

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

Silvia Bonacchi, Mariusz Mela

University of Warsaw

Institute of Anthropocentric Linguistics and Culturology,

ul. Szturmowa 4, 02-678 Warszawa, Poland

E-mail: s.bonacchi@uw.edu.pl; mariusz.mela@gmail.com

**Keywords:** interoperability, compatibility, spoken language, culturological analysis, suprasegmental analysis, multimodal analysis, annotation layers.

## 1. Description of the project MCCA

The computer programmes Praat (version 5834\_win64, [www.fon.hum.uva.nl/praat](http://www.fon.hum.uva.nl/praat), developed by Paul Boersma and David Weenink), Folker (version 1.2., [agd.ids-mannheim.de/folker.shtml](http://agd.ids-mannheim.de/folker.shtml), developed by Thomas Schmidt, Wilfried Schütte and Martin Hartung) and ELAN (Eudico Linguistic Annotator, version 4.7.3, <http://www.lat-mpi.eu/tools/elan>, developed by Han Slotjes) can be used in tandem for the purposes of pragmatolinguistic analysis of speech corpora. The above mentioned tools are used in our project MCCA (“Multimodal Communication: Culturological Analysis”, [www.mcca.uw.edu.pl](http://www.mcca.uw.edu.pl), in extended form: “Culturological and suprasegmental analysis of communicative interactions marked by (im)politeness”, supported by a grant from the Polish National Science Centre, UMO-2012/04/M/HS2/00551), for suprasegmental, multimodal and culturological analyses (see Müller, 1998; Ogden, 2006; Poggi, 2007; McNeill, 2005; Schmitt, 2005; Bonacchi 2013; Bonacchi & Karpiński, 2014), i.e. an analysis of vocal, verbal and kinetic displays (according to Sager, 2004: 123ff.) as relevant communicative behaviour in several cultural settings.

The project MCCA is based on collaboration between a Polish (the Institute of Anthropocentric Linguistics and Culturology at the University of Warsaw) and a German (the Institute of Slavistics and the Institute of Computational Linguistics and Phonetics at the Saarland University) scientific unit. Its aim 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 the annotation of the speech corpora we have chosen to use the ELAN computer programme for several reasons. Firstly, it offers the possibility to integrate the suprasegmental and verbal analysis with a multimodal analysis thanks to its technical characteristics and the high degree of interoperability with other programmes for speech analysis – for example it is interoperable to a high degree with EXMARaLDA ([www.exmaralda.org](http://www.exmaralda.org)), another programme we considered using. An important

reason for preferring ELAN as an “umbrella-tool” (for the creation of complex annotations for video and audio resources) to other programmes was that 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. Secondly, we wanted to work with an open system which could also be used by students and for teaching purposes. These possibilities were also available with ANVIL (Annotation of Video and Language Data, [www.anvil-software.de](http://www.anvil-software.de)), another programme we considered using. ANVIL interfaces well with Praat, because it allows the pitch contour and a waveform to be displayed. The announced compatibility in the forthcoming version of ANVIL with ELAN files will permit the user to switch between the programmes for optimized linguistic annotation and analysis. 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).

While working on the project, we have encountered some problems with the interoperability of these three tools, which we will describe in the poster. At the same time we will indicate from the point of view of users the possible (technical) solutions that could facilitate the work of linguistic annotators of multilingual data.

## 2. ELAN, Folker and Praat and their interoperability in speech analysis

A fundamental moment in the annotation work turns out to be the transcription of the display of verbal behaviour 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.

In order to have the transcription as a unique step, we have carried it out using the Folker programme, which makes it possible to modify the transcription at any moment during further elaboration and to read it again as a tier in ELAN without compromising time-alignment. 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). At the end of the transcription work, Folker files (\*.flk) can be exported as Elan-files (\*.eaf), in which a given tier refers to a given speaker. Folker files can also be exported as Praat-TextGrid files for further annotation (for example for the annotation of pitch, duration, intensity, intonation contours, characteristics of filled pauses, laughter, reductions, corrections, turn-taking etc.).

After importing the Folker-transcription into ELAN as separate tiers for every speaker, we defined further annotation layers related to further levels of analysis (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. 1)

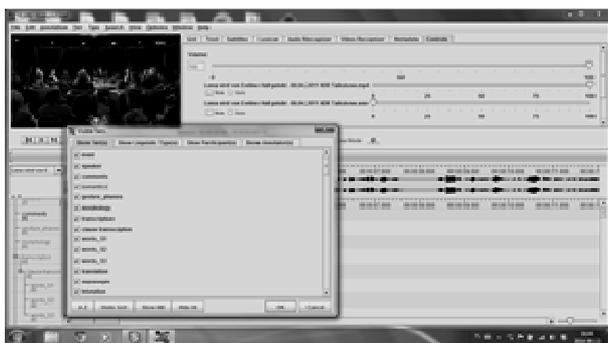


Figure 1: ELAN-view including MCCA-StandardTemplates

### 3. The annotation work steps in detail

For the transcription of speech data, we have used a system mainly based on GAT2 standard conventions for Basic Transcription (Selting et al., 2009), which has proved to be very suitable not only for German and Italian, but also for the Polish language. We chose to use GAT2 not only because it is a widely recognized convention, but also because it is compatible with the software used by us and because it is highly suitable for all languages we work with: German, English, Polish, and Italian. GAT2 conventions enable one to refine the transcription by indicating prosodic and acoustic phenomena like loudness, changes in intonation, pauses etc., and the turn taking dynamics.

The GAT2 conventions were developed for the German language, but they have also been successfully used for the transcription of recordings in other languages (English, French, Italian). We have also found them to be well-suited to transcribing Polish. Just a few small adjustments would facilitate work with Polish speech data:

1. Polish diacritic signs (ą, ę, ń, ś, ł) on the keyboard;
2. the introduction of special signs to mark sentence or word disruptions (anacoluthons), which are not always marked by a repair initiation;
3. We have also used a special mark up “\_ \_” for interruptions due to overlaps, as shown in the following Folker output as segment list:

```
{00:19} 0011 MO do jakiej poprawy przepraszam'
{00:21} 0012 AZ do (.) rekolek_
{00:22} 0013 [_cji patriotycznych ]
{00:22} 0014 MO [czy mam czy czy mam]
```

In the example above the sentence of speaker AZ „do rekolekcji patriotycznych” (“[I call on you] to become a better patriot”) was interrupted after the third syllable of the word “re-ko-lek-cji” (literally: “retreat”) by the speaker MO and concluded after her sentence “czy mam czy czy mam” (“should I should should I”) which indicates a high degree of fighting for interactional power through the control of turn-taking.

We have encountered some problems while adapting the GAT2 conventions to Polish or Italian. These problems included among others:

- different intonation patterns in various languages, which make the evaluation of default (not marked) values for utterances difficult;
- the lack of a consistent inventory of 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).

At the moment there is no binding convention system for the 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 modest. Mostly the transcription (Fig. 2) has to be converted to a text-file and time-alignment to the audio file is lost.

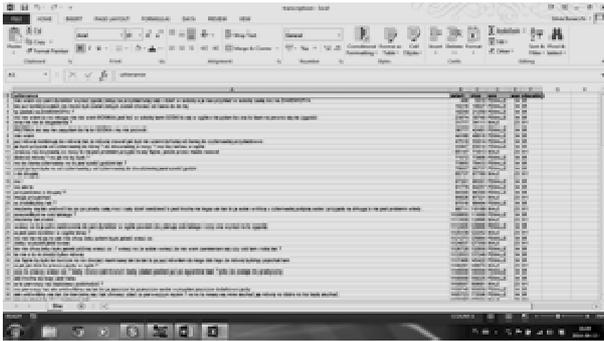


Figure 2: Transcription of a Polish dialogue using the Excel programme (by courtesy of Piotr Pęzik, Pelcra-Project, University of Łódź, <http://clarin.pelcra.pl/Spokes/>).

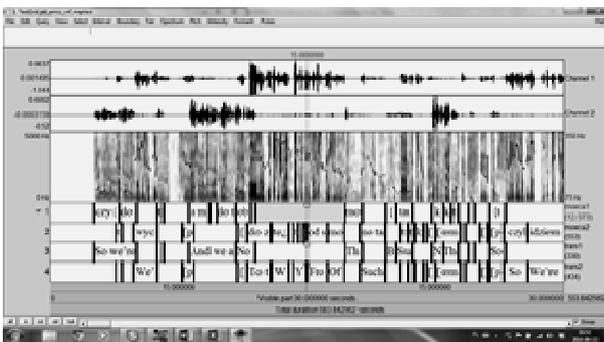


Figure 3: Transcription in PraatTextgrid (by kind courtesy of Maciej Karpiński, PoInt project, see Karpiński & Klešta, 2001)

At the end of the transcription work with Folker (see Fig. 4), 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).

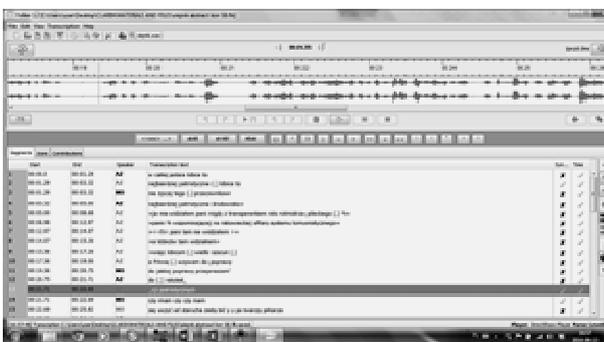


Figure 4: Folker-transcription of Polish data (<http://www.tvn24.pl/kropka-nad-i,3,m/niesiolowski-zulebili-policje-zawisza-pan-powiedzial-juz-wszystko,288294.html>), segments-view.

Folker makes it possible to show moments where the contributions of speakers overlap, 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. 5) that have to be annotated in the final tier-structures of ELAN.



Figure 5: Folker transcription of Polish data in score view, which shows clearly the turn-taking dynamics with interruptions and overlapping contributions.

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. 6), as a contribution list, as a contribution list with audio player, as a GAT basic transcript and as a quantification.



Figure 6: Output as a compact score with audioplayer

When exported as an ELAN file, the contributions in the Folker transcription are transformed into tiers for every speaker and additional levels of analysis (tiers) can be added. For the annotation of the vocal display (for the later suprasegmental analysis) we have added further tiers, but they are only descriptive. For a graphic visualisation of the vocal display we need to refer to Praat. It is used for measuring pitch, duration and intensity of voice and for showing intonation contour. 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. We would like to stress here

the importance of developing in the future the function of importing Praat-images into ELAN, in which we would be able to see, for example, the focus accent on the word “panią” in the Polish utterance “a Panią (.) wzywam to poprawy” (English: “and I call on YOU to change your behaviour!”) (see Fig. 7).

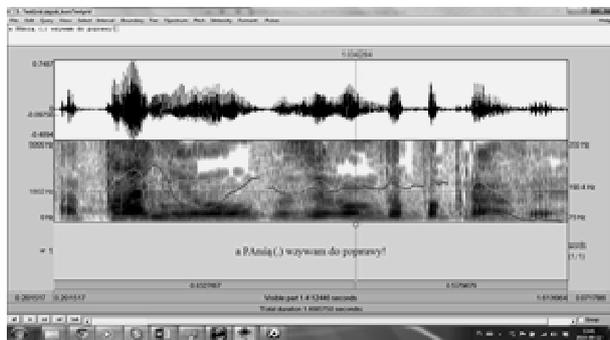


Figure 7: ELAN-files with the Polish word “Panią” (English: “YOU” with a focus accent) exported to Praat for further investigation.

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. 8). The list is however not final and universal. Our ultimate aim is to develop a tier-structure for multimodal analysis which would permit the preparation of compatible data in several European languages (see Bonacchi & Mela, 2014 in print).

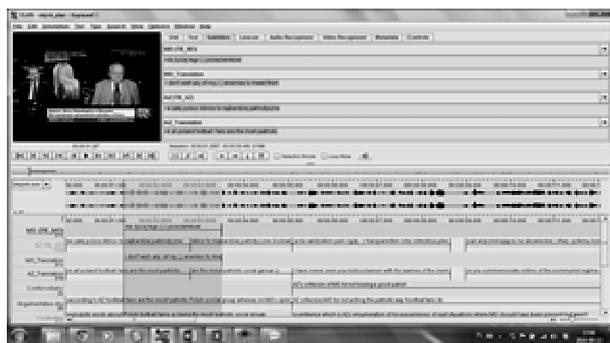


Figure 8: An example of a multimodal annotation in ELAN or the analysis of conflictual communicative behaviour.

#### 4. Concluding remarks

The implementation of the tools described above makes an thorough analysis of face-to-face-interactions possible. 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. For this aim the creation of widely flexible conventions

which can cover the variety of linguistic phenomena in various spoken languages is necessary.

- The advantages of more interoperable IT tools which can permit a deep analysis of communicative displays, from verbal ones through the prosodic aspects to nonverbal communication.
- The benefits of investigating multilingual corpora in order to develop second-order frameworks for the annotation, comparison and explication of communicative phenomena.

#### References

Bonacchi, S. (2013). (Un)Höflichkeit. Eine kulturologische Analyse Deutsch-Italienisch-Polnisch. Frankfurt et al.: Peter Lang.

Bonacchi, S., Karpiński, M. (2014). Remarks about the use of the term ‘multimodality’. In *Journal of Multimodal Communication Studies* 1/2014, pp. 1--7.

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

Müller, C. (1998). *Redebegleitende Gesten. Kultur – Theorie – Sprachvergleich*. Berlin: Arno Spitz.

McNeill, D. (2005) *Gesture and Thought*. Chicago: University of Chicago Press.

Ogden, R. (2006). Phonetics and social actions in agreements and disagreements. In *Journal of Pragmatics* 38, pp. 1752--1775.

Poggi, I. (2007). *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. Berlin: Weidler Buchverlag.

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-deliverable-D5C3\\_v1\\_5-finaldraft.pdf](http://www.clarin.eu/system/files/clarin-deliverable-D5C3_v1_5-finaldraft.pdf))

Schmitt, R. (2005). Zur multimodalen Struktur von turn-taking. In *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 6, pp. 17--71.

Selting, M. et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In *Gesprächsforschung-Online-Zeitschrift zur verbalen Interaktion* 10, pp. 353--402.