaborative construction of non-serious episodes of interaction by non-speaking children with cerebral palsy and their peers. *Clinical linguistics and phonetics* 23: 583–597.
- Collins, S. and I. Markova 1995 Complementarity in the construction of a problematic utterance in conversation. In: Markova, I., C.F. Graumann and K. Foppa (eds.): *Mutualities in Dialogue*, 238–263. Cambridge: Cambridge University Press.
- Duffy, J. R. 2005 *Motor Speech Disorders: Substrates, Differential Diagnosis and Management*. St. Louis, Missouri: Mosby.
- Ferm, U., E. Ahlsen and E. Björck-Akesson 2005 Conversational topics between a child with complex communication needs and her caregiver at mealtime. *Augmentative and Alternative Communication* 21: 19–40.
- Fiehler, R. 2006 Was gehört in eine Grammatik gesprochener Sprache? Erfahrungen beim Schreiben eines Kapitels der neuen Duden-Grammatik. In: A. Deppermann, R. Fiehler and T. Spranz-Fogasy (eds.): *Grammatik und Interaktion. Untersuchungen zum Zusammenhang von grammatischen Strukturen und Gesprächsprozessen*, 21–41. Radolfzell: Verlag für Gesprächsforschung.
- Goodwin, C. 1981 *Conversational Organization. Interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, C. 1995 Co-Constructing meaning in conversation with an aphasic man. *Research on Language and social Interaction* 28: 233–260.
- Goodwin, C. 2000 Gesture, aphasia and interaction. In: D. McNeill (ed.): *Language and Gesture*, 85–95. Cambridge: Cambridge University Press.
- Goodwin, C. 2003 Conversational frameworks for the accomplishment of meaning in aphasia. In: Charles Goodwin: *Conversation and Brain Damage*, 90–116. Oxford: Oxford University Press.
- Haddington, P. 2006 The organization of gaze and assessments as resources for stance taking. *Text and Talk* 26: 281–328.
- Kendon, A. 1967 Some functions of gaze-direction in social interaction. *Acta Psychologica* 26: 22–63.
- Laakso, M. and A. Klippi 1999 A closer look at the “hint and guess” sequences in aphasic conversation. *Aphasiology* 13: 345–363.
- Light, J., B. Collier and P. Parnes 1985 Communicative interaction between young nonspeaking physically disabled children and their primary caregivers. Part III – Modes of communication. *Augmentative and Alternative Communication* 1: 98–107.
- Pennington, L. and H. McConachie 1999 Mother-child interaction revisited: communication with non-speaking physically disabled children. *International Journal of Language and Communication Disorders* 34: 391–416.
- Rossano, F. 2010 Questioning and responding in Italian. *Journal of Pragmatics*, 1–16.
- Sacks, H., E. A. Schegloff and G. Jefferson 1974 A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735.
- Selting, M. 2000 The construction of units in conversational talk. *Language in Society* 29: 477–517.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock and S. Uhmman. 2009 Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.





- Sidnell, J. 2006 Coordinating gesture, talk, and gaze in reenactments. *Research on Language and Social Interaction* 39: 377–409.
- Thiele, A. 1999 *Infantile Cerebralparese. Zum Verhältnis von Bewegung, Sprache und Entwicklung – theoretische Grundlagen einer frühen Förderung verbaler und nonverbaler Kommunikation*. Berlin: Edition Marhold.
- Von Tetzchner, S. and H. Martinsen 1996 Words and strategies. In: S. von Tetzchner and M. H. Jensen (eds.): *Augmentative and Alternative Communication: European Perspectives*, 65–88. London: Whurr.
- Wöllstein-Leisten, A., A. Heilmann, P. Stepan and S. Vikner 2005 *Deutsche Satzstruktur. Grundlagen der syntaktischen Analyse*. Tübingen: Stauffenburg.





III Multimodal corpora







*Jens Edlund, David House and Jonas Beskow
Department of Speech, Music and Hearing, School of Computer Science and
Communication, KTH, Stockholm*

Gesture movement profiles in dialogues from a Swedish multimodal database of spontaneous speech

1. Introduction

While speech and conversation analysis has traditionally been relatively confined to the audio speech signal and transcripts, there is an increasing demand for multimodal data which can capture the broader components of talk in interaction including facial gestures, body posture, manual gestures, head movement and the ongoing activities in the particular setting. Video recordings of varying quality have been available, but these are often lacking in the temporal and spatial resolution needed to enable investigators to carry out a fine-grain analysis. The availability of affordable high-definition video and motion capture equipment has now made it feasible to collect large databases of spontaneous speech recorded in high-quality audio and video plus the addition of motion capture. Recordings of this type will enable us to further our understanding of human-human talk in interaction and also help in the development of dialogue systems that are intended to be modeled on human-human interaction, for example interaction between humans and humanlike robots.

Our own research context is rooted in the technical development of audio-visual synthesis and the deployment of talking head agents in experimental dialogue systems (Beskow, Granström, and House 2007; Granström and House 2005). Fundamental research goals for this kind of application are to increase our understanding of how visual expressions are used to convey and strengthen the functions of spoken language and to further our understanding of interactions between visual expressions, dialogue functions and speech acoustics. The technological application of such knowledge is to be able to create an animated talking agent displaying realistic communicative behavior using multimodal speech synthesis. There has been very little data with which we can measure with precision multimodal aspects such as the timing relationships between vocal signals and facial and body gestures, but also acoustic properties that are specific to conversation, as opposed to read speech or monologue, such as the acoustics involved in floor negotiation, feedback, grounding and resolution of misunderstandings. We have



now remedied this in part by the completion of the Spontal database project in which more than 60 hours of unrestricted conversation comprising in excess of 120 dialogues between pairs of speakers has been recorded.

The Spontal corpus is rich enough to capture important variations among speakers and speaking styles to meet the demands of current research of conversational speech. In addition to comprising a resource for traditional audio and video conversational and phonetic analyses, the addition of motion capture data is expected to contribute to new ways of analyzing conversational behavior on both the macro (global) and micro (local) levels. On the macro level, automatic data processing enables us to look at properties of entire dialogues such as general participant movement and gesture patterns and how movement interactions may change over the course of the dialogue. On the micro level the motion-capture data combined with high-definition video and high-quality audio enables us to measure multimodal interactions such as between gesture and intonation with a much greater temporal and spatial resolution than with previously available video data. Furthermore, the size of the database enables us to study variability in the macro and micro dimensions.

One example of the previous use of motion capture in the analysis of the interaction between facial gestures and prosody is reported in Beskow, Granström, and House (2006). In this study, the data was limited to scripted speech produced by an actor (see Beskow et al. 2004) rather than spontaneous dialogue, but the study is methodologically relevant for the Spontal data. Results of an analysis of markers located on the face indicated that all markers underwent greater movement in syllables having prosodic prominence than in non-prominence bearing syllables. We thus found a clear correlation between facial movement and prosody. Moreover, the speaker varied expressive mode, and this was also evident from the movement data, for example more lip corner movement was associated with the happy mode and more head movement was associated with the confirming mode.

The aim of this contribution is two-fold. First of all, the chapter presents the Spontal database project including experiences and lessons learned over the three years spent in planning, organizing and carrying out the data collection. Secondly, we will present some examples of analysis on the macro level in which average participant movement is profiled over entire dialogues. These gesture profiles may correlate to certain grammatical elements of talk-in-interaction such as turn-taking dynamics, but they may also reflect interactional dynamics through the course of a dialogue not present in or readily extractable from the audio signal.



2. The Spontal Corpus

2.1. Research program

Spontal: Multimodal database of spontaneous speech in dialogue is a Swedish speech database project which began in 2007 and was concluded in 2010. It was funded by the Swedish Research Council, KFI – Grant for large databases (VR 2006–7482). The point of departure for planning the project stemmed from the fact that both vocal signals and facial and body gestures are important for communicative interaction and that signals for turn-taking, feedback giving or seeking, and emotions and attitudes can be both vocal and visual. Our understanding of vocal and visual cues and interactions in spontaneous speech is growing, but there is a great need for data with which we can make more precise measurements. A large Swedish multimodal database freely available for research will serve to enable researchers to test hypotheses covering a variety of functions of visual and verbal behavior in dialogue, and the overall goal of the project was to create a Swedish multimodal spontaneous speech database rich enough to capture speaker and speaking style variation comprised of high-quality audio and video recordings (high definition) and motion capture for body and head movements for all recordings. Previous reports from the Spontal project can be found in Beskow et al. (2010) and Edlund et al. (2010).

In addition to the main corpus, a number of smaller, specialized corpora have been recorded using basically the same configuration. This includes non-Swedish corpora, such as Spontal-N (Sikveland et al. 2010), a number of English recordings and single recordings in several other languages (e.g. Arabic and Italian); a number of dialogues where additional sensors are used to capture more data are in the planning. Part of the configuration was also used in the recording of the d64 database in Dublin (Oertel et al. 2010).

2.2 Database design and recording scenario

The specification of subjects and conversational task is a balance and trade-off between planning and spontaneity. As we wanted to collect recordings that were as spontaneous as possible, no initial task was demanded of the participants. Subjects were told that they were allowed to talk about absolutely anything they want at any point in the session, including meta-comments on the recording environment and suchlike, with the intention to relieve subjects from feeling forced to behave in any particular manner. They were also told to not feel any pressure to keep the conversation going at all times, and that in case they felt like it, it would be fine to sit in silence as well. The recording studio is equipped with an intercom system which subjects





were instructed to use to contact the recording staff should they feel inconvenienced. They were also told that they could stand up and leave at any point, should they feel the need to. No subject felt the need to use the intercom system, nor did anyone leave prematurely.

To place a minimum structure on the dialogues, the recordings were formally divided into three 10 minute blocks, although the conversation was allowed to continue seamlessly over the blocks, with the exception that subjects were informed, briefly, about the time after each 10 minute block. After 20 minutes, they were also asked to open a wooden box which had been placed on the floor beneath them prior to the recording. The box contained objects whose identity or function was not immediately obvious. The subjects could hold, examine and discuss the objects taken from the box, but they could also choose to continue whatever discussion they were engaged in or talk about something entirely different.

The main corpus contains 60 hours of dialogue consisting of 120 half-hour sessions. The subjects are all native speakers of Swedish. We wanted to keep subject variability as open as possible but still exert some control to ensure subject and dialogue variation. We therefore balanced the subjects (1) for gender, (2) as to whether the interlocutors are of opposing gender and (3) as to whether they know each other or not. This balance resulted in 15 dialogues of each configuration: 15x2x2x2 for a total of 120 dialogues.



2.3. Subject recruitment and ethics and privacy

Subjects for the recordings were recruited from friends, colleagues and through advertisements. The creating and maintenance of a booking system for scheduling recording sessions was a necessity especially concerning the balance of the subjects. All subjects signed a written agreement (1) that the recordings are to be used for scientific analysis, (2) that the analyses will be published in scientific writings and (3) that the recordings can be replayed in front of audiences at scientific conferences for illustration and demonstration purposes. Each subject was rewarded with two cinema tickets per recording session.

2.4. Technical specifications and illustrations

In the base configuration, the recordings are comprised of high-quality audio, high-definition video, and motion capture data. Figure 1 shows an overview of the recording studio, Figure 2 shows the studio at the beginning of a session and the recording setup is illustrated schematically in Figure 3.





Gesture movement profiles in dialogues from a Swedish multimodal database 269

Each subject is recorded on two microphones – a Bruel & Kjaer 4003 omni-directional goose-neck at a distance of one meter, and a head-mounted Beyerdynamic Opus 54 cardioid, which was replaced by a head-mounted Sennheiser ME 3-ew cardioid for the final 40 recordings. This combination of microphones is used to achieve optimal recording quality (Bruel & Kjaer), while ensuring that we have recordings with a high degree of speaker separation (Beyerdynamic/Sennheiser). Mixer consoles are used as microphone pre-amplifiers and to supply phantom power to the microphones. The output of the consoles is connected to an M-Audio interface and recorded with the free audio software Audacity in 4 channels 48 KHz/24 Bit linear PCM wave files on a 4 x Intel 2.4 MHz processor PC.

Two JVC HD Everio GZ-HD7 high definition video cameras (1920x1080i resolution, 26.6 MBps bitrate) were placed with a good view of each subject, approximately level with their heads. The cameras are set so that they capture in full view a person with arms reaching out to each side.

In addition, six infrared OptiTrack cameras from Naturalpoint capture the participant's head, torso, and arm movements. The OptiTrack data consists of unsorted point clouds, and requires post processing to use. Figure 4 shows the placement of motion capture markers in detail.



Figure 1: The recording studio. Video cameras are on tripods to the left and right in the picture. The turntable used for synchronization is visible just right of center, with the B&K microphones in microphone stands on either side. Motion capture cameras are barely visible, mounted along the walls just under the ceiling.





As each of the three systems used is susceptible to error, we take measures to ensure that the recordings can be synchronized even if frames are dropped or recordings partially lost due to for example hardware failure. The synchronization mechanisms evolved during the recordings as new obstacles appeared, but were unchanged from about 50 % of the recordings and onwards, as follows. A record player is included in the setup. The turntable is placed between the subjects and to the side, in full view of both video and motion capture cameras. A marker is placed on the turntable, which rotates with a constant speed (33 rpm). An LP record is also placed on the turntable. The record has been deliberately scratched at such a place that the pickup will hit the scratch each time the marker passes the pickup. The sound from the turntable is recorded on a separate fifth channel on the M-Audio interface. This setup enables high-accuracy synchronization of the frame rate of each of the systems in post processing (Edlund and Beskow 2010).

A similar analogue system is used to synchronize the start and end of the recordings. A box containing a green diode and an infrared diode is placed next to the turntable. When powered, the diodes light up and can be detected by automatic means both in the video and the motion capture. The power switch is in the control room, and is also connected to a sine tone generator.



Figure 2: The recording studio during the instruction phase of a session. Subjects are wearing headmounted microphones and head bands with Optitrack markers, and additional markers are fixed to torsos, arms and hands (Edlund and Beskow 2010).





Gesture movement profiles in dialogues from a Swedish multimodal database 271

When the switch is flipped, the diodes light up and the sine tone is forwarded to a sixth channel on the M-Audio interface. We use three blinks at the start and end of each session, and two blinks to signify the start of each ten minute block. One blink is used to show that the recording staff has left audio comments on the sixth channel, which doubles as a synchronization channel and an acoustic notebook.

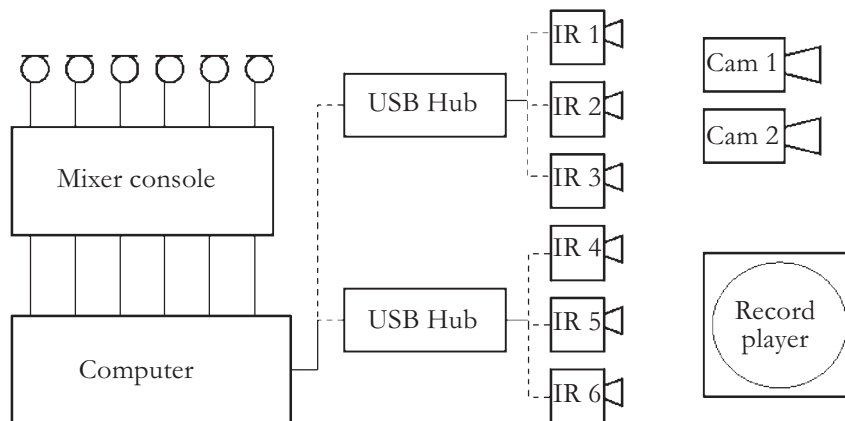


Figure 3. A schematic illustration of the recording setup.



Figure 4: Each subject wears markers on hands, wrists, elbows, shoulders, sternum, and three markers mounted on a head-worn tiara.





2.5. Annotation

The Spontal database is partially transcribed orthographically with the expectation of complete transcription in the near future. The orthographic transcription includes orthographic words as well as labels for events such as breathing, coughing, laughing, and interactional tokens for turn management and feedback, such as *eh*, *hm*, *okay*, *uh-huh*. Care is taken to make the transcriptions quick to do and at the same time as consistent as possible. In order to ensure speed and reliability, we use the in-house dialogue annotation tool seen in Figure 5, the Higgins Annotation Tool (<http://www.speech.kth.se/hat/>), which is based on the Snack Sound Toolkit (<http://www.speech.kth.se/snack/>).

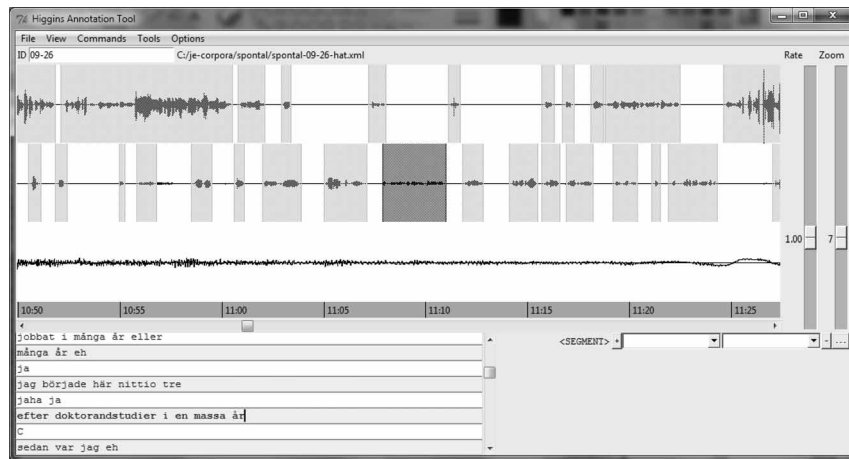


Figure 5. Screen shot of the Higgins Annotation Tool with a presegmented Spontal dialogue loaded.

Automatic methods are used wherever possible. The first step is to segment the audio into speech and non-speech segments in each of the speaker channels. We accomplish this automatically using the technique described in Lasowski (2011) with a step size of 100ms and minimum speech and non-speech durations of 200ms.

Annotators then transcribe the speech chunks. In the next stage, the transcriptions will be fed to a forced alignment system which provides rudimentary phonetic transcriptions with onset and offset times. These are more reliable and more precise than the original speech/non-speech labels, and will be used to modify the time stamps.





Gesture movement profiles in dialogues from a Swedish multimodal database 273

Video will also be segmented and mapped to the speech data using automatic methods developed within the Spontal project. The start and end times are found by automatically locating frames where the green diode is lit. The position of the participants' heads is pointed out manually, but in a manner that makes it very quick: the annotator is shown an average still picture of a number of pictures extracted from the video, which makes it simple to point out where a participant's head was during most of the time in the recording (see Figure 6). The head position information is fed to a video processor which extracts close-ups and smaller low-resolution videos that are used for overviews.



Figure 6. Averaged still picture from one of the video recordings from a Spontal dialogue. The average is based on roughly 30 stills, one from each minute of the recording.

The motion capture data will be treated in a similar manner. It is, however, not part of the original Spontal project, and its processing will take longer to finish. Higher-level manual annotation is outside the scope of the project. Parts of the data are however already being used in several newly inaugurated research projects, which will result in various annotations. Researchers using the data once it is available are naturally encouraged to share their annotations as well.





2.6. Experiences gained and some lessons learned

During our recording experience so far, there have been a number of more or less unexpected technical challenges which have been overcome. These include selection and installation of a suitable light source for the video recordings which does not interfere with the motion capture cameras; shiny objects, eyes and eyeglasses which create spurious reflections interfering with the motion capture data; problems with USB power for the motion capture system; and synchronization problems finally solved by the use of the turn-table. There have, fortunately, been no problems at all for our subjects to speak spontaneously and unplanned, or to come up with topics for their interactions. Our initial examinations of the data reveal that people speak about a great variety of things, from what seems like regular work meetings through surprisingly open-hearted gossip to common dinner table topics such as “where did you grow up” and “what do you do for a living”.

Dividing the half-hour sessions into 10 minute sections and informing the subjects about the elapsed time after each section reassures the subjects that all is proceeding well, and the introduction of the box after 20 minutes creates diversity in the conversational topics, but the general impression of our transcribers so far is that all the dialogues are natural-sounding and of high quality. It seems that the subjects quickly become rather unaware of the audio, video and motion capture equipment and busily proceed with their dialogues. As mentioned, subjects were informed that they may end a session at any time should they feel uncomfortable. They were also informed that should they feel that they had said something inappropriate, they could (at a later date) go through the material together with the recording staff and delete the offending utterance. Nobody chose to do either. Finally, a large portion of subjects asked without prompting if they could participate again. Taken together, we take these observations as an indication that our subjects felt relaxed and untroubled by the whole recording situation, and we have good hope that the recorded data is representative of conversations people have in their everyday lives.

3. Gesture movement profiles

Similarity or dissimilarity between interlocutors in terms of such characteristics as prosody, dialect, sociolect, gestures and posture is a process that can be established and that can develop during a dialogue. Keller and Tschacher (2007) employ an automatic approach to the analysis of video to quantify coincidental head movements (synchrony) between therapist and patient dur-





ing psychotherapeutic sessions. Their approach employs Motion Energy Analysis (MEA) which is based on an image differencing algorithm that takes into consideration differences in the grey-scale distribution changes between subsequent video frames. By using motion capture data collected in the Spontal project, we are able to perform a similar automatic analysis of movement and produce a number of interesting and useful measurements enabling us to graphically plot dialogue profiles over time. This type of visualization can contribute to new ways of analyzing dialogue behavior and new insights into dialogue dynamics by revealing how the degree and range of movement by interaction partners can converge or diverge over time.

3.1. Method

As an illustration of how motion capture data can be used – even with very little processing – we picked several Spontal dialogues at random and prepared the motion capture data as follows. For each frame, any data point whose position was within 0.5 meters from a point between the participants was discarded, creating a blind area of one meter width right between the speakers. This was done as a precaution to exclude the possibility that a motion capture track gets confused with another between the speakers. This precaution is well motivated for this type of study, as mixing tracks between speakers would instantly create artificial similarities. This treatment has a relatively small effect on the data, since the speakers are seated further apart than one meter, and the only frames that are lost are those where a speaker leans heavily forward.

The remaining data was divided into two tracks – the points to the left of the blind area constituting one track, and the points to its right the other. The points in each track were averaged to create a center point for each track and frame. For each sequence of two frames, the Euclidean distance between the center points in each track was calculated. The two resulting tracks were subjected to a five point median filter to remove extraordinary movements due to ghost markers showing up and then disappearing. The result does not capture certain types of movement – particularly it models symmetrical arm movements poorly, but it is nevertheless a measure of average marker movement in the time between the two adjacent frames, per speaker. As mentioned before, the motion capture equipment aims at 100Hz, but this varies somewhat. However, as the synchronization between the cameras is highly accurate, we do not need to worry – each frame contains data sampled at the same time, which is all we need know for a rough investigation of the movement profiles.





Next, the two resulting time series of speaker average marker movement were down-sampled to pick an average over 30 seconds every 30 seconds, resulting in heavily smoothed data. Finally, the data was plotted in graphs. Note that although the unit of the Y-axis is nominally cm/s, the absolute amount of movement has little meaning and is very hard to interpret. The metric is better viewed as a relative measure, interpreting the relationship between the amount of movement between the interlocutors and its variation over time.

3.2. Profile types

A visual analysis of the profiles of a few of the dialogues shows a rough grouping into three basic types of movement profiles. There frequently seems to be a marked asymmetry between the ranges of motion of the two interlocutors. One of the speakers generally exhibits more motion activity throughout the dialogue than the other one.

In Figure 7, speaker B (right) clearly displays more movement than does speaker A (left) who shows very little movement activity throughout the dialogue. In this profile, it is interesting to note that speaker B displays what could be an adaptation to speaker A's passivity of motion showing a marked decrease of motion as the dialogue progresses. Noteworthy is also the fact that after the box has been examined (the blank interval), speaker B displays more motion again comprising a kind of resetting, although here the range of motion does not reach the same level as at the beginning of the dialogue.

A qualitative auditive analysis of this dialogue reveals that speaker B is indeed much more active in the dialogue than speaker A. Speaker B is involved in the Spontal project and well acquainted with the technical setup. This is a dialogue between two males who have not met before. During the first five minutes of the dialogue, Speaker A mainly asks short technical questions to which speaker B responds explaining the technical details largely in a monologue style without much spoken response from speaker A. About five minutes into the dialogue, however, speaker B finishes his explanations and begins asking speaker A questions about his participation in research experiments. Speaker A responds with short answers and with increasing engagement in the conversation. During the time interval 15–20 minutes into the dialogue, the basic dialogue characteristics have changed from monologue to turn-taking with a more balanced speech activity between the two dialogue partners. This appears to correspond to the decreased motion exhibited by B which can be characterized as adaptation to A's lack of motion. After the box has been examined, speaker B increases his contribution to the





dialogue with explanations about the box and its contents in much the same style as at the beginning of the dialogue. This appears to correspond to the increase in motion activity by this speaker after the box has been examined.

In Figure 8, we see an alternation of asymmetry of motion between the two dialogue partners with speaker D (right) showing considerably more motion than speaker C (left) at the beginning of the dialogue and in the interval before the box is opened. In contrast to the dialogue in Figure 7, there are six peaks of motion which are very well synchronized between the two speakers. These are located roughly at 5 minutes, 7 minutes, 9 minutes, 15 minutes 18 minutes and 25 minutes into the dialogue. These areas of motion synchronization probably comprise dialogue transitional areas characterized by the laughter of both participants. However, until the motion capture data is finally processed and synchronized with the audio and video tracks we cannot be sure of this. At the end of this dialogue we can observe symmetry in degree of motion activity but asymmetry of motion peak timing following the examination of the box.

The dialogue displayed in Figure 8 is between two females participants who know each other. The participants contribute much more equally in this dialogue than in the previous one. The dialogue begins with general commentary about the recording situation with concomitant laughter, but quickly turns to planning a trip together and then (from 5 to 10 minutes into the dialogue) D asks C questions about her job and a recent Christmas party. C responds during these five minutes with an engaged account of the activities at the party accompanied by frequent feedback and comments by D. During this interval (from 5 to 10 minutes into the dialogue) C exhibits slightly more motion activity than D and there are three peaks of synchronization. From 10 to about 15 minutes into the dialogue, D continues to ask C questions with D being very responsive with feedback, follow up questions and turn-taking. At about 15 minutes into the dialogue, the roles switch when D starts talking about a friend who has just had a new baby. This passage corresponds to D's two peaks of motion at about 15 and 17 minutes which also show synchronization with two of C's motion peaks. Following the opening of the box, the participants talk about baking with sour dough, and the conversation is characterized by very rapid turn-taking, quick questions and responses, and lots of speaker overlap. This corresponds to the final passage in Figure 8 (following the blank interval) where both speakers have the same motion activity, but only one peak of motion synchronization.

In Figure 9, speaker F (right) shows generally higher levels of motion than speaker E (left). More noteworthy, however, is the fact that these two speakers show a great amount of motion synchrony with at least seven peaks of





motion coinciding approximately at minutes 8.5, 9.5, 13, 16.5, 18, 19, and 22.5. It is also interesting that the motion synchrony increases as the dialogue progresses showing evidence of adaptation between the interlocutors. There is even a correlation of 0.56 between the two curves displayed in Figure 9.

The dialogue plotted in Figure 9 is between two female participants who do not know each other. The dialogue begins with the participants introducing themselves to each other. E asks F about her job and F responds with an explanation about her job during the initial three minutes of the dialogue. Then they ask each other where they are from and discuss dialects. This part of the dialogue is characterized by polite questions and responses and lasts until about 8 minutes into the dialogue. At this point, F asks E about her studies, and they discover that they have a mutual acquaintance. This leads to more rapid turn-taking and seems to correspond to the first two peaks of synchronization at 8.5 and 9.5 minutes. This leads E to talk about her previous job at the airport, with F relating to this with a story about a pilot. After the box has been opened, the conversation focuses on the contents of the box and is characterized by a balance between the participants, questions and responses and turn-taking.

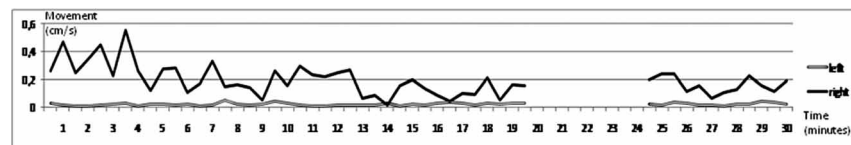


Figure 7: Average marker movement over time for speaker A (left) and speaker B (right) for one of the Spontal dialogues illustrating asymmetry of range of motion between speakers with some adaptation on the part of speaker B.

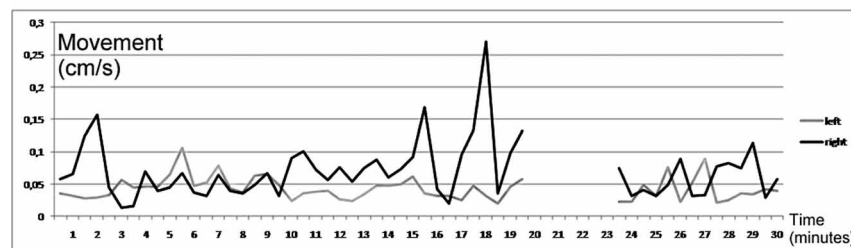


Figure 8: Average marker movement over time for speaker C (left) and speaker D (right) for one of the Spontal dialogues illustrating alternation of range of motion between speakers and peaks of synchronization.





Gesture movement profiles in dialogues from a Swedish multimodal database 279

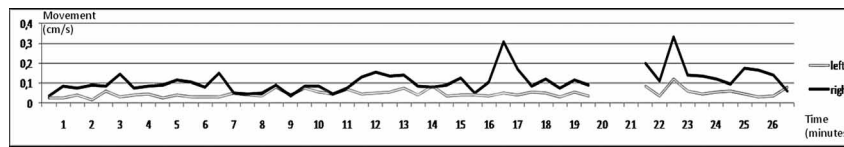


Figure 9: Average marker movement over time for speaker E (left) and speaker F (right) for one of the Spontal dialogues illustrating increasing movement synchronization between speakers over time.

4. Conclusions

The planning, organization and data collection carried out within the Spontal project has demonstrated the feasibility of collecting large multimodal databases of spontaneous dialogue which can be investigated in a variety of novel ways to help us expand our window of analysis on both the macro and micro levels. Face-to-face interaction is by its very nature multimodal with both speech and movement contributing as signals for interaction, and there has been a lack of substantial multimodal datasets. In particular, audio and video alone may not be sufficient to allow us to investigate in detail the relations between gesture, speech, and facial expressions, or to enable us to investigate variability and conventions in gestures in interaction. Adding motion capture provides precision data in three dimensions that we think will prove valuable in this respect and help us understand what aspects of gesture and motion belong to language and grammar and what aspects belong to the interactional organization of spoken language.

As the use of motion capture in speech research is as of yet uncommon, we have presented here some of our experiences in recording face-to-face interactions with audio, video and motion capture. When the data processing is completed, such visualization tools as the gesture movement profiles exemplified here will enable researchers to examine specific locations of motion coupled to the video and audio recordings. It will also, for example, be possible to investigate areas of movement phenomena across many dialogues opening up the possibilities of generating new research questions in studying talk in interaction.

Acknowledgments

The work presented here is funded by the Swedish Research Council, KFI – Grant for large databases (VR 2006–7482) and Humanities and Social Sciences (VR 2009–1764) “Question Intonation in Swedish.” It is performed at KTH Speech Music and Hearing (TMH) and the Centre for





Speech Technology (CTT) within the School of Computer Science and Communication.

References

- Beskow, J., L. Cerrato, B. Granström, D. House, M. Nordstrand and G. Svanfeldt 2004 The Swedish PF-Star Multimodal Corpora. In: *Proceedings LREC Workshop on Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces*. Lisboa, 34–37.
- Beskow, J., J. Edlund, B. Granström, J. Gustafson and D. House 2010 Face-to-face interaction and the KTH Cooking Show. In: A. Esposito, N. Campbell, C. Vogel, A. Hussain and A. Nijholt (eds.), *Development of Multimodal Interfaces: Active Listening and Synchrony*, 157–168. Berlin/Heidelberg: Springer.
- Beskow, J., B. Granström and D. House 2006 Visual correlates to prominence in several expressive modes. In: *Proceedings of Interspeech 2006*. Pittsburg, PA, 1272–1275.
- Beskow, J., B. Granström and D. House 2007 Analysis and synthesis of multimodal verbal and non-verbal interaction for animated interface agents. In: A. Esposito, M. Faundez-Zauny, E. Keller and M. Marinaro (eds.), *Verbal and Nonverbal Communication Behaviours*, 250–263. Berlin: Springer-Verlag.
- Granström, B. and D. House 2005 Audiovisual representation of prosody in expressive speech communication. *Speech Communication* 46: 473–484.
- Edlund, J. and J. Beskow 2010 Capturing massively multimodal dialogues: Affordable synchronization and visualization. In: M. Kipp, J.-C. Martin, P. Paggio and D. Heylen (eds.), *Proceedings of LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality (MMC 2010)*, Valetta, Malta, 160–161.
- Edlund, J., J. Beskow, K. Elenius, K. Hellmer, S. Strömbergsson and D. House 2010 Spontal: A Swedish spontaneous dialogue corpus of audio, video and motion capture. In: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, M. Rosner and D. Tapias (eds.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Valetta, Malta, 2992–2995.
- Keller, E. and W. Tschacher 2007 Prosodic and gestural expression of interactional agreement. In: A. Esposito, M. Faundez-Zauny, E. Keller and M. Marinaro (eds.), *Verbal and Nonverbal Communication Behaviours*, 85–98. Berlin: Springer-Verlag.
- Laskowski, K. 2011 Predicting, detecting and explaining the occurrence of vocal activity in multi-party conversation. PhD Thesis CMU-LTI-11-001, Carnegie Mellon University.
- Oertel, C., F. Cummins, N. Campbell, J. Edlund and P. Wagner 2010 D64: A corpus of richly recorded conversational interaction. In: M. Kipp, J.-C. Martin, P. Paggio and D. Heylen (eds.), *Proceedings of LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality (MMC 2010)*. Valetta, Malta.
- Sikveland, R.-O., A. Öttl, I. Amdal, M. Ernestus, T. Svendsen and J. Edlund 2010 Spontal-N: A corpus of interactional spoken Norwegian. *Proceedings of LREC*. Valetta, Malta.





Patrizia Paggio
University of Copenhagen
Centre for Language Technology (CST)

Towards an empirically-based grammar of speech and gestures

1. Introduction

The purpose of this article is to discuss how non-verbal behavior, in particular head movement and facial expressions, can be represented in a multimodal grammar. The term grammar is used here in a rather broad sense to indicate not only syntax, but all aspects of language structure, and we follow Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994) in conceiving of the grammar of a language as a system of constraints operating at various levels (phonology, morphology, syntax, semantics). We extend this notion by talking about multimodal grammar, which we define as the system of constraints that models the interaction of speech with non-verbal behavior in language. Still following HPSG, we use typed feature structures to model grammatical constraints. Only, in our case, constraints relate to the shape and dynamics of gestures, their possible interpretation and their relation to speech. In particular, we focus on three issues: i. the relation between non-verbal behavior and speech; ii. the expression of feedback through gestures and iii. the contribution of gestures to information structure. Our analysis is based on Danish multimodal data annotated according to the MUMIN gesture coding scheme¹. The scheme and its application to data in several languages, as well as the use of such annotated multimodal data for machine learning, are described in detail in Paggio and Diderichsen (2010). Here, we are interested in how the various gesture types in the annotated data can be represented in a grammar, and how the empirical findings comply with theoretical assumptions about how gestures interact with speech.

To start with, why should grammar be concerned with gestures, and why are gestures difficult to deal with?

Human communication is *situated* in the human body: we cannot avoid using our face, hands and body while we speak, and in face-to-face conversation we clearly react not only to our interlocutor's words but also to their

¹ MUMIN was originally the name of a Nordic network on Multimodal Interfaces, see <http://www.cst.dk/mumin/>.



gestures². A possible cognitive explanation of this close relationship between speech and non-verbal behavior may be that language emerged millions of years ago on top of our ancestors' ability to interpret and replicate gestures, so that speaking and gesturing partly depend on the same neurological mechanisms (Arbib 2005). From this interdependence between the two communication modalities follows the fact that considering speech in isolation, does not do full justice to the way human communication actually works. This is a conclusion that many researchers have already come to, indicating a growing interest for gesture studies as a discipline as witnessed in a wealth of recent publications (Kendon 2004; McNeill 2005; Duncan, Cassell, and Levy 2007; Poggi 2007; Cienki and Müller 2008; Gullberg and de Bot 2010), as well as specialized associations and conferences (the International Society for Gesture Studies, Multimodal-Corpora.org) and journals (*Gesture*, Benjamins publishers).

However, speech and gestures are very different in nature. This difference, paired with the complexity of their interaction, constitutes a challenge for the definition of a grammar of speech and gesture.

First of all, since gestures are largely non-conventionalized, a fact that in turn depends on their essentially indexical and iconic rather than symbolic nature (Allwood 2008), they are not only ambiguous at the content level, but also largely unpredictable at the expression level. In other words, they show more or less open-ended variation in terms of shape and dynamics. Attempts have been made to categorize hand gestures into meaningful types. Kendon (2004) describes for instance iconic hand gesture types – so-called gesture families – that share common physical features. Similarly, Kipp (2004) posits a lexicon of metaphoric and iconic hand gestures where the content of the gesture is associated with an array of possible realisations in terms of shape and dynamics. The assumption here is that prototypical gestures can and should be described together with the possible physical realisations they can receive due to individual variability, cultural factors, affect, etc. Promising as these approaches may be, it remains to be shown that developing a lexicon of hand gestures capturing the expressiveness of a large population of speakers is a feasible enterprise.

If we now turn our attention to head movements and facial expressions, which are the focus of this paper, the notion of a closed list of possible types of gestures that captures a potentially open-ended list of realisations in a

² We use *gesture* as a synonym of non-verbal behavior in general, not only hand gestures, and unless otherwise specified, we refer to the whole gesture phrase including preparation and retraction.





meaningful way, seems less problematic. Most people know intuitively what a nod, a head turn or a smile are, even though each of these gesture types can be realized with varying degrees of intensity, amplitude and so on. Our grammar and our coding scheme operate therefore with a list of possible expression categories: e.g. *nod*, *turn* for head movements and *laugh*, *smile* for facial expressions.

However, as we will see below, these categories are not necessarily easy for the annotators to use in practice, and the content of these gesture types varies depending on the context.

Another difficulty that a grammar of speech and gestures must cope with is the interaction between speech and non-verbal behavior. This interaction occurs at different levels, from prosody to pragmatics (McNeill 1992), and involves speech units of different granularity. In the simplest case, gestures overlap with single words or syllables. However, we also find combinations of more complex hand gestures³ as well as single facial expressions that overlap with longer linguistic contributions not always corresponding to syntactic phrases.

Last but not least, a multimodal grammar must be able to represent not only the shape and dynamics of gestures, but also their content or the pragmatic function they have in communication. Depending on their semiotic type, gestures contribute meaning at different levels in the semantics of the word, chunk, sentence or maybe discourse segment they relate to. In short, a multimodal grammar must cope with segmentation and representation problems. In other words, which segment of speech should a specific gesture be associated with? Is this relation constrained? How should the shape and content of gestures be modelled? And what representation should be given to the integrated multimodal contribution? We will discuss some of these issues based on our empirical data.

We start in Section (2) by proposing a typology of non-verbal behavior and showing how the contribution of each type to the grammar can be modelled by typed feature structures. In Section (3) we describe the recordings and the annotation procedure, and give an overview of the annotated gesture types in the data. Then in Sections (4), (5) and (6), we discuss what we can learn from the data with respect to what speech segments gestures relate to, how they are used to express feedback, and how they contribute to the information structure of sentences. Section (7) contains the conclusions.

³ In the literature also called gesture phrases, i.a (Kendon 2004; Kipp 2004).



2. Types of non-verbal behavior

The types of non-verbal behavior we use in our annotations and analyses are defined in the MUMIN annotation scheme. The first dimension along which we categorize non-verbal behavior, concerns the communication modality involved. Relevant modalities are *head (movement)*, *face (expression)*, *hand (gesture)* and *body (posture)*. Under each type, we define a number of categories corresponding to prototypical expression types. For head movements, possible types are *nod*, *turn*, *jerk*, *headBackward*, *headForward*, *tilt*, *sideTurn*, *shake*, *waggle* and *headOther*. For facial expressions, we posit the types *smile*, *laughter*, *scowl* and *faceOther*. In both cases, the *Other* feature is used when none of the other categories fits the data. Neither body posture nor hand gestures have been annotated in the data yet, so we will not propose a classification for these modalities.

In the annotation scheme, each gesture type comes with a number of possible features describing it in more detail. The use of features to represent properties of gestures is not new. Previous proposals have suggested for example that feature structures are a convenient and elegant way of representing the unimodal content of individual modalities as well as their integration, for instance for parsing purposes (Johnston et al. 1997; Paggio and Jongejan 2005; Navarretta and Paggio 2009). More recently, Alahverdzhieva and Lascarides (2010) have proposed typed feature structures to represent the shape of hand gestures. Here we adopt a similar approach for head movements and facial expressions. In the typology of modality types, each modality introduces a number of features that get instantiated in the representation of specific examples.

$$\left[\begin{array}{l} \textit{nod} \\ \text{HEAD-MOVEMENT } \textit{nod} \\ \text{HEAD-REPETITION } \textit{repeated} \end{array} \right]$$

Figure 1: Feature structure representation of a repeated nod

$$\left[\begin{array}{l} \textit{smile} \\ \text{FACE-EXPRESSION } \textit{smile} \\ \text{EYEBROWS } \textit{raise} \end{array} \right]$$

Figure 2: Feature structure representation of a smile with raised eyebrows

Figures (1) and (2) show two such instantiated feature types – the first step in the definition of our multimodal grammar. As is customary, square brackets delimit the list of features (or attribute-value pairs) that define a feature





Figure 3: Annotation of a nod

structure. The type of the feature structure is indicated in italics at the top of the structure. In Figure (1), the feature structure is of type *nod*, in Figure (2) it is of type *smile*. Each type is associated with a specific set of attributes. (Only the shape attributes are shown here). In the case of a nod, the attributes are HEAD-MOVEMENT with value *nod* and HEAD-REPETITION with value *repeated* (the alternative is *single*). In the case of a *smile*, the attributes are FACE-EXPRESSION with value *smile* and EYEBROWS with value *raise* (other possible values are *frown* and *browsOther*).

An example of a nod from our corpus is shown in Figure (3): the frame picture shows the woman to the left with her head lifted, just after having completed a nod, whilst the editing window to the right shows the annotated attributes, which in addition to HEAD-MOVEMENT and HEAD-REPETITION, include a link to the corresponding word (*okay*) and feedback attributes (we will return to these shortly).

Figure (4) shows a smile with raised eyebrows also performed by the woman to the left, with the relevant annotated attributes in the editing window.

There can of course be many more shape attributes, depending on the granularity of the description to be achieved. Dimensions concerned with the amplitude or velocity of a head movement, for example, or the position of the eyes or the mouth in a facial expression, are clearly relevant to communication and may be added. However, if we conceive of these attributes not only as dimensions in a grammar, but also as elements in an annotation, issues of reliability and cost-effectiveness make coarse descriptions based on a limited number of elements more appealing. The attributes mentioned above are those currently annotated in our corpus.



Figure 4: Annotation of a smile

In gesture studies, several classifications of gestures (most often in fact hand gestures) have been introduced, based on the relation of gestures to meaning, their relation to speech, their degree of arbitrariness or iconicity and their pragmatic functions. Following Allwood (2008), we have chosen to classify gestures using Peirce's three semiotic categories *indexical*, *iconic* and *symbolic* (Peirce 1931), which are defined by looking at the way the connection between a sign and its object is established. The type *symbolic* characterizes gestures whose denotations are established by means of an arbitrary conventional relation. *Iconic* is assigned to gestures that denote their objects by similarity – whether concrete or abstract (metaphoric). Finally *indices*, according to Peirce, have a real and direct connection with the objects they denote: for instance, smoke is an index of fire. In MUMIN we sub-divide indexicals into *deictic* gestures, which point to an entity in the conversational situation and *non-deictic* gestures. Non-deictics comprise i. *displays*, that are expressions of emotions and attitudes; ii. *beats* (also sometimes called *batonic*), rapid movements of the hands or head that underlie the rhythm of the corresponding utterance and are used to express emphasis; iii. other indexical gestures with *interactive* function, e.g. head movements used to give and elicit feedback or to support turn management and discourse segmentation.



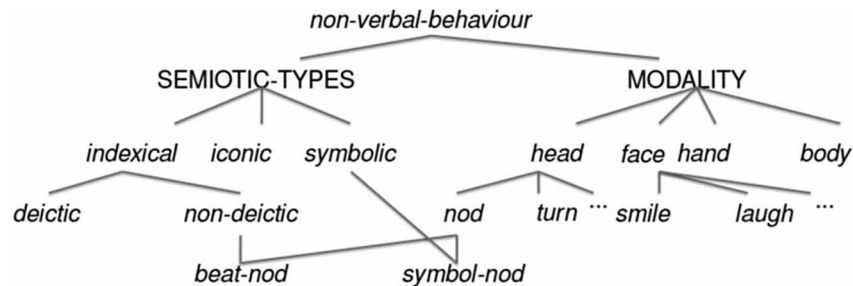


Figure 5: A hierarchy of non-verbal behavior types

The two dimensions along which non-verbal behavior can be characterized are brought together in a unified typology in Figure (5). Specific gesture types are to be found in the intersection between semiotic categories on the one hand, and modality types on the other. There is of course no one-to-one correspondence between the two dimensions: for instance, a nod can be a symbol when it corresponds to an acceptance or agreement act, or it can function as a beat that accompanies a stressed word. Both possibilities are indicated in the typology. An annotator will have to assign an interpretation taking the context into account. In the example shown earlier in Figure (3), for instance, the repeated nod functions like a beat rather than a symbolic sign of agreement, since it neither shows agreement to a statement nor positive evaluation of a question. The two dialogue participants are in fact introducing each other, and the woman's repeated nod comes just after the other speaker has told her his name. The short dialogue excerpt is shown below:

Speaker A: HEJ jeg hedder HANNE
 Speaker B: jeg hedder JESPER
 Speaker A: OKAY så fik vi LIGE det på PLADS

(hi my name is Hanne
 my name is Jesper
 okay so we got that right)

As she pronounces her second turn *okay so we got that right*, Speaker A nods twice. The nod strokes accompany the first two stressed syllables (in small caps) in the utterance.

In fact, one and the same gesture can sometimes be interpreted at different levels: an iconic hand gesture, for example, can be used as a beat at the same time. This is a problem for an annotator if the semiotic types are dis-





joint as proposed here. In other words, either iconic or beat will have to be chosen for the annotation in case of ambiguity between the two interpretations. A possible solution is to define a preference order between the values, e.g. always choose symbolic if appropriate, otherwise iconic and only as a last possibility indexical.



Figure 6: Deictic gesture in a TV debate

An interesting question is whether each semiotic type is associated with specific content features. To start with indexicals, according to Peirce they establish a relation with the objects they refer to not by interpretation, but by force of reality: when something burns, smoke indicates fire whether we realize what is happening or not. Similarly, a pointing gesture implies an entity as an object of the pointing event, whether the entity is visible to the listener or not. In Figure (6), we see a frame from a TV debate in which president Obama addresses an interlocutor somewhere in front of him with both hands while saying *I'm talking to you Joe*. Although we do not see Joe, we have no difficulty in accommodating the reference. The relevant issue here is how an index should be represented in the semantics. In the case of a deictic, it is not difficult to imagine that the content of the gesture would contain a referential index to some discourse referent. This could be a nominal index or an event index. But what are beats, displays and other indexicals indices of?

Let us start with beats. These are short gestures that underlie the accents in the speech stream, without, however, necessarily coinciding temporally with stressed syllables⁴. The hypothesis we want to put forward concerning

⁴ This is observed i.a. by McClave (1994), who notes that beats often begin well before the stressed syllable they are perceived to accompany.





beats is that they indicate focal elements in the speech modality, a relation that can be modelled by having the gesture introduce an index which is ultimately structure-shared with an element in the focus domain of the corresponding speech segment. To represent this relation, we will use the attribute COMM(UNICATIVE)-FUNCTION, which is used in the MUMIN coding scheme to annotate a number of communicative phenomena such as feedback, turn management and sequencing, and which we use here also to represent the function of beats.

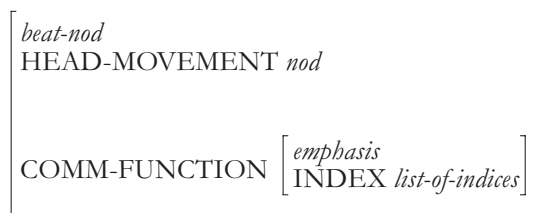


Figure 7: The content of a beat

Thus, in Figure 7, the semantic contribution of a head nod with a beat function is represented by means of the attribute COMM-FUNCTION. The value of this attribute is again a feature structure (enclosed by embedded square brackets) of type *emphasis*, and it introduces an attribute INDEX with a list of indices as its value. Note that the representation proposed here is intended as a generic framework: it shows the kind of relation that a nod can have with the related speech segment. We will see further below how it is applied to the analysis of specific examples of multimodal signs, where the indices are actually structure-shared with referents in the representation of the corresponding utterance.

Analyzing and representing displays in a unified manner is more difficult. Displays seem to indicate attitudes or reactions (e.g. feedback) towards whole states of affairs. In a way, they resemble adjectives and adverbials, which have scope over predications. In our data, we have especially focused on the analysis of feedback, the communicative mechanism through which participants in a conversation exchange information about the three basic communicative functions defined in Allwood et al. (1992: 1), i.e. *contact*, the fact that participants are willing and capable of continuing to interact; *perception*, the fact that they are willing and capable of perceiving what is being communicated; finally *understanding*, the fact that they are able to understand the message that is being communicated. Since in most cases we do not believe it is possible to distinguish one aspect of feedback from the others, we



use the term *cpu*, which stands for *communication, perception and understanding*, as the basic feedback value. Moreover, feedback also has a *direction* depending on whether it is being given or elicited, and it can be associated with an *emotion* or an *attitude*. For example, a nod can be used to give feedback and show agreement, a smile can be used to give or elicit feedback and show interest.



Figure 8: The content of a smile

Figure 8 again shows the generic representation of the content of a gesture, this time a smile that functions as a feedback-giving display, for instance the smile we saw in Figure 4. The FACE-EXPRESSION value is *smile*, and the value of COMM-FUNCTION is a feature structure of type *feedbackGive*, which is characterized first of all by the two features FEEDBACK *cpu* and F-DIRECTION *feedbackGive*. The value of the third attribute, ATTITUDE, is not instantiated: the type *attitude-value* stands for an open list which includes e.g. *interested*, *uninterested*, *agree*, *nonAgree*, *amused* etc. Finally, the attribute FEEDBACK-ARG is intended to express the relation between the gesture and the speech segment to which the same gesture provides feedback. It takes as its value a *handle*, a type used to link scopal elements to predications in Minimal Recursion Semantics (Copestake et al. 2005), a formalism for computational semantics that can easily be integrated with HPSG grammars. The idea, which will not be developed in detail here, is to use this *handle* as a kind of index to link the gesture to the interlocutor's utterance to which it constitutes feedback.

Let us now turn to the semantics of iconic gestures. Iconics are not the focus of this paper, since they essentially correspond to hand gestures, so we will refer to previous studies that have proposed a feature-based representation of their content. Alahverdzhieva and Lascarides (2010) call them *depictive* and suggest that they should be represented as a set of elementary predications including all the physical gesture features, e.g. *hand-shape-open*, *palm-orientation-flat*, etc. Paggio and Jongejan (2005) also use features, but the





values they propose constitute an interpretation of the gesture rather than a list of physical descriptions. For a gesture expressing the concept of *smallness*, for example, the content would be something like [CONT|PRED *small*]. The advantage of this type of solution is that it can be unified with the semantics of the corresponding word, a process that can be used to limit ambiguity of interpretation in both modality channels. Certainly, since an iconic gesture usually constitutes a direct comment to some semantic aspect of the message communicated in the speech modality, the grammar must be able to integrate the content features from gesture and speech into one unified structure.

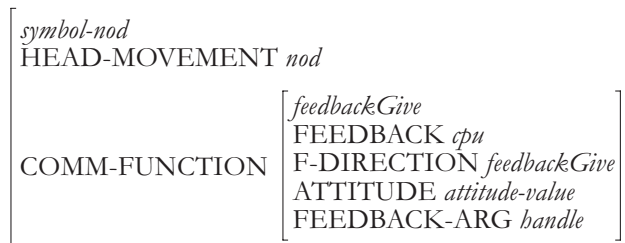


Figure 9: Feature structure representation of an affirmative nod

If iconic gestures typically add meaning to single words and predications, symbolic gestures (also called emblems) often correspond to entire propositions, perhaps speech acts. A symbolic gesture that occurs often in our corpus is the affirmative nod. Its counterpart is the denying shake. An affirmative nod can have several meanings, e.g. it can express a positive answer to a question, an acceptance of a request, agreement to a statement (Paggio and Navarretta 2010; Navarretta and Paggio 2010), or receipt of a prior utterance as news (Whitehead 2009). We show in Figure 9 how the agreement meaning can be represented by instantiating the value of the ATTITUDE attribute to *agree*. Apart from this, the COMM-FUNCTION feature is the same as for the smile in Figure 8.

Figure 10 sums up the different types of communicative function discussed so far in a feature structure hierarchy. The hierarchy shows the relevant attribute-value pairs for each type. The type *emphasis* was used above to represent the content of beats, while the types *feedbackGive* and *feedbackElicit* were used for various kinds of indexical and symbolic head movements and facial expressions. Other communicative functions foreseen by the MUMIN coding scheme, which, however, are not discussed here (and hence do not appear in the hierarchy), are *turn management* and *sequencing*.



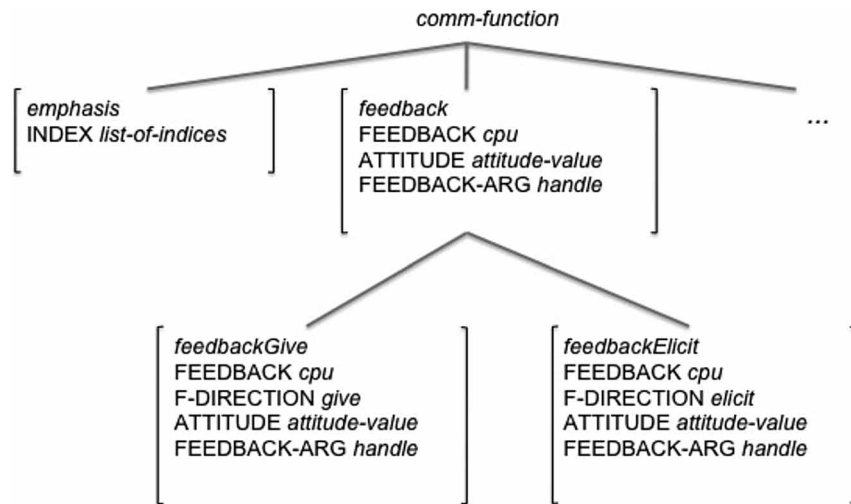


Figure 10: A hierarchy of communicative functions for gestures

Having introduced our gesture typology and proposed generic representations for some of the types, in the next section we look at their occurrence in naturally-occurring data.

3. The multimodal corpus

3.1. The recordings



Figure 11: Recordings from the Danish NOMCO dialogues: total and split views

Our multimodal corpus is part of a common collection of video-recorded interactions for Swedish, Danish, Finnish and Estonian which is being developed within the Nordic NOMCO project (Paggio and Diderichsen 2010).





The full name of NOMCO is *Multimodal Corpus Analysis in the Nordic Countries*⁵. The project aims at providing comparative annotated data on which to base investigations of communicative phenomena, especially feedback, turn taking, sequencing and information structure. The Danish NOMCO data consist of a set of 12 first encounter dialogues of a duration of approximately 5 minutes each. The way in which the recordings were carried out is described in detail in Paggio and Diderichsen (2010). The participants were six males and six females, all native speakers of Danish aged between 21 and 36, either university students or people with a university education. They did not know each other beforehand and were unaware of the purpose of the recordings. They were told to stand in front of each other in the studio and try to get to know each other in the space of about 5 minutes. The videos were recorded in the TV studio of the Faculty of Humanities at the University of Copenhagen. For each dialogue, two versions were produced, one showing a long shot of the two participants facing each other, the other combining two mid shots taken from different angles in a split video. The two views are shown in Figure (11).

A questionnaire was given to each participant to collect information on how they experienced the conversations. The subjects were asked to rate their experience on a number of parameters including *enjoyment*, *intimacy*, *naturalness*, *tenseness*, *awkwardness*, etc. The results are quite homogeneous for all 12 participants and indicate that the subjects were not too affected by the artificial setting even though they were aware of it. In particular, since the variables perturbedness, tenseness and awkwardness were all below average, we consider the corpus a relatively valid exemplification of natural interaction.

3.2 The annotation

The first step in the annotation process was to produce an orthographic transcription of the audio signal. This was done using Praat (Boersma and Weenink 2009). The transcription includes word boundaries as well as word stress, indicated by a “,” before the stressed vowel. Pauses are represented by a “+”, and filled pauses glossed with English words, e.g. *laugh*, *breath* or expressions such as *oh*. The Praat transcriptions were then imported into the ANVIL tool (Kipp 2004), which was used for gesture annotation. Figure (12) displays the orthographic transcription of a short dialogue together with a co-occurring gesture element in the ANVIL annotation board.

⁵ The project is funded by the NOS-HS NORDCORP programme, see <http://sskkii.gu.se/nomco/> for more details.

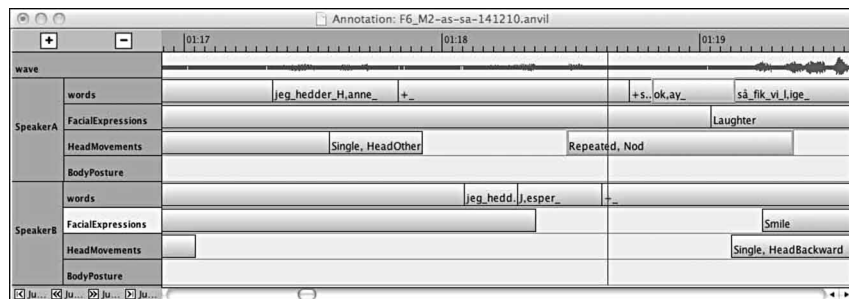


Figure 12: Orthographic transcription and gesture element in the ANVIL annotation board

The dialogue was already discussed earlier and is repeated below for readability's sake:

Speaker A: HEJ jeg hedder HANNE
Speaker B: jeg hedder JESPER
Speaker A: OKAY så fik vi LIGE det på PLADS

(hi my name is Hanne
my name is Jesper
okay so we got that right)

One of the purposes of this study is to investigate the relation between gestures and focusing. To this end, first of all the attribute “boundary true” was added in conjunction with sentence boundaries in the orthographic transcription. Secondly, for each sentence, topic and focus were identified and the attributes “topic true” and “focus true” were added to the corresponding words. We used here a method of annotation developed for an earlier study, where we looked at how pauses and focus interact in spoken Danish (Paggio 2006a; Paggio 2006b). In short, *topic* indicates the presupposed entity about which the sentence predicates something new, while *focus* indicates non-presupposed information. Not all sentences have a topic, whereas the focus is always present. The annotation guidelines include principles for how to assign topic and focus in general, as well as in specific syntactic constructions.





Table 1: Topic and focus annotation example

Word token	topic	focus	boundary
+	false	false	true
jeg	true	false	false
hedder	false	true	false
H,anne	false	true	false
+	false	false	true

In Table (1) we show in table format the assignment of topic, focus and clause boundary attributes to the utterance *jeg bedder Hanne* (lit: ›I call Hanne‹, or ›My name is Hanne‹) from the dialogue under consideration. Boundaries are placed together with the pauses that precede and follow the sentence, *jeg* (I) refers to topic, i.e. the entity about which the sentence predicates something new, whilst *bedder Hanne* (lit: call Hanne), which contains the only stressed word, is the focus, i.e. the new information.

Let us now turn to gestures. As already mentioned, gesture annotation follows the MUMIN coding scheme. In particular, we selected a number of attributes that describe the shape and function of head movements and facial expressions. We have seen that the shape categories are rather coarse-grained since they are intended to distinguish between functionally different types rather than describe gesture shape and dynamics in detail. As for function, the annotation scheme foresees features relating to *feedback*, *turn management*, *sequencing* and *information structuring*. At present, however, only feedback attributes are being coded. The semiotic types mentioned earlier are also left aside for the moment.

In addition to shape and function features, for each gesture under consideration, a relation with the corresponding speech expression is also defined.

The relation, called *MMRelationSelf* (where MM stands for multimodal, and Self refers to the fact that the relation concerns the speaker's own speech), establishes a link between the gesture under consideration and the word or words in the orthographic transcription of the gesturer's speech that the annotator judges to be semantically related. If the subject is silent while making the gesture (for example if they smile at something the interlocutor has said or done), the relation links the gesture with the corresponding pause in the transcription. In Figure (13), we show again how the repeated nod performed by Speaker A in the dialogue above is annotated in the ANVIL editing window in which the annotator specifies the various attributes of a gesture element.

Three annotators, all of them students of linguistics that had never worked with gesture analysis before, created the annotation. To ensure relia-

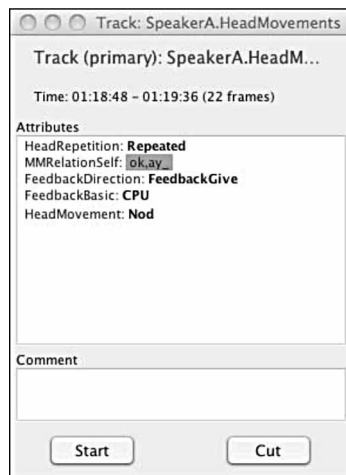


Figure 13: Annotation of a repeated nod in the ANVIL editing window

bility, each of them started by coding an entire interaction, the results were discussed and corrected and a set of written guidelines were developed based on these discussions. In this preliminary exercise, the *cohen's kappa* figures obtained were on average for the three pairs of coders in the range 0.5–0.6 for face attributes and 0.6–0.8 for head movements. Considering the fact that the agreement measure calculated in ANVIL reflects agreement of segmentation as well as labelling, and the fact that the annotators had only had limited training, these figures are quite satisfactory.

Each of the remaining videos was subsequently annotated by one of the coders and corrected by the other. Disagreements were again discussed and reconciled. If the two coders still could not agree, a third annotator made the final decision. Throughout this process, the guidelines were continually improved with examples and explanations. After having annotated five videos following this procedure, we repeated the inter-coder agreement exercise between the two annotators who had shown most disagreement the first time and noted an improvement of about 10 % for both face and head gestures.

The annotation work is still ongoing. In the following section, we provide figures on the annotated data from five of the twelve videos the corpus consists of.



3.3. Gestures in the data

The five videos annotated and analyzed so far make up for about 30 minutes of interaction (about 6 min. per video). The total number of word tokens (including filled pauses) is about 6000. The total number of gestures identified is 1919.

Table 2: Gesture types in the Danish NOMCO dialogues

Face		Eyebrows		Head Movement		Head Repeat	
Smile	330	Frown	44	Nod	249	Single	928
Laughter	143	Raise	222	Jerk	70	Repeated	280
Scowl	0	BrowsOther	1	HeadBackward	101		
FaceOther	29			HeadForward	139		
				Tilt	214		
				SideTurn	182		
				Shake	136		
				Waggle	31		
				HeadOther	86		
Face total	502	Brows total	267	Head total	1208	Repeat total	1208

Table (2) shows how the gestures are distributed according to the four shape attributes. Note that these are not mutually exclusive: *Eyebrows* attributes may occur on their own, but also in conjunction with a *Face* or a *Head movement* attribute; while *Single* and *Repeated* always occur together with *Head movement*⁶. Head movements constitute the majority of the gestures and most of them are single movements. To have an idea of how frequently they occur, we can relate the number to the duration of the interactions (30 min.) as well as to the number of words that occur in the dialogues (about 6000). Thus, there is one head movement per 0.9 seconds and there are 0.3 head movements per word.

4. Interaction between gestures and speech

An important aspect that a grammar of speech and gestures must cope with is the interaction between speech and non-verbal behavior. As already mentioned, gestures may relate to speech units of different granularity. In the simplest case, they overlap with single words or syllables. This is in general true of batonic gestures, or beats, that tend to occur with stressed words. However, in our data we also find a large number of facial expressions that overlap with

⁶ A more detailed description of the individual expressions and movements is given in the Appendix. The descriptions are taken from the annotation guidelines.



longer linguistic contributions not always corresponding to syntactic phrases. For instance, repeated nodding accompanied by intense gazing towards the speaker, or a smile, may start in the middle of the speaker's utterance and continue up to a breathing pause. A general question is whether temporal overlap is the best criterion on which to anchor the speech-gesture relation, or whether a more semantics-oriented approach should be adopted.

The issue of synchrony between gestures and speech has been studied by many (McClave 1994; McNeill 1992; Loehr 2004; Kipp 2004; Chui 2005; McNeill 2005; Ferré 2010). In general, these studies converge in showing that gestures tend to overlap temporally with the words they are semantically related to, also called *lexical affiliates*, a term introduced by Schegloff (1984). But the same studies also point to the fact that gesture strokes tend to occur before the onset of their lexical affiliates. However, these results build mainly on analyses of hand gestures. Since we are dealing with head movements and facial expressions, we decided to ask the annotators to indicate explicitly the word or words that each gesture is semantically related to. In this way, we will be able to study how this semantic relation maps onto temporal alignment. In linking gestures to related words, annotators were not instructed to make sure that the words chosen for each gesture should form syntactically well-formed constituents. In fact, direct adjacency between the linked words was not a constraint either. For example, repeated nods are often linked to a series of stressed words, while the intervening non-stressed words are left out of the multimodal relation. The purpose was again that we wanted to be able to study afterwards to what degree syntactic well-formedness played an implicit role in the annotators' assessment of what constituted a semantically meaningful multimodal combination of gestures and speech.

Having indicated in the annotation which words are related to the gestures allows us to investigate what types of speech segments participate in this relation. The first issue is whether gestures always relate to stressed words (or segments containing stressed words). Alahverdzhieva and Lascarides (2010), for example, claim that for gestures to attach to words in a multimodal lexical sign, the word in question must be prosodically prominent. A quick look at the material shows that this is largely, although not entirely, true. In fact, 27 % of about 1900 speech segments linked to gestures by means of the *MMRelationSelf* attribute are unstressed. However, 7 % are pauses (in these cases there may be a stressed segment in the interlocutor's speech that relates to the gesture), while 12 % are filled pauses. It could be argued that the presence or absence of stress is not relevant in these cases. This leaves 8 % unstressed words, which is arguably a low percentage. The words in question are mostly conjunctions, interrogative pronouns and adverbs.





Table 3: Unstressed speech related to gestures

Speech	Face (%)	Head (%)	All gestures (%)
Pause	8	6	7
Filled pause	16	9	12
One unstressed word	7	8	8
One stressed word	21	30	27
Several words (at least one stress)	48	47	46
Total	100	100	100

Table (3) shows the distribution of the various speech segment types in percentages. The same distribution is shown in counts in Figure (14).

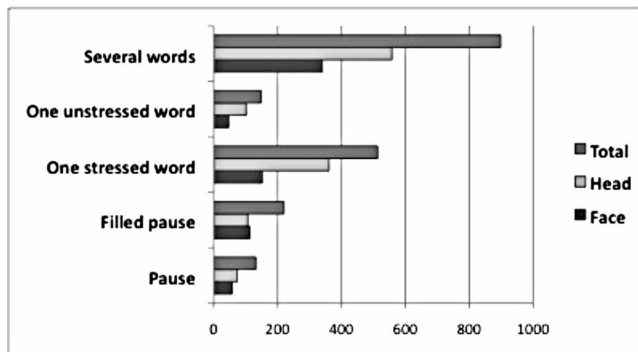


Figure 14: Speech segment types in multimodal links

Thus, although it seems reasonable to constrain the attachment of gestures to stressed lexical heads, the grammar should also allow for gestures to attach to unstressed interjections and the like.

In Figure (15), we show how a smile with feedback function can form what we call a multimodal lexical sign together with a word of type *filler*. The representation should be compared with the feature structure in Figure (8), which represents a generic smile not yet combined with speech.

In general, the feature structures that correspond to linguistic signs in HPSG contain at least the two attributes PHON, which represents the sign's sound, and SYNSEM, which represents syntactic and semantic information. Signs can be lexical if they correspond to words, or phrasal if they correspond to phrases. Here, we introduce a more complex kind of sign, which combines speech with non-verbal behavior. We still distinguish between lexical signs, that have no internal syntactic structures, and phrasal structures, that do.



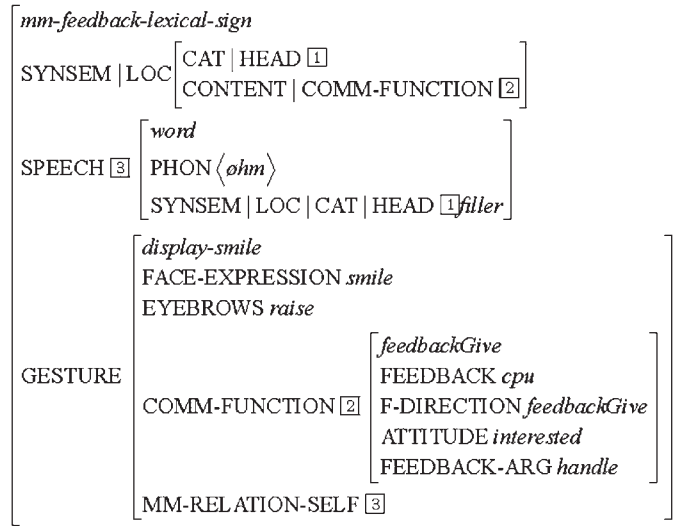


Figure 15: Feature structure representation of feedback multimodal lexeme

In the figure, then, our multimodal lexical feature structure consists of three top-level features: as customary in HPSG, `SYNSEM|LOC` represents the syntax and semantics of the local sign, here the combination of speech and gesture; in addition, `SPEECH` represents the word, and `GESTURE` the smile. As expected, there is no stress constraint on the word. The smile has been interpreted as showing an *interested* attitude. Tags (boxed numerals) are used to show structure sharing between features in the overall structure. In particular, the tags here ensure that the syntactic representation of the combined multimodal sign gets its values from the speech segment (through the `HEAD` attribute), and that the communicative function of the same combined sign gets its values from the gesture (through the `COMM-FUNCTION` attribute). Of course, it is not only the gesture that contributes to the content of the multimodal sign. Therefore, the `CONTENT` attribute will contain additional attributes the values of which are to be found in the speech segment, see Copestake et al. (2005) for more details.

To complete our explanation of the feature structure in Figure (15), the value of `MM-RELATION-SELF` in the `GESTURE` part of the sign is structure-shared with the value of the whole `SPEECH` attribute. The *handle* feature is not instantiated. It should point to the feature structure representing the utterance to which the smile and the *ohm* give feedback. This, however, requires a formal model of a whole discourse, which is beyond the scope of this paper to provide.





Table 4: Number of words in multimodal links

Number of words	No.
1 word	936
2 words	543
3 words	274
4 words	114
5 or more words	45
Total	1912

Another issue concerning the relation between gestures and speech is whether the speech segments correlating with gestures need to form well-formed syntactic constituents. Again, the data allow us to investigate the issue in a quantitative way. A first rough measure concerns the length of speech segments that the annotators have linked with the gestures. Relevant counts are shown in Table (4).

Table 5: Number of grammatical phrases in 3-word multimodal links

Phrase status	No.
Phrases	181
Non phrases	93
Total	274

In order to investigate the issue of syntactic well-formedness, we looked at the 3-word segments, which constitute good candidates for phrasehood, and counted how many were well-formed syntactic phrases or clauses, i.e. NPs, PPs, ADVPs, APs, VPs, Ss or lists of such. The results in Table (5) show that only about two thirds of the multimodal links to 3-word groups refer to grammatical phrases. We have not yet extended the analysis to groups of less or more than three words. However, even these limited results clearly indicate that the annotators did not consider syntactic well-formedness a requirement when assigning words to gestures. In fact, this observation confirms what other researchers have already noted. Alahverdzhieva and Lascarides (2010) propose that gestures that do not attach lexically should be made to attach to prosodic rather than syntactic constituents. Their proposal is mainly based on evidence from Giorgolo and Verstraten (2008). In that study, subjects were asked to evaluate synchronous and slightly asynchronous versions of the same video clips. In some clips, the original audio was kept, in others it was replaced by a hummed version of it retaining the same prosody. The results show that subjects were more willing to accommodate asynchrony if the original audio was kept than if they only heard humming.





It seems reasonable to ascribe this to the fact that semantic coherence overrules prosodic parallelism in perceiving multimodal signals.

We will not attempt to model the relation between gestures and prosody, but we will be concerned with how gestures combine with speech at the semantic level. In order to accommodate cases in which the multimodal links refer to fragments and discontinuous words, we suggest that the grammar should contain rules where the semantic contribution of gestures can be combined with the semantics of speech not only at the phrasal level, but also at the sentence or even utterance level. Let us look at an example from the NOMCO corpus:

Speaker A: nå men så er du også fra FIRS jo
Speaker B: jeg er fra FIRS

(well so you are also from 1980 then
I am from 1980)

As she replies *I am from 1980* confirming Speaker A's statement, Speaker B nods. The stroke of the head movement coincides more or less with the stressed syllable of the final word, but semantically the nod relates to the speech act expressed by the entire clause.

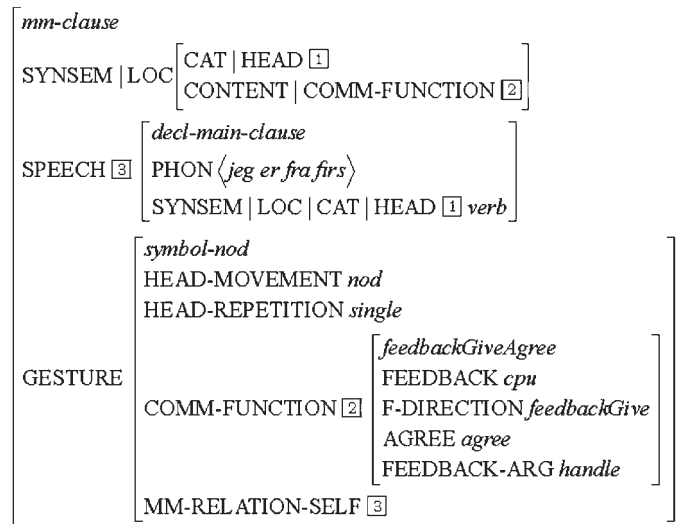


Figure 16: Feature structure representation of a multimodal clause





Figure (16) shows a feature structure representation of the analysis of Speaker B's turn. It should be compared with Figure 9, which accounts for a generic affirmative nod. In the combined speech-gesture representation, the shape and function attributes of the nod make up an embedded feature structure that constitutes the value of the top-level attribute GESTURE. The attribute SPEECH, on the other hand, contains a very lean representation of the syntax and phonology of the clause. To be complete, it should also have features that account for the internal syntactic structure of the clause, as well as its content. Finally, the top-level attribute SYNSEM contains syntax and semantics of the combined multimodal sign: the syntactic values come from the speech segment (the clause), whilst the communicative feedback function is provided by the gesture.

5. Feedback in gestures

In the preceding section we saw how feedback gestures can be represented together with the words or clauses they relate to. Here, we look at how feedback is expressed by means of facial expressions and head movements in our corpus. As already pointed out, we understand feedback in the sense of unobtrusive signs from the interlocutor to show attention, understanding, perhaps agreement or disagreement without trying to take the floor, or from the speaker to elicit an interlocutor reaction along the same lines. Feedback in this limited sense is similar to what others have called *backchanneling* (Yngve 1970).

Table 6: Feedback annotation in the Danish NOMCO dialogues

Basic		Direction		Agree	
Cpu	803	Underspecified	1	Agreement	43
None	1116	GiveElicit	13	NonAgree	1
		Give	650	None	1875
		Elicit	139		
		None	1116		
Basic total	1919	Direction total	1919	Agreement total	1919

Table (6) shows how many of the gestures have been judged to have a feedback function, how these gestures group based on the direction of the feedback, and how many of them also have an agreement function, i.e. show that the interlocutor agrees or disagrees with what is being stated. Almost 42% of the gestures have a feedback function, and of these, the great majority are used to give rather than elicit feedback. Unsurprisingly, non agreement is al-



most absent: it is a behavior we would not expect from two strangers who are making an effort to get to know each other. Perhaps more interestingly, only 43 gestures have been judged to express agreement. To understand this, one must note that in our annotation guidelines, agreement is defined as a reaction indicating acceptance of the truth of a statement. Very often in these dialogues, the interlocutor shows participation and attention (coded “Feedback cpu”), but there is not much discussion of factual statements. Earlier we saw an example of agreement from our corpus, and we discussed how the combination of speech and gesture in the example could be represented in a feature structure in Figure (16). Here is one of the few examples of disagreement:

Speaker A: jeg tror faktisk de var MEGET værre end vi FORESTILLER OS
Speaker B: det ved jeg IKKE

(I think actually they were much worse than we imagine
I don't know (lit: that know I NOT))

The two subjects are talking about monks and nuns in medieval Denmark. As he replies *I don't know* to Speaker A's statement, Speaker B tilts his head once. From the context, the annotator has interpreted the head movement as expressing disagreement with the preceding statement. Speaker B, although he says he doesn't know, actually means he doesn't agree. In fact, he goes on to explain how much we actually know about the topic from old church paintings.

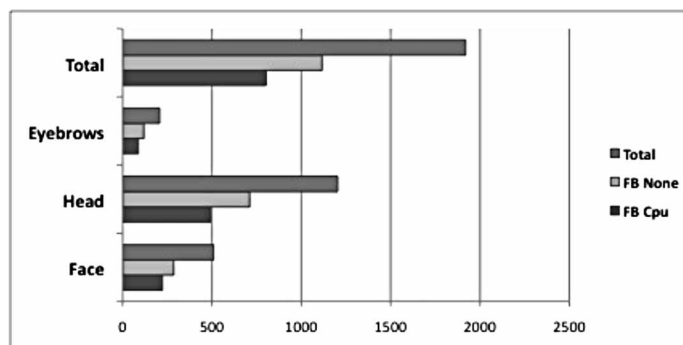


Figure 17: Feedback behavior in different gesture types

Let us now look at how feedback is expressed in different types of gestures. This is shown in the chart in Figure (17): head movements are the most represented type, followed by facial expressions and eyebrows attributes. How-





ever, the distribution of feedback is proportionally more or less the same in the three categories. In other words, it cannot be said that any of the three modalities is somehow “specialized” for feedback expression.

Table 7: Feedback annotation in the Danish NOMCO dialogues

Gesture type	Cpu (%)	None (%)	Total
Face	44	56	100
Head	41	59	100
Eyebrows	42	58	100

The same conclusion can be drawn from Table (7), which shows for each of the three modalities in what percentage they express a feedback reaction. The distribution between Cpu (feedback) and None (which means that the behavior has a function different from feedback) is more or less constant. Looking at specific head movement and facial expression types, however, shows that feedback is especially associated with head nods and smiles.

6. Multimodal information structure

The final issue we want to discuss is whether gestures relate to information structure. Our hypothesis is that they do, based on two observations. Firstly, we have seen that gestures, when they do not accompany filled pauses, tend to relate to prosodically prominent speech. Secondly, we know that focus is also associated with prosodic prominence in many languages, including Danish (Paggio 2009).

We have tested our hypothesis by counting the proportion of gestures that are related to one or more focused words, and the proportion of focused phrases that are accompanied by gestures.

Table 8: Gestures accompanying focus phrases

	gesture (no.)	gesture (%)
focus	1020	53
non-focus	899	47
total	1919	100

Table 9: Focus phrases accompanied by gestures

	focus (no.)	focus (%)
gesture	1020	79
non-gesture	275	21
total	1295	100



The results are shown in Tables (8) and (9), respectively. Gestures are more or less evenly distributed between focused phrases and other phrases. *Other phrases* in this context may be sentence topics, phrases that belong to the background part of the sentence, or filled pauses. We have not yet investigated what the function of the gestures that overlap with non-focused phrases is. Some of them are feedback gestures, other possible functions not yet coded in our corpus could be turn management or discourse sequencing.

Table 10: Focus phrases accompanied by gestures

Gesture type	No.
Head single	492
Head repeated	158
Smile	192
Laughter	44
FaceOther	14
Eyebrows	120
Total	1020

The reverse picture, that is the proportion of focused phrases that are accompanied by gestures, shows a more interesting tendency in that almost 80 % of the focus phrases in the corpus are related to gestures. We have not yet analyzed the remaining 20 % to search for an explanation of why they are not accompanied by gestures. To see the significance of correlation between focused phrases and gestures, remember that if we look at words in general, without worrying about information structure, it is only 30 % of them that are related to a gesture⁷. To sum up, the data show a significant and quite strong tendency for focus phrases to be accompanied by gestures. Which types of gestures is displayed in Table (10): single head movements are the most used to underline focus, followed by repeated head movements and smiles. Thus, we believe several factors contribute together to the expression of focus in Danish, i.e. word order, syntax, prosody (Paggio 2009) as well as gestures. More research is needed, however, to understand how gestures interact with syntactic and prosodic features and to investigate whether different speakers orient themselves more towards specific features.

We conclude this section by showing how a repeated batonic nod underlining the prosodically prominent words in a clause can be represented in a feature structure. The example reads:

⁷ A chi-square test on a contingency table containing the two sets of data gives chi-square=279, probability less than 0.0001.





det KUNNE man kunne man jo GODT mærke
 (one could, could indeed really feel it
 (lit: that COULD one could indeed REALLY feel))

The accented words, *kunne* and *godt*, are accompanied by the two strokes of the repeated nod. They constitute the (discontinuous) focus. The utterance here spans over a grammatical sentence containing a self-repair. The intonation clearly marks the sequence as a prosodic unit, and the two strokes come so quickly after each other that it seems reasonable to consider them as one complex gesture. Semantically, they relate to the words that constitute the focus of the sentence.

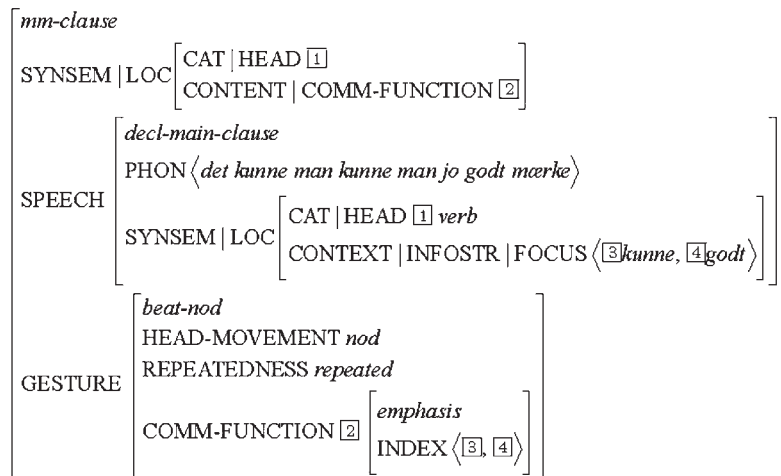


Figure 18: Multiple focus in a multimodal sign

This is expressed in the feature structure in Figure (18) by letting the FOCUS attribute in the representation of the clause contain two elements corresponding to the contents of the two focal words (see Paggio (2009) on the analysis of the information structure of sentences). In the GESTURE part of the feature structure, then, the indices in the INDEX attribute are structure-shared with the same two focal words. The representation should be compared with Figure 7, where the beat nod is not combined with speech, and the indices it introduces are not instantiated.





7. Conclusion

In this paper we have introduced a typology of non-verbal behavior types based on semiotic categories and coarse-grained shape attributes, and proposed typed feature structures to represent the form and content attributes of the various types. Based on analyses of data from a multimodal corpus of Danish conversations, it was then discussed how head movements and facial expressions interact with speech. It was noted that the speech segments to which gestures relate are mostly stressed, but a large number of filled pauses also participate in multimodal relations. The syntactic nature of the speech segments has also been analyzed, showing a large number of cases in which the speech segments are not syntactically well-formed. Consequently, grammar rules should be able to attach gestures at different syntactic levels. Examples of lexical and clausal attachments were presented.

Two of the possible communicative functions that gestures may have in conversations were discussed, namely feedback and emphasis. It was shown that although head movements make up the largest group in the expression of feedback, and in fact the largest group altogether, the proportion of gestures used in each modality to give or elicit feedback is the same. As for emphasis, a significantly large portion of all focus domains in our data are associated with non-verbal behavior, making it reasonable to claim that non-verbal behavior plays an important role in the expression of the information structure of sentences. Feature structure representations of multimodal signs expressing feedback and emphasis were proposed to show how the two phenomena can be modelled in a multimodal grammar.

To sum up, the grammar we have sketched out consists of feature structure types modelling i. different gestures distinguished by their shape characteristics ii. different communicative functions that gestures can have in specific contexts iii. different multimodal signs, i.e. combinations of gestures and speech segments accounting for the way in which gestures and speech relate semantically.

Although we feel that we have learnt a lot from the annotated data, the corpus contains in fact more information than it was possible to analyze here. Thus, in future we want to look in more detail at the issue of how gestures and speech relate in terms of time overlap, at the notion of prosodic phrases and their role with respect to gestures, and finally at how not only focus but also topic is associated with gestures.





Towards an empirically-based grammar of speech and gestures

309

Acknowledgments

I would like to thank my colleague Costanza Navarretta, my partners in the NOMCO project Elisabeth Ahlsén, Jens Allwood and Kristiina Jokinen, the annotators Sara Andersen, Josephine B. Arrild, Anette Studsgård og Bjørn Wesseltolvig, and the two anonymous reviewers of this paper.

This research was funded by the Danish Council for Independent Research in the Humanities and by the NOS-HS NORDCORP programme.

Appendix

Description of attributes for the annotation of facial expressions and head movements in the NOMCO project

Face

Face attributes refer to general facial expressions.

Smile	The facial expression shows pleasure, favour, amusement, or sometimes derision and scorn. Smile is characterized by an upturning of the corners of the mouth and usually accompanied by a brightening of the face and eyes.
Laughter	The facial expression shows merriment, amusement, or derision or nervousness and it is accompanied by an audible vocal expulsion of air from the lungs that can range from a loud burst of sound to a series of chuckles.
Scowl	An angry expression, like a frown. The term “frown” is reserved for the attribute Eyebrows.
FaceOther	To be used if none of the other values are appropriate. A comment *must* be added.

Eyebrows

Eyebrows attributes refer to eyebrow movements.

Frown	The eyebrows contract and move towards the nose.
Raise	The eyebrows are lifted.
BrowsOther	To be used if none of the other values are appropriate. A comment *must* be added.

HeadMovement

Head movements are defined similarly to gaze wrt. the subject's own body. Gaze and head movement are related. For instance, if the head is up, gaze is also up if the subject does not move their pupils.



310

Patrizia Paggio

Nod	A head movement down-up.
Jerk	A quick head movement backwards up.
HeadBackward	A movement of the head backwards (up), this can either be a movement of the head only or can be a movement of the whole trunk. This movement occurs often as a turn accepting signal.
HeadForward	A movement of the head forwards (down), this can either be a movement of the head only or can be a movement of the whole trunk. This movement occurs often as a turn elicit signal.
Tilt	A movement of the head leaning on one side.
SideTurn	A rotation of the head towards one side.
Shake	A rotation of the head first to one side and then to the other.
Waggle	A movement of the head back and forth, side to side, like a mixture of shaking and moving backwards or forwards. Usually produced to show uncertainty, doubtfulness.
HeadOther	To be used if none of the other values are appropriate. A comment <i>*must*</i> be added.

HeadRepetition

It indicates whether the movement is repeated or not. A repeated gesture is a sequence of similar gestures in rapid succession.

Single	The movement is not repeated.
Repeated	The movement is repeated.

FeedbackBasic

Feedback basic is used to specify whether feedback is of type contact, perception and understanding (the default) or other (that is one value out of C/P/U).

CPU	The subject gives or elicits signs that the message is being perceived and understood (CPU stands for contact, perception, understanding). The feature corresponds to ›Understand‹ in earlier versions of this scheme.
BasicOther	The subject gives or elicits signs that the message is being perceived, but perhaps not understood. The coder may want to indicate a more specific value choosing from C, P and U, in the comment.



FeedbackDirection

The feedback direction feature is used to distinguish between feedback giving, eliciting, giving-eliciting. It can be underspecified.

Feedback This value should be chosen if the coder is not sure whether the subject is giving or eliciting feedback.

Underspecified This value should be chosen if the coder thinks that the subject is giving and eliciting feedback at the same time.

FeedbackGiveElicit The subject gives feedback by showing that they have perceived the message and are willing to maintain contact and continue with the communication.

FeedbackGive The subject elicits signs that the interlocutor has perceived the message and is willing to maintain contact and continue with the communication.

FeedbackAgreement

Feedback can also be accompanied by signs of agreeing or not agreeing.

Agree The subject gives or elicits signs of agreement. In earlier versions the type was called Accept.

NonAgree The subject gives or elicits signs of lack of agreement. In earlier versions the type was called NonAccept.

Emotion/Attitude

Emotions and attitudes can co-occur with any of the communicative features. They include i.a. Ekman's six basic emotions: happiness, surprise, anger, sadness, fear, disgust/contempt.

Happy, Sad, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Interested, Uninterested, Disappointed, Satisfied, Other

SemioticType

Based on Peirce's theory, three types are defined: indexical, iconic and symbolic. In cases where a gesture has multiple semiotic values, the one perceived to be most evident or strongest should be chosen.

IndexDeictic Indexical gestures express a relation of cause-effect between the sign (the gesture) and its meaning. In particular, Indexical Deictics locate aspects of the discourse in the physical space (e.g. by pointing). They can also be used to index the addressee.



312

Patrizia Paggio

IndexNon-deictic	Indexical gestures express a relation of cause-effect between the sign (the gesture) and its meaning. In Indexical Non-deictic the indexical relation is between the gesture and the effect it establishes. Baton or beat gestures fall into this category.
Iconic	Iconic gestures express a semantic feature by similarity or homomorphism. Examples are gestures that express size (length, height, etc.) of an object mentioned in the discourse. Included in this category are also gestures that are sometimes called metaphoric.
Symbolic	Symbolic gestures (emblems) are gestures in which the relation between form and content is based on social convention (e.g. the okay gesture). They are culture-specific.
Semiotic	
Other	To be used if none of the semiotic values apply. A comment must be added.

MMRelationSelf

Underspecified multimodal relation, gesture and speech by the same person.



References

- Alahverdzhieva, K. and A. Lascarides 2010 Analysing speech and co-speech gesture in constraint-based grammars. In: S. Müller (ed.), *Proceedings of the HPSG10 Conference*, 6–26. CSLI Publications.
- Allwood, J. 2008 Dimensions of embodied communication – towards a typology of embodied communication. In: I. Wachsmuth, M. Lenzen and G. Knoblich (Eds.), *Embodied Communication in Humans and Machines*. Oxford: Oxford University Press.
- Allwood, J., L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio 2007 The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. In: J.-C. Martin, P. Paggio, P. Kuehnlein, R. Stiefelbogen, and F. Piansi (Eds.), *Multimodal Corpora for Modelling Human Multimodal Behaviour*, Volume 41 of *Special issue of the International Journal of Language Resources and Evaluation*, 273–287. Berlin: Springer.
- Allwood, J., J. Nivre, and E. Ahlsén 1993 On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9: 1–26.
- Arbib, M. A. 2005 From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* (28): 105–124.
- Boersma, P. and D. Weenink 2009 Praat: doing phonetics by computer (version 5.1.05) [computer program]. Retrieved May 1, 2009, from <http://www.praat.org/>.





- Chui, K. 2005 Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics* 37: 871–887.
- Cienki, A. and C. Müller 2008 *Metaphor and Gesture*. Amsterdam: John Benjamins.
- Copestake, A., D. Flickinger, C. Pollard and I. Sag 2005 Minimal recursion semantics: An introduction. *Research on Language & Computation* 3: 281–332.
- Duncan, S., J. Cassell and E. Levy 2007 *Gesture and the Dynamic Dimension of Language*. Amsterdam: John Benjamins.
- Ferré, G. 2010 Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. In: *Proceedings of LREC: Workshop on Multimodal Corpora Language Resources and Evaluation, Workshop on Multimodal Corpora*, Volume W6, Malte, 86–91.
- Giorgolo, G. and F. A. Verstraten 2008 Perception of »speech-and-gestures« integration. In: *Proceedings of the International Conference on Auditory-Visual Speech Processing 2008*, 31–36.
- Gullberg, M. and K. de Bot (eds.) 2010 *Gestures in Language Development*. Amsterdam: John Benjamins.
- Johnston, M., P. R. Cohen, D. McGee, S. L. Oviatt, J. A. Pittman and I. Smith 1997 Unification-based multimodal integration. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 281–288.
- Jokinen, K., C. Navarretta and P. Paggio 2008 Distinguishing the communicative functions of gestures. In: *Proceedings of the 5th Joint Workshop on Machine Learning and Multimodal Interaction*, Volume 5237 of *LNC3*, Utrecht, The Netherlands, 38–49. Berlin: Springer.
- Kendon, A. 2004 *Gesture*. Cambridge: Cambridge University Press.
- Kipp, M. 2004 *Gesture Generation by Imitation – From Human Behavior to Computer Character Animation*. Boca Raton, Florida: Dissertation.com.
- Loehr, D. P. 2004 *Gesture and Intonation*. Ph. D. thesis, Georgetown University.
- McClave, E. 1994 Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research* 23: 45–66.
- McNeill, D. 1992 *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- McNeill, D. 2005 *Gesture and thought*. Chicago: University of Chicago Press.
- Navarretta, C. and P. Paggio 2010 Classification of feedback expressions in multimodal data. In: *Proceedings of the ACL 2010 Conference Short Papers*, 318–324.
- Paggio, P. 2006a Annotating information structure in a corpus of spoken Danish. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation LREC2006*, Genova, Italy, 1606–1609.
- Paggio, P. 2006b Information structure and pauses in a corpus of spoken Danish. In: *Conference Companion of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 191–194.
- Paggio, P. 2009 The information structure of Danish grammar constructions. *Nordic Journal of Linguistics* 32: 137–164.
- Paggio, P., J. Allwood, E. Ahlsén, K. Jokinen and C. Navarretta 2010 The NOMCO multimodal Nordic resource – goals and characteristics. In: *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Paggio, P. and P. Diderichsen 2010 Information structure and communicative functions in spoken and multimodal data. In: P. J. Henriksen (ed.), *Linguistic Theory*





- and Raw Sound*, Volume 49 of *Copenhagen Studies in Language*, Samfundslitteratur, 149–168.
- Paggio, P. and B. Jongejan 2005 Multimodal communication in virtual environments: Communicating with the Staging virtual farm. In: O. Stock and M. Zancanaro (eds.), *Multimodal Intelligent Information Presentation*, 27–47. Kluwer Academic Publishers.
- Paggio, P. and C. Navarretta 2009 Integration and representation issues in the annotation of multimodal data. In: C. Navarretta, P. Paggio, J. Allwood, E. Ahlsén, and Y. Katagiri (eds.), *Proceedings of the NODALIDA 2009 workshop Multimodal Communication From Human Behaviour to Computational Models*, Volume 6 of *NEALT Proceedings Series*, Northern European Association for Language Technology (NEALT), 25–31.
- Paggio, P. and C. Navarretta 2010 Feedback in head gestures and speech. In: M. Kipp, J.-C. Martin, P. Paggio, and D. Heylen (eds.), *Proceedings of the LREC2010 workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 1–5.
- Peirce, C. S. (1931) *Elements of Logic. Collected Papers of Charles Sanders Peirce*, Volume Two. Cambridge: Harvard University Press.
- Poggi, I. 2007 *Hands, Mind, Face and Body: A Goal and Belief View of Multimodal Communication*. Berlin: Weidler.
- Pollard, C. and I. A. Sag 1994 *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press.
- Schegloff, E.A. 1984 On some gestures' relation to talk. In: J. M. Atkinson and J. Heritage (eds.), *Structures of Social Action*, 266–298. Cambridge: Cambridge University Press.
- Whitehead, K. 2009 Some uses of head nods in third position in talk-in-interaction. Paper presented at the annual meeting of the International Communication Association, Marriott, Chicago, IL, May 21
- Yngve, V. 1970 On getting a word in edgewise. In: *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 567–577.





Keywords

- attention getting device 189
- Augmentative and Alternative
Communication (AAC) 237–239
- Autosegmental-metrical Phonology /
Prosodic Phonology 104–107, 138
- boundary cue, phonetic-prosodic
114–115, 131–133
- construction 24–27, 30, 36
- Conversation Analysis 239–240
- deictic «here» 183
- determiner selection 63–66
- embodiment 2, 4–5, 7
- extension/intension 20 f., 23, 27–30,
36
- facial expression 282–283, 308
- feedback 289–290, 299, 303–305,
309
- gaze 194, 211 ff. 242 ff.
- head movement 282–284, 308
- information structure 305, 307–308
- Interactional Grammar 4–7
- Interactional Linguistics 9, 105–107,
138
- intonation phrase 110–111, 116–118
- intonation pattern 42–43, 50 ff., 61 ff.
- motion capture 269
- Multimodal Corpora 9
- multimodal grammar 281, 283, 308
- multimodality 173
- nucleus accent 82, 89 ff.
- overlap 207–211
- pointing 175, 226–230
- projection 47–49, 67–68, 75, 86, 89, 98
- prosodic (dis-)integration 19, 32, 34–36
- prosody 2–3, 5, 7
- recipient design 209, 211
- relative clause 24 ff
- restrictive/appositive 19, 21–25, 27–32,
34–37
- retraction 95
- self-repair 7, 40–43
- sequence organization 142, 152,
155–156, 157–161
- synchronization 270
- syntactic co-construction 74 ff., 98
- syntax-prosody interface 111
- terminal item completion 76, 80, 97
- transcription 146, 163, 272
- turn constructional unit (TCU) 144,
148–150, 152, 237
- turn-taking 146, 151–154, 161, 207–209

