

## Future prospects of research on prosody The need for publicly available corpora

### Comments on Margret Selting “Prosody in interaction: State of the art”\*

Arnulf Deppermann

Institute for the German Language (IDS), Mannheim, Germany

In her overview, Margret Selting makes the case for the claim that dealing with authentic conversation necessarily lies at the heart of an interactional-linguistic approach to prosody (see Selting this volume, Section 3.3). However, collecting and transcribing corpora of authentic interaction is a time-consuming enterprise. This fact often severely restricts what the individual researcher is able to do in terms of analysis within the scope of his or her resources. Still, for dealing with many of the desiderata Margret Selting points out in Section 5 of her extensive overview, the use of larger corpora seems to be required. In this commenting paper, I want to argue that future progress in research on prosody in interaction will essentially rest on the availability and use of large public corpora. After reviewing arguments for and against the use of public corpora, I will discuss some upshots regarding corpus design and issues of transcription of public corpora.

#### 1. Publicly available corpora: Pros and cons

Although in a rudimentary fashion, the need for a larger corpus already becomes visible in the first steps of a close sequential analysis of single cases. In research on prosody, we find it especially hard to follow the basic conversation-analytic maxim that it is our duty as an analyst to show that conversationalists orient to some putative device (cf. Selting this volume, Section 4.2). This is because in analyzing single cases, it often is not plainly evident that interactants orient to prosodic features as such – in contrast to, e.g. sequential actions or lexical categorizations. It is a major challenge to find out

---

\* I thank Dagmar-Barth-Weingarten and Elisabeth Reber for helpful comments on an earlier version of this paper.

and to show that they react specifically to some prosodic feature – and not to some other contingently co-occurring feature in the turn, patterns of multimodal participation and/or the sequential environment. We often need larger collections of cases in order to see and to show that they exhibit systematic practices of using and responding to prosodic features which cannot be accounted for on the grounds of, e.g., lexical items or action type properties. Equally, the single case often does not show clearly if and how some prosodic feature might resonate with other co-occurring features which can be said to necessarily accompany the prosodic device in question in order to achieve some interactional effect or which can be seen to determine systematically the interactional import of the prosodic device.

In other words, the size and the composition of the researcher's own corpus often severely limits what can be achieved. Larger corpora are needed in order to discover the range of variants of some device in question, their factual import (or, just to the contrary, the fact that some variation does not matter) and the generic, context-free properties of prosodic practices in contrast to their context-sensitive adaptations. While single case analysis may not even allow for noticing a possibly distinct phenomenon and for developing testable hypotheses, corpora often exhibit patterns which inspire the inductive discovery of some phenomenon (which then in turn needs close single case analysis).

As Margret Selting makes clear, co-occurring grammatical constructions, sequential environments, (e.g., institutional) genres, communities of practice, kinds of speakers and languages/language types will be the major candidates for comparative and contrastive studies (cf. Selting this volume: Sections 5.1.7, 5.2, 5.4, 5.5), which look for context-sensitive contrasts and specializations as well as for commonalities and generalizations which transcend the individual context. Turning to larger corpora thus will not only lead to a differentiation of findings, but may as well open our eyes to even more general structures than can become obvious in a small corpus. Larger corpora are thus indispensable for refining analyses, testing the robustness of findings in different contextual environments and checking how general some phenomenon or mechanism actually is. Finally, only large corpora allow for providing statistical evidence. According to conversation-analytic standards, however, codings necessary for such quantification should only be assigned after having carefully worked through a large collection of cases by hand. This is needed as the basis for discovering the relevant emic categories which can serve as codes (cf. Schegloff 1993).

Yet, many researchers with a conversation-analytic background refuse to work with public corpora. I will deal with two major objections and possible counter-arguments against them:

1. The researcher does neither have comprehensive knowledge of the interactional context nor of its ethnographic background. This may lead to an inadequate analysis. If we do not know exactly conversational histories and ethnographic

prerequisites for action and interpretation, we may not be able to arrive at a comprehensive analysis of the single instance, i.e., we may miss some delicate details of what exactly is indexed by the single token in its context (e.g., disappointment because of the deception of hopes that had priorly been inspired by the interlocutor; prosodic escalation in overlap because of claims to superior expertise because of professional history...). However, the availability of the recording of the whole encounter and ethnographic background knowledge seem to be relevant only if culture-specific contextualization conventions (Gumperz 1982), which do not exhibit direct sequential repercussions, matter or if phenomena relating to larger sequences, like contrastive prosodic packaging of different larger activities, are at issue.

In most cases we will still be able to identify generic, “context-free” functional uses and potentials which matter for the prosodic practice as such and which account for the possibility to use it in an infinite range of individual contexts. An example is glottal closure at the end of candidate turn constructional units. While motivations for the use of glottal closure in this position of the construction of a turn are multifaceted and highly context-dependent, glottal closure is invariably used to contextualize the unit as an unfinished fragment (see Selting 2001). Admittedly, if the corpus does not allow for more than a KWIC-concordance with severely limited access to the sound surrounding the focal phenomenon, it will be impossible to carry out a serious sequential analysis. Therefore, a public corpus must provide access to the whole communicative event or at least “big packages” within complex speech events. This is a requirement which is, e.g., in contrast to corpora suited for applications in speech technology (cf. Wichmann 2008).

2. Another objection against public corpora is the indexicality of transcription practices (Mondada 2007). Transcripts inevitably vary with respect to the research question at hand and also with regard to theoretical commitments about impressionistic vs. selective transcription, the nature of units and boundaries, the adequacy of categories like ‘(focal/nucleus) accent’, the relevance of formal vs. functional features in transcripts, etc (cf. Selting this volume: Section 6.2). The objection against a public corpus thus is that researchers are faced with transcripts, which may be too coarse, too selective, too fine-grained or laden with theoretical commitments and therefore distort possibly relevant features of the original sound or make them inaccessible. Still, I will argue below that it is possible to define a base standard for searchable public data bases which are useful for all different concerns of prosodic research in Interactional Linguistics.

In what follows, I will sketch some considerations for the design of public corpora which comply with the needs of Interactional Linguists.

## 2. Criteria for the design of publicly available corpora

There are some basic corpus-linguistic criteria which apply to spoken corpora in general (see Merkel and Schmidt 2009, Bird and Simons 2002, Baude et al. 2006). They are also relevant for corpora suited for research on prosody. Among the most important criteria are:

- The corpus needs to be publicly available via internet or on CD/DVD as a machine-readable and searchable data-base,
- it should use interoperable data-formats (preferably XML for text and WAV for sound).
- it needs to include audio- and possibly video-files,
- they should be aligned with the transcript so that the sound can be played immediately from the transcript,
- it should include meta-data concerning speakers and speech events,
- the corpus should be managed by a data base which allows for browsing the corpus and for searches using (combinations of) meta-data and features of the transcripts.

For interactional research on prosody, the development of corpora which are equipped with search tools which address aspects of sequential structure of interactions, such as turn-beginnings and -endings, turn-transition or overlap, is a major task.<sup>1</sup> Data-bases for research on prosody in interaction should allow for reading transcripts and listening to recordings of whole interactions. The download of segments of sound and transcript needs to be possible in order to submit them to subsequent analysis in other programs such as PRAAT. In this way, the corpus should allow for a systematic overview of the corpus, multi-criteria searches, the exhaustive retrieval of candidate tokens for building a collection, the possibility of refining transcripts according to the researchers' needs and for analyzing the original sound data.

## 3. The adequate level of detail of transcription in publicly available corpora

I will now return to the issue of transcription. Wichmann (2008: 205) makes a plea for richly annotated corpora: "One hopes that future spoken corpora will provide linguistically sophisticated syntactic, pragmatic and discourse annotation together with an equally sophisticated prosodic annotation that can then be complemented by automatic analysis of global trends, such as pitch, pause, loudness and voice quality." Prosodic annotation allows for the possibility to search prosodic features directly in the

---

1. A model in this respect is the French data-base *Corpus de langues parlées en interaction* (CLAPI): <http://clapi.univ-lyon2.fr/>.

transcripts.<sup>2</sup> This seems to be a major requirement for a systematic prosodic analysis of large data-bases. However, a closer look at the problems associated with the search for a “generic level” of prosodic transcription in a public corpus suggests that prosodic transcription causes too many problems to be advisable:

- There is no common agreement on the degree of granularity of prosodic transcription advisable: Which phenomena are to be included? In how much detail?
- There are competing standards, conventions, and philosophies of transcription (see Selting this volume). IPA, GAT, TOBI, and CA (just to name the most commonly used systems) focus on different phenomena, presuppose different categories for coding and “lump together” what are considered to be the “same phenomena”, make different theoretical assumptions regarding relevant distinctions and prosodic systems in speech and interaction. In sum, not one of these systems can be taken as an undisputed base representation complying with all needs of Interactional Linguistics, being compatible with all possible theoretical commitments. Each transcript inevitably implies some kind of categorization and thus interpretation of observable phenomena, which may be disputed or refused on theoretical grounds.
- The reliability and validity of prosodic transcripts is quite poor in practice (see, e.g., the London-Lund corpus; cf. also O’Connell and Kowal 1999). In part, deficits arise from problems inherent in the transcription systems, such as lacking, imprecise or polysemous definitions of transcription conventions, multiplication of signs for the same phenomena, etc. Still more important are problems associated with the transcribers, i.e., differences in individual styles of transcription, lack of training, lack of knowledge in phonetics and prosody, and lack of time and of runs of correction and validation. These problems are aggravated if large corpora have to be transcribed by students in a relatively short time, with little training and with no specific research goal in mind.

In sum, problems of searchability (most importantly, resulting in false negatives) and dependency on non-shared theoretical decisions taken by the corpus-designers and transcribers are inevitable when relying on prosodic transcriptions. Most serious researchers on prosody probably would not rely on pre-transcribed prosodic transcripts, but they would adapt the transcription according to their own research interests and theoretical convictions.

Because of the various problems of prosodic transcripts, orthographic transcripts offer the best basis for searches. At first sight, this may seem paradoxical because orthographic transcripts do not contain any prosodic annotation at all. But it is precisely this lack which makes them suitable for many kinds of comprehensive searches which

---

2. I will not discuss Wichmann’s suggestion that corpora should contain pragmatic and discourse annotation, which seems to me at least as problematic as prosodic annotation from a conversation-analytic point of view.

do not discard relevant phenomena because of transcription decisions. Automatic searches in orthographic transcripts inevitably result in false positives, which have to be discarded by the researcher by hand, but they run the risk of false negatives (exclusion of relevant phenomena) to a much lesser extent than all other kinds of transcripts (prosodic, eye dialect, etc.). Still, many interesting phenomena cannot be searched using orthographic transcripts. If we are, e.g., interested in identifying interactional, prosodic and phonetic properties of different kinds of intonation phrases, we cannot resort to orthographic transcripts for creating collections via automatic searches. Using prosodically annotated transcripts, however, would presuppose that we already know most of the details relevant to the boundaries of intonation phrases, whose detection is precisely the goal of the study. So, while the possibility to dispose of large publicly available corpora will serve many research interests, the benefit of automatic procedures of retrieval and analysis will depend crucially on research goals.

Today, public corpora suited for Interactional Linguistics are still rare (for overviews see Merkel and Schmidt 2009, Wichmann 2008, McCarthy and O'Keefe 2008).<sup>3</sup> Moreover, not one of the few corpora that are available does conform to all of the criteria outlined in this paper. Thus, the construction of public corpora of interactional data and the advancement of their standards is a major task. If it is accomplished, it can also foster future research in Interactional Linguistics.

## References

- Baude, Olivier et al. 2006. *Corpus oraux*. Paris: CNRS.
- Bird, Steven and Simons, Gary 2002. "Seven dimensions of portability for language documentation and description." *Language* 79: 557–582.
- Gumperz, John J. 1982. *Discourse strategies*. London: Cambridge University Press.
- McCarthy, Michael and O'Keefe, Anne 2008. "Corpora and spoken language." In: *Corpus Linguistics. Vol.2*, Anke Lüdeling and Merja Kytö (eds.), 1008–1024. Berlin: de Gruyter.
- Merkel, Silke and Schmidt, Thomas 2009. "Korpora gesprochener Sprache im Netz – eine Umschau." *Gesprächsforschung* 10: 70–93 (<http://www.gespraechsforschung-ozs.de/heft2009/px-merkel.pdf>).
- Mondada, Lorenza 2007. "Commentary: transcript variations and the indexicality of transcribing practices." *Discourse Studies* 9 (6): 809–821.
- O'Connell, Daniel C. and Kowal, Susanne 1999. "Transcription and the issue of standardization." *Journal of Psycholinguistic Research* 28: 103–120.

---

3. From 2011 onwards, a large public corpus (Forschungs- und Lehrkorpus Gesprochenes Deutsch: FOLK) of conversations in German with audio files and transcripts aligned will be available via internet: <http://agd.ids-mannheim.de/html/folk.shtml>.

- Schegloff, Emanuel A. 1993. Reflections on quantification in the study of conversation. In: *Research on Language & Social Interaction* 26 (1): 99–128.
- Selting, Margret 2001. “Fragments of units as deviant cases of unit-production in conversational talk.” In: *Studies in Interactional Linguistics*, Margret Selting and Elizabeth Couper-Kuhlen (eds.), 229–258. Amsterdam: Benjamins.
- Wichmann, Anne (2008) “Speech corpora and spoken corpora.” In: *Corpus Linguistics. Vol. I*, Anke Lüdeling and Merja Kytö (eds.), 187–207. Berlin: de Gruyter.