Annotating Modality Interdependencies

Eva Reimer[1], Bianka Trevisan[1], Denise Eraßme[1], Thomas Schmidt[2], Eva-Maria Jakobs[1]
[1]RWTH Aachen University
`{e.reimer, b.trevisan, e.m.jakobs}@tk.rwth-aachen.de`
[2]Institut für Deutsche Sprache (IDS) Mannheim
`thomas.schmidt@ids-mannheim.de`

Abstract

This paper discusses computational linguistic methods for the semi-automatic analysis of modality interdependencies (the combination of complex resources such as speaking, writing, and visualizing; MID) in professional cross-situational interaction settings. The overall purpose of the approach is to develop models, methods, and a framework for the description and analysis of MID forms and functions. The paper describes work in progress—the development of an annotation framework that allows annotating different data and file formats at various levels, to relate annotation levels and entries independently of the given file format, and to visualize patterns.

1 Objective

Professional verbal interaction settings are often characterized by *modality interdependencies* (*MID*): the combination of complex resources such as speaking, writing, and visualizing. In the research project ModiKo (funded by the DFG, JA1172/3-1), we investigate MID in professional interaction settings. The aim is to develop methods and a framework for the systematic description and analysis of MID forms and functions. Both the quantity of data and the complexity of phenomena to be examined require novel methods. Ultimately, we envisage an annotation framework that allows to annotate different data (text, video, sketches) and file formats (.txt, .mpeg, .jpeg) at various levels, to relate annotation levels and entries independently of the given file format, and to visualize patterns of relations.

2 Case study

The project ModiKo uses data generated in a former research project (IMIP, 2008-2011, BMBF) investigating professional interactions as part of industrial process modelling. In the case study, experts interview employees in a company aiming to understand how production processes are organized. Figure 1 shows a typical situation and example of our data. A process modeler (on the right) interviews an employee (on the left) to get details about the clearance of goods.

Figure 1. Case Example

During the interview, the actors interrupt each other, speak simultaneously, make notes on a clipboard to fix information, or use the clipboard to visualize parts of the production process. By doing so, they combine resources (speaking, writing, visualizing).

3 Corpus

The original database consists of video files (548 minutes, .mpeg), transcripts of verbal interactions (266 pages, .docx) and sketches (89 sheets, .jpeg). The transcription of the verbal parts follows GAT 2 (Selting et al 2009).

The project's research interests—to investigate MID forms and functions—require a more sophisticated solution because of the broad range of phenomena to be considered. For a precise and flexible handling of data, the ModiKo corpus distinguishes three types of documents: (1) *primary documents* (video files of the observed professional interactions and files of the sketches produced by the actors involved), (2) *secondary documents* (multimodal transcripts of the primary documents), and (3) *tertiary documents* (multi-level annotations of the secondary documents). The tertiary document allows the researcher to annotate different phenomena such as speech-

accompanying gestures (e.g. to point at sth), objects used in the observed situations (e.g. a clipboard), or contextual and linguistic information. The transformation of primary into secondary documents is an interpretative work.

4 Approach

In the following, a first approach for the multi-level annotation (tertiary documents) is outlined. The approach builds up on existing tools for multimodal annotation (e.g. ELAN, ANVIL) (Sloetjes et al 2011; Kipp 2014). However, these tools only allow restricted linking of annotations, particularly, the linking of text and video data annotation with sketches (Kipp 2014).

For the implementation of MID annotation, ModiKo aims at uncovering challenges and requirements for the development of a holistic annotation tool. The approach is driven by both the ModiKo research objective and the data richness of the case example. The analysis of MID forms and functions and the identification of related patterns require an efficient annotation method allowing to annotate different data and file formats at various levels, to relate annotation levels and entries independently of the given file format, and to visualize patterns.

The approach is innovative because of the complex annotation tool combining different tasks or objectives: the annotation of speech-related phenomena, a linguistic multi-level description of the primary document (Trevisan 2014), and the annotation of MID forms and functions. The annotation of speech-related phenomena and the linguistic multi-level annotation are a pre-condition for the detection of patterns indicating MID forms and functions.

The development of the annotation system is a highly challenging task for different reasons: it requires pioneering in the development of annotation levels, categories, and rules for the description and detection of MID forms and functions. Part of the approach is a quantitative analysis: How often does a MID form or function occur in the interaction scheme? What are typical indicators for a certain form or function?

Despite the quantity and heterogeneity of the data, the annotation system must fit criteria such as efficiency, flexibility, and coherence, i.e., minimized effort needed to achieve the intended annotation, possibility to add, correct, modify categories and annotation levels, as well as consistency and reliability of annotations.

In ModiKo first steps have been completed: indicators for MID forms were identified manually. A prominent example is the verb *schreiben* (*to write*) in the phrase *ich schreib hier mal rein.* The term indicates the MID form speaking/writing (Ullrich et al forthcoming). The existing corpus was transformed into a form suitable for manual and (semi-)automatic analysis and annotation. To this end, the selected text files (transcripts) were transferred to EXMARaLDA (Schmidt and Wörner 2014) and bundled with related video files and sketches. Videos and transcripts were aligned, the diagrams' content modeled in XML, and metadata was systematically recorded for all data types. The first half of the corpus in this form is now ready for tokenization and PoS-tagging. Finally, for each file linguistic annotation levels were defined and created. At the moment, the linguistic annotation follows Trevisan (2014). The annotation levels for MID forms are work in progress.

5 Outlook

Future work will focus on the development of suitable visualization methods for the interplay of data from different modalities, an optimization of the semi-automatic annotation, and a more systematic approach for the annotation of videos. Concluding, the transformation in tertiary documents is a first step to overcome the challenges for a formal and systematic description of MID.

References

Michael Kipp. 2014. ANVIL: The Video Annotation Research Tool. *Handbook on Corpus Phonology.* Oxford University Press: 420-436.

Thomas Schmidt and Kai Wörner. 2014. EXMARaLDA. *Handbook on Corpus Phonology.* Oxford University Press: 402-419.

Margret Selting et al. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung-Online-Zeitschrift zur verbalen Interaktion* (10).

Han Sloetjes, Aarthy Somasundaram, and Peter Wittenburg. 2011. ELAN–Aspects of Interoperability and Functionality. *ISCA*: 3249-3252.

Bianka Trevisan. 2014. Bewerten in Blogkommentaren. Mehrebenenannotation sprachlichen Bewertens. Dissertation. RWTH Aachen University.

Anna Ullrich, Eva-Maria Jakobs, and Denise Eraßme. Forthcoming. „ich schreib hier mal rein ähm". Modality-taking–Schreibhinweise in professionellen mündlichen Interaktionssituationen.