

ELAN development, keeping pace with communities' needs

Han Sloetjes, Aarthy Somasundaram

Max Planck Institute for Psycholinguistics

P.O. Box 310, 6500 AH Nijmegen, The Netherlands

E-mail: Han.Sloetjes@mpi.nl, Aarthy.Somasundaram@mpi.nl

Abstract

ELAN is a versatile multimedia annotation tool that is being developed at the Max Planck Institute for Psycholinguistics. About a decade ago it emerged out of a number of corpus tools and utilities and it has been extended ever since. This paper focuses on the efforts made to ensure that the application keeps up with the growing needs of that era in linguistics and multimodality research; growing needs in terms of length and resolution of recordings, the number of recordings made and transcribed and the number of levels of annotation per transcription.

Keywords: multimedia annotation, multimodality, transcription

1. Introduction

ELAN [1] is a versatile multimedia annotation tool that is used in a variety of research areas in linguistics and beyond. Areas such as sign language research, field linguistics, gesture studies, or more general, multimodal interaction research. The direction of developments has largely been determined by the input and feedback of users from these communities. Long-term commitment and dedication are absolutely necessary for continued use and acceptance by the communities. General descriptions of the tool and its data model have been presented at previous LREC editions (Wittenburg et al., 2006, Crasborn et al., 2006). This paper outlines the most important changes and additions made in order to meet with the increased requirements of the user communities. These requirements have grown both qualitatively and quantitatively. Qualitatively in the sense of growing demands on usability/ease-of-use and transcription speed, quantitatively in the sense that an increase can be observed in media file sizes, in the number and length of the recordings and in the number of annotation levels created.

The rapid developments in media recording equipment in terms of supported resolutions (high definition) and increased storage capacity, has led to the production of large quantities of data. Not only in lab/studio situations, but also in the field it has become much easier to record for many hours and sometimes even with more than one camera. The original files are often too big and too unpractical to be used in ELAN directly, therefore often smaller files, of still high quality, are extracted. Especially the mp4/H.264 format has gained popularity and particularly on Windows this is still problematic. Additional codecs are necessary; only Windows 7 has built in support for mp4 and attempts have been made to make this support available in (the Java based) ELAN as well.

Not only more and larger media files are created, we can also observe an increase in the number of levels of

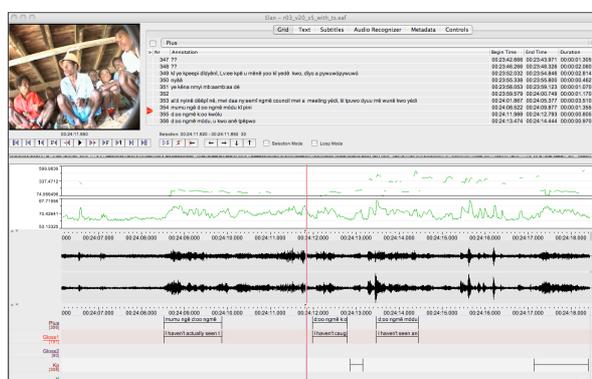


Figure 1 ELAN window in the default working mode, the "Annotation mode"

annotation that are added to the primary data. Many layers of analysis are added, resulting in transcription files with, in some cases, more than dozens of tiers. As a result of these changes two main requirements of ELAN are to improve the efficiency of common tasks and to allow to perform tasks on multiple files in a single action, where possible. The first task has been tackled by introducing new, task-based working modes and by adding more tier-based operations and more options for auto-creating annotations. The second task by increasing the number of operations that can be performed on multiple files.

2. Task oriented working modes

To accommodate a more efficient workflow new "working modes" have been implemented that are optimized for a specific task. In general the transcription process consists of two major components: segmentation and text input. In the main window of ELAN, as it has been for many years, both actions can be performed, while the interface is optimized for precision (accurate alignment) rather than speed. This traditional mode is still the default, now named "Annotation mode".

2.1 Segmentation mode

This mode has been designed for fast creation of segmentations, possibly while the media is playing. This functionality was previously available, in a rudimentary form, in a separate window. Now it has been enhanced and integrated as the working mode tailored to the segmentation task. It allows to quickly switch between tiers, create annotations with one or two keystrokes and conveniently adjust the alignment of annotations with the mouse. A segment can be easily split into two segments and two segments can be merged into one, as part of the fine tuning process after a coarse segmentation during play back. The behaviour of the segmentation mode can be customized in several ways. A delay can be specified to compensate for keystroke latency. Instead of marking both begin and end time of a segment, it is possible to create a segment of a fixed duration by a single keystroke. In conjunction with controlled vocabularies containing mappings of values to keyboard keys, the segmentation mode turns into a scoring tool that allows for segmenting and labeling at the same time by using different keys for marking segments.

2.2 Transcription mode

This mode has been introduced early 2011 (version 4.1.0). The Transcription mode is geared to rapid text entry into existing segments (annotations). In contrast to the previous two modes, transcription mode displays the annotations of selected tiers in a spreadsheet like table with a vertical orientation (the timeline goes top-down). Work in this mode is highly keyboard oriented; after typing in some text and hitting the Enter key, the next annotation is activated: the cell becomes editable and the media player automatically plays the corresponding

segment. The contents of the table is customizable; the user can decide which tier types or individual tiers are on display. The navigation between cells can be modified to either stay within one and the same column or between columns (jump to the next column in the same row). The introduction of this mode has proven to be very useful. It reduces the number of mouse clicks and the overall time needed to transcribe an hour of recording. Also, in field situations the ease of use of this mode has made it less problematic to engage local assistants, who often have limited computer experience, in the transcription process. This new transcription interface is discussed in more detail in Dingemanse et al. (2012). They claim that the new interface significantly reduces annotation time and that it improves on the workflow of some other commonly used tools.

2.3 Synchronization mode

The media synchronization mode has been available for many years already and has recently been updated to also visualize the waveform of audio files, making synchronization of different types of media (video, audio and timeseries data) more convenient. Although it is best to clip media files beforehand such that they are in sync, this mode allows to set an offset per file in the case they are not.

2.4 Interlinearization mode (in preparation)

The Interlinearization mode is designed for the task of semi-automatic, semi-supervised interlinearization of one layer of annotation into subordinate layers of annotation (parses and glosses). This functionality depends on the availability of a lexicon [2], a lexicon

No.	Type1 : po (practical orthography)	Type2 : tl (literal translation)	Type3 : tf (free translation)	Type4 : tn (translation in lingua f...)
1	àtítábi àna wiriwírí.	the maize too, smaller, smaller smaller.		
2	È! síkpákpa ðe!	yes! olden language!		
3	È! àwiriwíríá wàà mmo ne!	Yes! small ones that are there!		
4	tā ñsè láà piè ñdù lo!	let me go and bath!		
5	ñjñhù!	like when something unusual happen and the incident has been narrated to you and you don't really concern.		
6	ne gò kadé òsè ne!	and when day breaks,		
7	ne gò teè ò pié gò áta ò pié ñdù sasa ne!	since like you would have bath , when you wake up you will bath worm water!		
8	ímèè loò piè íó!	like i would have bath!		

Figure 2 ELAN window in “Transcription mode”

that at the same time is being build and extended in the process of annotation.

3. Scaling up: multiple file operations

Although multiple documents can be opened simultaneously in ELAN (each in its own window, in a so called Single Document Interface), most operations perform on the one document in the active window. Today a number of functions are available that execute on multiple files selected by the user. The files have to be available locally, on the local hard drive or network. These operations have in common that the user can define a “domain” containing folders and/or annotation files (.eaf files, ELAN’s native format). These domains can be named and saved for later use, allowing the user to specify different views on the local corpus. Especially now that many users are working with hundreds of annotation files, being able to edit, convert or search them simultaneously makes a huge difference. The time needed for a task is diminished to an extent that it wouldn't be possible to perform it when it has to be done on a file-by-file basis. Two categories of operations can be distinguished: editing and non-editing. The latter don't change the files in the domain (e.g. export functions), while the former do change the files and can in general not be undone easily.

3.1 Multiple file creation

It is possible to create multiple annotation files for a collection of media files in one action. The user can select a folder, possibly including sub-folders, containing media files, select a template for the tier setup and generate eaf files for the recordings. Video and audio files are combined into a single transcription based on a “same name, different extension” principle. Videos of multiple camera recordings can be combined in a single transcription on the basis of a different prefix or suffix in the file name.

3.2 Multiple file data export

Currently ELAN provides three “export from multiple files” options. Firstly it is possible to export annotation documents as tab-delimited text. The result is in fact an accumulated tabbed single file export. This export has been improved by the option to create a separate column for each tier (resulting in annotations with the same time boundaries appearing in the same row).

Furthermore multiple files can be exported to a single wordlist or a single file containing all unique annotation values. In both cases tiers can be selected and an occurrence count can be included in the export.

For corpora holding transcriptions containing several dozens of tiers, it can be advantageous to be able to export a selection of the tiers of each file as new eaf files. Depending on the targeted audience, different views on the transcriptions can be represented by the choice of included tiers.

Common conversions for many ELAN users are those to and from Toolbox and Praat. These conversions are now

available as multiple file operations, most other import and export options will become available as such operations as well. The user interface for customizing the import/export can, in most cases, be re-used. Here again the purpose is reducing the number of clicks and repetitions of the same action.

3.3 Multiple file find-and-replace

The find-and-replace in multiple files is an example of the “editing” type of multiple file operations. Since all files potentially can be changed by this action, the user should take care of keeping save copies; a replace can usually not simply be undone by a reverse find-and-replace.

Once the files have been added to the domain, it is possible to specify a selection of tiers to work on, the search query and the replace value can be entered and the type of matching (regular expressions are supported) can be chosen.

A powerful search engine for structured search in local corpora is included in the distribution as well. It allows for constructing and executing complex queries based on structural and temporal relations of annotations within and over tiers (Stehouwer & Auer, 2011). Annotation content can be disclosed, patterns detected and hypotheses tested. The results can be exported to tab-delimited text for further processing in spreadsheet and statistical analysis applications.

3.4 Multiple file edits of tiers and linguistic types

Even when annotation documents have been based on a template, the need may emerge to add new tiers and linguistic types at a later stage. In fact, it is most likely that while working with the data additional transcription layers are necessary for types of events that were either not anticipated or not planned for transcription initially. The flexible document handling in ELAN allows for adding any number of tiers to any individual document. But adding tiers and types, changing tier and type names and changing names and codes used for participants and annotators can be done for complete local corpora as well. Collaborators from the Radboud University in Nijmegen [3], have created a friendly yet powerful utility for this purpose.

3.5 Multiple file annotations from overlaps, merging and subtraction

A number of tier based operations can be performed on a collection of files, with the common options available when performing these operations in a single file. The rationale behind these operations is discussed in the next section. Currently available are new annotations from overlaps of annotations on different tiers, annotations by merging tiers and annotations by subtracting annotations on different tiers.

4. Tier-based operations

Manipulation of annotations and tiers has been improved and extended considerably. As already mentioned,

annotation creation is a tedious, time consuming task therefore many functions have been added to speed up the transcription process and to reduce the number of mouse clicks and keystrokes to get to a certain result.

Copying and pasting an annotation with or without depending annotations, duplication of an annotation, shifting all annotations of a document or just the annotations of a single tier, left or right of the current media time or only those within the selected time interval, deletion of annotations left or right of the crosshair, deletion of annotations randomly selected by mouse clicks, are some others. It is now possible to fill a tier with same length annotations in a single action and to label and number annotations. The examples above all represent fairly common edit operations that are mainly annotation oriented.

Tier oriented operations include a function to copy complete tiers, possibly including their dependent tiers. The gaps between annotations on a tier can be filled with new annotations; when the tier contains speech annotations, the gaps represent silence. By creating the new annotations on a different tier than the speech tier, the transcription is conveniently and efficiently enriched with a silence tier. The hierarchical nature of tier structures can be turned to advantage by the options to automatically create child annotations on depending tiers, either per parent annotation when it is created, or for all annotations of selected tiers in one go.

The triple consisting of the functions to create annotations based on the overlaps of annotations of two tiers, to merge annotations of two tiers onto a third tier and to create annotations by subtracting the annotations of one tier from those of another, together represent the logical AND, OR and XOR (exclusive or) operations applied to time segments. Although e.g. overlaps can be found and extracted by the search facilities, there can be good reasons to use the overlap function to (permanently) create a new tier. The new tier itself can hold and represent linguistically meaningful units and it can in turn become a source tier for following tier operations. Combination of several of these tier based operations with the option to concatenate annotation values can add value to a corpus on the one hand and circumvent possible limitations of the search mechanism on the other (Johnston, 2010).

A combination of the overlap and merge routines is applied in the non-standardized “compare annotators”

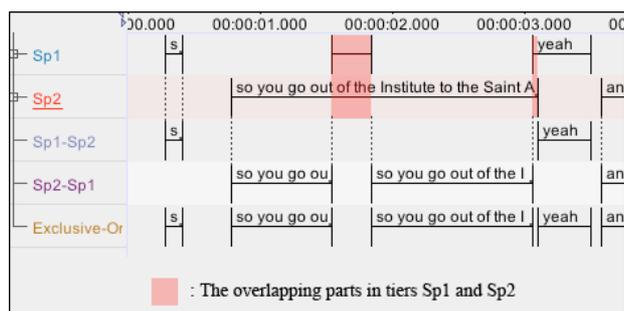


Figure 3 Illustration of subtraction of annotations.

feature (Lausberg & Sloetjes, 2009). Implementations of more commonly accepted interrater agreement algorithms are expected to be added in the future.

4.1 Automatic annotation

Given the (growing) amount of data and given the fact that manual annotation is time intensive, the need for automatic annotation by applying pattern recognition software is emerging (Auer et al., 2010) [4]. ELAN has implemented an extension mechanism for this type of software components and has been extended with a "silence recognizer" that can automatically segment silence versus non-silence for audio files of reasonable quality.

First results of more sophisticated recognizers are now being evaluated and will be published soon. It is expected that automatic processing of audio, video and text (parsing, tagging) will reduce the time needed to produced full fledged transcriptions of recordings considerably.

5. Conclusion

ELAN has been extended and improved in the past few years, often in close cooperation with users and inspired by their input. It remains a challenge to adapt the application continuously in order to keep up with the changing requirements of user communities.

6. References

- Auer, E., Russel, A., Sloetjes, H., Wittenburg, P., Schreer, O., Masnieri, S., Schneider, D., Tschöpel, S. (2010). ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*
- Crasborn, O., Sloetjes, H., Auer, E., Wittenburg, P. (2006). Combining video and numeric data in the analysis of sign languages with the ELAN annotation software. In C. Vettorelli (Ed.), *Proceedings of the 2nd Workshop on the Representation and Processing of Sign languages: Lexicographic matters and didactic scenarios* (pp. 82-87). Paris: ELRA
- Dingemans, M., Hammond, J., Stehouwer, H., Somasundaram, A., Drude, S. (In press). A high speed transcription interface for annotating primary linguistic data. In *Proceedings of the EACL 2012 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*.
- 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

- Stehouwer, H., Auer, E. (2011). Unlocking language archives using search. In C. Vertan, M. Slavcheva, P. Osenova, & S. Piperidis (Eds.), *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*. Hissar, Bulgaria, 16 September 2011 (pp. 19-26). Shoumen, Bulgaria: Incoma Ltd.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.

- [1] <http://www.lat-mpi.eu/tools/tools/elan>
- [2] <http://www.lat-mpi.eu/tools/lexus>
- [3] <http://www.ru.nl/sign-lang>
- [4] <http://www.mpi.nl/avatech>