Measuring by marking; the multimedia annotation tool ELAN

Han Sloetjes, Olaf Seibert

Max Planck Institute for Psycholinguistics
P.O. Box 310, 6500 AH Nijmegen, The Netherlands
Han.Sloetjes@mpi.nl, Olaf.Seibert@mpi.nl

Introduction

ELAN is a multimedia annotation tool developed by "The Language Archive" [1, 2], a unit of the Max Planck Institute for Psycholinguistics. It is applied in a variety of research areas in which audio and/or video recordings are the basis for qualitative and/or quantitative analysis of different modalities of communication. This paper aims at presenting a general overview of the tool with an emphasis on the most recent developments.

General overview of the tool

ELAN is a tool for manually adding textual annotations to audio and/or video recordings. The annotations are stored separately from the source media in XML files. The data model is tier based, allowing for multi-level, multi-participant annotation of time based media. A tier, in this context, is a kind of layer object, a container for grouping annotations that share the same topic or target and the same coding scheme. For example, one tier could contain annotations transcribing the speech of participant AA, another tier could record the left-hand gestures of participant BB etc. The user defines the tiers and determines what to use each for, which codes or categories to apply to each of them. The basic form of an annotation consists of a start time and end time, marking a specific segment of the media stream, possibly enriched by a text string. Annotations can also be linked to other annotations and be part of a hierarchically structured group of annotations. By marking segments of interest and storing them as annotations the user produces datasets that represent a measurement of the scored behavior. In many cases these measurements are the basis for further qualitative or quantitative analysis performed by other tools.

It is possible to link multiple video and multiple audio streams (e.g. in the case of recordings with multiple cameras and microphones) and the user interface allows to specify which videos and which sound channels to view and hear at any given time. ELAN is available for Windows, Mac OS X and Linux, is mostly written in the Java programming language, is free for anyone to use and is open source.

Linguistic Annotator

The "LA" in the ELAN acronym (originally: **E**udico **L**inguistic **AN**notator, where Eudico was the name of a project) indicate that the tool initially was conceived as a tool for annotating verbal human communication. And indeed an important part of the user base consists of (field) linguists. Considerable funding contributions came from a big project on endangered languages, DoBeS [3]. Within this project more than 60 teams were funded to document a language that is in danger of becoming extinct. The efforts of these teams were supported and facilitated by the development of software for archiving language data, producing metadata descriptions and for transcription of primary data. The aim was to produce transcription and translation annotations for most of the recordings and more elaborate morpho-syntactic glossing annotations for a smaller portion of the data. Nevertheless, most of the functionality of ELAN and the design of the data format are not particularly linguistic by nature but rather generic. To illustrate this, it is only to the present day that serious efforts are made to integrate a lexicon component (which adds linguistic functionality). That's how almost from the beginning the tool was also used for non-verbal human communication, with strong user communities in sign language research and in gesture and multimodal interaction research. Apart from that, but to a lesser extent, the application is also used for coding or scoring animal behavior.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

What these user groups have in common in terms of required functionality is segmenting and labelling (although labelling is not always required, e.g. in some multimodal interaction projects and animal behavior scoring segmentation on multiple levels suffices). Almost always annotation takes place on multiple tiers (e.g. for multiple participants and multiple modalities (speech, gesture, facial expressions)) and in many cases the feature of creating tier hierarchies (parent-child relations between tiers) is harnessed. Linguists often use dependent tiers for translations and for morpho-syntactical analysis layers and some gesture researchers subdivide gesture units into smaller units on dependent tiers.

But there are also requirements that are specific to certain user groups. Many linguists also use tools like Toolbox or FLEx [4] for transcription and morpho-syntactic glossing and for building the lexicon and they rely on the import and export functions for these formats for interoperability. An exhaustive description of Toolbox and FLEx functionality and on how to transfer data between these tools and ELAN is given by [5]. It is fairly common in endangered languages projects to create subtitled videos as one of the outcomes, therefore the export to subtitled text is important for such projects (ELAN does not create hardcoded subtitled videos).

Gesture and multimodal interaction researchers seem more likely to benefit from tier-based operations, features that allow to combine annotations on two or more tiers to create new annotations, e.g. annotations from overlaps, or annotations by merging or subtracting. These functions are useful for combining already created, existing segmentations for the purpose of creating new annotations representing a specific desired (sub-) result.

Interrater agreement or reliability

In some projects it is important to assess the reliability of annotations created by different annotators (inter-annotator) and/or annotations created by the same annotator but at different times (intra-annotator). For a long time ELAN did not have built-in functionality to calculate agreement except for a simple algorithm for comparing annotations on two tiers without chance correction. For more elaborate assessment of agreement or reliability other tools were (and are) available and exporting as .csv text suffices to feed the data into these tools.

One of these tools is EasyDIAg [6] implemented in Matlab in the context of NEUROGES [7] (gesture) research. The algorithm is based on Cohen's kappa, extended with a routine to match annotations based on the amount of overlap (therefor the segmentation decisions are part of the agreement assessment). Recently this algorithm has been (re-)implemented in Java and included in ELAN, with some additional flexibility in pairing tiers to compare. There is no predefined prefix or suffix on the basis of which tiers are combined and tiers do not have to be in the same file in order to be compared.

Another tool is Staccato [8] which calculates a Degree of Organization for segmentations (ignoring the labels) while applying a Monte Carlo simulation for randomization. This tool was already implemented in Java and was added to ELAN as a separate library, following a collaboration within a CLARIN-D [9, 10] working group for multimodality.

Manual vs automatic annotation

ELAN is tool for manual annotation of recordings; the usual workflow is that the user inspects the media stream, identifies relevant segments (depending on the area of interest and research) and creates annotations using the mouse and/or the keyboard. The level of precision completely depends on the user and on the nature of the research project. In some types of research it is not necessary that the segments are frame-precise and an overshoot of a few hundred milliseconds at start and end boundary are acceptable. The segmentation task can then be performed in the Segmentation Mode, in which segments are created by one or two key strokes while the media is playing. In other projects it is crucial that segment boundaries are as exact as possible, frame precise. This requires close inspection with a lot of frame stepping back and forth and playing a selection repeatedly before deciding on the segment. Most likely this is done in Annotation Mode. Especially the latter approach consumes a lot of the annotator's time and is therefore expensive.

Yet, this is still the default workflow in ELAN. Although we have been involved in several projects that aimed at development and integration of tools for (semi-)automatic segmentation and/or labelling, successful application thereof in real world scenario's proved to be problematic. Some algorithms perform fairly well when applied to "clean", good quality data, but perform poorly on naturalistic lab or field data.

We have been involved in two projects concerning automatic audio and video recognition, AVATecH and AUVIS [11], both in collabaration with two Fraunhofer institutes. A first version was included in 2010 and within the AUVIS project the algorithms were improved and the user interface further streamlined. These technologies can not replace manual creation of annotations yet, but can be applied assistively, in a scenario in which the automatic recognizer creates a segmentation which can then be manually corrected.

In the context of the big CLARIN project ELAN has been extended with client functionality for the text-processing tool-chaining framework WebLicht. Text can be uploaded to that webservice and the user can select one of the available tools for e.g. morpho-syntactic parsing and tagging of the text. The returned data are converted back to ELAN's data format and added as new tiers. This facility is mainly of interest for users who are working on (communication in) languages for which such taggers exist (so-called major or bigger languages).

Sharing Comments with team members

ELAN does not support true collaborative annotation, where multiple users simultaneously work on the same file. But to improve collaboration of team members who work on the same transcripts, either successively or in different files that are later merged, a commentary framework was implemented. A Comment is text that the user links to a segment of the media and possibly to a tier. Comments can be used as notes, remarks or questions concerning a fragment for later or for discussion with colleagues. This is quite similar to the way comments in word processors are applied. There are several ways in which comments can be shared: via email, via a file sharing (cloud) service (such as Dropbox) or via the DASISH Web Annotator (DWAN) framework [12, 13]. Comments can search and filtered based on one or several of their properties.

Export to Theme format

One of the more recently added export formats is that of Theme [14] data files. Theme is an application for detection and analysis of hidden patterns in time (so called t-patterns) in behavioral data. Annotations of (a selection of) tiers can be stored in a set of files that can be imported into a Theme project. A Theme project requires at least two text files, a Category or Variable-Value-Table File (.vvt, a file containing a listing of the coding Classes and the Items of each class) and one or more Raw Data files. The VVT is similar to the idea of Controlled Vocabularies and their entries in ELAN; when exporting the user can choose to export the entire CV's or just the values that were actually applied to annotations. The data text files, in which each line represents a record of a time-stamped event, contain the annotation data. Time-alignable annotations in ELAN have a duration (ELAN does not support single point annotations), therefore each annotation occurs twice in the output, once as a start ('b') event and once as an end ('e') event. The notion of Actor in Theme corresponds to either the tier name or the participant label of a tier (if it is there). This is a multiple file export, facilitating transfer of the data of a (sub-) corpus to Theme in one action.

Conclusions

ELAN is a multimodal annotation tool that is used in many behavioral research projects. It offers many functions that are indispensable to measure behavior on the basis of media recordings of (human) interaction. Despite participation in projects that aim at a shift towards automatic recognition of segments of interest, ELAN is still primarily an application in which the user has to mark the segments of interest manually. The user also has to determine the code or category for each segment, if any, resulting in a data set that represents a measurement of behavior. Manual annotation is a time consuming and expensive procedure, therefore it is inevitable that there

will be more initiatives to improve the algorithms for automatic recognition. Both in the case of manual and automatic annotation the quality of the results has to be assessed therefore the new features for calculating inter-rater reliability are a useful extension of the program. The export to the Theme data format improved the level of interoperability with other tools that are relevant to the field. Some features that are in preparation are a lexicon component, cross-referencing between annotations and simplified user interfaces that expose only a limited set of functions, necessary for specific tasks. A simplified interface with reduced functionality will hopefully make the learning curve less steep so that it will become easier to involve more people in annotation tasks.

References

1. ELAN annotation tool, <http://tla.mpi.nl/tools/tla-tools/elan/>. Accessed 27 January 2016.
2. The Language Archive, <http://tla.mpi.nl/>. Accessed 27 January 2016.
3. DoBeS, Dokumentation bedrohter Sprachen, <http://dobes.mpi.nl/>. Accessed 27 January 2016.
4. FLEx, FieldWorks Language Explorer, <http://www.sil.org/resources/software_fonts/flex>. Accessed 20 January 2016.
5. Pennington, R. (2014). Producing time-aligned interlinear texts: Towards a SayMore–FLEx–ELAN workflow. <http://www.academia.edu>. Accessed 25 March 2016.
6. Holle, H., Rein, R. (2014). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior Research Methods, 46(3).*
7. Lausberg, H., Sloetjes, H. (2015). The revised NEUROGES-ELAN system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior Research Methods, DOI: 10.3758/s13428-015-0622-z.*
8. Lücking, A., Ptock, S., Bergmann, K. (2011). Staccato: Segmentation Agreement Calculator. In *Proc. of the 9th International Gesture Workshop, May 25-27, 2011, Athens, Greece*
9. Common Language Resources and Technology Infrastructure, <http://clarin.eu/>. Accessed 20 January 2016.
10. CLARIN Germany, <http://www.clarin-d.de/de/>. Accessed 20 January 2016.
11. Audiovisuelles Data Mining in multimodalen Sprachdaten, <https://tla.mpi.nl/projects_info/auvis/>. Accessed 27 January 2016.
12. Data Service Infrastructure for the Social Sciences and Humanities, <http://dasish.eu/>. Accessed 27 January 2016.
13. Lenkiewicz, P., Shkaravska, O., Goosen, T., Broeder, D., Windhouwer, M., Roth, S., Olsson, O. (2014). The DWAN Framework: Application of a Web Annotation Framework for the General Humanities to the Domain of Language Resources. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), May 26-31, 2014, Reykjavik, Iceland*
14. Theme, T-Pattern Analysis, <http://patternvision.com/>. Accessed 20 January 2016.