

Expanding and connecting the annotation tool ELAN

Han Sloetjes, Aarthy Somasundaram, Sebastian Drude, Herman Stehouwer,
Kees Jan van de Looij

Max Planck Institute for Psycholinguistics
Wundtlaan 1, 6525XD Nijmegen, The Netherlands

E-mail: {han.sloetjes, aarthy.somasundaram, sebastian.drude, herman.stehouwer, keesjan.vandelooij}@mpi.nl

Abstract

The annotation tool ELAN allows for adding time-linked textual annotations to digital audio and video recordings. It is applied in various disciplines within the humanities, with linguistics, sign language and gesture research represented most prominently in its user base. This paper highlights new developments in ELAN with an emphasis on those features that introduced new technological and methodological approaches to analysing both audio/video and derived textual data.

1. Introduction

Annotation of audio and video recordings, be it manual or (semi-)automatic, is a crucial step in many types of research within the humanities. ELAN¹, developed at The Language Archive² / Max Planck Institute for Psycholinguistics, is a tool for manual annotation that has been around for more than a decade and has user communities in language documentation, sign language and gesture studies but also in psychology, education, dance and human and animal behaviour research projects. This application enables the user to create multi-levelled, multi-participant, time-linked annotations to one or more media streams, including timeseries streams. Both qualitative and quantitative research is supported; arguably the qualitative oriented application his predominant but the quantitative utilization is gaining popularity. In this paper we focus on recent developments that aim at improving the workflow of researchers by introducing task oriented modes and at expanding the scope of the program by implementing a framework for computational annotation creation modules and by connecting to web services that likewise apply computational techniques to create annotations. ELAN is free and open source software that runs on Windows, MacOS and Linux.

2. The Interlinearization mode and text processing modules in Lexan

One of the recent and still ongoing developments concerns the introduction of the Interlinearization mode. This mode, on the one hand, provides a user interface optimized for the task of adding linguistically relevant layers to an orthographical transcription of the media. Layers for morphological break down, part of speech tags and glossing are part of the common repertoire of documentary linguists (Bow, 2003). On the other hand this mode is the hub to Lexan, the extensible framework for annotation and text processing modules. Such modules can perform a variety of tasks, from simple to

complex, from word segmentation to interlinearization based on machine learning algorithms. Some modules are expected to produce multiple suggestions for new annotation layers and to improve their suggestions based on interactive user feedback accommodated by the user interface of, primarily, the interlinearization mode. The name "Lexan" indicates that this framework interconnects ELAN with the TLA multimedia lexicon tool LEXUS³. This architecture allows to build and enrich a lexicon while annotating and at the same time to use information in the lexicon in the annotation suggestions process. This combination of NLP (Natural Language Processing) techniques with manual media annotation marks a new line of development in ELAN and brings together technologies that normally seem to develop apart.

For this sort of work other tools are and have been around for a long time and providing interoperability with these tools (often implemented on the level of file format conversion) is highly important for many users.

3. Interoperability with FLEx

The FieldWorks Language Explorer⁴ is a prominent example of such tools, therefore import and export facilities for the FLEx format have been implemented and revised with the goal to make repeated transfer of data ("round-tripping") between the tools as seamless as possible. Importing FLEx files was possible since ELAN version 3.8 (2009) but because the FLEx format at that time did not support time alignment and speaker information an export function was not implemented at the same time. That has been added recently, after the introduction of the "begin-time-offset", "end-time-offset" and "speaker" attributes at the phrase level of the flextext format (2012, FLEx 7). The import has been updated such that a group of tiers is created per speaker. Additionally efforts are made to retain punctuation and font information where possible. Punctuation elements are linked to an ISOcat data category so that on export these elements receive the correct attribute again. Exporting an ELAN document that is the result of a

¹ <http://tla.mpi.nl/tools/tla-tools/elan>

² <http://tla.mpi.nl>

³ <http://tla.mpi.nl/tools/tla-tools/lexus>

⁴ <http://fieldworks.sil.org/flex>

FLEX import is fairly straightforward. Exporting just any ELAN document to FLEX remains a challenge; where ELAN is very flexible and allows to have any number of tiers without predefined content designation, is FLEX more rigid providing a fixed set of layers of known categories.

4. Connecting to web services and online resources

ELAN is a standalone desktop application that in principle works with locally available (local hard drive, local network) resources. Audio and video files are (more and more) often very large, up to several Gb. per file, and high accuracy annotation is still problematic when using media streaming, even in situations with high speed internet connections. For the vast majority of features of ELAN an internet connection is not required, but recently several options have been added that allow the user to connect to online services and resources. In 2008 association of tiers and annotations with ISOcat⁵ data categories was introduced and this feature has recently been improved and made more relevant. By default tiers are generic annotation "containers", oblivious of the type of content of the annotations; there are no predefined tiers, e.g. "translation tier" or "gesture phases tier". By associating a tier with a concept registered and described in ISOcat it acquires an explicit content designation.

In the CLARIN-NL SignLinc⁶ project support for external controlled vocabularies was added, enabling collaborators to share vocabularies over the network (Crasborn and Sloetjes, 2010). This feature improves consistency within a team and prevents team members from making (unchecked) changes to the vocabularies.

In the context of several CLARIN⁷ projects and of the AVATech⁸ project extensions were developed that call web services which produce segmentations and/or annotation content taking audio, video or text, or a combination thereof, as input. The WebLicht⁹ tool chaining framework is a core service in CLARIN-D¹⁰ (Hinrichs et al., 2010) and preliminary support for calling services registered with this framework is now available to users of ELAN. Tiers can be uploaded (in the required XML format, TCF) in order to be processed by well known parsers and taggers; the results are added as new tiers and thus enrich the annotation document.

No matter how useful these web services are or will become, for many field linguists, and other researchers who are working offline a lot, these provisions will not be available. Therefore the core functions will always be independent of online services. For some services, like ISOcat, it is possible to work with a local cache; a selection of categories is stored on the local machine for

use while offline.

Most parsers and taggers are only available for a small number of "major" languages, linguists who study lesser described, let alone, endangered languages usually don't have similar mature, well tested and well trained systems at their disposal. The Lexan approach, stepwise building up "personalized" computational assistance based on the input and feedback of the user, can come to their rescue.

5. Local corpus enrichment and exploration

Though ELAN has been a multiple document application almost from the start, most functions of ELAN allow the user to interact with one, the current active, document. But in recent years more and more functions have been introduced that operate on multiple files e.g. on an entire local corpus. The urge for such functions emerged with the growing number of recordings and transcriptions research teams nowadays are working on (Johnston, 2010).

A shortlist of multiple files functions contains creation of transcriptions for a set of recordings, editing the collection of tiers in transcriptions, creating annotations by applying logical operations (AND, OR, XOR) on annotations of selected tiers (Lausberg and Sloetjes, 2009), extracting information by executing search queries, generating simple statistics, converting multiple files to and from specific formats etc.

For some types of research assessing the quality of the annotations and the skills of the annotators is crucial. How the inter-annotator reliability best is assessed is still under discussion (Gut, 2004; Holle and Rein, 2012; Lücking, 2011) and the best approach can differ depending on the properties of the data and the focus of the research. A few algorithms for calculating the inter-annotator agreement have been implemented in ELAN and are available for application on multiple files. Especially concerning time-alignment (the segmentation step of the annotation process) there seems to be no generally accepted algorithm for assessing agreement. By offering several alternatives, the choice remains to the user while some of the hassle of exporting data to other tools is taken away from her/him.

6. Conclusion

In this paper we show how researchers working with digital audio/video materials across disciplines can apply new technologies as a result of connections established between ELAN and local or online modules and services. Features that allow to enrich and explore a local corpus are introduced and briefly discussed.

7. References

- Bow, C., Hughes, B. and Bird, S. (2003). Towards a general model of interlinear text. In *Proceedings of EMELD Workshop 2003: Digitizing and Annotating Texts and Field Recordings*. Lansing MI, USA.
- Crasborn, O., Sloetjes, H. (2010). Using ELAN for annotating sign language corpora in a team setting. In *Proceedings of the 4th Workshop on the*

⁵ <http://www.isocat.org>

⁶ <http://www.ru.nl/sign-lang/@673229/pagina/>

⁷ <http://www.clarin.eu>

⁸ <http://www.mpi.nl/avatech>

⁹ <http://www.clarin-d.de/language-resources/weblicht-en.html>

¹⁰ <http://www.clarin-d.de>

Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (pp. 137-142). Paris:ELRA.

- Gut, U., Bayerl, P. (2004). Measuring the Reliability of Manual Annotations of Speech Corpora. In *Proceedings of Speech Prosody, Nara*.
- Hinrichs, M., Zastrow, T. and Hinrichs, E. (2010). WebLicht: Web-based LRT Services in a Distributed eScience Infrastructure. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta.
- Holle, H., Rein, R. (forthcoming) Assessing interrater agreement of movement annotations. To appear in *H. Lausberg (ed.), NEUROGES The Neuropsychological Gesture Coding System*. Peter Lang: Berlin.
- Johnston, T. (2010). Adding value to, and extracting of value from, a signed language corpus through secondary processing: implications for annotation schemas and corpus creation. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (pp. 137-142). Paris:ELRA.
- Lausberg, H., Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, 41(3), 841-849.
- Lücking, A., Ptock, S. and Bergmann, K. (2011). Staccato: Segmentation Agreement Calculator. In *Proc. of the 9th International Gesture Workshop, May 25-27, 2011, Athens, Greece*.