A Tool for Feature-Structure Stand-Off-Annotation on Transcriptions of Spoken Discourse

Kai Wörner
SFB 538 Mehrsprachigkeit
Max-Brauer-Allee 60
22765 Hamburg
E-mail: kai.woerner@uni-hamburg.de

Abstract

This paper presents an annotation tool and format for the stand-off annotation of transcriptions of spoken discourse like they are produced in a conversion analysis or pragmatic framework. It was developed at the Collaborative Research Centre on Multilingualism in Hamburg, where many suchlike corpora from different research projects exist. It transfers findings from the field of the so-called "annotation science" into a practical application for researchers in these areas.

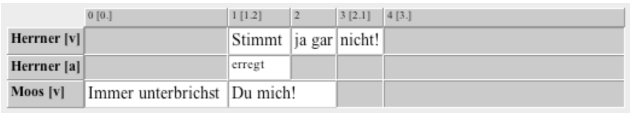
1. Annotation Science

Annotation Science (Ide 2007), a discipline dedicated to developing and maturing methodology for the annotation of language resources, is playing a prominent role in the fields of computational and corpus linguistics. While progress in the search for the right annotation model and format is undeniable, these results only sparsely become manifest in actual solutions (i.e. software tools) that could be used by researchers wishing to annotate their resources right away, even less so for resources of spoken language transcriptions.

The paper presents a solution consisting of a data model and an annotation tool that tries to fill this gap between „annotation science" and the practice of transcribing spoken language in the area of discourse analysis and pragmatics, where the lack of ready-to-use annotation solutions is especially remarkable.

2. Transcription

Transcriptions of discourse vary from other language resources in many ways: Compared to written texts, transcriptions of discourse strongly depend on the temporal organization of the speaker's contributions. To analyze phenomena like turn-taking, it is crucial to exactly record what happens when. In discourse, there is not one primary data like in most written texts: every speaker's utterances make up for one stream of primary data, and these streams of primary data can and frequently do overlap.



	0 [0.]	1 [1.2]	2	3 [2.1]	4 [3.]
Herrner [v]		Stimmt	ja gar	nicht!	
Herrner [a]		erregt			
Moos [v]	Immer unterbrichst	Du mich!			

Figure 1: Overlapping speaker contributions in a transcription and inline annotation.

Furthermore, transcriptions of discourse are not always based on standard orthography or on standard phonetic transcription conventions that use phonetic alphabets like IPA or SAMPA: some transcription conventions aim to represent specifics in articulation by departing from the rules of standard orthography. While this sometimes makes it easier for the reader to oversee these specifics of the transcriptions, these transcriptions get far less useful as a source for automatic taggers, since these rely on a standard orthography in the first place.

These characteristics of transcriptions in discourse analysis, pragmatics and related domains also sets them apart from transcriptions that are used in computational linguistics for the purposes of training speech analysis or speech synthesis systems. To be useful for their purposes, these transcriptions are always produced using standard orthography and the temporal relations between the utterances of different speakers, as well as potential overlap, are ignored (the temporal structure inside one speaker's utterances, on the other hand, is important, too).

The transcription process, too, varies between the different disciplines: While transcriptions for speech technology are only produced and corrected once to be used for training afterwards, transcription in the focused research area is mostly an iterative process that is potentially endless, especially when the underlying method follows a corpus driven (Tognini-Bonelli 2001) approach.

3. Modeling Transcriptions

Departing from these and other observations, a data model for the transcription of spoken discourse has to:

- supply a timeline to represent the temporal properties of the speaker's utterances
- allow for overlap of speaker's utterances and overlap of annotations
- be prepared for different options for the segmentation of the transcribed material
- support most existing writing systems

Existing data models are distributed on an axis between two poles. The one pole consists of a model where language is seen as a hierarchical structure in which the downmost layer represents the actual textual content (OHCO, Ordered Hierarchy of Content Objects) (DeRose et al. 1990). These models always have to introduce special "workarounds" to deal with overlapping hierarchies or speaker contributions. The other pole is represented by a model that is based on directed, acyclic graphs and is centered on the aspect of temporal relations of the respective units.

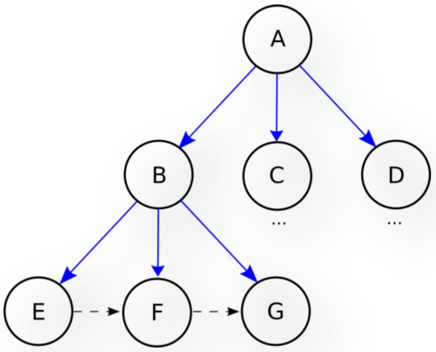


Figure 2: Hierarchical Model.

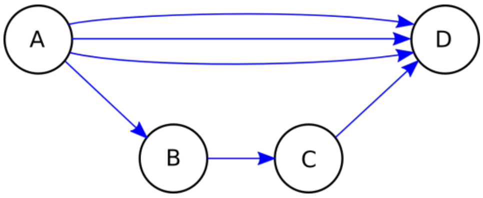


Figure 2: Directed Acyclic Graph.

One model for the transcription of spoken language is provided by the TEI-Guidelines (Burnard & Bauman 2007) that have departed from a model for editions of text and have special mechanisms to deal with transcriptions of spoken language, belonging more in the direction of the hierarchical pole. The NITE Object Model (Evert et al. 2003) and the Linguistic Annotation Framework (LAF) (ISO WD24611 2008), combine elements from hierarchical and graph based models, while the Annotation Graph Formalism (Bird & Liberman 2001), is located more towards the graph-based pole. Instantiations exist for some, but not all of these models.

Models based on the Annotation Graph Formalism are best suited to represent transcriptions of spoken language since they do not need any workarounds to cope with speaker overlaps and do not necessarily rely on any kind of basic segmentation of the text – something that, for example, limits the usefulness of the LAF for the purpose of modeling spoken language resources.

4. EXMARaLDA

Transcription graphs, on which the EXMARaLDA transcription format and toolset is built, specify and substantiate the annotation graph formalism. The EXMARaLDA data format (Schmidt 2004), in particular, introduces the possibility to segment the textual content of the transcriptions according to transcription conventions. Departing from a so-called „basic transcription", that in most cases only contains segments motivated by changes in the constellation of the discourse (speaker changes, interruptions) that are directly linked to the explicit timeline, the EXMARaLDA system allows for the automatic segmentation into linguistically motivated segments (like words, utterances, sentences, turns) that are defined through rules laid down in the transcription conventions and do not necessarily link to distinct points on the timeline. While it is possible to add annotations to the former type of segments by simply adding additional annotation-tiers to the transcription in the transcription tool (cf. http://www.exmaralda.org/tools), there is no mechanism, neither in software nor in the data format, to add annotations to the latter.

5. Annotation

Based on these premises, an annotation format and software tool were developed to facilitate annotation of these segments. Since tools and large and valuable corpora for the EXMARaLDA format already exist, one indispensable premise was that no changes on the tools and the existing format should be necessary.

This precondition determined the usage of stand-off-annotation: stand-off-annotation stores annotations and the annotated content in different locations (i.e. files), and connect them through pointers. Since all segments in EXMARaLDA transcriptions are easily identifiable through unique IDs, pointing to annotatable segments was easy to accomplish.

For the actual pointers from annotations to the annotated content, XLink and XPointer are utilized. Since these are both established XML-technologies, this part of the solution does not leave standardized XML terrain, is well documented and can be utilized by means of existing XML tools.

Annotations themselves are modeled as feature structures, following the TEI's recommendations for features structures that are also an ISO standard (ISO 2006).

Feature structures model information as attribute/value pairs, where the value can either be atomic or another attribute/value pair. That way, they allow for the easy modeling of simple attribute/value combinations, but also allow for much more complex annotation structures like trees. Furthermore, feature structures offer an established method of creating libraries of frequently used features that can be utilized by pointing at them.

The model is thus capable of annotating all segments that can exist in all types of EXMARaLDA transcriptions. To utilize the model, a prototypical program, Sextant, was developed that enables the user to annotate segmented EXMARaLDA transcriptions manually through a comfortable user interface.

By using all available linking features of the XPointer framework, it would also be possible to annotate any spans of characters inside of existing segments. By extending the framework to include pointers to absolute time points on the timeline, it would be possible to catch phenomena that don't have an expression in the character data of the transcription. These two extensions to the existing solution would then cover the annotation of all possible entities in an EXMARaLDA transcription without needing to change anything in the existing data model and in actual corpora.

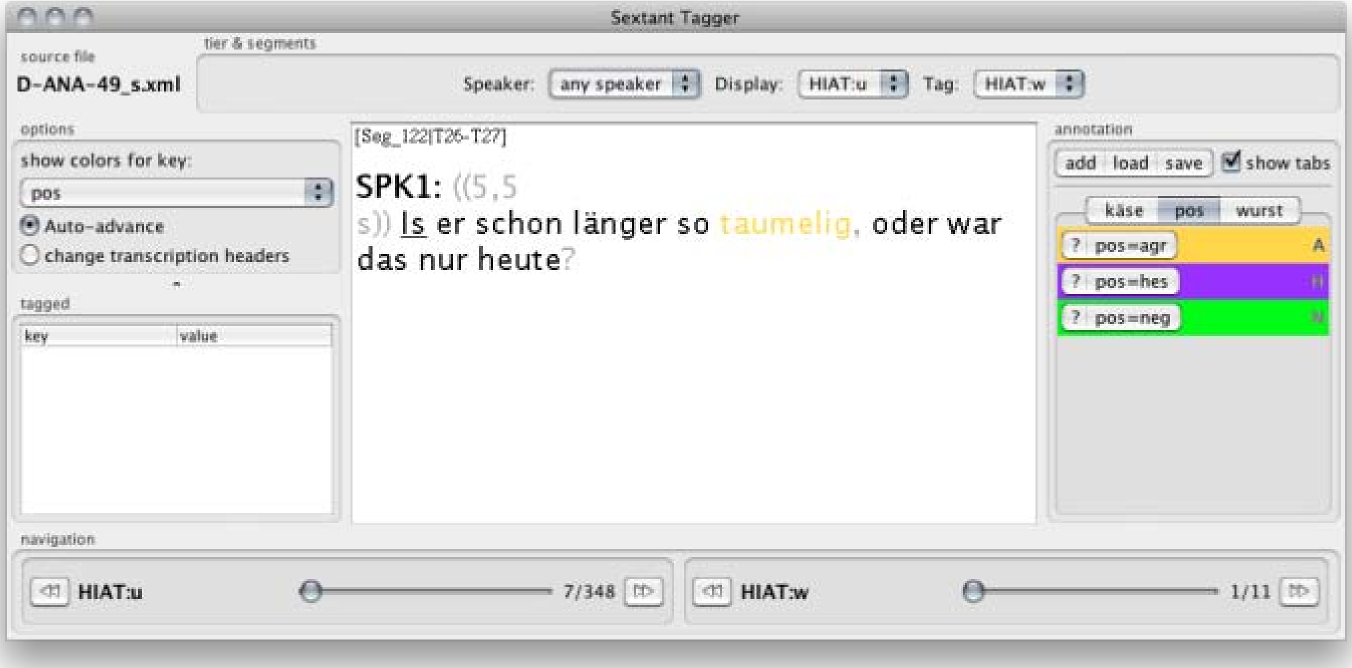
6. Open Issues

There are still open issues, though, especially on the side of tools that utilize the newly created possibilities. The most important step would be to extend EXMARaLDA's corpus analysis tool EXAKT to include standoff-annotations in its search routines.

To facilitate that, it would be necessary to have some linking mechanism between transcriptions and their annotations. Standoff annotation files already contain metadata about the transcription they are annotating, but there would have to be some linking from transcriptions to their annotations for a search tool to know about its existence. To keep the original transcriptions untouched, this information would have to be stored with EXMARaLDA's corpus management tool.

One further step would be the development of more sophisticated visualization methods for transcriptions and their annotations.

Using standoff annotation also generates certain problems that still have to be coped with: since standoff annotations rely the on ids of the transcriptions not to change, a mechanism will have to be introduced that arranges for this steadiness when transcriptions change.



source file D-ANA-49_s.xml

Speaker: any speaker Display: HIAT:u Tag: HIAT:w

show colors for key: pos

● Auto-advance
○ change transcription headers

☑ show tabs

[Seg_122[T26-T27]

SPK1: ((5,5 s)) Is er schon länger so taumelig, oder war das nur heute?

? pos=agr A
? pos=hes H
? pos=neg N

HIAT:u 7/348

HIAT:w 1/11

Figure 2: Annotation of word-segments
in the Sextant tool.

7. Conclusion

The presented approach bridges the gap between annotation science and the exercise of transcription in the fields of pragmatics and conversation analysis.

The chosen model combines feature structures in standoff-annotation and a data model based on annotation graphs, combining their advantages: The transcription model is ideally fitted for the transcription of spoken language by centering on the temporal relations of the speaker's utterances and their reference to a timeline and is implemented in time-tested and reliable tools that support an iterative workflow, while standoff annotation allows for more complex annotations than would be possible in the annotation graph formalism alone, while relying on an established and well documented model.

8. References

Bird, Steven/Liberman, Mark (2001). A formal framework for linguistic annotation. In: *Speech Communication 33*, pp. 23-60.

Burnard, Lou & Bauman, Syd (2007): TEI P5: Guidelines for Electronic Text Encoding and Interchange. [http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html]

DeRose, S. J., Durand, D. G., Mylonas, E., & Renear, A. H. (1990). What is text, really. *Journal of Computing in Higher Education, 3,* 3-26.

Evert, Stefan, Carletta, Jean, O'Donnell, Timothy J, Kilgour, Jonathan, Vögele, Andreas & Voormann, Holger (2003). The NITE Object Model. *Proceedings of the EACL Workshop on Language Technology and the Semantic Web*.

Ide, Nancy (2007). Annotation Science: From Theory to Practice and Use. In: Rehm, Georg/Witt, Andreas/Lemnitzer, Lothar (Eds.): *Data Structures for Linguistic Resources and Applications – Proceedings of the Biennial GLDV Conference 2007*. Gunter Narr Verlag, Tübingen.

ISO/TC 37/SC 4 (2008): Language resource management - Linguistic Annotation Framework (WD24611). ISO.

ISO/TC37/SC 4 (2006): Language resource management - Feature structures - Part 1: Feature structure representation. ISO.

Johansson, Stig (1995): The approach of the Text Encoding Initiative to the encoding of spoken discourse. *Spoken English on Computer: Transcription, Markup and Application*. Longman, Harlow.

Schmidt, T. (2005). Time-based data models and the Text Encoding Initiative's guidelines for transcription of speech. *Arbeiten zur Mehrsprachigkeit, Folge B, 62*.

Tognini-Bonelli, Elena (2001): Corpus linguistics at work. Benjamins, Amsterdam.