The case for *Språkbanken Dialog*

Staffan Larsson   **Christine Howes**   **Eleni Gregoromichelaki**
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Sweden
staffan.larsson@ling.gu.se
christine.howes@gu.se
eleni.gregoromichelaki@gu.se

Abstract

We argue that the National Language Bank of Sweden should be extended with an additional infrastructure supporting research on linguistic interaction. Our main argument is that dialogue is not (just) text or speech, and consequently, that studying dialogue requires dialogue-specific infrastructure.

1   Introduction

One of Lars Borin's main achievements is, of course, his very successful development and management of Språkbanken, now a part of the National Language Bank of Sweden under the name Språkbanken Text. Following his lead and inspiring pioneering example, we would like to take this opportunity to argue the case for a Swedish Dialogue Bank - *Språkbanken Dialog*.

2   Språkbanken

Quoting from the homepage of the National Language Bank of Sweden, "The purpose of The National Language Bank of Sweden is to develop a national e-infrastructure supporting research in language technology, linguistics and other fields of study where research is conducted based on language data." Currently, the National Language Bank of Sweden has three parts – Språkbanken Text, Språkbanken Tal and Språkbanken Sam. These all provide important infrastructures for researchers in various fields. Språkbanken Text contains text corpora that can be searched for occurrences of words and phrases, including longitudinal data. Språkbanken Tal contains (or will contain when it is launched in 2023) recorded speech aligned with text for use in research and development of speech technologies. Språkbanken Sam contains text and some speech recordings focusing on (1) official multilingual texts and terminology for research in official communication and social conditions, and (2) folk narratives, as well as other text and speech material from the dialect and folklore archives.

The National Language Bank of Sweden is a significant achievement and a valuable resource for language technology purposes. However, a considerable lacuna, in our view, remains: these resources do not provide a comprehensive collection of spoken, written, and/or multimodal interactions in Swedish (and/or minority languages) that are available and searchable in the way that is needed to explore the *interactive* aspects of language use and structure. This is what we argue is still needed.

3   Dialogue

It is now widely accepted that human conversation does not consist of a sequence of sentences simply placed one after the other. There are specific phenomena that only become visible at the level of dialogical interaction, for example, so-called "grounding processes" (Clark, 1996), turn-taking (Sacks et al., 1974), repair (Schegloff et al., 1977), and multimodal input and output (Bavelas & Gerwing, 2007), which are

the features of dialogue that make it so much easier to process and engage in than monologue. On the other hand, phenomena which have been considered sentence-internal and requiring specialised syntactic/semantic mechanisms can be seen under a new, more illuminating, light when considered in the context of conversation. For example, phenomena like anaphora, ellipsis, syntactic/semantic dependencies, and speech act recognition/production can extend across turns and participants (see (3), below). In fact, it can be shown that such puzzling phenomena rely more crucially on interactive mechanisms for their resolution than individual processing capacities, a case of 'computational offloading' to the social environment (Gregoromichelaki, 2017). Across linguistics, psychology, philosophy, and cognitive science, it is now recognised that the primary ecological niche of language use is face-to-face interaction. Therefore, it has now become common to talk about the human 'interaction engine' (Levinson, 2020) to refer to the evolutionarily and culturally shaped linguistic skills and social capacities that are involved in language processing and general action coordination. Formal grammars, computational implementations, and linguistic/psycholinguistic theories now attempt to model formally and test experimentally these interactive processes to explain human linguistic cognition and behaviour (Ginzburg, 2012; Gregoromichelaki et al., 2020; Healey et al., 2018; Cooper, 2022).

In the field of language technology and AI, it is also becoming a familiar theme to address human interaction and conversation as the source of invaluable data. Many current architectures take advantage of training data from dialogue and multimodal corpora, whether annotated or not, and there is a recognition in recent work that large-scale language models – even those which make use of visual data – lack sufficient training data of conversational strategies such as repair (Lemon, 2022). Additionally, models increasingly seek to leverage interactive processes with human-in-the-loop teaching and supervision as a means of extending the capabilities of Large Language Models and artificial agents like social robots developing their trustworthiness, reliability, and alignment with human values.

As an illustration, let us look at an example of a dialogue with the sort of annotations we envision for Språkbanken Dialog:

(1) STANLEY: Louis, I$_{[ref:\text{STANLEY}]}$ just didn't$_{[\text{NPI}-licensor]}$ think

     [[assertion; change of turn: *split utterance*]]

   LOUIS: you$_{[ref:\text{STANLEY}]}$'d ever$_{[\text{NPI}]}$ hear from me$_{[ref:\text{LOUIS}]}$?

     [[continuation & clarification & confirmation request & quotation]]

               [BBC Transcripts, *Dancing to the Edge*, Episode 5, example from: Gregoromichelaki (2017)]

Here the annotation needs to indicate the dialogue-act multifunctionality of subsentential turns. We also need to have information about the dependency between the Negative Polarity Item (NPI) *ever* and its licensor *n't* that occur in different turns by different speakers even though no single surface string can be syntactically reconstructed. In confirmation of this, it needs to be indicated in the annotation how the incremental change of speaker within a quotative clause reporting the first speaker's mental state ('Stanley$_{speaker}$ did not think || that Stanley$_{addressee}$ will hear from Louis$_{speaker}$') results in incremental switches in the interpretation of indexicals. This evidence of dependencies crossing turns and speakers render untenable any simple analysis of the shared string as a joined surface syntactic form with respect to the semantics:

(2) #Louis I just didn't think you'd ever hear from me.

In addition, it is demonstrated that grammatical analyses need to incorporate semantic and, crucially, pragmatic factors, e.g., turn-taking in dialogue, in order to provide a coherent and unified analysis of syntactic/semantic phenomena. Moreover, understanding both human psychological processes and the functioning of end-to-end models and AI architectures with respect to linguistic behaviour requires becoming aware and modelling such interactions of what have been standardly taken as separate modules of linguistic/non-linguistic knowledge in standard monological accounts.

4 Språkbanken Dialog

With this in mind, let us try to explain in more detail why Språkbanken Dialog is needed, and how we envision it.

Språkbanken Dialog is (would be) a large collection of linguistic interactions, including video recordings of face-to-face interactions, audio recordings of spoken interactions, transcribed interactions (aligned with the source video or sound recordings), and written interactions taken e.g. from social media and chat applications. It is possible to view, annotate and analyse individual interactions across multiple turns – something not currently offered by any Språkbanken resources. It is also possible to relate individual interactions to each other, e.g. temporally, spatially, or with respect to the speakers involved (while keeping to GDPR restrictions).

What about overlap with existing Språkbanken resources? It is true that other Språkbanken resources already contain linguistic interactions. In fact, as far as possible, such material should also be included in Språkbanken Dialog. However, none of the existing resources offer the possibility of adequately exploring the interactive aspects of these dialogues. In Språkbanken Text, interactions are treated as any other text, and it is not possible to see full interactions across several turns, nor to annotate or analyse them. The argument for Språkbanken Dialog rests on the fact that linguistic interaction is not reducible to, or analysable in terms of, individual words or phrases.

So maybe Språkbanken Dialog could just be a different interface to existing Språkbanken resources? Such a thing would certainly be useful, but there are also reasons to include additional resources not covered by other Språkbanken infrastructure. Currently, linguistic interactions are collected by researchers and students working on dialogue in the course of their research activities. This data can be in the form of text, audio, video, or some combination thereof. Currently, a lot of these resources never become available to other researchers. We believe that Språkbanken Dialog could offer infrastructure that would enable and encourage low-effort sharing, annotation and analysis of dialogue data (including multimodal data), thus boosting research on linguistic interaction in Swedish and other languages.

5 Future work

We leave for future work to fund, organise and implement Språkbanken Dialog. In this, we hope to follow Lars Borin's inspiring example.

Acknowledgements

References

Janet B. Bavelas & Jennifer Gerwing. 2007. Conversational hand gestures and facial displays in face-to-face dialogue. *Frontiers of social psychology: Social communication*, pages 283–307.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.

Robin Cooper. 2022. *From perception to communication: An analysis of meaning and action using a theory of types with records (TTR)*. Oxford University Press, Oxford. to appear.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.

Eleni Gregoromichelaki, Gregory J. Mills, Christine Howes, Arash Eshghi, Stergios Chatzikyriakidis, Matthew Purver, Ruth Kempson, Ronnie Cann, & Patrick G. T. Healey. 2020. Completability vs (in)completeness. *Acta Linguistica Hafniensia*.

Eleni Gregoromichelaki. 2017. Quotation in Dialogue. In *The Semantics and Pragmatics of Quotation*, pages 195–255. Springer.

Patrick G. T. Healey, Gregory J. Mills, Arash Eshghi, & Christine Howes. 2018. Running Repairs: Co-ordinating Meaning in Dialogue. *Topics in Cognitive Science*, 10(2):367–388.

Oliver Lemon. 2022. Conversational grounding in emergent communication–data and divergence. In *Emergent Communication Workshop at ICLR 2022*.

Stephen C Levinson. 2020. On the human "interaction engine". In N. J. Enfield & S.C. Levinson, editors, *Roots of human sociality: Culture, cognition and interaction*, pages 39–69. Routledge.

Harvey Sacks, E.A. Schegloff, & Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735.

E.A. Schegloff, G. Jefferson, & H. Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.