# EXMARaLDA meets WebAnno

Steffen Remus<sup>\*</sup> Hanna Hedeland<sup>†</sup> Anne Ferger<sup>†</sup> Kristin Bührig<sup>†</sup> Chris Biemann<sup>\*</sup>

\***Language Technology, MIN** Universität Hamburg, Germany

{lastname}@informatik.uni-hamburg.de

<sup>†</sup>Hamburg Centre for Language Corpora (HZSK)

Universität Hamburg, Germany

{firstname.lastname}@uni-hamburg.de

#### Abstract

In this paper, we present an extension of the popular web-based annotation tool WebAnno, allowing for linguistic annotation of transcribed spoken data with time-aligned media files. Several new features have been implemented for our concomitant current use case: a novel teaching method based on pair-wise manual annotation of transcribed video data and systematic comparison of agreement between students. To enable annotation of spoken language data, apart from technical and data model related issues, the extension of WebAnno also offers a partitur view for the inspection of parallel utterances in order to analyze various aspects related to methodological questions in the analysis of spoken interaction.

### 1 Introduction

We present an extension of the popular web-based annotation tool WebAnno<sup>1</sup> (Yimam et al., 2013; Eckart de Castilho et al., 2014) which allows linguistic annotation of transcribed spoken data with time aligned media files.<sup>2</sup> Within a project aiming at developing innovative teaching methods, pair-wise manual annotation of transcribed video data and systematic comparison of agreement between annotators was chosen as a way of teaching students to analyze and reflect on authentic classroom communication, and also on linguistic transcription as a part of that analysis. For this project, a set of video recordings were partly transcribed and compiled into a corpus with metadata on communications and speakers using the EXMARaLDA system (Schmidt and Wörner, 2014), which provides XML transcription and metadata formats. The EXMARaLDA system could have been further used to implement the novel teaching method, since it allows for manual annotation of audio and video data and provides methods for (HTML) visualization of transcription data for qualitative analysis. However, within the relevant context of university teaching, apart from such requirements addressing the peculiarities of spoken data, several further requirements regarding collaborative annotation and management of users and data became an increasingly important part of the list of desired features: a) proper handling of spoken data (e.g. speaker and time information) b) playback and display of aligned audio and video files c) visualization of the transcript in the required layout d) complex manual annotation of linguistic data e) support for collaborative (i.e. pair-wise) annotation f) support for annotator agreement assessment g) reliable user management (for student grading). Furthermore, a web-based environment was preferred to avoid any issues with installation or differing versions of the software or the problems that come with distribution of transcription

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/

<sup>&</sup>lt;sup>1</sup>https://webanno.github.io

<sup>&</sup>lt;sup>2</sup>https://github.com/webanno/webanno-mm

and video data. Another important feature was to use a freely available tool to allow others to use the teaching method developed within the project using the same technical set-up.

While WebAnno fulfills the requirements not met by the EXMARaLDA system or similar desktop applications, it was designed for the annotation of written data only and thus required various extensions to interpret and display transcription and video data. Since there are several widely used tools for the creation of spoken language corpora, we preferred to rely on an existing interoperable standardized format, the ISO/TEI Standard Transcription of spoken language<sup>3</sup>, to enable interoperability between various existing tools with advanced complementary features and WebAnno.

In Section 2, we will further describe the involved components, in Section 3 we will outline the steps undertaken for the extension of WebAnno, and in Section 4, we will describe the novel teaching method and the use of the tool within the university teaching context. In Section 5, we present some ideas on how to develop this work further and make various additional usage scenarios related to annotation of spoken and multimodal data possible.

### 2 Related work

**The EXMARaLDA system:** The EXMARaLDA<sup>4</sup> transcription and annotation tool (Schmidt and Wörner, 2014) was originally developed to support researchers in the field of discourse analysis and research into multingualism, but has since then been used in various other contexts, e.g. for dialectology, language documentation and even with historical written data. The tool provides support for common transcription conventions (e.g. GAT, HIAT, CHAT) and can visualize transcription data in various formats and layouts for qualitative analysis. The score layout of the interface displays a stretch of speech corresponding to a couple of utterances or intonational phrases, which is well suited for transcription or annotations spanning at the most an entire utterance, but an overview of larger spans of discourse is only available in the visualizations generated from the transcription data. The underlying EXMARaLDA data model only allows simple span annotations of the transcribed text; more complex tier dependencies or structured annotations are not possible. When annotating phenomena that occur repeatedly and interrelated over a larger span of the discourse, e.g. to analyze how two speakers discuss and arrive at a common understanding of a newly introduced concept, the narrow focus and the simple span annotations make this task cumbersome.

**WebAnno – a flexible, web-based annotation platform for CLARIN:** WebAnno offers standard means for linguistic analysis, such as span annotations, which are configurable to be either locked to (or be independent of) token or sentence annotations, relational annotations between two spans, and chained relation annotations. Figure 1 (left) shows a screenshot of the annotation view in WebAnno. Various formats have been defined which can be used to feed data into WebAnno.

For analysis and management, WebAnno is also equipped with a set of assistive utensils such as *a*) web-based project management; *b*) curation of annotations made by multiple users; *c*) in-built interannotator agreement measures such as Krippendorff's  $\alpha$ , Cohen's  $\kappa$  and Fleiss'  $\kappa$ ; and *d*) flexible and configurable annotations, including extensible tagsets. All this is available without a complex installation process for users, which makes it particularly suitable for research organizations and a perfect fit for the targeted use case in this work.

**The ISO/TEI Standard for Transcription of Spoken Language** The ISO standard ISO 24624:2016 is based on Chapter 8, Transcriptions of Speech, of the highly flexible TEI Guidelines<sup>5</sup> as an effort to create a standardized solution for transcription data. As outlined in Schmidt et al. (2017), most common transcription tool formats, including ELAN (Sloetjes, 2014) and Transcriber (Barras et al., 2000), can be modeled and converted to ISO/TEI. The standard also allows for transcription convention specific units (e.g. utterances vs. phrases) and labels in addition to shared concepts such as speakers or time information, which are modeled in a uniform way.

<sup>&</sup>lt;sup>3</sup>http://www.iso.org/iso/catalogue\_detail.htm?csnumber=37338

<sup>&</sup>lt;sup>4</sup>http://exmaralda.org

<sup>&</sup>lt;sup>5</sup>http://www.tei-c.org/Guidelines/P5/

#### 3 Adapting WebAnno to spoken data

**Transcription, theory and user interfaces** A fundamental difference between linguistic analysis of written and spoken language is that the latter usually requires a preparatory step; the transcription. Most annotations are based not on the conversation or even the recorded signal itself but on its written representation. That the creation of such a representation is not an objective task, but rather highly interpretative and selective, and the analysis thus highly influenced by decisions regarding layout and symbol conventions during the transcription process, was addressed already by Ochs (1979).

It is therefore crucial that tools for manual annotation of transcription data respect these theory-laden decisions comprising the various transcription systems in use within various reserach fields and disciplines. Apart from this requirement on the GUI, the tool also has to handle the increased complexity of "context" inherent to spoken language: While a written text can mostly be considered a single stream of tokens, spoken language features parallel structures through simultaneous speaker contributions or additional non-verbal information. In addition to the written representation of spoken language, playback of the aligned original media file is another crucial requirement.

**From EXMARaLDA to ISO/TEI** The existing conversion from the EXMARaLDA format to the tool-independent ISO/TEI standard is specific to the conventions used for transcription, in this case, the HIAT transcription system as defined for EXMARaLDA in Rehbein et al. (2004). Though some common features can be represented in a generic way by the ISO/TEI standard, for reasons described above, several aspects of the representation must remain transcription convention specific, e.g. the kind of linguistic units defined below the level of speaker contributions.

Furthermore, metadata is handled in different ways for various transcription formats, e.g. the EXMARaLDA system stores metadata on sessions and speakers separated from the transcriptions to enhance consistency. The ISO/TEI standard on the other hand, as any TEI variant, can make use of the TEI Header to allow transcription and annotation data and various kinds of metadata to be exported and further processed in one single file, independent of the original format.

**Parsing ISO/TEI to UIMA CAS** The UIMA<sup>6</sup> (Ferrucci and Lally, 2004) framework is the foundation of WebAnno's backend. UIMA stores text information, i.e. the text itself and the annotations, in so-called CASs (Common Analysis Systems). A major challenge is the presentation of time-aligned parallel transcriptions (and their annotations) of multiple speakers in a sequence without disrupting the perception of a conversation, while still keeping the individual segmented utterances of speakers as a whole, in order to allow continuous annotations. For this, we parse the ISO/TEI<sup>7</sup> XML content and store utterances of individual speakers in different views (different CAS of the same document) and keep time alignments as metadata within a CAS.

We use the annotationBlock XML element as a non-disruptive unit since we can safely assume that ISO/TEI span annotations are within the time limits of the utterance. Note that annotations, such as incidents, which occur across utterances, are not converted into the WebAnno annotation view, but are present in the partitur view. Other elements, such as utterances, segments, incidents, and existing span annotations are converted to the main WebAnno annotation view.

**New GUI features** In order to show utterances and annotations in a well known and established parallel environment similar to EXMARaLDA's score layout of the partitur editor, we adapt the existing online show case demos<sup>8</sup> and call this view the partitur view henceforth. Figure 1 (right) shows a screenshot of the adjustable partitur view. Both views, the annotation view and the partitur view are synchronized, i.e. by clicking on the correct marker in the particular window, the focus changes on the other.

Also, the partitur view offers multiple media formats for selection, viewing speaker or recording related details and a selectable width of the partitur rows. In the annotation view, we use zero width span annotations for adding time markers. Each segment starts with a marker showing the respective speaker. All markers are clickable and trigger the focus change in the partitur view and start or pause the media.

<sup>&</sup>lt;sup>6</sup>Unstructured Information Management Architecture: https://uima.apache.org/

<sup>&</sup>lt;sup>7</sup>Since ISO/TEI is too powerful in its general form, we restrict ourselves to the HIAT conventions.

<sup>&</sup>lt;sup>8</sup>available at http://hdl.handle.net/11022/0000-0000-4F70-A



Figure 1: Screenshot of the WebAnno-EXMARaLDA plugin. Left: WebAnno's annotation view; Right: approximate EXMARaLDA partitur view. Both sides are synchronized by clicking the correct markers.

For media management, we added a media pane to the project settings, where we included support for uploading media files, which implies hosting them within the WebAnno environment, benefitting from access restrictions through its user management. Additionally, we added support for streaming media files that are accessible in the web by providing a URL instead of a file. Furthermore, multiple media files can be mapped to multiple documents, which allows proper reuse of different media formats for multiple document recordings.

## 4 WebAnno goes innovative teaching

As part of a so-called "teaching lab" the extended version of the WebAnno tool was used by teams of students participating in a university seminar to collaboratively annotate videotaped authentic classroom discourse. Thematically, the seminar covered the linguistic analysis of comprehension processes displayed in classroom discourse. The seminar was addressed to students in pre-service teacher training and students of linguistics. Students of both programs were supposed to cooperate on interdisciplinary teams in order to gain the most from their pedagogic as well as their linguistic expertise. The students had to choose their material according to their own interest from a set of extracts of classroom discourses from various subject matter classes. Benefitting from the innovative ways to decide on units of analysis such as spans, chains, etc., different stages of the process of comprehension. This approach made single steps of analysis transparent for the students, and thus allowed for their precise and explicit discussion in close alignment with existing academic literature. Compared to past seminars with a similar focus, but lacking the technological support, these discussions appeared more thoughtful and more in-depth. The students easily developed independent ideas for their research projects. Students remarked on this very positively in the evaluation of the seminar.

# 5 Outlook

By implementing an extension of WebAnno, we showed that it is possible to repurpose a linguistic annotation tool for multimodal data, in this case transcribed according to the HIAT conventions using the EXMARALDA transcription and annotation tool. The ISO/TEI standard, which can model transcription data produced by various tools according to different transcription conventions, was used as an exchange format. Obvious next steps would therefore be to extend the interoperability to include full support and transcript visualization for further transcription systems, as well as a generic fallback option. Other important tasks to take on are extensions of the ISO/TEI standard to model both metadata in the TEI Header and the complex annotations generated in WebAnno in a standardized way.

#### References

- Claude Barras, Edouard Geoffrois, Zhibiao Wu, and Mark Liberman. 2000. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication Special issue on Speech Annotation and Corpus Tools*, 33(1–2).
- Richard Eckart de Castilho, Chris Biemann, Iryna Gurevych, and Seid Muhie Yimam. 2014. WebAnno: a flexible, web-based annotation tool for CLARIN. In *Proceedings of the CLARIN Annual Conference 2014*, pages 1–3.
- David Ferrucci and Adam Lally. 2004. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. *Natural Language Engineering*, 10(3–4):327–348.
- Elinor Ochs. 1979. Transcription as theory. In E. Ochs and B.B. Schieffelin, editors, *Developmental pragmatics*, pages 43–72. Academic Press, New York.
- Jochen Rehbein, Thomas Schmidt, Bernd Meyer, Franziska Watzke, and Annette Herkenrath. 2004. Handbuch für das computergestützte Transkribieren nach HIAT. Arbeiten zur Mehrsprachigkeit, Folge B, 56:1 ff. DE.
- Thomas Schmidt and Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, editors, *Handbook on Corpus Phonology*, pages 402–419. Oxford University Press.
- Thomas Schmidt, Hanna Hedeland, and Daniel Jettka. 2017. Conversion and annotation web services for spoken language data in clarin. In *Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016*, number 136, pages 113–130. Linköping University Electronic Press, Linköpings universitet.
- Han Sloetjes. 2014. ELAN: Multimedia annotation application. In Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, editors, *Handbook on Corpus Phonology*, pages 305–320. Oxford University Press.
- Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. 2013. WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–6, Sofia, Bulgaria.