

The use and interoperability of computer programmes Folker, ELAN and Praat for multimodal linguistic annotation



Silvia Bonacchi, Mariusz Mela
University of Warsaw
Institute of Specialised and Intercultural Communication
ul. Szturmowa 4, 02-678 Warszawa, Poland
s.bonacchi@uw.edu.pl; mariusz.mela@gmail.com
www.mcca.uw.edu.pl



Introduction

The joint analysis of the verbal and vocal display in speech corpora for the purposes of pragmalinguistic analysis requires specific annotation tools with high degree of interoperability (Schmidt et al. n. y.). In our project MCCA ("Multimodal Communication: Culturological Analysis", www.mcca.uw.edu.pl, in extended form: "Culturological and suprasegmental analysis of communicative interactions marked by (im)politeness") we have annotated speech data in Polish and German creating description layers (tiers) of the verbal and vocal displays (according to Sager, 2004: 123ff.) as relevant communicative behaviour in different cultural (national, situational, contextual) settings. The aim of the project MCCA is not only to transfer specialist knowledge and produce new knowledge about intra- and intercultural dialogue and mechanisms that disturb effective face-to-face communication, but also to develop standards of linguistic annotation for Polish data that would be compatible with international tools for the description and analysis of speech data.

For these purposes we chose to use the following programmes:

- PRAAT (version 5834_win64, www.fon.hum.uva.nl/praat, developed by Paul Boersma and David Weenink),
 - FOLKER (version 1.2., agd.ids-mannheim.de/folker.shtml, developed by Thomas Schmidt, Wilfried Schütte and Martin Hartung)
 - ELAN (Eudico Linguistic Annotator, version 4.7.3, http://www.lat-mpi.eu/tools/elan, developed by Han Slotjes)
- While working on the project, we have encountered some problems with the interoperability of these three tools, which we will describe. At the same time, we will indicate the desirable (from the point of view of users) (technical) solutions that could facilitate the work of linguistic annotators of multilingual data.

ELAN, Folker and Praat and their interoperability in speech analysis

The reasons for choosing ELAN as an "umbrella-tool" were the following ones:

- 1) it offers the possibility to integrate the suprasegmental and verbal analysis with a multimodal analysis due to its technical characteristics and a high degree of interoperability with other programmes for speech analysis – for example it is interoperable to a high degree with EXMARaLDA (www.exmaralda.org);
- 2) it offers a very high degree of flexibility when it comes to defining the tier-structures (annotation tracks or layers) and thus defining the levels of linguistic analysis;
- 3) it is an open system, which can also be used by students and for teaching purposes.
- 4) The conclusive reason for choosing ELAN for our project was that it was already used in other Polish research centers (for example by the Centre for Speech and Language Processing in Poznań for the DiaGest Corpus, http://cslp.wa.amu.edu.pl, a scientific unit which we work closely with).

INTEROPERABILITY OF FOLKER AND ELAN

A fundamental moment in the annotation work turns out to be the transcription of verbal display and its connection with tiers related to the description of other displays, for example annotation levels related to the use of voice, turn-taking, nonverbal behaviour etc. Even though transcriptions in other Polish research groups are carried out directly in ELAN, we have found it necessary to do the transcription as a separate step.

Folker proved to be a very good tool for transcription of German and Polish speech data. It was created for the transcription of German and it permits a good realization of the GAT2 standard conventions for Basic Transcripts (see Selting et al. 2009). GAT2 conventions enable one to refine the transcription by indicating prosodic and acoustic phenomena like loudness, changes in intonation, pauses, and the turn-taking dynamics. At the moment there is no binding convention system for transcription of Polish. There are several very good convention systems and software supporting them (for example AnnotationPro, developed at the University of Poznań, http://annotationpro.org/, see Klessa, Karpiński & Wagner, 2013), which were developed by Polish research groups, but the degree of their interoperability with current programmes for annotation is still very modest. Mostly the transcription (Fig. 1) has to be converted to a text-file and time-alignment to the audio file is lost.



Figure 1: Transcription of a Polish dialogue using the Excel programme (by courtesy of Piotr Pepek, Pelera-Project, University of Łódź, http://clarin.pelera.pl/Speakers)

We have found the GAT2 conventions to be well-suited to transcribing Polish. Just a few small adjustments to Folker would facilitate work with Polish speech data:

- Polish diacritic signs (ą, ę, ń, ś, ł) on the keyboard;
- the introduction of special signs to mark sentence disruptions and word truncations (anacoluthons), which are not marked by a repair initiation;
- a special mark up for repair initiation;
- a special mark up " _/_" for interruptions due to overlaps, as shown in the following Folker output as a segment list:

```
{00:19} 0011 MO do jakiej poprawy przepraszam?  
{00:21} 0012 AZ do (...) rekołek_  
{00:22} 0013 [_c]i patriotycznych ]  
{00:22} 0014 MO [czy mam czy czy mam]
```

We have encountered some problems while adapting the GAT2 conventions to Polish, among others:

- > different intonation patterns, which make the evaluation of default (not marked) values for utterances difficult (Karpiński & Klešta, 2001);
- > the lack of a consistent inventory of the so-called 'filled pauses', which are different in various languages (e.g. the German 'ähm' vs. the Polish 'hmm' or the English 'umm');
- > the annotation of non-lexical, quasi-lexical units and paralinguistic sounds (like clicks, percussives, see Karpiński, 2012).

Folker makes it possible to show moments where the contributions of speakers overlap (see Fig. 2), which helps a conversation analyst to notice, among other things, regularities and interdependencies between speakers (e.g. which of the speakers and when/how often interrupts his/her interlocutor), and so it offers a preview of the linguistic phenomena (see Fig. 3) that have to be annotated in the final tier-structures of ELAN.

A very important improvement for Folker could be the possibility to change some annotation-units automatically, for example through the "search and replace" function – this function is available only in ELAN. We would like to stress here that the modifications of the transcription layers do not result only from later corrections, but also from possible changes in the used conventions (for example for filled pauses, hesitation phenomena, repairs).

The program also offers the possibility of producing an output of the transcription in *.html-format as a segment list or score, as a compact score with audio player file (Fig. 4), as a contribution list, as a contribution list with audio player, as a GAT basic transcript and as a quantification.



Figure 2: Folker-transcription of Polish data (http://www.mca.uw.edu.pl/eng/naid-1.3.m/niesielowski-cate-bili-polije-zawisza-pan-powiedzial-juz-wszystko.288294.html), segments-view.



Figure 3: Folker transcription of Polish data in score view, which shows clearly the turn-taking dynamics with interruptions and overlapping contributions.



Figure 4: output as a compact score with audio player

At the end of the transcription work with Folker (see Fig. 2), we get a final transcription format (*.flk), which can be exported to different formats compatible with the current programmes for further annotation: as EXMARaLDA Basic Transcription (*.exb, *.xml), as an ELAN annotation file (*.eaf), as PRAAT TextGrid (*.textGrid), as F4 Transcript (*.rtf, *.txt), as an Audacity label file (*.txt), as a TEI file (*.xml), or as Plain text subtitles (*.txt).

After importing the Folker-transcription in eaf.format into ELAN, in which the scores of the speakers are displayed in separate tiers, we defined further annotation layers related to further levels of analysis (i.e. speech acts, vocabulary, types of sentences, PoS, voice, gestures, facial movements, etc.). To standardize the work of analysis within the research team we have created some templates (MCCA-StandardTemplates) for linguistic annotation (see Fig. 5 and Fig. 6)



Figure 5: ELAN-view including MCCA-StandardTemplates

INTEROPERABILITY OF FOLKER AND PRAAT

Folker files can also be exported as Praat-TextGrid files for further annotation. Praat makes possible not only the investigations of pitch, duration, intensity, intonation contours, characteristics of filled pauses, laughter, reductions, corrections, turn-taking etc., but it also allows a graphic visualization of them in Praat images and the extrahation of values related to the vocal realization (Fig. 6). Furthermore, Praat makes a fine phonetic transcription possible, which is fundamental for such parameters as duration, lengthenings and hesitation phenomena, which are very important in (im)polite communication (Bonacchi 2013).

INTEROPERABILITY OF PRAAT AND ELAN

Analogically, it is possible to export tiers from ELAN as Praat-TextGrid files, but it is not possible to import Praat-images back to ELAN. For the annotation of the vocal display (for the later suprasegmental analysis) you can create in ELAN description layers, but they are only descriptive. We would like to stress here the importance of developing in the future a user-friendly function of importing Praat-images into ELAN, in which we would be able to see, for example, the stress accent on the word "panią" in the Polish utterance "a Panią (.) wzywam do poprawy" (English: "and I call on YOU to change your behaviour!") (see Fig. 7).



Figure 6: Praat TextGrid file



Figure 7: ELAN-files with the Polish word "Panią" (English: "YOU" with a focus accent) exported to Praat for further investigation.

For further multimodal analysis of the conflict interaction shown below we have defined the following tiers: words, translation, conflict initiator, argumentative structure, vocabulary, man's signals, woman's signals, gestures, face, voice, and type of situation (see Fig. 7). The list is however not final and universal. Our ultimate aim is to develop a tier-structure for multimodal analysis (Bonacchi & Karpiński 2014) which would permit the preparation of compatible data in several European languages (see Bonacchi & Mela 2014, in print).



Figure 7: An example of a multimodal annotation in ELAN or the analysis of conflictual communicative behaviour.

Concluding remarks

The central problem addressed in our paper concerns the proper interoperability of Folker and ELAN and the relatively limited interoperability of ELAN and Praat. We have tried to show which implementations could facilitate our work as researchers and could be developed within the CLARIN-infrastructure.

The authors of the paper would like to draw the attention of software developers to the following aspects:

- The introduction of a unitary standard of transcription for European languages based on the Latin alphabet requires the creation of widely flexible conventions which could cover the variety of linguistic phenomena in various spoken languages. It should be supported by user-friendly IT-tools.
- A deep analysis of communicative displays, from verbal ones through prosodic aspects to nonverbal communication, requires interoperable IT-tools with a high degree of compatibility in the exporting and importing functions.
- The development of tools for annotation of multilingual corpora, in which the verbal layer can be easily exported for further elaboration (for example automatic translation or segmentation) and imported again to IT tools.

References

- Bonacchi, S. (2013). (Un)Höflichkeit. Eine kulturologische Analyse Deutsch-Italienisch-Polnisch. Frankfurt et al.: Peter Lang.
- Bonacchi, S., Mela, M. (2014, in print). Multimodal Analysis of Conflict: A proposal of a Dynamic Model. In I. Poggi et al. (Eds.), Conflict and Communication. Multimodal Social Signs of Conflict and Negotiation in Humans, Animals, and Machines. Berlin: Springer Verlag.
- Karpiński, M., Klešta, J. (2001). The Project of Intonational Database for the Polish Language. In St. Puppel, G. Demenko (Eds.), Prosody 2000. Poznań: Faculty of Modern Languages and Literature, UAM.
- Karpiński, M. (2012). The Boundaries of Language: Dealing with Paralinguistic Features. In Lingua Posnaniensis, vol. LIV (2)/2012, pp. 37-54.
- Klessa, K., Karpiński, M., Wagner, A. (2013). Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features. In D. Hirst, B. Bigi (Eds.), Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, pp. 51-54.
- Sager, S. F. (2004). Kommunikationsanalyse und Verhaltensforschung. Grundlage einer Gesprächsethologie. Tübingen.
- Schmidt, Th., Elenius, K., Trilsbeek, P. Multimedia Corpora (Media encoding and Annotation) (http://www.clarin.eu/system/files/clarin-deliverableDSC3_v1_5finaldraft.pdf)
- Selting, M. et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In Gesprächsforschung-Online-Zeitschrift zur verbalen Interaktion 10, pp. 353-402.

Aknowledgements

This research was supported by Research Grant UMO-2012/04/M/HS2/00551 from the NCN (Polish National Research Center) and based on collaboration between the Institute of Specialised and Intercultural Communication at the University of Warsaw, the Institute of Slavistics and the Institute of Computational Linguistics and Phonetics at the Saarland University.