

Extending ELAN into Variationist Sociolinguistics

Naomi Nagy (*University of Toronto*) & Miriam Meyerhoff (*Victoria University Wellington*)

Prior to the implementation of ELAN, it was common for sociolinguists to use at least four different software applications, and consequently four file formats, along the route from recording participants to conducting statistical analyses of the data. We present a method which allows for transcription, extracting, coding, preparation for statistical analysis, calculation of some basic frequency statistics, and creation of a concordance all within one program, ELAN.

ELAN (tla.mpi.nl/tools/tla-tools/elan, Wittenburg et al. 2006) has established itself as a valuable tool for language documentation and is frequently used for transcription and multi-tier mark-up illustrating levels of linguistic structure as well as translations and glosses. In this workshop, we illustrate an extension to its utility: extracting and coding tokens of linguistic variables for quantitative analysis in the variationist sociolinguistic framework. This approach improves consistency, reliability and the accountability of our coding to the original recording. In this workshop we illustrate the following benefits:

- seamless connections between recording, transcript and coding of dependent and independent variables. This facilitates revision and intercoder reliability tests.
- reuse of contextual factor coding (e.g., style, topic, interlocutor) as well as some structural (morphological, syntactic) tags
- wide-ranging exportability (to, e.g., R, Rbrul, Praat, Goldvarb, SPSS, and forced-alignment and automated formant extractor programs)
- importability of transcripts from Word/text files and many other transcription formats (e.g., text, Transcriber, Shoebox/Toolbox, CHAT, Praat ...)
- complex searches and concordance capabilities to speed up token “extraction”
- archivability of mark-up related to each data file in a consistent and small-file-size format
- smoother workflow for collaborative research teams and shared data

Our demonstration will illustrate ELAN’s functionality for variationist research with ongoing work on variation in subject pronoun presence in Faetar and N’kep, endangered languages spoken in southern Italy and Vanuatu (respectively).

The hands-on portion of the workshop will allow participants to code a brief excerpt in English for a well-known sociolinguistic variable and/or work on their own data.

Participants should bring a laptop with ELAN installed (download from <http://tla.mpi.nl/tools/tla-tools/elan> – Mac, Windows and Linux versions available) and an (unzipped) copy of the workshop files, available at http://projects.chass.utoronto.ca/ngn/pdf/NWAV2014_ELAN.zip. It will be helpful to have a spreadsheet program such as Excel installed as well.

Potential participants should note that this will cover the same content as the NWAV 2013 workshop, though all are welcome for a “refresher.”