

# Digital Tools Summit for Linguistics: Issue Statements

Thorsten Trippel, Bielefeld University, Germany, thorsten.trippel@uni-bielefeld.de

From my perspective there are two basic issues in digital tools and cyberinfrastructure development in linguistics namely the interoperability of existing tools and promoting the (correct) use of existing tools. This starts from the premises that though tools for almost all wanted levels of annotation, archiving and description of resources are available. That does not mean that the existing tools are seen as perfect, but the development of these tools is on a good way, details of individual tools need to be enhanced for stability, usability, and the Seven Dimensions of Portability introduced by Bird and Simons for long time access to resources. If this is true and the tools are available then we have to ask ourselves why there are not more resources available in a good or even best practice format, being well documented. If this hypothesis is true then there are only a few possible reasons: first people are tool illiterate, second linguists just do not care for the use of these tools because pen and paper seems to be the best way and third even if linguists work with one piece of software, they cannot interact with software used by others.

## ***Software Illiteracy***

Due to the electronic revolution, linguists in the so called developed world are able to use a computer and even train themselves in the use of new software such as a new version of their favorite office package. However, this seems not to be the case for linguistic tools. Colleagues working on transcription of English radio speech still use tape recorders and when saying they are using a computer for transcription they talk about a text processor. Hence software illiteracy may be an issue with some but not as a general problem.

## ***Ignorance of the need***

Some linguists do not yet believe in archiving their data for future reuse or reuse by others. They even believe that a quick paper transcription, or annotation in a text processor is faster and more efficient than an electronic version. This perspective, as shortsighted it may be, is rather common and a real problem as the argument may be valid. For one time use the tool they are most familiar with may be the most efficient, but the same is not necessarily true when using larger amounts of data. However, it is not clear if an existing annotation of some special example will ever be wanted by someone else. But if it is not available there is no choice. The task will be addressed by funding organizations pressing the linguists to use archiving methods, however, the archive may be worthless if an annotation is available as a picture or in an incomprehensible text processors format, hence lost as much as a dirty piece of paper.

## ***Incompatible tools***

Using different annotation tools can be an issue of personal preference and a matter of taste, next to the selection of a tool for hard reasons such as special features of one product needed for a certain task. However, using the different annotation tools does not result in incompatible data, but someone has to sit down and write a conversion program. The fact that this is possible, shows that the tools could work together. though this process may involve additional logic or default insertion as well as losing some information in the process. Current standardization initiatives work on annotations schemes, data categories, ontologies, etc. (see the International Organization for Standardization working group on Language Resources, ISO TC 37 SC 4), but do not work yet on the interoperability of tools with generic formats based on the existing tools. The software programmers on the other hand do not consider it as their users file them with feature requests, trying to put a feature seen in another software into their favorite.

From my perspective the workshop needs to work on a marketing strategy for the use of tools and on a marketing strategy on creating standards and forcing developers enable interoperability from their tools