

Temporality in Interaction

Edited by
Arnulf Deppermann
and Susanne Günthner

Studies in Language and Social Interaction

27

John Benjamins Publishing Company

Temporality in Interaction

Edited by

Arnulf Deppermann

Institute for the German Language (IDS)

Susanne Günthner

University of Münster



John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/slsi.27

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2014044587 (PRINT) / 2014045525 (E-BOOK)

ISBN 978 90 272 2637 2 (HB)

ISBN 978 90 272 6899 0 (E-BOOK)

© 2015 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Temporality in Interaction

Studies in Language and Social Interaction (SLSI)

ISSN 1879-3983

Studies in Language and Social Interaction is a series which continues the tradition of *Studies in Discourse and Grammar*, but with a new focus. It aims to provide a forum for research on grammar, understood broadly, in its natural home environment, spoken interaction. The assumption underlying the series is that the study of language as it is actually used in social interaction provides the foundation for understanding how the patterns and regularities we think of as grammar emerge from everyday communicative needs. The editors welcome language-related research from a range of different methodological traditions, including conversation analysis, interactional linguistics, and discourse-functional linguistics.

For an overview of all books published in this series, please see
<http://benjamins.com/catalog/slsi>

Editors

Sandra A. Thompson
University of California, Santa Barbara, USA

Elizabeth Couper-Kuhlen
University of Helsinki, Finland

Editorial Board

Peter Auer
University of Freiburg, Germany

Barbara A. Fox
University of Colorado, USA

Paul Drew
Loughborough University, UK

Marja-Liisa Helasvuo
University of Turku, Finland

Cecilia E. Ford
University of Wisconsin, USA

K.K. Luke
Nanyang Technological University, Singapore

Volume 27

Temporality in Interaction

Edited by Arnulf Deppermann and Susanne Günthner

Table of contents

| | |
|--|-----|
| Introduction: Temporality in interaction <i>Arnulf Deppermann and Susanne Günthner</i> | 1 |
| Section I. Mechanisms of temporality in interaction | |
| The temporality of language in interaction: Projection and latency <i>Peter Auer</i> | 27 |
| Retrospection and understanding in interaction <i>Arnulf Deppermann</i> | 57 |
| Ephemeral Grammar: At the far end of emergence <i>Cecilia E. Ford and Barbara A. Fox</i> | 95 |
| Section II. Temporally-structured constructions: A temporal perspective on syntactic constructions | |
| Temporality and the emergence of a construction: A discourse approach to sluicing <i>Paul J. Hopper</i> | 123 |
| Temporality and syntactic structure: Utterance-final intensifiers in spoken German <i>Wolfgang Imo</i> | 147 |
| Grammar, projection and turn-organization: <i>(il) y a NP</i> 'there is NP' as a projector construction in French talk-in-interaction <i>Simona Pekarek Doehler</i> | 173 |
| Word order in time: Emergent Hebrew (N _s)V/VN _s syntax <i>Yael Maschler</i> | 201 |
| A temporally oriented perspective on connectors in interactions: <i>und zwar</i> (‘namely/in fact’)-constructions in everyday German conversations <i>Susanne Günthner</i> | 237 |

Section III. Temporal organization of multimodal interaction

| | |
|--|-----|
| Multimodal completions | 267 |
| <i>Lorenza Mondada</i> | |
| Coordinating the temporalities of talk and dance | 309 |
| <i>Leelo Keevallik</i> | |

Section IV. Appendix

| | |
|---------------------------|------------|
| Transcription conventions | 337 |
| Index | 341 |

Video clips (.mov) of numerous examples in this volume can be found online, at <http://dx.doi.org/10.1075/slsi.27.video>



This symbol marks the availability of a video clip of an example.

INTRODUCTION

Temporality in interaction

Arnulf Deppermann and Susanne Günthner

The authors establish a phenomenological perspective on the temporal constitution of experience and action. Retrospection and projection (i.e. backward as well as forward orientation of everyday action), sequentiality and the sequential organization of activities as well as simultaneity (i.e. participants' simultaneous coordination) are introduced as key concepts of a temporalized approach to interaction. These concepts are used to capture that every action is produced as an inter-linked step in the succession of adjacent actions, being sensitive to the precise moment where it is produced.

The adoption of a holistic, multimodal and praxeological perspective additionally shows that action in interaction is organized according to several temporal orders simultaneously in operation. Each multimodal resource used in interaction has its own temporal properties.

1. The need for a temporal understanding of linguistic structures

As Paul Hopper (2006) once stated, for most linguists, “the mention of time evokes questions of verb morphology, [...] tense and aspect.” This volume, however, does not address semantic aspects of time, instead it deals with the real-time dimension that all spoken language inhabits. It aims at bringing together a body of research which demonstrates that studies of linguistic structure can gain profound insights once they take the temporality of linguistic production and reception into account and, thus, do justice to the irreducibly temporal nature of the situated use of language.

Various linguists have mentioned the fact that as Jespersen (1924:26) put it “...a sentence does not spring into a speaker's mind all at once, but it is framed gradually as he goes on speaking”. However, theories of linguistic structure, in general, are still based on a detemporalized notion of ‘language’, reducing “a temporal medium to a fixed, stable, and timeless one” (Hopper 2011:22). Consequently, grammatical features and constructions are still treated as an inventory of forms, instead of studying them in the ongoing temporal flow of time (Auer 2005, 2006).

Only recently and mainly thanks to approaches such as ‘on-line syntax’ (Auer 2009) and ‘emergent grammar’ (Hopper 1987, 1998), studies of language in interaction have begun to look systematically at linguistic structures as they unfold in real time (cf. Auer 2000; Goodwin 2002; Günthner/Hopper 2010; Hopper 2011; Auer & Pfänder 2011).

This volume is dedicated to issues of temporality and language. It addresses questions of how participants handle temporal processes and exigencies in everyday spoken interaction and how grammar – viewed as a dynamic, context-sensitive system – can be adapted to temporal constraints and emerges in the processes of interaction. As language-in-interaction is bodily and temporally situated in the mutual presence of face-to-face interaction, its meaning is inevitably produced and interpreted in time (Schütz & Luckmann 1979: 83; Psathas & Waksler 1973).

Thus, the empirical analyses in this volume show how the course of the ongoing production of linguistic structure is shaped by co-participants’ local verbal and non-verbal (re-)actions as well as the opportunities and restrictions provided for by the linguistic structures which are accomplished at a given moment of interaction. The collected papers provide evidence of the sensitivity and adaption of linguistic structures to the collaborative temporal unfolding of an ongoing interactional sequence. Furthermore, they reveal that once temporality and the temporal processes of producing and interpreting language in interactions are taken into account, many new questions of language usage and of the interactive emergence of grammatical phenomena arise. Viewing grammar from a temporal perspective leads to a radical change in the conceptualization of language: Instead of treating grammar as a mental structure, the authors recontextualize grammar in its actual usage in everyday social life. Analyzing linguistic structures “as these unfold in realtime discourse” (Silverstein 1984: 182) represents a shift to a new perspective of dealing with language use and thus raises fundamental questions of theory and method.

2. The temporal constitution of experience and action

The key role temporality plays in the accomplishment of both linguistic and interactional structure stems from the fact that it is basic for all human action and experience. Every lived moment is a fleeting, continuously renewing present. It is never self-contained, but an ever-moving, non-locatable, ephemeral point of continuous transition of immediate future and lived present into past (Husserl 1928). There is no sheer, static presence. Lived temporality is characterized by this ever-moving, future-directed now. The lived present is brought about by continuous passive synthesis of impressions, i.e., sensations and perceptions, which become retentions of ongoing experience (Husserl 1928: 385–395). This also

includes “protentions” (Husserl 1928) of the immediately impending future in terms of more or less open expectations and possibilities turning to factual experience (Gurwitsch 2010: Part IV). This continuous synthesis accounts for the unity and continuity of experience. Every attempt at grasping the present moment of experience necessarily destroys the identity of it, because it can only be captured as a remembered past moment, which is not identical with the present of the subject reflecting on it (Merleau-Ponty 1945: Part III, Section 1–2). Lived temporality therefore is not to be identified as a series of points in linear, chronological time observed from a “God’s eye” point of view beyond it. Rather, chronological time is perceived with reference to lived temporality. Lived temporality is relevance-structured time. Time figures in actions in terms of expectations, opportunities and exigencies for action. It is experienced as *durée* (Bergson 1970 [1889]; Schütz 1974: 62–70; Schütz & Luckmann 1979: 80–87), which is qualitatively structured in terms of temporal units, episodes, and cycles of activities, events, perceptions and experiences.

These fundamental features of the temporal constitution of experience and action are also basic for the accomplishment of action and linguistic structures in interaction.

3. Retrospection and projection

Action in interaction is inevitably temporal practice. Participation in interaction requires online production and understanding of linguistic and pragmatic structure as they unfold in time (Auer 2009). The intersubjectivity of the “we-relationship” in face-to-face interaction is rooted in temporality: It rests on the possibility that subjects mutually synchronize their consciousness by perceiving one another in shared time (Schütz & Luckmann 1979: 90–97) from the point of view of the ever-moving, fleeting present. In this process, action and understanding are always oriented both backwards by retrospection and forwards by projection. As Heritage (1984: 241ff.) shows, every turn in interaction (and we could add: every part of a turn as well) is at once context-bound and context-renewing. Turns are context-bound in being retrospectively tailored to prior context, i.e., just that interactional moment at which they are produced. Most notably, they are designed by reference to the current participation framework and the immediately prior turn in the interaction (Deppermann 2013). Prior context is both a restriction and a resource for building a next turn. It accounts for choices of formulation in action formation. It supplies presuppositions of previously accomplished meanings, linguistic and interactional structures, which can be used as common ground to build on. Furthermore, prior linguistic context provides “structural latencies”

(Auer this volume), which can be built upon for analepsis (cf. also Hopper this volume), or which can be re-indexed by anaphorical means, i.e., pronouns, lexical substitution, (partial) repetition and reformulation. Further, both variation and contrast assume their local indexical functions to a high degree only by retrospective reference to prior talk (Du Bois 2010). Thus, next interactional moves make heavy use of reassembling and recombining resources that prior interaction has provided, supplying the grounds for the self-referential emergence of interactional structure as participants construct interactional sequences (Goodwin 2013). Most of the manifold practices of retrospection and its basic role for producing formulations in interaction only become obvious once we extend the scope of analysis beyond the traditional unit of research in grammar, i.e., beyond the boundaries of the clause (or sentence). Retrospective practices create coherence between current and past activities, and they are indispensable for displaying understanding and accomplishing intersubjectivity (Deppermann this volume). This involves confirming, building on and expanding prior context, but equally includes practices of retraction (Auer 2009), which modify and substitute prior talk by repair (Schegloff et al. 1977) and various kinds of reformulation (specification, generalization, exemplification, etc.; see Gülich & Kotschi 1996; Günthner this volume). While the construction of turns which immediately build on prior talk either does not need to represent its retrospective bases explicitly or can do so with very economical, highly indexical resources (such as pronouns), retrospective recontextualization of non-adjacent prior talk context has to make use of more explicit and effortful means (Deppermann this volume). Practices of back-linking and skip-connecting (Mazeland & Huiskes 2001), quotation (Clift & Holt 2006) and reformulation (Heritage & Watson 1979) are then needed in order to recall prior talk and reinstate its current topical relevance. An inquiry into how people accomplish retrospection in interaction in their formulatory choices shows that it is not simply the amount of physical time having elapsed between current talk and the relevant retrospective context which determines which resources are used for recontextualization. Practices of retrospection are sensitive to qualitative time in terms of ongoing pragmatic relevance and cognitive salience of prior context and of projections (see below), which are still lingering, making newly produced bits of talk understandable as their fulfilment.

While participants' anaphoric practices index more or less precisely the retrospective point of reference that current talk relates to, 'retrospection' may also be understood in a more general fashion. As Bakhtin (1986) pointed out, every current use of a word builds on a discursive history of prior uses of the same word, both on an individual, biographical and on a diachronic, socio-historical scale. "Our speech [...] is filled with others' words [...]. The words of others carry with them their own expression, their own evaluative tone, which we assimilate,

rework, and re-accentuate.” (Bakhtin 1986: 89). Therefore, each occurrence of a word can be associated with its prior contexts of use, with styles and genres and the social milieus, identities of users and goals of interaction they have been tied to, being the most important sources for their connotative meanings. Building on Bakhtin, Hopper (2011) argues that the online construction of emerging grammatical structures in talk is not informed by abstract categorical grammatical knowledge. Rather, speakers use their memory of prior usage, which provides them with knowledge about sedimented, routine ways of speaking (Hopper 1987, 1998; see also Günthner et al. 2014). These are used as constructional bits and pieces to produce grammatical structure by speakers as they go about building up their turns at talk in a manner which is sensitive to the situated contingencies of the interactional moment.

Turns are context-renewing in adding to and transforming prior context: What a turn does in interaction is not defined by its meaning as a ‘speech act’ in isolation, but by the precise situated and indexical ways it transforms the given interactional context at that moment in ways that are consequential for interactional organization and sense-making. ‘Context’ thus needs to be understood temporally as a reflexive and dynamic notion (cf. Gumperz 1982; Auer 1992): Context is not a container of variables with static values determining interactional practice, but it is continuously renewed, i.e., both reproduced and redefined, by practice itself in the temporal process of the interaction (cf. Heritage & Clayman 2010). A major part of the interactional import of every turn lies in its projective properties. “By projection I mean the fact that an individual action or part of it foreshadows another. In order to understand what is projected, interactants need some kind of knowledge about how actions (or action components) are typically (i.e., *qua* types) sequenced, i.e. how they follow each other in time.” (Auer 2005: 8). Projection operates on various levels of interactional practice, most notably on the levels of grammar, turn-construction (Günthner 2011a; this volume), turn-taking (Pekarek Doehler this volume) and sequencing of actions (Schegloff 2007), but also with regard to lexical co-occurrence, topical development (Maschler this volume), choices of styles and codes and the coordination of various multimodal resources (see below). Local projections established by the production of some turn or turn-component provide a range of expectations about “what comes next?” Projections are basic for the anticipation and coordination of action in interaction (Goodwin 2002): The structure of the emerging turn establishes projections concerning (a) the kind of action the turn is to perform, (b) possible points of turn-completion allowing or calling for turn-transition, and (c), often-times, expectations about the next action to be performed by next speakers.

As for point (a). The grammar of many languages allows for “early projection” in action formation (Levinson 2013). Think, e.g., of question-pronouns and

verb-fronted imperative or interrogative syntax, which index from the very beginning of the turn the kind of (or at least a very restricted range of) action(s) the upcoming turn is to perform. This allows recipients to understand its pragmatic import early. Similarly, projector constructions are specialized in enabling topic change, structuring extended descriptive and argumentative multi-unit turns, and securing the floor for the speaker (Pekarek Doehler 2011, this volume; Günthner 2008, 2011a, 2011b).

As for point (b). Emerging syntactic structure, but also, with a narrower scope, prosody enable hearers to “monitor the structure of emerging talk prospectively in order to locate unit completions, upcoming moments where it will be possible for them to take the position of the speaker” (Goodwin 2002: 26). The relative smoothness of turn-taking with few delays rests decisively on the capacity of the emerging turn to project possible moments of its completion. This enables recipients to orient prospectively to upcoming transition-relevance places and to prepare their responses accordingly (Schegloff 1996). Grammatical projection of turn-completion lays the foundation for collaborative completions (cf. Lerner 1991, 2004; Auer this volume), because recipients can infer on their behalf which kinds of structures are needed and may be adequate for completing the structure which was recognizably started by what the speaker has already produced.

As for point (c). The most powerful mechanism of projection of next actions is conditional relevance (Schegloff 1968). Conditional relevance describes the expectation that upon the production of some first pair part action of some type A (e.g., a question) by speaker 1 some second pair part action of a matching type B (e.g., an answer) is due as next action by speaker 2. Conditional relevance, thus, accounts for the production of adjacency pairs, which can be said to be the basic building blocks of sequences in interaction and thus of interactional structures in general (Schegloff 2007).

Adjacency pairs are paramount instances of how projection is intrinsically tied to what interactional structures mean: The conversational action performed by a first pair part cannot be described without referring to its projective properties. Projections are almost always more or less schematic (except for ritualized cases such as having to respond by “I will” at a wedding ceremony). Furthermore, they usually do not define just one option for continuation or response, but they allow for a type-defined range of options, which may be more or less restricted dependent on the source of projection. E.g., while polar interrogatives (*yes/no*-questions) set up a narrow set of type-conforming options, which are obligatory for a response or at least the start of it (Raymond 2003, 2013), *wh*-questions, especially if they are “telling questions” (Thompson et al. 2015), allow for a wide range of possible kinds of responses that fit. Just as there are linguistic devices to co-refer anaphorically in retrospection, there are linguistic devices which work

cataphorically. Goodwin (1996) terms them 'prospective indexicals.' They may operate with a narrower temporal scope, like cataphoric pronouns, but they may also have a much more wide-ranging scope such as story prefaces, which project what the climax or moral of a story will be about. While the full meaning of what prospective indexicals refer to is only revealed in subsequent talk, they act as indices to establish expectations about that talk. In this way, they enable recipients to coordinate their interactional participation with what is projected.

4. Sequentiality and simultaneity

The discussion of retrospection and projection leads us to distinguish two aspects in the temporality of activities, which, however, are closely related. On the one hand, there is temporality in terms of the *temporal perspectives* inherent in turns-at-talk. Retrospection refers to the past in the present, it is the term for how a current structure relates to the interactional past. Projection, in contrast, refers to the future in the present, it captures how a current structure anticipates what the interactional future might be. On the other hand, there is temporality in terms of the *temporal ordering of activities*, i.e. their beginning, extension, and completion, their timing, their sequential organization and their simultaneous coordination. Also in this case, retrospection and projection are basic principles for interactional ordering: While retrospection concerns how a turn indexes its position with respect to the development of the interactional sequence so far by its design, projection establishes expectations for what is to come next.

We shall now consider issues of temporal order in interaction in more detail.

The basic insight of CA, which distinguishes it from other theories of social action, is that actions are sequentially organized. The point of departure for analyzing what actions are and do is not the individual action – as assumed by Speech Act Theory, (Neo-)Gricean Pragmatics, Relevance Theory and also Social Phenomenology – but action sequences (Atkinson & Heritage 1984; Schegloff 2007). This amounts to a radically temporalized approach to action: Both the design and the meaning of an action are tied to its position in a temporally ordered sequential organization of actions. 'Sequentiality' (and 'sequential organization') refers to the universal and irreducibly temporal fact that every action is produced as a step in the succession of adjacent actions, being sensitive to the precise moment where it is produced. With CA's insistence on the sequential context of an action being the most immediate and most important dimension of context both for the production and the analysis of turns at talk, the notion of 'context' becomes temporalized and emergent (Schegloff 1992a). Interactional context is thus seen as a reflexive, self-organizing accomplishment by interactants. Social structures

(identities, social relationships and institutions) are revealed to exist as temporal and pragmatic structures, i.e., as “Vollzugswirklichkeit” (‘realities by performance’, Bergmann 1985). This aspect of “Vollzugswirklichkeit” is nicely captured by the notion of “doing being X” (see Heritage & Clayman 2010). The existence and relevance of social structures, however, depends on their continuous reproduction and enactment by practical action, which, in turn, confirms and indexically interprets their relevance for current practical interactional tasks and activities.

In contrast to ‘sequentiality’, ‘sequence’ does not refer to just any succession of actions. Sequences are temporally complexes of actions with recognizable beginnings and ends, engendered to deal with some specific joint interactional business, which is schematically projected in the initiation of the sequence (see Schegloff 2007). The adjacency pair is the prototypical basic organizational format of sequences. It can be extended by pre- and post-sequences, which gain their function with respect to the core adjacency pair (Schegloff 2007). The notion of ‘sequence’ again points to how interactional time is qualitatively structured in terms of practical relevancies. These account for what interactants conceive of as boundaries, units, and finished vs. unfinished structures in talk. Sequences rest on generic, type-defined relationships between actions in first, second, third and fourth position vis-à-vis each other, which are constitutive of their pragmatic meaning and the accomplishment of intersubjectivity (Schegloff 1992b; Deppermann this volume). The notion of ‘sequentiality’, on the other hand, leads us to focus on the irreducibly indexical character of the production of action in interaction, and, thus, to the fact that we have to deal with situated contingencies in order to account for its “unique adequacy” (Psathas 1995: 17) as a contribution to just this collaborative activity at just this moment of interaction.

Sequentiality also entails that meaning in interaction is a processual phenomenon. It is not to be located at the spot where a turn is produced, and it is not defined by speaker’s prior intentions (Haugh 2008). Interactionally relevant, intersubjective or disputed meanings emerge only via temporally extended sequences of displays of understanding and consecutive negotiation of meaning. It is only by such temporally extended processes that they become part of the interlocutors’ common ground (Clark & Brennan 1991) and consequential for the future of the interaction.

Sequentiality is not the only property of temporal ordering between activities that matters. Timing and simultaneity are equally important. In addition to their sequential relationship, the timing of actions has interactional origins and functions of its own. Precision-timing of speaker-change and turn-transition with transition spaces of less than a second (cf. Jefferson 1988) relies heavily on the projectability of transition relevance places (Sacks et al. 1974). Delays in turn-production may be produced to convey interactional meaning, such as

indexing that a dispreferred action is about to be produced (Pomerantz 1984), but they may also function to coordinate talk with body movements in order to create interactionally shared space with respect to the projected joint activity (Mondada 2009).

In classic mono-modal accounts of conversational structure by CA, which only deal with the vocal-acoustic mode of action, simultaneity essentially surfaced under the guise of overlapping talk (Sacks et al. 1974; Schegloff 2000; Jefferson 2004). It has been shown how the turn-taking machinery provides a mechanism to minimize overlap and how interactants treat overlap as a phenomenon to be resolved (Schegloff 2000; Jefferson 2004). Furthermore, analyses revealed how overlaps are used as resources to convey affiliation and sharedness (Goodwin & Goodwin 1992), to support and complete partners' turns under construction (Oloff 2009) or to build competition and conflict (Kotthoff 1993).

Analyses of video-taped multimodal interaction, however, soon made clear that simultaneity of activities is a basic feature of all face-to-face interaction (Goodwin 2000). Participation in face-to-face interaction extends far beyond the vocal-acoustic modality. Talk, gaze, gestures, bodily posture, facial expression, orientation in space, walking and the manipulation of objects have to be coordinated both sequentially and simultaneously in interactional cooperation (see Keevallik this volume, Mondada this volume). Both intra-personal and inter-personal coordination of activities (Deppermann & Schmitt 2007; Deppermann 2014) involve the simultaneous use of multimodal resources of various kinds. There is no 'floor' and no turn-taking mechanism used to organize the deployment of all the multimodal resources that participants (not just "speakers"!) use in face-to-face interaction. The pragmatic properties of simultaneous activities mobilizing different multimodal resources therefore cannot be accounted for as "overlap", even if they are directly related to floor management (Schmitt 2005). The issue of how multimodal simultaneity of activities links up with sequential organization of talk and other multimodal resources still remains largely to be explored (see below).

5. Multimodal temporalities

The use of videotaped data and the adoption of a holistic, constitution-theoretic, multimodal and praxeological perspective reveal that action in interaction is regularly organized according to several temporal orders simultaneously in operation (cf. Goodwin 1981, 2002; Streeck et al. 2011). Each multimodal resource used in interaction has its own temporal properties (see Keevallik this volume, Mondada this volume). It may consist of discrete entities which are produced bit by bit (like words) or which get their shape via suprasegmental, analogical

trajectories (like prosodic contours), it may allow for static configurations (like body postures) or be fleeting in nature (like talk), it may require preparation and retraction phases of varying temporal scope (compare eye movements vs. relocation of the body), it may be produced as punctual, rhythmic or sustained structure. A few examples:

- While the orientation of the lower parts of the body is essential for establishing an enduring f-formation constitutive of a focused encounter (Kendon 1990: 209–238) and indexes participants' primary involvement, more fleeting, interpolated involvements are accomplished by body torque of the head and the trunk (Schegloff 1998).
- Whereas beat gestures are punctual activities to be produced in close synchrony with focal accents in many languages (McNeill 1992), pointing gestures precede the verbal reference forms they are coordinated with in situated acts of referring (Kendon 1972; Schegloff 1984) and they may be held long after their apex is reached and the verbal reference has been formulated.
- While prosodic contours are gestalts emerging from the temporal differences in pitch movement, tied to and constrained by cycles of human respiration, grammatical structures are produced by discrete units open to incrementation and extension across several TCUs and speakers.

All multimodal resources in interaction are sensitive to the basic properties of temporal structuration of praxis by projection and retrospection, sequentiality and simultaneity. Language, however, is unique in making temporality also a referential object. The grammatical categories of tense and aspect and both the various past-oriented narrative genres (Linde 1993; Ochs & Capps 2001; Weinrich 1971) and future-oriented linguistic activities, such as planning, forecasting and announcing, allow for a separation of the there-and-then of the referential plane of the interaction from its pragmatic here-and-now (cf. Bühler 1982). Building on the linguistic achievement of creating a situation-transcendent referential world of talk, other modalities may also participate in the construction of meanings denoting temporally remote events, like gestures in the case of “Deixis am Phantasma” (Bühler 1982; Stukenbrock 2014).

The particular temporal constraints and affordances each modal resource implies are consequential for how different resources are coordinated to produce actions. E.g., the opportunity to freeze gestures can be used to index continuing claim to speakership gesturally if the speaker has lost her turn after overlap (Oloff 2013). To produce synchronized multimodal action, participants make use of the temporal logics of the different resources, taking into account the requirements each modality places on attention and perceptibility both intra- and inter-personally. Participants establish projections which cross-cut modalities: