



Friday October 26 **Reaching out to the users**

We have to make a real effort to find new users, convincing them to participate and train them; especially since a huge amount of digital data are becoming available. Also changing their attitude: there is still the lexicographer's angst to deviate from the traditional approach of manual extraction of excerpts from agreed sources, manual sorting of cards, etc.

We could develop a European user survey, possibly web based. In the CLARIN handbook should be practical, make possible DH workflows transparent. Target: "progressive" humanities researchers, students.

Many humanities researchers are hesitant or even afraid of pragmatic changes in their workflow. Many of them are largely ignorant as to how NLP applications could help them. We can only educate them with regard to these applications.

Some good lexicographers searched corpora automatically with very old tools. They feared the moment when their working routines wouldn't work anymore with new hard- and software.

In CLARIN-D a list of 100 questions to which users expect answers, was compiled.

In Poland task analysis for tools has been designed: tools are used by specific user groups for specific purposes. The problem was that the tool was treated as another view of the topic; algorithms and annotations had to be explained beforehand, and accuracy and completeness as well. People didn't trust the tool for that reason. We shouldn't come with speculative, risky high-edge research but focus on basic tools we have a lot of experience with. On widely needed tools and their combinations, and which we can make trainable for adaptation. Small use cases based on concrete research questions make researchers curious. Work bottom-up instead of top-down. Show researchers that this tool gives them better research. We have to keep in mind if our solutions correspond to their problems, if the tools offered by language and speech technology is an answer to the questions of HSS scholars as they see them.

Technology doesn't solve all answers, but it already improved research a lot. Make the researchers partners in the struggle for improvement. Focus on the additional value provided already.

There are portals in which research questions are already answered. They may be less good, but CLARIN could help them technically and could find experts who know what researchers want. This way we could find potential users. What about ownership? We are basically technology-providers. We need to change the attitude and provide more service and don't leave the researchers to deal directly with all tool providers. A gearbox alone doesn't solve problems. Look at small accessible tools instead of big shiny solutions.

Branding: if we want to provide solutions, we have to market it with the CLARIN corporate identity, instead of individual institutions. We have a lot of tools which are good for very specific purposes. If we want to reach out to the end users as customers, we have to provide services.

In CLARIN-D there are a lot of tools and models based on a variety of program languages, but with our interface, students trained in 90 minutes can create their own annotated corpora. The way from visualization to interactivity to user interfaces is very complicated. Point of departure: what can be done with users' data. This is the first step. People notice that their data are slowly disappearing.

We have to be careful: CLARIN was successful because of its clear focus on language data.

Branding of CLARIN, by a 4 minute YouTube film, which explains in a simple way to an unknown audience the use of a tool. Or 30 seconds teasers as a summary for the 45 minutes lectures on metadata etc. Subtitles not only in English and German, but in more languages would be useful. We want to impose things from the international level, but should also keep in mind that the humanities are very locally and nationally oriented. We should make use of the national projects to interact with their national humanities associations. Go to associations of translators, historians, art historians, psychology, etc. and present ourselves as a keynote speaker on their annual meetings. Annual meetings are the must-be-there for researchers. This is a good way of popularizing CLARIN.

It is not that hard getting to newspapers but we have to make an effort. They are not interested in long stories but in news. CLARIN NL published the launch of their 18th century letters during the national science week, and got a lot of exposure that way. You have to look for those opportunities and we need more specially trained PR people for that.

At the Dutch Institute for Lexicology guest lectures on language resources are given, and they could do the same in CLARIN: lectures and tutorials for students, lecturers, and researchers.

We can combine local portals with the portals of other countries.

CLARIN D is planning a PhD training session with the agenda being set by the students themselves.

Two different camps: users vs. providers. Has anyone written a cross-disciplinary paper? E.g. corpus of causally spoken Polish. Paper: how do women vs. men use the word "home"? Psychologists and computational linguists could cooperate.

In Austria, at the beginning there was not much to see. By word of mouth CLARIN became known in the communities, bottom-up with gatekeepers in the communities. Multidisciplinary teams are bottom-up and successful. Research platform: people hook up their proposals and CLARIN-AU gives support.

We could put in place an annual award of e.g. € 1000 for the best cross-disciplinary paper.