

Increasingly, as field linguists we are called upon to be experts in a range of disciplines: linguistic documentation and analysis, data digitization and archiving, and the production of language products that are directly useful to other linguists and to speaker communities in their language pedagogy efforts. Furthermore, we feel responsible for ensuring the long-term preservation of our work by following good practice recommendations and protocols by expert panels like E-MELD and OLAC. In order to decrease the burden on field linguists to be conversant in an ever-widening range of technologies and standards, a most pressing need is for suites of user-friendly interoperable digital tools that seamlessly integrate related phases of linguistic work.

The need for interoperability is exemplified in the realm of annotated text, audio and video corpora. Here we have the potential to develop an integrated workflow for three steps in language work: analysis, archiving, and community-oriented presentation. Currently available tools require distinct manipulation of working, archival, and presentation versions of digital language data. For example, for transcription and analysis (working version) we need a tool like Shoebox or Transcriber. For producing archival alignment files (archival version), we can use Elan to create a manipulable XML file. Finally, for producing user-friendly language products (presentation version) we rely on a wide variety of technologies, including Flash, HTML, and XSLT—or we may even contract the work out to editors or professional software developers.

Present technologies require a significant commitment to self-training on the part of the linguist: individual tools must be learned thoroughly enough to be useful, and transporting data between them often requires knowledge of other software or hacks. Linguists must also keep abreast of progress in available cyberinfrastructure and standards of long-term data preservation. Many linguists find such a commitment daunting, and may be tempted to instead use more familiar “less-than-best-practice” procedures (Salffner 2006).

An ideal suite of tools would be designed to keep the burden on the individual linguist to an absolute minimum. Such a suite would be easy to learn and use, making use of similar user interfaces between programs. Files would be easily transferable between tools. The suite would be cross-platform and open-source, and would include instruction manuals written in clear language addressed to beginning users. It would be based on general good practice standards to ensure long-term accessibility, and would facilitate archiving of language materials. Resultant products would be customizable and easy to use. Finally, the suite would be flexible enough to keep up with changing standards, as the field of computer-assisted linguistics matures.

Reference:

Salffner, Sophie. 2006. Software tools in linguistics—a user survey. Presentation at the Leipzig Students’ Conference in Linguistics, March 25.