

Dagmar Barth-Weingarten | Richard Ogden

“Chunking” spoken language: Introducing weak cesuras

Suggested citation referring to the original publication:
Open Linguistics (2021) , pp. 531 - 548
DOI <https://doi.org/10.1515/opli-2020-0173>
ISSN 2300-9969

Postprint archived at the Institutional Repository of the Potsdam University in:
Zweitveröffentlichungen der Universität Potsdam : Philosophische Reihe 174
ISSN: 1866-8380
<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-536259>
DOI: <https://doi.org/10.25932/publishup-53625>



Research Article

Dagmar Barth-Weingarten* and Richard Ogden

“Chunking” spoken language: Introducing weak cesuras

<https://doi.org/10.1515/opli-2020-0173>

received March 29, 2021; accepted July 28, 2021

Abstract: In this introductory paper to the special issue on “Weak cesuras in talk-in-interaction”, we aim to guide the reader into current work on the “chunking” of naturally occurring talk. It is conducted in the methodological frameworks of Conversation Analysis and Interactional Linguistics – two approaches that consider the interactional aspect of humans talking with each other to be a crucial starting point for its analysis. In doing so, we will (1) lay out the background of this special issue (what is problematic about “chunking” talk-in-interaction, the characteristics of the methodological approach chosen by the contributors, the cesura model), (2) highlight what can be gained from such a revised understanding of “chunking” in talk-in-interaction by referring to previous work with this model as well as the findings of the contributions to this special issue, and (3) indicate further directions such work could take starting from papers in this special issue. We hope to induce a fruitful exchange on the phenomena discussed, across methodological divides.

Keywords: Conversation Analysis, Interactional Linguistics, prosody, phonetics, intonation units, talk-in-interaction, syntax, kinetics

1 Background

1.1 This special issue

With this special issue, we would like to present current work in Conversation Analysis (CA) and Interactional Linguistics (IL) (see also Section 1.3) on the “units” of naturally occurring talk to a broader linguistic public.¹ While there are methodological (and theoretical) differences to other, earlier approaches, we all share an interest in how language works in human interaction. One aspect of the workings of human language is its being produced in spurts or chunks. Yet, especially when applying chunking models, such chunks are not always easy to separate and thus to identify. Moreover, when studying the fuzzier instances of boundaries of such chunks in more detail, it turns out that participants can use them for interactional purposes. This is what this special issue aims to topicalize thus also pursuing further issues highlighted in Szczepiek Reed and Raymond (2013).

¹ A more detailed introduction to, and discussion of, the relevant issues can be found in Barth-Weingarten (2016).

* **Corresponding author: Dagmar Barth-Weingarten**, Department of English and American Studies, University of Potsdam, Potsdam 14469, Germany, e-mail: dbarth@uni-potsdam.de

Richard Ogden: Department of Language and Linguistic Science, Centre for Advanced Studies in Language and Communication, University of York, York YO10 5GG, England, e-mail: Richard.ogden@york.ac.uk

ORCID: Dagmar Barth-Weingarten 0000-0002-7219-0013; Richard Ogden 0000-0002-5315-720X

In this introduction, we aim to lay out the focus of this special issue in more detail and the methodological approach chosen by the authors in dealing with its various aspects. In addition, we summarize their findings and indicate further directions such work could take. We hope that this special issue serves as a starting point for a fruitful exchange on the phenomena discussed, across methodological divides, to further our common understanding of natural human interaction.

1.2 The issue

Spoken language is usually (perceived, and analyzed, as being) produced in smaller spurts, or chunks. The following excerpt illustrates this kind of “chunking”.² The relevant audio snippet can be obtained from the supplementary material to this special issue. Transcription conventions follow the GAT2 system, which depicts one intonation unit per line. For further details, please see the Appendix as well as Couper-Kuhlen and Barth-Weingarten (2011). Line numbers are based roughly on seconds into the recording at which the relevant stretch occurs.

(1) Glass tiles (UAK 10004, 2:32–2:34) (overlap simplified)

((AE everyday face-to-face conversation. Marge and Frank are talking about renovating one of their bathrooms. Frank has already said that he is in favor of a certain kind of tiles: “silverish”, “light”, “aluminumlooking”, “glass tiles”, or “something similar” down the shower walls; “small ones on the sides” (not shown here), while Marge does not like them and suggested alternatives. Frank is coming back to his suggestion, though, after Marge’s mother (in the background) mentioned mosaic tiles.))

```

1  FRA:  moSAics:-
2         'i think moSAics would look goo::d-
3  MOT:  absoLUt(ely)

```

There is no doubt that this stretch consists of three chunks of speech, each depicted in one line in the transcript. In linguistics, this observation is usually captured with a unit model. The chunks are conventionally referred to as “intonation units” (IUs; also variously referred to as “tone group” in Halliday 1970, 1985, “tone-unit” in Crystal 1969, and “intonation-group” in Cruttenden 1997 among others; for a more detailed survey, see Szczepiek Reed 2010). They are identified on the basis of a number of external and internal criteria. The external cues include final lengthening, pausing, as well as a change in pitch level and/or pitch direction and anacrusis, i.e., the fast delivery of unaccented syllables at the beginning of a next unit. The internal cues mainly consist in a prominent syllable (the nucleus) embedded in a “coherent” pitch contour, which usually also embraces a number of further stressed and unstressed syllables. The latter are usually less prominent than the nucleus and form the pre-head, head, and tail of the IU (see, for instance, Cruttenden 1997, 27–8, 42–4). At times, some external boundary criteria are added to these, such as aspiration and a possible change in paralinguistic features, such as creaky voice quality (Crystal 1969; Halliday 1970, 1985; also Brazil et al. 1980; Brown et al. 1980 for the so-called British school; also du Bois et al. 1992, 100; Schuetze-Coburn 1992, 1994; du Bois 2008 for Discourse-Functional Linguistics). Similar criteria are used in the autosegmental-metrical approach (see, for instance, Beckman and Pierrehumbert 1986; Gussenhoven and Jacobs 2005; Nespor and Vogel 1986; Selkirk 1986). In essence, then, identifying IUs is mainly a matter of applying the list of criteria. In our sample excerpt, for instance, all three candidate

² I would like to thank Uwe-Alexander Küttner for allowing me access his data. They were recorded by a student during her stay in the US, with consent of the family members.

chunks contain a focus accent (-SA, -SA, -LU-). Moreover, while the end of the third chunk is less audible due to the overlap, the first two are delimited, among other features, by an audible high level tone and lengthening (last syllables of lines 1 and 2) as well as a small pitch upstep and anacrustic syllables (beginning of line 2).

However, already early accounts of the chunking of spoken language highlighted difficulties in any exhaustive analysis of spontaneous speech: Brown et al. (1980) noted that there are “many occasions where no clear indication is given of tone group boundary markers” (41). Cruttenden (1997) even admitted the identification of intonation-groups to be “something of a circular business” (29), and he highlighted sentence-final adverbials, vocatives, and reporting clauses in particular as notoriously difficult in terms of intonation-group delimitation (35–7). Other authors attribute any exceptions to the “unit rule” to emphasis, grammatical complexity, and speech rate (Quirk et al. 1985, 135, 1602, also Tench 1990). Yet, even that basically just acknowledges the problem: not all prosodic units are easily identifiable.

This also holds for our sample excerpt. When we try to transcribe how it continues after line 3 (for the sound file, see the supplementary material), we may end up with different transcript versions, especially for the stretch of talk 7–12 s into the excerpt. We depict two of the possible versions below with the relevant lines in minimal GAT2 transcript only, for argument’s sake.

(1’a) Glass tiles (UAK 10004, 2:34–2:46) – transcript version a (lines 3–5 in basic transcript, lines 7–12 in minimal transcript, background noise simplified).

```

3     MOT:  absoLUt[(ely) ]
4     FRA:  [i think] GLASSti:l:es==
5           =i think little glass TI:L:ES,
7  ->     (.) mosaic glass tiles might look good depending on what kind3
10    MAR:  (1.0)
11  -> FRA:  you know you can even put them on the floor en' you don't have
           to worry about (0.9) (about) slipping

```

(1'b) Glass tiles (UAK 10004, 2:34–2:46) – transcript version b (lines 3–5 in basic transcript, lines 7–12 in minimal transcript, background noise simplified).

```

3     MOT:  absoLUt[(ely) ]
4     FRA:  [i think] GLASSti:l:es==
5           =i think little glass TI:L:ES,
7  ->     (.) mosaic glas|s tiles
8  ->     might look good
9  ->     depending on what kind3
10    MAR:  (1.0)
11  -> FRA:  you know you can even put them on the floor
12  ->     en' you don't have to worry about (0.9) (about) slipping

```

If we focus on lines 7(-9) of the excerpt, we can see that transcript version (a) represents these as one chunk, while version (b) does so as three chunks. Both these versions can be considered correct, for different reasons. In favor of version (a) speaks that prominent unit-ending cues, like those at the end of line 5, occur only shortly before second 10 again. There is noticeable final lengthening and a high-rising final intonation on “kind” (Figure 1). Because of its marked pitch and duration, “kind” can be considered to bear the focus accent of this unit. However, we can also hear some – though less extensive – pitch and

³ The very end of this word is overlapped by a single clatter of a pot.

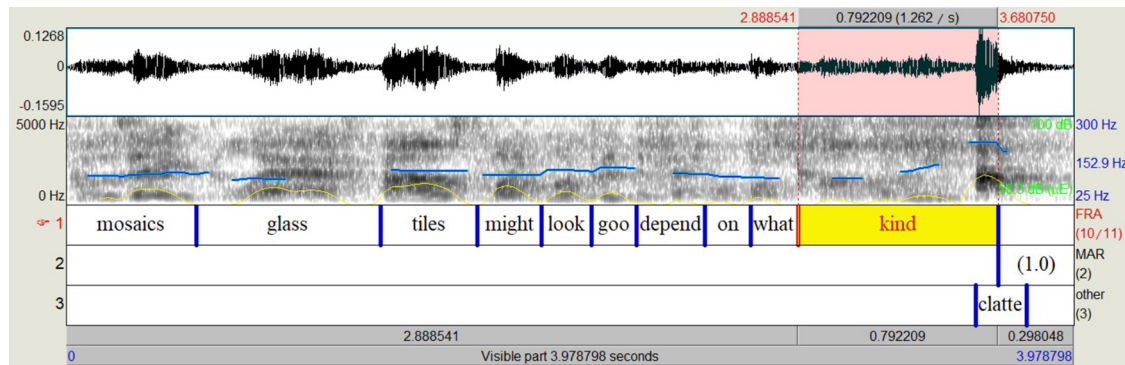


Figure 1: Waveform (top), spectrogram (middle), with overlaid pitch (blue) and intensity (yellow) of lines 7–9, annotated with wording, pausing, and other sounds (generated in PRAAT, Boersma and Weenink (2020)).

tempo changes *within* this candidate chunk: “tiles” is lengthened and ends with a small final rise, followed by a small pitch downstep to “might”. Something similar can be heard with “good” and “depending” a little later in this stretch. It is exactly these observations which transcript version (b) is based on. There these prosodic details serve as the justification for the line breaks notating this stretch of talk as three separate IUs with focus accents on “glass”, “might”, and “kind” (lines 7–9). Yet, these chunks are less clearly marked than the unit endings in line 5 as well as that preceding the pause in line 10. And, in retrospect, “glass” and “might”, while sufficiently prominent from an on-line perspective, are also less prominent than “kind”.

As a result, these units and in particular their alleged boundaries are somewhat fuzzy. A similar phenomenon can be found in lines 11–12. There we hear some parameter changes with “floor”, but, in retrospect, “slipping” carries the focus accent (Figure 2).

So, within the clearly identifiable chunks, further relevant parameter changes are audible, but are they sufficiently prominent to justify an IU boundary?

Scholars have suggested various solutions to the problem of identifying IUs. These have included, for instance, taking recourse to pausing (Brown et al. 1980, 46; but see, e.g., the clearly unit-internal pause in line 12) as well as syntactic and semantic criteria (Cruttenden 1997, 30; also Crystal 1969, 207; but see, e.g., the syntactically incomplete prosodic units in lines 1 as well as 4–5 in our sample excerpt). Others have suggested assuming a third category (“intonation subunits” in Du Bois et al. 1992, 68; also the intermediate phrase “ip” in the ToBI system, Beckman and Pierrehumbert 1986, for instance), but what if there are even more degrees, or if we only observe some of the parameter changes relevant for these unit types?

Hence, all these suggestions are variously problematic, for reasons of circularity and lack of granularity when it comes to fuzzy boundaries, i.e., instances of units that have something that is less than a full

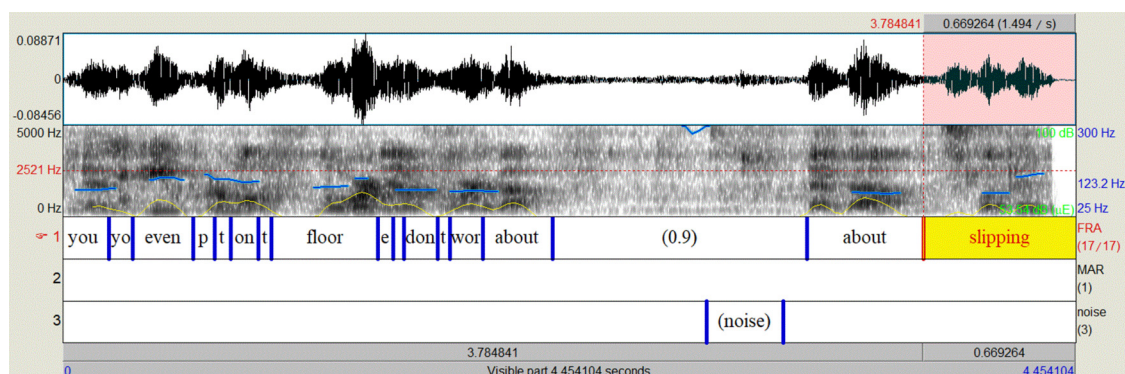


Figure 2: Waveform (top), spectrogram (middle), with overlaid pitch (blue) and intensity (yellow) of lines 11–12, annotated with wording, pausing, and other sounds (generated in PRAAT, Boersma and Weenink (2020)).

boundary but more than no boundary at all – an unsatisfactory situation when such cases are apparently not infrequent. And indeed, as Auer (2010) points out, the situation as such does not come as a surprise. The main reason for this is that in the unit approach it remains unclear whether one or several of the criteria mentioned for identifying IUs need to be met. Instead of an operationalization, the unit model suggests a complex heuristic. Its application to the data requires “erhebliches Geschick [considerable skill]” (8, see also Du Bois et al. 1992, 100) and even that does not always lead to clear decisions (Auer 2010, 8). Hence, the structuralist, unit-based approach to analyzing spoken language as such is an analytic problem, and as a result, the notion of “unit” becomes a rather questionable one (Deppermann and Proske 2015) and a challenge for novices to prosodic-phonetic chunking, but also for more experienced scholars, in particular, when their analyses go beyond idealized versions of talk.

Solving such analytical issues may then indeed, as Schegloff (1996a) put it, require “stretching the old linguistics to meet the challenge of talk-in-interaction [...] search with fresher eyes and ears, in the details of the talk with which we must, in the end, come to terms” (114). Rather than seeing talk-in-interaction as in some way defective or imperfect, we would then understand it as emerging, with all its specifics, from many indigenous practices to deal with the challenges (“contingencies”) of talking in real time, with one or more co-participants, and in a local environment that is “in flux” and brought about by the very fact that talk is produced in interaction.

1.3 The methodological approach: Conversation analysis and Interactional linguistics

Emanuel A. Schegloff’s advocating “fresher eyes and ears” comes from his particular interest in talk-in-interaction. Together with Harvey Sacks and Gail Jefferson, he developed Conversation Analysis (CA) – an approach that is influenced by sociology but interested in how the *in situ* social order of everyday interaction is accomplished by the participants themselves on the basis of their shared understanding of social order (see also Heritage 1984). CA starts from the assumption that the participants’ main aim is to maintain intersubjectivity and accomplish action – hence *inter*-action (for detailed introductions to CA, see, e.g., Hutchby and Wooffitt 1998, Sidnell 2010, also the papers in Sidnell and Stivers 2013, among many others). “Why that now?” summarizes CA’s research interest. To solve this question, CA has, for one, mainly considered the sequential organization of talk, i.e., the orderliness of the organization of utterances and actions and described a range of generic orders of organization to solve the “contingencies” participants face regularly in everyday and institutional talk. These include the issues of turn-taking, action formation, sequence organization, overall structural organization, word selection, and the practices of repair (Schegloff 2007, xiv). Second, in studying these, CA adheres to a number of basic methodological assumptions, among them:

- Conversational interaction is “the basic and primordial environment for the use and development (both ontogenetically and phylogenetically) of natural language” (Schegloff 1996a, 54; also Heritage 1984, 234–8). Also therefore, CA studies are intrinsically empirical and investigate “raw” data, i.e., audio and video recordings of naturally occurring interactions in their natural environment.
- In collections of instances of a phenomenon (Schegloff 1996b), “the locus of order is the single case” (Schegloff 1987, 102). Hence, CA cannot disregard deviant cases (Hopper 2008; Wootton 1989). Moreover, quantitative analysis, if needed, may only follow thorough and exhaustive qualitative analysis (Schegloff 1993a; Zimmerman 1993; but cf. Heritage 1999; also Mayring 2001; Barth-Weingarten 2003; Heritage 2004; Clayman et al. 2006).
- In its interactional exigencies, everyday talk is systematically organized; this order is displayed in the participants’ systematic use of shared sets of methodic practices and procedures for talking (or interacting), whose sharedness and use they rely upon and for which they hold each other mutually accountable. By their using these practices, the latter also become visible to the analyst. Next to microanalysis, the analyst can therefore also draw on sequential analysis and the “next turn proof procedure” to

examine participants' own displayed treatment and understanding of one another's talk through features of linguistic design and the next action embedded in the sequential context.

- There is order at all points (Schegloff and Sacks 1973, 290), so we cannot disregard any observation, even if it seems irregular at first sight. “[N]o order of detail can be dismissed, a priori, as disorderly, accidental or irrelevant” (Heritage 1984, 241).

In sum, then, CA seeks to explain both the single case and its details in its sequential context as well as the aggregate to identify robust interactional practices across specific instances. As such, it also presents itself as an ideal methodological starting point for tackling the details and challenges of prosodic chunking. The more so, as CA has – since the 1990s – been joined by interactional linguistics (IL) – an approach which combines CA's methodological tools and assumptions with a specific interest in linguistic phenomena and adds further linguistic methods where useful (for details, see, for instance, Couper-Kuhlen and Selting 2018). From the very start, this has also included a strong interest in prosody (see, e.g., Barth-Weingarten et al. 2010; Couper-Kuhlen and Selting 1996; Couper-Kuhlen and Ford 2004; Selting and Couper-Kuhlen 2001), inspired by the York Phonology/Phonetics for Conversation approach, which advocated impressionistic listening and a parametric approach already in the 1980s (Kelly and Local 1989a,b; also Local and Walker 2004, 1400).

This special issue is adopting the CA/IL approach to chunking spoken language. All papers discuss issues of chunking of natural talk-in-interaction and they consider the issues raised in a qualitative way. Where there is quantification, it is done only after detailed single-case analyses and in consideration of the potentially relevant details of the sequential placement of the target phenomenon. This approach is guided by the desire to understand the details of everyday talk, in everyday communicative situations, without the restrictions of, for instance, elicitation or laboratory equipment. The investigation includes the “order” of the details of prosody and phonetics. Thus, while acoustic analyses with PRAAT (Boersma and Weenink 2020), for instance, serve for illustrative purposes, prosodic-phonetic details are first studied auditorily in their indigenous environment and grounded in a close interactional and sequential analysis of the unfolding talk.

It was on this basis that Barth-Weingarten (2016) developed an alternative approach to the chunks audible in longer stretches of spoken language in order to solve the problems mentioned in Section 1.2.

1.4 The cesura model

More promising attempts at solving the problems of chunking began attributing more relevance to the boundaries between the chunks, rather than the chunks themselves (Peters et al. 2005; Mertens 2006; Degand and Simon 2009; also Auer 2010; Barnwell 2013). Yet, these boundary approaches still arrive at chunks in the end (but see Mertens and Simon 2013). The cesura approach (Barth-Weingarten 2016), in contrast, “radicalizes” the focus on the “boundary signals” and considers the chunks which the unit approach, and in essence also the boundary approach, focus on (Figure 3a and b), mere epiphenomena, the perceptual result of prosodic-phonetic parameter changes. In the cesura approach, in contrast, the focus lies on the prosodic-phonetic parameter changes themselves (changes in pitch, loudness, tempo, voice quality, vocal-tract configurations, etc.), but also their extent as well as where and how they cluster and thereby create “cesuras” (Figure 3c). The term “cesura”, then, refers to *all* kinds of discontinuities in the flow of talk produced by prosodic *and* phonetic parameter changes.

This understanding has two consequences. For one, the term “cesura” is not restricted to what has been studied under the heading of IU boundaries but can be employed to studying the variously sized “chunks” of the unit approach, including, but not restricted to, the concept of IU. Hence this concept also embraces the “beginnings” and “endings” of chunks larger than the IU (see, for instance, “spoken paragraphs” in Lehiste 1980, “paratones” in Yule 1980, “pitch sequences” in Brazil 1997) as well as of chunks smaller than the IU (see, for instance, “intonational phrase” [IPh] and “phonological utterance” in Beckman and

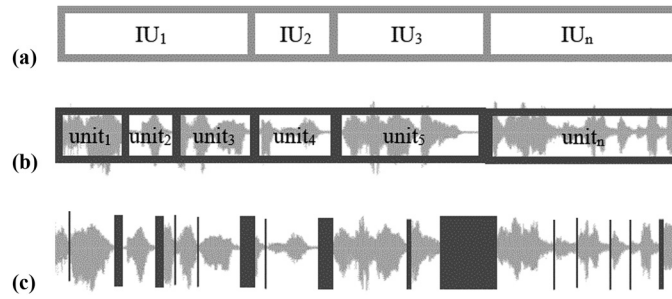


Figure 3: (a) Research focus of the unit approach. (b) Research focus of the boundary approach. (c) Research focus of the cesura approach (schematized).

Pierrehumbert 1986; Gussenhoven and Jacobs 2005; Nespor and Vogel 1986; Selkirk 1986) down to word, and even segment, boundaries.

Second, focusing on what comes between the “chunks” introduces a greater granularity in studying “boundaries” and how they are created. Parameter changes such as pitch and loudness changes may occur on their own, but more frequently they cluster – consider, for instance, the range of changes in the vocal-tract configurations involved in moving from one phonetic segment to the next (conventionally referred to as “segment boundaries”), but also the set of prosodic-phonetic cues, i.e., parameter changes, between (phonological) words, phrases, and IUs for instance (conventionally referred to as “word”, “phrase”, and “IU boundaries”, respectively). In general here, the more parameter changes occur at one point in talk, the stronger the perception of the cesura will be. As a result, the concept of “cesura” includes all sorts of variously strong “breaks” (junctures), ranging from stronger to “regular”, weaker, and weakest cesuras (see Figure 3c again).

This focus on what comes between chunks, rather than the units as such, and the resulting greater granularity are then also what in the end justifies speaking of a shift of perspective with the cesura approach as compared to the earlier unit, and also the boundary, approaches. And it is this focussing on “what comes between, and creates, the chunks” that also helps us explain, and capture, phenomena such as that of fuzzy IU boundaries: They are understood in terms of weak(er) cesuras, created by changes of fewer parameters and/or smaller extents of their changes at certain points in talk.

1.5 Notating weak(er) cesuras

Depending on the research focus, the parameter changes can be notated in variously granular fashions. A parametric notation (Figure 4a) shows greatest granularity, while a *Gestalt* notation is least granular but most efficient (Figure 4b). The latter uses “|” (a single vertical bar (character code: U + 007)) to notate weak(er) cesuras and can be further adapted to the granularity required by a certain study by distinguishing “|” and “||” etc.

Awareness of the phenomenon and being able to notate it allow us to take note of all cesuring phenomena, no matter whether big clusters or single parameter changes – or anything in between for that matter. Hence, the cesura approach also allows us to systematically attend to weak and fuzzy “boundaries”, i.e., points which previously were often either ignored in transcription and analysis or forced into one of the two categories (“IU” vs “non-IU”, for instance). The advantage of attending to them is obvious: once we can not(at)e such points, we can also study them, moving closer to answering questions such as: How are weak(er) cesuras produced? Which parameter changes are decisive? What functions do weak(er) cesuras have?

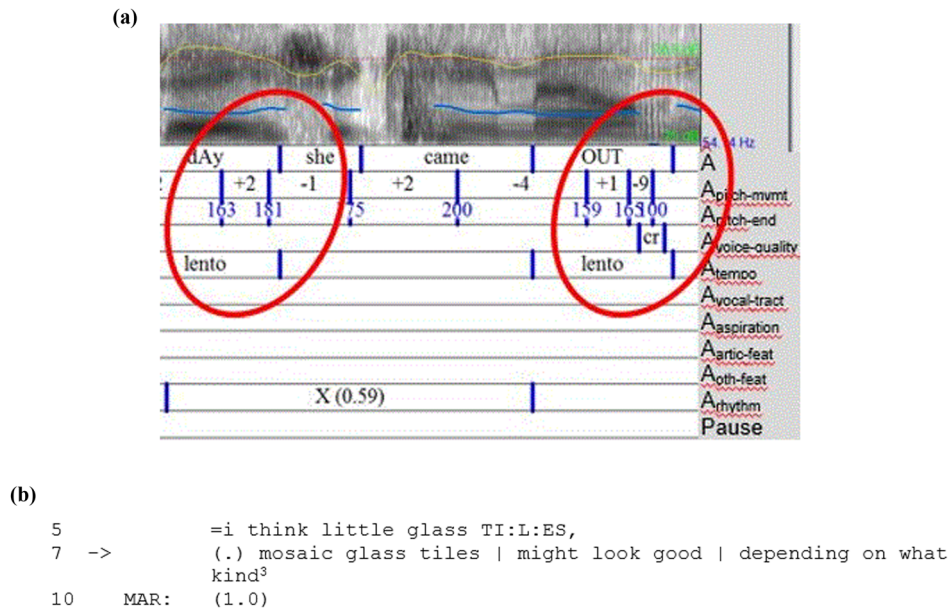


Figure 4: (a) Zooming in: Parametric notation of cesuring parameter changes by means of PRAAT (example not shown). (b) *Gestalt* notation of weak cesuras in example (1'), lines 5–10 (no individual parameter changes notated).

2 Gains from “taking a closer look” into prosodic-phonetic chunking

Employing a more granular approach allows further insight into prosodic-phonetic organization, as will be briefly illustrated in the following on the basis of previous work (Section 2.1) as well as papers in this special issue (Section 2.2).

2.1 Previous work on prosodic-phonetic “cesuring”

The cesura approach already helped complement earlier observations. Barth-Weingarten (2016, 93–177), for instance, ascertained cesuring parameters on the level of turn-final IUs in American English with CA/IL-type evidence. Most minimally these cues include

- a prominent syllable which is sufficiently marked by pitch, loudness, and/or duration,
- a noticeable final pitch movement rising or falling across at least 2ST, often more.

With falling final pitch movements this is often accompanied by

- glottalization, predominantly on the final (two) syllable(s).

This cluster is, moreover, regularly complemented by

- lengthening of mainly the utterance-final syllable to an extent that is often greater than the lengthening of the prominent syllable occurring in non-final position,
- a loudness/intensity decrease, often to lower than before in the Turn-Constructional Unit (TCU; see Sacks et al. 1974, for instance),

often accompanied by

- the audible release of the utterance-final plosive, if applicable,
- an open vocal tract configuration,
- some outflow of air, and/or
- a bit of silence.

(For the validity of these cues for other varieties of English and other languages, see Barth-Weingarten 2016, 169.)

Moreover, Barth-Weingarten (2016, 179–218) also showed that the end of turn-internal chunks as well as chunks within such chunks regularly exhibit clusters of features which resemble those for turn-final IUs in kind but are reduced in number and/or extent. Hence, we now also have CA/IL-type evidence that participants have prosodic-phonetic cues as to “where they are” in a longer exchange.

This already highlights that the chunking issue is not just an analytic one, but that participants can use weak(er) cesuras to accomplish certain *functions* in talk-in-interaction (for more details, see Sections 2.2 and 3). Moreover, when we notate weak(er) cesuras more systematically, we lay the foundation for not only extending our understanding of the prosodic-phonetic dimension but also for living up to the CA/IL principle of assuming order-at-all-points and pursuing the question: Are weak(er) cesuras interactionally relevant, i.e., do they constitute, or contribute to, practices that speakers use to pursue particular interactional ends? These latter issues then are exactly the questions this special issue intends to pursue further.

2.2 Papers focusing on prosodic-phonetic “cesuring” in this special issue

The paper by **Dagmar Barth-Weingarten**, **Uwe-Alexander Küttner**, and **Chase Raymond**, for instance, revisits what previous research labelled “**pivots**” – syntactic units that are connected by using one and the same lexico-syntactic item at their common juncture such that it constitutes both the end of the first unit and the beginning of the next at the same time. The component units of pivots are typically considered to be prosodically smoothly through-produced, such that also in this dimension the completion of the first unit simultaneously launches the second, allowing the speaker to continue past a possible transition-relevance place (TRP, Sacks et al. 1974). However, the paper shows that pivot constructions exhibit systematically varying degrees of prosodic-phonetic integration – i.e., stronger and weaker cesuras – which can be ordered on a cline. What is more, also the action dimension exhibits a cline, and it parallels the one in the prosodic-phonetic dimension. Tighter prosodic-phonetic integration of the actions accomplished by the components of the pivot construction co-occurs with rather retrospectively oriented post-pivot actions, looser prosodic-phonetic integration is associated with a more prospective orientation of the post-pivot’s action.

The paper by **Elizabeth Couper-Kuhlen** investigates weak(er) cesuras in the particle combination **OH+OKAY**. For this, she focusses on a specific sequential context: informings in recent everyday American-English talk-in-interaction (telephone and face-to-face). Here, the particles OH and OKAY are regularly used together in so-called “third position” to respond to information solicited by the OH/OKAY speaker in a prior turn. The paper shows, first, that different from simple OH and simple OKAY, the integrated OH+OKAY combinations regularly respond to the informing as not only supplying the information solicited but also accomplishing another action made relevant by the previous turn, such as dealing with a complaint, correcting a misbelief etc. Second, such OH+OKAY combinations can occur with varying degrees of prosodic-phonetic integration, ranging from juxtaposition of OH and OKAY with a strong cesura between them to fusion of the components with weak and weakest cesuras to occurrence as an indivisible whole. As long as the two parts are still distinguishable, both OH and OKAY make their individual contribution to the turn action-wise. Weak(er) cesuras here then seem to be a symptom of language change, the emergence of the particle combination as a relatively recent phenomenon, with OH and OKAY being integrated more closely due to the more frequent combined use (see also Bybee 2001, for instance). Nevertheless, differing cesuring strength also seems to be usable strategically: with OH+OKAYs with a strong cesura, OKAY still carries its full weight as a sequence-closing third. OH+OKAYs with a weak cesura, in contrast, may be deployed to construct multi-unit turns, while those with weakest cesuras serve as turn- or TCU-initial prefaces. Moreover, fused OH+OKAY can be used as a means to ambiguate epistemic positioning with a turn and thus buy the participants more space to manoeuvre. Differing cesuring strengths thus may indeed be a relevant design feature with OH+OKAY.

Also **Simona Pekarek Doehler** studies weak(er) cesuras in a process of language change. With French conversational data, the paper examines a recurrent pattern in which a **question-word question** is **followed by a candidate answer**. The candidate answer, however, follows its question with different degrees of proximity. Pekarek Doehler observes a continuum of synchronic usage, ranging from (i) the pattern being

implemented as two TCUs separated by a gap, to (ii) cases which display weak(er) cesuras due to speakers' more or less rushing into the next unit, to (iii) a variant in which both components are integrated in one TCU with one prosodic contour. As with the previous papers, the different degrees of integration also have repercussions on the action dimension: While cases of type (i) are used when participants pursue a response, the TCUs of type (ii) serve to propose the candidate answer as tentative – a guess, accomplished “on the fly”. With type (iii), the speaker delivers just one action, proffering a highly tentative guess. Pekarek Doehler argues that the three variants are connected in that the integrated format (iii) originates in the repeated interactional sequencing of the two originally subsequent actions of type (i) via type (ii). As with the particle combination in Couper-Kuhlen's paper, weakening of the cesuras is, thus, the routinized product of a frequent combination of language items in use. In Pekarek Doehler's case, it emerges from an interactionally motivated two-unit sequential pattern. Her findings support an understanding of interaction as a driving force for the routinization (or: grammaticization) of patterns of language use, which becomes even more visible when we study the fine details of talk in the prosodic-phonetic dimension. It also strikes us as interesting that in both these papers dealing with instances of language change, the patterns studied still show traces of the merger of the original components in the form of weakest cesuras between the original parts. Not(at)ing weak cesuras may thus also be considered a tool for identifying such cases of language change (for a similar phenomenon with the emergence of a final particle, see also Barth-Weingarten 2016, 221–40, for instance).

Finally, it needs to be noted that both Couper-Kuhlen's and Pekarek Doehler's studies also open up our view to other linguistic dimensions. For one, their observations highlight the divergence of lexical and prosodic units (fused version of OH+OKAY in Couper-Kuhlen's paper) as well as syntactic and prosodic units (type iii of the question-word question+candidate response pattern in Pekarek Doehler's paper). Second, Pekarek Doehler brings the kinetic dimension into the picture in that in her types ii and iii body movements and gaze behavior contribute to delivering the component parts as integrated. And indeed, extending the cesura approach to other linguistic dimensions (verbal and non-verbal) would be a useful next step, as will be argued in Section 3 and shown by these and other papers in this special issue.

3 Beyond prosodic-phonetic cesuras

3.1 Lexico-syntax

While the cesura model as such has been developed on the basis of, and for, the prosodic-phonetic dimension, it may also suggest a way of approaching other dimensions relevant in human interaction. Thus, we may, for instance, fruitfully expand the picture to other verbal dimensions. **Couper-Kuhlen's** paper in this special issue on the OH+OKAY particle combination already highlights an aspect of the role of weak(er) cesuras in the **lexical** dimension (see Section 2) (see also Szczepek Reed 2015).

Pekarek Doehler's paper, in turn, highlights cases where two **syntactic** units occur under one prosodic contour. Hence, specific syntactic and prosodic units may have a more intricate relationship than has hitherto been acknowledged – see also, for instance, the syntactically incomplete prosodic units in lines 1 and 4–5 in our sample excerpt. A systematic approach to “**syntactic** cesuring” based on CA/IL principles then, for instance, may also take Auer's (2009) “online syntax” as a starting point. Further relevant phenomena include turn-offering conjunctions (Barth-Weingarten 2016, 221–40; Koivisto et al. 2011; Mulder and Thompson 2008) and the questioning of syntactic categories in Ford et al. (2013, also Ono and Thompson 2020).

Xiaoting Li's paper, on the other hand, contributes to the discussion of weak cesuras by investigating **syntactically incomplete TCUs/turns** creating fuzzy boundaries with negative assessments of the recipient and non-present parties. Particularly, she explores how the speaker, at points of syntactic incompleteness, uses bodily-visual conduct to engage the recipient in negatively assessing others (see Section 3.2 for further details). This speaks to the fact that the kinetic dimension, too, needs to be added to complete the picture of how (dis)alignment of the dimensions can help the participants accomplish interactional goals together.

3.2 Kinetics

Kinetics is a dimension of human interaction that is available to the participants due to their co-presence and consequent mutual perceptibility in face-to-face interaction. Here, too, the cesura approach could help pave the way for dealing with the complexities of this dimension in that it guides us to see also the various bodily behaviors in a parametric way.

Research has shown that non-verbal bodily behaviors align with talk, while adding other dimensions of meaning to it (see, for instance, Kendon 1972, 2004; McNeill 1992, 2005; Mondada 2006; Mondada and Oloff 2011; Streeck 2009; Li 2014). This alignment, however, is complex, not least because bodily behaviors are also complex, embracing e.g. facial expression, manual gestures, body movements, etc. One fundamental issue in such studies is the “binding” between speech and non-speech events. For example, Streeck and Jordan (2009) highlight that gestures often start before the verbal formulation of their “referent”. Kaukomaa et al. (2013, 2014) explore how smiles and frowns work in turn-initial position. Sikveland and Ogdén (2012) show that gestures may be held across several turns at talk, until the action that coincided with the start of the gesture is closed off. Loehr (2012) shows that gesture peaks tend to align with intonation peaks, and gesture phrases with intermediate phrases. Holler and Levinson (2019) explain the binding problem in terms of multimodal *Gestalts*, where gesture and speech are bound together and enhance comprehension: multimodal speech is processed more rapidly than unimodal speech. What this shows is that – while there are still lots of uncharted terrain in this dimension – the alignment of important events in different modalities is – while close enough in time so that they are often perceived as belonging together, as part of a cluster of related events – not so precise as to be exactly synchronous. Discontinuities of movements, where they can be identified, are not all convergent, and certainly not so across all modalities. This is also illustrated by our sample excerpt, which, after all, comes from face-to-face talk-in-interaction. Multimodal transcription conventions follow Mondada (2018).

(1”) Glass tiles (UAK 10004, 2:38–2:40) (multimodal transcript)

5	FRA:	=i think+ litt•le† Raised eyebrows--->†	glass	*+##TI:L:ES, #	
5a		+right hd moves up		+#"little" gesture (fig.2)‡	
5b	mar	*disaffil. face, fig.1•			
					
		<u>fig. 1</u>		<u>fig. 2</u>	
7	Fra:	+* (.) +left hd moves up	mosaic*+§	G•L\$ASS +ti:les,=+	
7a	mar	*\$"down the walls"-gesture, fig.3\$+fold up hds+ *folds arms across chest*		*disaffil. face, fig.4 ->	
					
		<u>fig. 3</u>		<u>fig. 4</u>	

As line 5 shows, gesture peaks can also occur later than their “referent” (see Frank’s “little” gesture, which occurs on “tiles”). And body movements are not tied to verbal behavior as they can also occur instead of the latter (see Marge’s reactions to Frank’s suggestion in lines 5b and 7a).

Indeed, scholars studying talk-in-interaction have rather felt the need to introduce specific terms for those points in talk where linguistic dimensions converge (see “complex completion point” in Ford and Thompson 1996; “Zäsur” in Auer 2010; see also Ford et al. 1996, Selting 1996, for instance). And Ford (2004) already suspected that “different facets of turn projection may be deployed to be *precisely* non-convergent” (31, our emphasis; see also Clayman 2013, 158), in other words, that cases of divergence may be functional and thus interactionally relevant.

Also the papers in our special issue reflect this. **Verónica González Temer and Richard Ogden** set out to explore the (non)convergence of cesuras on the various verbal and non-verbal dimensions in Chilean Spanish conversation. With the example of the particle **MM** in face-to-face tasting sessions – in particular MMs marking incipient speakership and gustatory MMs – they show that vocal and non-vocal unit boundaries often disalign and argue that this serves as a semiotic resource in action ascription. As a possible multiple-meaning particle, MM is, for one, suited to show that various linguistic and non-linguistic dimensions contribute to its meaning, including sequential position, phonetic design, and kinetic behavior. Second, its embodied boundaries do not always coincide with the verbal ones – resulting in weak cesuras. The non-aligned boundaries of verbal and non-verbal phenomena are connected by larger temporal domains – stretches, rather than points, of talk – and thereby help create windows of opportunity. The latter provide participants with chances to do different actions at the same time, such as marking incipient speakership, indicating what the topic of that will be, securing reciprocity, displaying a stance, and mobilizing response, with what at first sight seems to be the same token. Hence, divergent boundaries, Temer and Ogden argue, should be considered a source of informational richness.

Li, too, focusses on the divergence of these main dimensions. Based on an examination of some 5 h of everyday Mandarin face-to-face conversations, she shows how such divergence can be used as a resource by the participants when speakers make negative, and thus socially inappropriate, assessments of the recipient and a non-present party. They leave the assessments syntactically incomplete, while at the same time deploying prosodic cues, such as lengthening, pause, and bodily-visual resources, such as facial expressions (lip pursing, eyebrow furrows), head shakes, and postural shifts, to display their evaluative stance and to solicit, and thus manage, recipient participation to collaboratively produce the assessment. Li argues that this displays the speaker’s orientation to the social-interactional inappropriateness of the assessments, which is an interactional, rather than a production, problem. The divergence of the dimensions, thus, has a clear function. Disregarding it for the sake of focusing on interactional units as such would make us miss out on the ways participants accomplish such subtle interactional tasks.

In the light of studies such as these and their findings, we would like to propose that a better way to see major events of conversation, such as TRPs, might indeed be as “windows of opportunity”. They may emerge “fuzzily”, with weak cesuras within and across modalities (see the prosodic-phonetic cesural areas in Barth-Weingarten (2013, 2016, 6, 75–6) and the multimodal phenomenon studies by Temer and Ogen in this special issue). As these “windows” emerge, they make (conditionally) relevant particular kinds of next actions (in terms of action accomplishment, but also in terms of “discourse/structural” organization, such as turn-taking). At the same time, such events are also evanescent, i.e. the opportunity for action fades after some time. Thus, with TRPs, in hearing a turn at talk in real time, a coparticipant monitors the talk, and other bodily conduct in face-to-face interaction, for cues which mark a place in time when they may come in with their own talk. This “place” is not a precise point, but rather a short stretch of time when the relevance of incoming talk and speaker transition is high. It builds up gradually, and there are also practices for a current speaker to close, or even prevent, this window and project their own continued talk while the TRP is building up (see the phenomenon of “rush-through” in Schegloff 1982; also Local and Walker 2004; Walker 2010, for instance). For multimodal talk, see the phenomenon studied by Temer and Ogden. At the same time, once that window is in place, it is not permanent and after a while (typically a fraction of a second), unless work is done to keep it open, it “closes”, and the process of speaker selection, for instance, starts again.

Pushing this approach even further, **Oliver Ehmer and Daniel Mandel** come to understand the verbal and non-verbal details of **transitions to joint enactments** in German face-to-face story tellings as kinetic kinds of “cesural areas”. Such enactments involve both a verbal quote and prosodic and bodily conduct that depicts the enacted figures. Looking into the precise realization of the transitions to the actual enactment, Ehmer and Mandel observe that, often, the gist of such enactments is already foreshadowed with prosodic and bodily cues in the talk preceding the actual quote in specific ways. For one, the cues of the various dimensions – prosodic and bodily parameter changes – occur at various places across the relevant stretch of talk (see also Temer and Ogden’s as well as Li’s paper). Moreover, just like with prosodic-phonetic “cesural areas” also the various multimodal cues (head movement, gaze etc.) are not necessarily aligned in their onset. Ehmer and Mandel argue, for one, that the spread of the phonetic and kinetic cues may lead to the perception of fuzzy boundaries, or weak cesuras, of the enactments. Second, these contribute to the building up of an “opportunity space” (Lerner 1991, see also “window of opportunity” above) for the enactment being accomplished jointly by the co-participants, which allows the latter to mutually take up, and elaborate on, and thus develop further, each other’s prior contributions. The fuzzy boundaries of the foreshadowings of enactments, for instance, help to jointly reconstruct what has happened or has been said (epistemic), negotiate a stance (deontic), and create involvement (emotional alignment). Ehmer and Mandel thus expand the “cesura approach” to multimodality even more radically. At the same time, they provide very practical examples of how this phenomenon could be analyzed, and how the results of such an analysis could be depicted.

3.3 Other “chunks” in talk-in-interaction

Last but not least, how we tackle these phenomena may also have consequences for our modelling of talk-in-interaction in general. So far we employ “action” (for instance, Levinson 2013), “practice” (for instance, Couper-Kuhlen and Selting 2018, 29), “turn”, and “TCU” (for instance, Clayman 2013; Drew 2013) as basic concepts in the description of how participants organize talk-in-interaction to accomplish action. Many of these concepts, however, are based on unit concepts. A TCU, for instance, is based on syntactic units (“sentences or clauses more generally, phrases, and lexical items” (Schegloff 2007, 3)), “intonational ‘packaging’” (4), i.e. IUs, and “it constitutes a recognizable action in context” (4; also Selting 1995; Thompson and Couper-Kuhlen 2005; Levinson 2013, 107; Couper-Kuhlen and Selting 2018, 29, for instance). While we need labels to refer to these features, already their labelling – *sentence*, *clause*, *phrase*, *item*, *package*, *action* – as such suggests that we are dealing with clear-cut entities, which, once we look at things more closely, may not typically be the case. Some papers in this special issue demonstrate that the cesura approach can, for instance, also lead to insights concerning further relevant “chunks”. **Couper-Kuhlen’s** paper highlights a composite particle, **Pekarek Doehler’s** paper focusses on a composite clause format, both with specific functions. And the paper by **Barth-Weingarten, Küttner, and Raymond**, for instance, demonstrates how a more fine-grained prosodic-phonetic analysis may lead to an appreciation of “in-between” concepts such as that of “subsidiary **actions**” (see also Auer 2007), i.e. an action that, rather than being clearly separable from the surrounding ones, retrospectively elaborates the main action preceding it, and thus neither represents a different version of “the same action”, nor a “fully independent” or “discrete” next action. The idea of describing “action” in terms of a cline connects this research to other scalar reconceptualizations of longstanding interactional phenomena in CA/IL, such as conditional relevance and the mobilization of response.

4 Conclusion

In this way of seeing things, then, non-convergent boundaries and weak cesuras are not problematic, but a phenomenon relevant for both scholars and participants. No doubt, such detail in the production of talk is a

challenge to describe, but at the same time it is a tool for research and it can apparently even be a resource for social action. Shifting our attention from identifying boundaries (and with boundaries, units), toward a developing understanding of the orderliness of the (fine) details of talk helps us to appreciate the processes and emerging relevancies of talk-in-interaction. With this special issue, we extend an invitation to collaborate in, and contribute to, this analytic endeavor.

The editors of this special issue would like to thank the participants of the international workshop “Divergent units”, held at Potsdam (Germany) in September 2018 and the panel “Divergent units and fuzzy boundaries” at IEMCA 2019 in Mannheim (Germany) for their willingness to embark on this expedition with us, sharing their thoughts and ideas on fuzzy boundaries and weak cesuras. This holds in particular for the contributors to this special issue, who also commented on first versions of the papers published here. Moreover, our thanks is due to colleagues who were willing to share insights from their perspectives when commenting on the papers submitted for this special issue (in alphabetical order): Michael Ashby, Pia Bergmann, Steven Clayman, Arnulf Deppermann, Davydd Gibbon, Marja-Liisa Helasvuo, Elliott Hoey, Jill House, Lorenza Mondada, Beatrice Szczepek Reed, Giovanni Rossi, Ann Wichmann, Nick Williams, and Margret Zellers.

Acknowledgements: The authors would like to thank Elizabeth Couper-Kuhlen and Uwe-A. Küttner as well as two anonymous reviewers for their helpful comments on earlier versions of this introductory paper.

Funding information: The workshop “Divergent units”, held at Potsdam (Germany) in September 2018, was funded by the German Research Foundation DFG (GZ BA 3495/5-1).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: All data used in this introductory paper are included in this published article (and its supplementary information files).

References

- Auer, Peter. 2007. “Why are increments such elusive objects? An afterthought.” *Pragmatics* 17(4), 647–58. doi: 10.1075/prag.17.4.03aue.
- Auer, Peter. 2009. “Online syntax: Thoughts on the temporality of spoken language.” *Language Sciences* 31, 1–13. doi: 10.1016/j.langsci.2007.10.004.
- Auer, Peter. 2010. “Zum Segmentierungsproblem in der gesprochenen Sprache.” *InLiSt – Interaction and linguistic structures* 49. <http://www.inlist.uni-bayreuth.de/issues/49/InList49.pdf>.
- Barnwell, Brendan. 2013. “Perception of prosodic boundaries by untrained listeners.” In: *Units of talk – Units of action*, eds. Szczepek Reed, Beatrice and Geoffrey Raymond, p. 125–65. Amsterdam: John Benjamins. doi: 10.1075/slsi.25.05bar.
- Barth-Weingarten, Dagmar. 2003. *Concession in spoken English*. Tübingen: Narr.
- Barth-Weingarten, Dagmar. 2013. “From “intonation units” to cesuring – an alternative approach to the prosodic-phonetic structuring of talk-in-interaction.” In: *Units of talk – units of action, Series “Studies in Language and Social Interaction”*, eds. Szczepek Reed, Beatrice and Geoffrey, Raymond, p. 91–124. Amsterdam: Benjamins.
- Barth-Weingarten, Dagmar. 2016. *Intonation units revisited: Cesuring in talk-in-interaction*. Amsterdam: John Benjamins. doi: 10.1075/slsi.29.
- Barth-Weingarten, Dagmar, Elisabeth Reber, and Margret Selting, eds. 2010. *Prosody in interaction*. Amsterdam: John Benjamins. doi: 10.1075/sidag.23.
- Beckman, Mary E. and Janet B. Pierrehumbert. 1986. “Intonational structure in Japanese and English.” *Phonology* 3, 255–309. doi: 10.1017/S095267570000066X.
- Boersma, Paul and David Weenink. 2020. *Praat: doing phonetics by computer* [Computer program]. Version 6.1.37, <http://www.praat.org/>.

- Brazil, David. 1997. *The communicative value of intonation*. Cambridge: Cambridge University Press.
- Brazil, David, Malcolm Coulthard, and Catherine Johns. 1980. *Discourse intonation and language teaching*. London: Longman.
- Brown, Gillian, Karen L. Currie, and Joanne Kenworthy. 1980. *Questions of intonation*. London: Croom Helm.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Clayman, Steven E. 2013. "Turn-constructional units and the transition-relevance place." In: *The handbook of conversation analysis*, eds. Sidnell, Jack and Tanya Stivers, p. 150–66. Chichester: Wiley-Blackwell. doi: 10.1002/9781118325001.ch8.
- Clayman, Steven E., Marc N. Elliott, John Heritage, and Laurie L. McDonald. 2006. "Historical trends in questioning presidents 1953–2000." *Presidential Studies Quarterly* 36(4), 561–83. doi: 10.1111/j.1741-5705.2006.02568.x.
- Couper-Kuhlen, Elizabeth and Dagmar Barth-Weingarten. 2011. "A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 12, 1–51. <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2011/px-gat2-englisch.pdf>.
- Couper-Kuhlen, Elizabeth and Cecilia A. Ford, eds. 2004. *Sound patterns in interaction*. Amsterdam: John Benjamins. doi: 10.1075/tsl.62.
- Couper-Kuhlen, Elizabeth and Margret Selting, eds. 1996. *Prosody in conversation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511597862.
- Couper-Kuhlen, Elizabeth and Margret Selting. 2018. *Interactional linguistics*. Cambridge: Cambridge University Press.
- Cruttenden, Alan. 1997. *Intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166973.
- Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- Degand, Liesbeth and Anne-Catherine Simon. 2009. "Mapping prosody and syntax as discourse strategies: How basic discourse units vary across genres." In: *Where prosody meets pragmatics*, eds. Barth-Weingarten, Dagmar and Nicole Dehé, Anne Wichmann, p. 79–105. Bingley: Emerald. doi: 10.1163/9789004253223.
- Deppermann, Arnulf and Nadine Proske. 2015. "Grundeinheiten der Sprache und des Sprechens." In: *Handbuch 'Satz, Schema, Äußerung'*, eds. Dürscheid, Christa and Jan Georg Schneider, p. 17–47. Berlin: de Gruyter. doi: 10.1515/9783110296037-003.
- Drew, Paul. 2013. "Turn design." In: *The handbook of conversation analysis*, eds. Sidnell, Jack and Tanya Stivers, p. 131–49. Malden, MA: Wiley-Blackwell. doi: 10.1002/9781118325001.ch7.
- Du Bois, John W. 2008. *Rhythm and tunes: The intonation unit in the structure of dialogic engagement. Presentation held at the International Conference "Prosody and interaction" in Potsdam, 15–17 September 2008*.
- Du Bois, John W., Stephan Schuetze-Coburn, Danae Paolino, and Susanne Cummings. 1992. "Discourse transcription." *Santa Barbara papers in linguistics* 4. Santa Barbara: Dept. of Linguistics, University of California.
- Ford, Cecilia E. 2004. "Contingency and units in interaction." *Discourse Studies* 6, 27–52. doi: 10.1075/prag.17.4.02cou.
- Ford, Cecilia E., Barbara A. Fox, and Sandra A. Thompson. 1996. "Practices in the construction of turns: The "TCU" revisited." *Pragmatics* 6(3), 427–54. doi: 10.1075/prag.6.3.07for.
- Ford, Cecilia E., Barbara Fox, and Sandra A. Thompson. 2013. "Units and/or Action Trajectories? The language of grammatical categories and the language of social action." In: *Units of talk – Units of action*, eds. Szczepek Reed, Beatrice and Geoffrey Raymond, p. 13–55. Amsterdam: John Benjamins. doi: 10.1075/slsi.25.02for.
- Ford, Cecilia E. and Sandra A. Thompson. 1996. "Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns." In: *Interaction and grammar*, eds. Ochs, Elinor, Emanuel A. Schegloff, and Sandra A. Thompson, p. 134–84. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620874.003.
- Gussenhoven, Carlos and Haike Jacobs. 2005. *Understanding phonology*. Cambridge: Cambridge University Press.
- Halliday, Michael A. K. 1970. *A course in spoken English: Intonation*. London: Oxford University Press.
- Halliday, Michael A. K. 1985. *An introduction to functional grammar*. London: Arnold.
- Heritage, John. 1984. *Garfinkel and ethnomethodology*. Cambridge: Polity Press.
- Heritage, John. 1999. "Conversation analysis at century's end: Practices of talk-in-interaction, their distributions, and their outcomes." *Research on Language and Social Interaction* 32(1–2), 69–76. doi: 10.1080/08351813.1999.9683609.
- Heritage, John. 2004. "Conversation analysis and institutional talk." In: *The Handbook of language and social interaction*, eds. Fitch, Kristine L. and Robert E. Sanders, p. 103–47. Mahwah, N.J.: Lawrence Erlbaum.
- Holler, Judith and Levinson, Stephen C. 2019. "Multimodal language processing in human communication." *Trends in Cognitive Sciences* 23(8), 639–52. doi: 10.1016/j.tics.2019.05.006.
- Hopper, Robert. 2008. "Speech, for instance: The exemplar in studies of conversation." In: *Methods in language and social interaction, vol. 2: Ethnography, ethnomethodology and conversation analysis*, ed. Hutchby, Ian, p. 300–19. London: Sage.
- Hutchby, Ian and Robin Wooffitt. 1998. *Conversation analysis. Principles, practices and applications*. Cambridge: Polity Press.
- Kaukoma, Timo, Anssi Peräkylä, and Johanna Ruusuvoori. 2013. "Turn-opening smiles: Facial expression constructing emotional transition in conversation." *Journal of Pragmatics* 55, 21–42. doi: 10.1016/j.pragma.2013.05.
- Kaukoma, Timo, Anssi Peräkylä, and Johanna Ruusuvoori. 2014. "Foreshadowing a problem: Turn-opening frowns in conversation." *Journal of Pragmatics* 71, 132–47. doi: 10.1016/j.pragma.2014.08.002.
- Kelly, John and John Local. 1989a. *Doing phonology. Observing, recording, interpreting*. Manchester, New York: Manchester University Press. doi: 10.1017/S0047404500016201.

- Kelly, John and John Local. 1989b. "On the use of general phonetic techniques in handling conversational material." In: *Conversation. An interdisciplinary perspective*, eds. Roger, Derek and Peter Bull, p. 197–212 Clevedon: Multilingual matters.
- Kendon, Adam. 1972. "Some relationships between body motion and speech." In: *Studies in dyadic communication*, eds. Siegman, Aron Wolfe and Benjamin Pope, p. 177–210. New York: Pergamon Press. doi: 10.1016/B978-0-08-015867-9.50013-7.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572.
- Koivisto, Aino, Ritva Laury, and Eeva-Leena Seppänen. 2011. "Syntactic and actional characteristics of Finnish *että*-clauses." In: *Subordination in Conversation. A cross-linguistic perspective*, eds. Laury, Ritva and Ryoko Suzuki, p. 69–102. Amsterdam: John Benjamins. doi: 10.1075/slsi.24.
- Lehiste, Ilse. 1980. *Phonetic characteristics of discourse*. Tokyo: Acoustical Society of Japan. doi: 10.1111/j.1467-9582.1982.tb00716.x.
- Lerner, Gene H. 1991. "On the syntax of sentences-in-progress." *Language in society* 20, 441–58. doi: 10.1017/S0047404500016572.
- Levinson, Stephen C. 2013. "Action formation and ascription." In: *The handbook of conversation analysis*, eds. Sidnell, Jack and Tanya Stivers, p. 103–30. Chichester: Wiley-Blackwell. doi: 10.1002/9781118325001.ch6.
- Li, Xiaoting. 2014. *Multimodality, interaction and turn-taking in Mandarin conversation*. Amsterdam: John Benjamins. doi: 10.1075/scl.3.
- Local, John and Gareth Walker. 2004. "Abrupt-joins as a resource for the production of multi-unit, multi-action turns." *Journal of Pragmatics* 36, 1375–403. doi: 10.1016/j.pragma.2004.04.006.
- Loehr, Daniel P. 2012. "Temporal, structural, and pragmatic synchrony between intonation and gesture." *Laboratory Phonology* 3(1), 71–89. doi: 10.1515/lp-2012-0006.
- Mayring, Phillip. 2001. "Kombination und Integration qualitativer und quantitativer Analyse." *Forum Qualitative Sozialforschung* 2(1). <https://www.qualitative-research.net/index.php/fqs/article/view/967/>.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, David. 2005. *Thought, imagery, and language: How gestures fuel thought and speech*. Chicago: University of Chicago Press.
- Mertens, Piet. 2006. "A predictive approach to the analysis of intonation in discourse in French." In: *Prosody and syntax: Cross-linguistic perspectives*, eds. Kawaguchi, Yuji, Ivan Fonagy, and Tsunekazu Moriguchi, p. 64–101. Amsterdam: John Benjamins. doi: 10.1075/ubli.3.
- Mertens, Piet and Anne Catherine Simon. 2013. "Towards automatic detection of prosodic boundaries in spoken French." In: *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013). Leuven, 11–13 September 2013, Leuven, Belgium*, eds. Mertens, Piet and Anne Catherine Simon, p. 81–7. https://www.arts.kuleuven.be/ling/cohista/conference/idp2013/documents/proceedings_idi2013.
- Mondada, Lorenza. 2006. "Participants' online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence." *Discourse Studies* 8, 117–29. doi: 10.1177/1461445606059561.
- Mondada, Lorenza. 2018. *Conventions for multimodal transcription*. https://franzoestik.philhist.unibas.ch/fileadmin/user_upload/franzoestik/home/Personen/Mondada/Unterordner/Mondada_conv_multimodality.pdf.
- Mondada, Lorenza and Florence Oloff. 2011. "Gestures in overlap: the situated establishment of speakership." In: *Integrating gestures. The Interdisciplinary Nature of Gesture*, eds. Stam, Gale, and Mika Ishino, p. 321–38. Amsterdam: John Benjamins. doi: 10.1075/g.4.
- Mulder, Jean and Sandra A. Thompson. 2008. "The grammaticization of *but* as a final particle in English conversation." In: *Crosslinguistic studies of clause-combining: the multifunctionality of conjunctions*, ed. Laury, Ritva, p. 179–204. Amsterdam: John Benjamins. doi: 10.1075/tsl.80.
- Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Ono, Tsuyoshi and Sandra A. Thompson eds. (2020). *The 'Noun Phrase' across Languages. An emergent unit in interaction*, p. 316–27. Amsterdam: John Benjamins. doi: 10.1075/tsl.128.
- Peters, Benno, Klaus J. Kohler, and Thomas Wesener. 2005. "Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache." In: *Prosodic structures in German Spontaneous Speech. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 35a*, eds. Klaus J. Kohler, Felicitas Kleber, and Benno Peters, p. 143–84. Kiel: Institut für Phonetik und digitale Sprachverarbeitung. https://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35a/aipuk35a_6.pdf.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A simplest systematics for the organisation of turn-taking for conversation." *Language* 50, 696–735. doi: 10.1016/B978-0-12-623550-0.50008-2.
- Schegloff, Emanuel A. 1982. "Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences." In: *Analyzing discourse: Text and talk*, ed. Tannen, Deborah, p. 71–93. Washington: Georgetown University Press.

- Schegloff, Emanuel A. 1987. "Analyzing single episodes of interaction: an exercise in conversation analysis." *Social Psychology Quarterly* 50(2), 101–14. doi: 10.2307/2786745.
- Schegloff, Emanuel A. 1993. "Reflections on quantification in the study of conversation." *Research on Language and Social Interaction* 26(1), 99–128. doi: 10.1207/s15327973rlsi2601_5.
- Schegloff, Emanuel A. 1996a. "Turn-organization: one intersection of grammar and interaction." In: *Interaction and grammar*, eds. Ochs, Elinor, Emanuel A. Schegloff, and Sandra A. Thompson, p. 52–133. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620874.002.
- Schegloff, Emanuel A. 1996b. "Confirming allusions: Toward an empirical account of action." *American Journal of Sociology* 102(1), 161–216. doi: 10.1086/230911.
- Schegloff, Emanuel A. 2007. *Sequence organization in interaction: a primer in conversation analysis*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791208.
- Schegloff, Emanuel A. and Harvey Sacks. 1973. "Opening up closings." *Semiotica* 8(4), 289–327. doi: 10.1515/semi.1973.8.4.289.
- Schuetze-Coburn, Stephan. 1992. "Prosodic phrase as prototype." *Proceedings of the IRCS workshop on prosody in natural speech*, vol. 1, p. 171–80. Philadelphia, PA: Institute for Research in Cognitive Science, Univ. of Pennsylvania.
- Schuetze-Coburn, Stephan. 1994. *Prosody, syntax, and discourse pragmatics: Assessing information flow in German conversation*. PhD thesis. Los Angeles, CA: University of California, USA.
- Selkirk, Elizabeth. 1986. "On derived domains in sentence phonology." *Phonology* 3, 371–405. doi: 10.1017/S0952675700000695.
- Selting, Margret. 1995. "Der 'mögliche Satz' als interaktiv relevante syntaktische Kategorie." *Linguistische Berichte* 158, 298–325.
- Selting, Margret. 1996. "On the interplay of syntax and prosody in the constitution of turn-constructive units and turns in conversation." *Pragmatics* 6(3), 357–88. doi: 10.1075/prag.6.3.06sel.
- Selting, Margret and Elizabeth Couper-Kuhlen eds. 2001. *Studies in interactional linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Sidnell, Jack. 2010. *Conversation Analysis. An introduction*. Chichester: Wiley and sons.
- Sidnell, Jack and Tanya Stivers eds. (2013). *The handbook of conversation analysis*. Oxford: Wiley Blackwell. doi: 10.1002/9781118325001.
- Sikveland, Rein Ove and Richard Ogden. 2012. "Holding gestures across turns: Moments to generate shared understanding." *Gesture* 12(2), 167–200. doi: 10.1075/gest.12.2.03sik.
- Streeck, Jürgen. 2009. *Gesturecraft. The manufacture of meaning*. Amsterdam: Benjamins. doi: 10.1075/gs.2.
- Streeck, Jürgen and J. Scott Jordan. 2009. "Projection and anticipation: The forward-looking nature of embodied communication." *Discourse processes* 46(2–3), 93–102. doi: 10.1080/01638530902728777.
- Szcepek Reed, Beatrice. 2010. "Intonation phrases in natural conversation: A participants' category?." In: *Prosody in interaction*, eds. Barth-Weingarten, Dagmar, Elisabeth Reber, and Margret Selting, p. 191–212. Amsterdam: John Benjamins. doi: 10.1075/sidag.23.
- Szcepek Reed, Beatrice. 2015. "Managing the boundary between 'yes' and 'but': Two ways of disaffiliating with German 'ja aber' and 'jaber'." *Research on Language and Social Interaction* 48(1), 32–57. doi: 10.1080/08351813.2015.993843
- Szcepek Reed, Beatrice and Geoffrey Raymond. 2013. "The question of units for language, action and interaction." In: *Units of talk – Units of action*, eds. Szcepek Reed, Beatrice and Geoffrey Raymond, p. 277–312. Amsterdam: John Benjamins. doi: 10.1075/slsi.25.01szc.
- Tench, Paul. 1990. *The roles of intonation in English discourse*. Frankfurt/Main: Lang.
- Thompson, Sandra A. and Elizabeth Couper-Kuhlen. 2005. "The clause as locus of grammar and interaction." *Discourse Studies* 7(4–5), 481–505. doi: 10.1177/1461445605054403.
- Walker, Gareth (2010). "The phonetic constitution of a turn-holding practice: Rush-throughs in English talk-in-interaction." In: *Prosody in interaction*, eds. Barth-Weingarten, Dagmar, Elisabeth Reber, and Margret Selting, p. 51–72. Amsterdam: John Benjamins. doi: 10.1075/sidag.23.
- Wootton, Anthony J. 1989. "Remarks on the methodology of conversation analysis." In: *Conversation: An interdisciplinary perspective*, eds. Roger, Derek and Peter Bull, p. 238–58. Clevedon: Multilingual Matters.
- Yule, George. 1980. "Speakers' topics and major paratones." *Lingua* 52, 33–47. doi: 10.1016/0024-3841(80)90016-9.
- Zimmerman, Don H. 1993. "Acknowledgment tokens and speakership incipency revisited." *Research on Language and Social Interaction* 26(2), 179–94. doi: 10.1207/s15327973rlsi2602_4.

Appendix: Transcription conventions

The transcription follows the GAT2 conventions (for a concise survey of the transcription symbols, see Couper-Kuhlen and Barth-Weingarten 2011, 37–9). Multimodal transcription follows Mondada (2018). These notations were complemented by the following symbols:

- | weak cesura
- ↓ larger pitch downstep
- ↓ smaller pitch downstep