EXMARaLDA meets WebAnno

Steffen Remus, Hanna Hedeland, Anne Ferger, Kristin Bührig, Chris Biemann

Hamburg Centre for Language Corpora
Language Technology Group
Universität Hamburg
Agenda

• Introduction: project context
• Use case: analysing discourse in teacher education
• Requirements
• Implementation
• Experiences
• Outlook
Introduction: project context

• funding through ‘teaching lab’ within the project ProfaLE (Professional teachers’ actions to promote subject-based learning under changing social conditions)

• Universität Hamburg is part of the joint ‘Qualitätsoffensive Lehrerbildung‘ (teacher education quality campaign) run by the German federal government and the German states using funds from the German Federal Ministry of Education and Research (BMBF)
Teacher education: needs and claims

needs
• students demand practical relevance in education
  • “how to act as a teacher in my own future lessons?”
• students expect modern i.e. digital forms of information and communication
• students suffer from time pressure

claims
• teachers are expected to instruct pupils in linguistic reflection
• teachers have to perform a daily diagnosis of pupils’ actions and comprehension
• teachers are expected to provide state of the art teaching in their specialist disciplines
• teachers are expected to participate in advanced training programmes
A linguist’s suggestion to teacher education

- analyses of classroom interaction
  - well structured interaction
  - extensive research literature
- comprehension processes
  - different dimensions of language use
    (verbal, kinesic, prosodical: intonation, pauses)
  - interrelation between different actants
  - quite relevant for learning and teaching
Analysing discourse in education

- challenges in teaching students in analysing authentic discourse data
- authentic discourse data – vs. ’vignettes’
  - different conventions in documenting discourse data (protocols, scripts, transcriptions)
  - different conventions of notation
  - different traditions in analysing discourse
  - different approaches and methodologies
- no recipes for teaching discourse analysis within education
Requirements

• **transcripts** are crucial for qualitative analysis
  • theory-laden models of discourse

• **annotation** makes students’ analyses explicit
  • should allow for assessment (gold standard) and comparison (**intra-/inter-annotator agreement**)

• **web-based tools** are suitable for teaching

• **open source software** is important for re-use
  • reliability?
  • sustainability?
Implementation

- **WebAnno** fulfils most requirements, but lacked support for spoken data:
  - representation of spoken features (e.g. speakers, non-verbal behaviour)
  - media playback
  - transcript visualization
Solution:

• **WebAnno plugin** development by the *Language Technology Group (LT)* in cooperation with the *Hamburg Centre for Language Corpora (HZSK)*, both at the University of Hamburg

• development of a suitable **teaching corpus** of classroom discourse
Plugin functionality

- parsing of transcriptions (HIAT) in the TEI-based ISO standard for Transcription of Spoken Language format created with the EXMARaLDA transcription and annotation tool
- mapping using the Common Analysis System (CAS) within Apache’s Unstructured Information Management Architecture (UIMA)
- transcription convention specific visualizations of the annotation and transcript views
- playback of several audio and video formats
Modelling spoken data

• several **speakers** exist and are described by their individual metadata
  • speaker identity and metadata refers to *some parts of the texts only*, i.e. those belonging to a certain speaker

• speaker contributions **overlap**
  • *serialisation of speaker contributions* and their units is possible but non-trivial – time information remains crucial, overlaps must be recognised
Modelling spoken data

• **non-phonological** material (e.g. breathing, laughter, coughing) occurs in the transcribed text

• there are **non-verbal** speaker contributions (e.g. waving, nodding), possibly parallel to transcribed text

• non-verbal behaviour or **relevant background noise** may span across speaker contributions

• **time-based annotations** may start and end within tokens
<table>
<thead>
<tr>
<th>Line</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>((0.5s)) But when you first met him when you first met, the two/ yes, the two of you, I mean was it like ((0.4s))...</td>
</tr>
<tr>
<td>15</td>
<td>I mean did you think? Good God this/ this is the guy, this is the girl I'm gonna get married to and live the rest of my life with&quot; ((0.4s)) David?</td>
</tr>
<tr>
<td>16</td>
<td>I felt that, yeah.</td>
</tr>
<tr>
<td>17</td>
<td>Definitely.</td>
</tr>
<tr>
<td>18</td>
<td>You did I hope, didn't you?</td>
</tr>
<tr>
<td>19</td>
<td>Did you?</td>
</tr>
<tr>
<td>20</td>
<td>((0.3s)) Yeah.</td>
</tr>
<tr>
<td>21</td>
<td>((laughing))</td>
</tr>
<tr>
<td>22</td>
<td>((laughing)) Was it really like that?</td>
</tr>
<tr>
<td>23</td>
<td>((laughing)) Yeah.</td>
</tr>
<tr>
<td>24</td>
<td>Yeah, yeah, I was/ I was/ it was definitely.</td>
</tr>
<tr>
<td>25</td>
<td>As soon as I see her.</td>
</tr>
<tr>
<td>26</td>
<td>You know, I didn't even have to meet her.</td>
</tr>
<tr>
<td>27</td>
<td>I saw her on the telly and I just wanted to meet her.</td>
</tr>
<tr>
<td>28</td>
<td>You know, I just fancied her.</td>
</tr>
</tbody>
</table>
WebAnno Annotation view

Speaker assignment
Non-phonological/non-verbal material
WebAnno Annotation view

Time aligned media playback
Beckhams

PAR: But when you first met him, when you first met, the two of you, the two of you, I mean was it like a bow, wham, a smack?

DAV: I mean did you think “Good God this is the guy, this is the girl I’m gonna get married to and live the rest of my life with.”

VIC: Did you?

DAV: Definitely. You did hope, didn’t you?

VIC: Yeah. (0.3s) Was it really like that?

(0.3s)

DAV: (laughing) Yeah. Yeah, yeah, I was. I was. It was definitely. As soon as I see her.

VIC: (laughing)

DAV: You know, I didn’t even have to meet her. I seen her on the telly and I just wanted to meet her. You know, I just fancied her.

VIC: And you obviously weren’t disappointed when you got to gather. Ah, similar for you, Victoria?

DAV: (0.6s) So, it was just a case of getting together then.

VIC: (0.6s) You know he was sitting there with his family and I really liked that. And he is a very kind

DAV: What? Disappointed?

VIC: (laughing)

DAV: (laughing)

VIC: (0.3s) he shared the same sort of family values as me. (0.3s) You know he was sitting there with his family and I really liked that. And he is a very kind

PAR: (0.5s) person. And we’re we are really good friends, on top of everything.

VIC: What’s been the lowest part of the marriage? There is at there is a piece in the book, which is quite quite moving actually.

PAR: And it’s about when the media accused David of having an affair.

VIC: That was probably one of the most difficult things to actually write in the book. Obviously, it’s lovely writing all the nice things that’ve happened
First annotation results
Annotation Layers
working with transcripts
• students have to look for a unit of description
  • extent
  • character
• relation to comprehension
• team discussion

steps in using WebAnno
• annotation unit
• consideration of layers, tag sets and tags
• inter annotator agreement
Benefits

• increased student interest in the field of linguistics (data analyses) through data-driven discussions
• starting points for reflection of teachers’ role in classroom communication
• more explicit input for the lecturer on students’ understanding of concepts taught in the lecture
• students learned to use tools of DH while focusing on the content of the lecture
Outlook

- the teaching method can be used in further contexts and/or for other topics
- using the ISO/TEI standard, the WebAnno plugin can be more easily extended to provide support for further transcription conventions
Thanks for your attention

- Demo version (guest/guest): ltdemos.informatik.uni-hamburg.de/webanno-mm

- Plugin code: github.com/webanno/webanno-mm