ELAN: a Professional Framework for Multimodality Research

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, Han Sloetjes

Max Planck Institute for Psycholinguistics
P.O. Box 310, 6500 AH Nijmegen, The Netherlands
Peter.Wittenburg@mpi.nl

Abstract

Utilization of computer tools in linguistic research has gained importance with the maturation of media frameworks for the handling of digital audio and video. The increased use of these tools in gesture, sign language and multimodal interaction studies has led to stronger requirements on the flexibility, the efficiency and in particular the time accuracy of annotation tools. This paper describes the efforts made to make ELAN a tool that meets these requirements, with special attention to the developments in the area of time accuracy. In subsequent sections an overview will be given of other enhancements in the latest versions of ELAN, that make it a useful tool in multimodality research.

1. Introduction

Different sub-fields in linguistics that study multimodality phenomena have very different requirements with respect to the time accuracy of their annotations. At the MPI for Psycholinguistics intensive studies about the usage of gestures in particular communicative circumstances (Enfield, 2005), the synchronization between gestures and speech (de Ruiter et al., 2003), multimodal interaction [1] and the differences between different sign languages [2] are carried out. Collaborators from the University of Nijmegen are studying Sign Languages [3] interaction patterns. This research work is putting strong requirements on the flexibility, the efficiency and in particular the time accuracy of supporting computer tools. Gestures and Sign Language data in daily communicative situations is normally recorded with standard video equipment operating at a frame rate of 50 frames per second (i.e. 20 ms per frame). In all research where the timing relationship between gestures and signs and communicative acts have to be studied or where special signs in various languages have to be compared it is of crucial importance to rely on the timing accuracy of all components involved. A shift by one frame can already lead to wrong conclusions and to annoying and inefficient work. In a speech signal a shift of 20 ms can already point to another phoneme, when comparing signs a shift of 20 ms can already point to a different articulator movement. Further, it is of great importance for the research work to be able to compare different synchronized video streams, determine offsets for each stream and its annotations, easily find and visualize comparable patterns in different annotated recordings and to easily modify annotations. Although a large project at the Nijmegen university about bilingualism [4] is primarily interested in verbal communication, they also have similar requirements. This is mainly due to the partly small segments (phoneme level) that have to be coded and the many errors that are made during the coding process due to ambiguities. Therefore, the new version of ELAN (2.4.2) was developed in close collaboration with the involved scientists. Due to the strong requirements in terms of time accuracy we also asked SPEX [5] as an independent and neutral organization to test ELAN 2.4.2 carefully. In older ELAN versions we found out that the used components such as Quicktime [6] or Java Media Framework [7] in combination with the chosen MPEG codecs where not sufficient. Codecs differ in the way how they relate video and audio information during the encoding and decoding processes and software modules differ in the quality of the implementation and cause jitters and instabilities.

Even worse is the timing accuracy when rendering media streams across the Internet. Although the new web-version of ELAN called ANNEX [8] is receiving high acceptance due to its direct access to archival content without having to download and install software and data, ELAN will remain the tool for accurate and efficient work on local computers. ANNEX offers the additional functionality of easy web-based operation with normal HTML browsers. In the following figure the relation between ANNEX and ELAN is indicated.

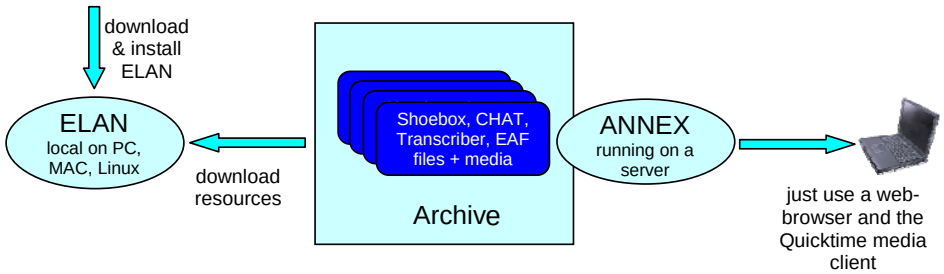

Figure 1 shows the relation between ELAN and ANNEX. While ELAN remains a highly accurate tool working on local computers, ANNEX is operating via the web.

2. ELAN Annotation Tool

ELAN is a linguistic annotation tool that was designed for the creation of text annotations for audio and video

	Elan < v 2.2		Elan > v2.2		MediaTagger (Mac)	
Obstr./vocal	**dev. out of 25**	**dev. in ms**	**dev. out of 25**	**dev. in ms**	**dev. out of 25**	**dev. in ms**
/f/	13	30	0	0	10	30
/t/	21	31	0	0	4	40
/b/	4	27	3	7	20	25
/k/	17	20	0	0	12	20
/k/	23	26	0	0	18	22

Figure 2: Excerpt from the SPEX test results

files of language use (Brugman & Russel., 2004). Annotations are grouped on layers, in ELAN referred to as "tiers". Tiers can have different relationships to each other, e.g. independent, aligned, or embedded.

From its very beginning ELAN has been an application available for Windows and Mac OS systems. Later a version for Linux was added as well.

This choice for platform independency in combination with the eminent need, from a developers perspective, for maintainability and portability, has led to the decision to use Java as the programming language. This choice comes with a cost, especially in the area of media handling performance. Since there is no pure Java, high performance media solution available, ELAN always had to rely on existing technologies that bridge the gap between a Java application and a native media framework such as QuickTime (with QuickTime for Java) or Java Media Framework. Both solutions having limitations with respect to accuracy and reliability.

3. Media handling in ELAN

3.1. Time accuracy

To overcome these shortcomings on Windows platforms we have developed a lightweight software layer that enables the integrated use of standard Windows DirectX [9] libraries in ELAN. Through this bridging layer ELAN can fully take advantage of the optimizations realized in the native media framework. As a result the overall performance, accuracy and stability improved significantly.

This was confirmed by the outcome of performance tests executed by SPEX, an independent organization specialized in speech and language technology. In these tests two versions of ELAN, the older JMF based version and the new DirectX based version, and MediaTagger (on Mac) were compared. Praat [10], the leading application for speech analysis and synthesis, was used for the segmentation of a speech fragment and it also served as the reference application in the test. The segments (n = 100) were played back repeatedly (n = 25) in the tested applications, resulting in a matrix of measurement values. Human perception played an important role in the evaluation of deviations of played segments.

The DirectX based ELAN version showed the best results with very few glitches and disturbances in segment play back. When the main media file is an MPG file, playback of short segments (less than 40 ms) is not always accurate (overshoot of 10 ms in 20% of the cases), but when the main file is a WAV file, playback of short fragments run without problems. Both the JMF based ELAN and MediaTagger displayed far more serious deviations in playback of fragments (Fig.2).

In ELAN it is possible to create annotations with a maximum precision of 1 ms. In combination with the accurate fragment play back capability, linguists can achieve the high precision in their annotations that they need for their research.

3.2. Media formats

For a long period of time ELAN's support was limited to the use of media files in MPEG-1 and WAV format. As of version 2.4 this restriction has been lifted which has meant another important improvement to the usability of ELAN in the mentioned fields of research. The underlying media framework, be it DirectX, QuickTime or JMF, now determines whether or not a certain file type can be used, adding enormously to the user's flexibility and freedom of choice.

The level of detail that can be achieved in MPEG-1 is often not sufficient for multimodal research, making support for MPEG-2 with it's high resolution increasingly important. Several codecs have been compared on their quality, most of which, unfortunately, have a few disadvantages for the system as a whole. Therefore at the MPI we are continuously searching for new and stable codecs that can be used in combination with ELAN.

To fully exploit the advantages of high resolution video, each movie can be detached from the main ELAN window into its own, resizable window, revealing any detail present in the video to the researcher.

3.3. Multiple video's

In the fields of research described in the introduction the use of multiple video camera's is becoming more and more widespread. Situations are recorded from different angles and from multiple distances. Details hidden in one recording often are visible in another. Close inspection of all available material improves the quality of research.

Therefore the number of video's that can be associated with an ELAN transcription first was increased from 1 to 2 and later from 2 to 4. By default one video will have the lead and will occupy the largest area in the ELAN window while the other videos will be displayed as thumbnails. A double click on one of the thumbnails will make that video the one occupying the largest area (Fig. 3).

Videos can be added to or removed from the document at any time, at any stage of transcription, providing the utmost flexibility to the user in the annotating process.

A special mode has been invented to allow for the synchronization of media files in case they are out of sync. A relative or absolute offset can be determined and stored per media file. This can be done even after annotations

have been created; there is an option to realign all annotations in one operation after changing the offset of media files.

4. Search options in ELAN

Searching is another area where ELAN has been extended considerably to meet the needs of researchers. Users can define several patterns to be matched on tiers based on temporal and structural constraints between them, resulting in complex structured queries. These queries can now be saved and reloaded again, for use in the same or in another document. The search results can be exported to a tab-delimited text file, that in turn can serve as input to applications that are able to perform statistics on the results.

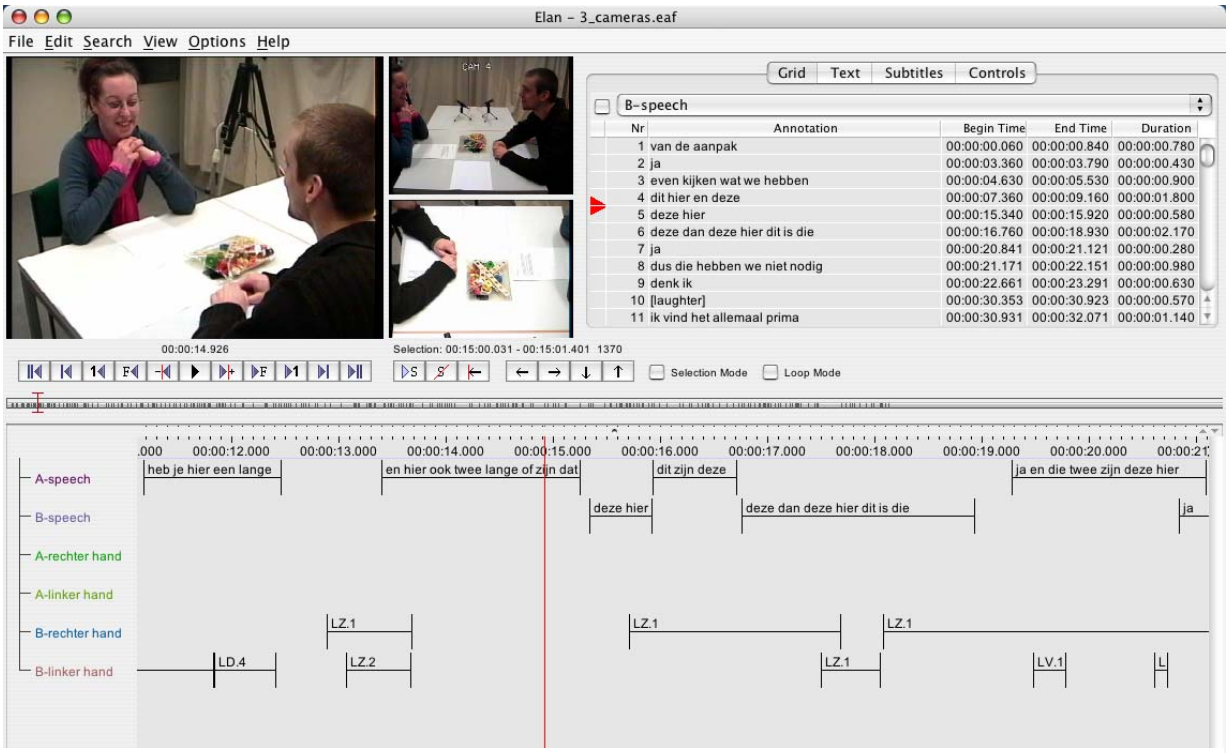

Figure 3 shows a typical screen layout for ELAN. Multiple synchronized video streams are embedded, different viewers for the annotations are available. In the lower part the time line viewer presents the annotations, created on user-defined tiers, time aligned with media time. In the upper right part the grid viewer renders the annotations of one tier in a more readable, tabular way.

This search mechanism works within a single annotation document at a time. Above that, searching can be extended to multiple files that the user can add to a custom search domain, consisting of individual files and/or complete directories containing multiple annotation files. Currently this type of search is limited to unstructured, free text search, however, in the next version full structured search in multiple files will be supported as well. Since no assumptions can be made on the number of files and the number of annotations in the domain, an indexing technique is applied to ensure high performance. An index with the appropriate information has to be created only once, as long as the files in the domain don't change. While searching in ELAN will be limited to EAF files, ANNEX which is a web-based utilization tool for a complete archive will support interoperability features, i.e. other annotation formats, such as Shoebox [11] and CHAT [12], will be included in the search.

5. Getting data in and out

5.1. Import and export

Linguists often prefer to utilize more than just one tool in the line of their research. Therefore within a project several annotation formats can co-exist. In order to accommodate a smooth and streamlined workflow, import and export modules for some of the main annotation formats have been implemented in ELAN.

The use of Shoebox/Toolbox and CHAT is widespread and these files can be imported as well as exported. In addition there are converters or import modules for Signstream [13] and Transcriber [14] files.

5.2. Printing and interlinear text output

Creating a full-fledged, multi layered transcription is a tedious and time consuming task. In addition to ELAN's default, archival EAF format, users like to be able to get their annotations out in other ways and in other formats.

Creating a hardcopy of the annotation file by printing on paper is a traditional and much requested output option. For this purpose a routine has been developed to transform annotation graphs (i.e. a groups of related and depending annotations) into blocks of interlinear text, where vertical alignment reflects the hierarchical relationships. This is a non-trivial process taking into account that several options for line- and block-wise

wrapping and for presentational characteristics are provided.

This routine is also used for interlinear text export, only this time vertical alignment is based on the number of characters in each annotation and is the result an editable text file.

6. Productivity

A tool's ease of use and its ergonomic qualities greatly contribute to the experience and appreciation of the every day user. Also the level of productivity that can be reached is of utmost importance. Several major improvements have therefore been made in this very area.

Semi automatic segmentation (creation of annotation with single keystrokes while the video is running), tier tokenizing (create new annotations on a depending tier for each token in annotations on a super-ordinate tier) and the use of user-specified vocabulary sets are just a few examples. Multiple undo and redo let the user reverse unintentionally made changes.

Many researchers develop coding systems in the process of annotation itself, which makes new options to change the relationships between tiers and to copy complete tiers very valuable.

7. Conclusion

All these new functions together with longer existing functionality and its tested time accuracy make the current ELAN version a professional tool to be used for the very time consuming manual annotation work for studying multimodal interaction. Its source code will remain Open Source and the latest version can be downloaded from MPI's web-site (www.mpi.nl/tools) together with an elaborate manual.

ELAN will continuously be enhanced, but it will remain a tool designed to work locally on personal computers, since media control via the network will not be as smooth and precise for some time. ELAN's functionality will be enhanced in parallel with ANNEX where possible. The following major extensions are planned for future releases: structured search on multiple files as mentioned, interaction with the LEXUS [15] lexicon component, connection to the ISO DCR [16] incorporation of time series data stemming from movement tracking devices such as eye trackers and cyber gloves, and new structural constraints between tiers.

8. References

Brugman, H. & A. Russell (2004). Annotating Multi-media / Multi-modal resources with ELAN
. In: *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*

Enfield, N. J. (2005). The body as a cognitive artifact in kinship representations: hand gesture diagrams by speakers of Lao. *Current Anthropology*, 46.1, 51-81.

J.P de Ruiter et al (2003) SLOT: A Research Platform for investigating multimodal communication. *Behavior Research Methods, Instruments and Computers*, 35 (3), 408-419

[1] http://www.mpi.nl/research/projects/multimodal_interaction/
[2] http://www.mpi.nl/research/publications/AnnualReports/MPI_AR04_print
[3] http://www.let.ru.nl/sign-lang/
[4] http://corpus1.mpi.nl/ds/imdi_browser -> DBD
[5] http://www.spex.nl/
[6] http://www.apple.com/quicktime/
[7] http://java.sun.com/products/java-media/jmf/
[8] http://www.mpi.nl/annex/
[9] http://www.microsoft.com/windows/directx
[10] http://www.praat.org/
[11] http://www.sil.org
[12] http://childes.psy.cmu.edu
[13] http:// www.bu.edu/asllrp/SignStream
[14] http:// www.etca.fr/CTA/gip/Projets/Transcriber
[15] http://www.mpi.nl/lexus/
[16] http:// www.cs.vassar.edu/~ide/papers/LREC2004-DCR.pd