



FOLKER



An Annotation Tool For Efficient Transcription Of Natural, Multi-Party Interaction

Thomas Schmidt (SFB Multilingualism, Hamburg) / Wilfried Schüttele (IDS, Mannheim)

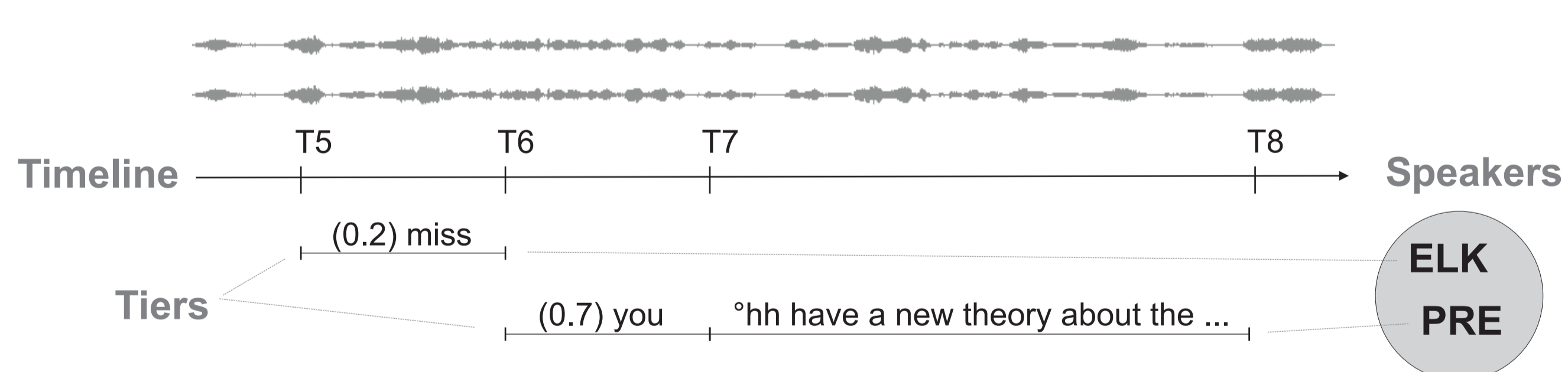
FOLKER - the FOLK Editor - is an annotation tool for the efficient transcription of natural, multi-party interaction in a conversation analysis framework. FOLKER is being developed at the Institute for German Language in and for the FOLK project, whose aim is the construction of a large corpus of spoken present-day German, to be used for research and teaching purposes. FOLKER builds on the experience gained with multi-purpose annotation tools like ELAN and EXMARaLDA, but attempts to improve transcription efficiency by restricting and optimizing both data model and tool functionality to a single, well-defined purpose.

GAT Transcription Conventions

0001 **PRE** good evening
 0002 **AUD** ((laughter))
 0003 **PRE** i have with me (.) tonight (0.3) ann elk
 0004 mistress ann elk
 0005 **ELK** (0.2) miss
 0006 **PRE** (0.7) you
 0007 °hh have a new theory the brontosaurus
 0008 **ELK** well eh can i just eh say here chris for one moment
 0009 that i have a new theory about the brontosaurus
 0010 **AUD** ((laughter))
 0011 **PRE** exactly
 0012 **AUD** ((laughter))
 0013 **PRE** what is it

FOLKER is optimised for transcription according to the GAT transcription conventions - a de-facto standard in Germany for the transcription of natural interaction in conversation analytic research. GAT defines three transcription levels which correspond to different degrees of prosodic detail. FOLK and FOLKER make use of the first of these levels – the minimal transcript – which provides rules for the transcriptions of words in a modified orthography, for the description of pauses, breathing and non-phonological phenomena (coughing, laughing etc.) and for the handling of uncertain or incomprehensible passages. The figure above gives an example of a GAT minimal transcription as a plain text file.

Data Model and Format



FOLKER's basic data model is derived from the **single-timeline-multiple-tiers (STMT)** model, i.e. it conceptualizes a transcription as a set of annotations which are assigned via a start and an end point to a single, fully ordered timeline, and which are partitioned into a number of tiers such that no two annotations within a tier overlap. While this data model represents the temporal structure of events, including possible overlaps between different speakers, further structure is added by combining adjacent annotations into contributions and re-segmenting them into the entities defined in the transcription convention (i.e. words, pauses etc.) with the help of a finite state transducer. The resulting data structure can be serialized into a TEI-like XML format. Since the temporal information in the data model is structurally compatible with the data models of tools like **ELAN** or **EXMARaLDA**, FOLKER can provide export filters for these tools.

```
<folker-transcription>
<speakers>
<speaker id="PRE"/>
<speaker id="ELK"/>
<speaker id="AUD"/>
</speakers>
<recording path="/MyTheory.wav"/>
<timeline>
<!-- [...] -->
<timepoint id="T5" time="18.79"/>
<timepoint id="T6" time="20.18"/>
<timepoint id="T7" time="23.08"/>
<timepoint id="T8" time="25.99"/>
<!-- [...] -->
</timeline>
<!-- [...] -->
<contribution speaker="ELK" start="T5" end="T6">
<pause duration="0.2"/>
<w>miss</w>
</contribution>
<contribution speaker="PRE" start="T6" end="T8">
<pause duration="0.7"/>
<w>you</w>
<time reference="T7"/>
<breathe type="in" length="2"/>
<w>have</w>
<w>a</w>
<w>new</w>
<w>theory</w>
<w>about</w>
<w>the</w>
<w>brontosaurus</w>
</contribution>
</folker-transcription>
```

Different Editable Views

FOLKER has three editable views of the data each of which is optimized for a specific step in the transcription workflow. Users can switch between them at any time in the transcription process. The **Segment View** (top) is most efficient for initial transcription. It displays individual annotations in a vertical list. Speaker assignment, annotation text and temporal assignment can be freely modified in this view and individually for each annotation. The **Partitur (musical score) View** (middle) displays the same transcription in a horizontal layout. This view is best suited for editing temporal relations, most prominently speaker overlap. The **Contribution View** (bottom) also uses a vertically organised layout, but, instead of individual annotations, displays adjacent annotations of speakers as contributions. This view is best suited for final proof-reading and corrections of a transcript.

Navigation in all three views is synchronized with navigation in (a waveform visualization of) the audio recording.

Syntax and time control

	Start	End	Speaker	Transcription text	Syntax	Time
7	00:20.18	00:23.08	PRE	°hh have a new theory about the brontosaurus	✓	✓
8	00:23.08	00:25.99	ELK	well eh can i just eh say here chris for one moment	✓	✗
9	00:25.99	00:28.12	ELK	that i have a new theory about the brontosaurus	✓	✗
10	00:28.12	00:31.74	AUD	((laughter))	✗	✓
11	00:31.74	00:32.20	PRE	exactly	✓	✓

Using a regular expression, the syntax of the annotation text is checked during input for conformance with the transcription conventions. Errors (as in line 10) are indicated in the column 'Syntax'. Likewise, the temporal integrity of annotations is verified. If two annotations belonging to the same speaker overlap (as in lines 8 and 9), this is indicated in the column 'Time'.

Comparison with other tools

In contrast to many other widely used transcription tools (like **ELAN**, **EXMARaLDA**, **ANVIL**, **CLAN** or **Transcriber**), FOLKER is explicitly designed not as a multi-purpose tool, but rather as a tool which supports one specific (albeit widely used) annotation scenario in a maximally efficient way. Thus, while FOLKER shares much of its base functionality and its basic data models with these tools, it is less comprehensive in terms of the functions it offers, but also easier to learn and use. Consequently, we do not claim that FOLKER is generally superior or inferior to other transcription tools. Quite to the contrary, we believe that researchers should be encouraged to exploit the growing interoperability between different solutions and use tools with different specializations for different tasks in their corpus construction workflows. In such a workflow, we see FOLKER as a tool for the creation of base transcriptions which can later be supplemented with additional annotations through other tools. The fact that FOLKER's data model is a subset of EXMARaLDA's data model and that the tools also share a fair proportion of their code base, makes it especially easy to use FOLKER and EXMARaLDA side-by-side.