Interview Andrea Rapp - TextGrid

5-7 minutes

Prof. Dr. Andrea Rapp on Tools, the Added Value of Networked Data and the Purpose of Quality Control

What does your center contribute to TextGrid?

Rapp: Here in Trier, we develop the essentials, the foundational elements, for example the simple, direct access to the software – the interfaces that you see when you access TextGrid. We also develop tools for specialists, such as tools that can be used to evaluate dictionaries: We have digital dictionaries in which individual parts of entries, such as "keyword", "grammatical description", or the "explanation of meaning", are marked up. As a result of this, one can search for these items in any of these dictionaries. We are planning to link together different dictionaries, so that a user can switch from one dictionary to another in order to look for a particular word in Middle High German or Old High German or in different dialects. In order to consistently find the correct word, you need tools that link the appropriate passages with each other.

Could you give an example?

Rapp: Take the multiple names for the German word "Brombeere" (blackberry, bramble): everybody knows the plant, but it is known by different words in every dialect: "Schwarzbeere", "Dornbeere", "Maulbeere", etc. If I do not speak a particular dialect, I cannot search for the word in that dialect, but when the dictionaries are interlinked with one another, I can conduct a search using just one of the dialectical forms and nevertheless retrieve all the alternative forms too. In addition to single words, we also want to mark up phraseologisms. They are often used in dictionaries but they cannot be searched for as keywords. Take, for example, a fixed phrase such as the German expression "vor sich her". I will not find an entry for it under "vor", "sich", or "her" in the dictionary. If I have marked the expression a common construction, a praseologism, then I can see how it is used in the dictionary, and we can directly evaluate and utilise dictionaries that contain a lot of previously unused information in their entries.

Isn't it still necessary to have someone who recognizes the relationship as such and marks it up?

Rapp: Ultimately, if something is to be precise, a person must make the decision. The computer can help by making suggestions, however. It can mathematically determine that certain words very often occur next to each other or in the

immediate vicinity, and then a person can review the proposed lists and decide which cases are indeed fixed idioms. We are now developing tools that support researchers by taking over routine tasks such as searching large amounts of data and statistically evaluating the results.

Is TextGrid just a gigantic search engine, then?

Rapp: No, it's more, because the information contained in it consists not just of random results, but is philologically sound, abstracted and of very high quality. And then there's something that only occurred to us over time: dictionaries in and of themselves contain text, that is, data. If we don't just access this data by reading it, but enhance it, then it can become a tool by itself. In working with a literary text, for example, we can ask whether or not a particular word is specific to this writer, if it is limited to a particular region or is very widespread. The networked, enhanced dictionaries can give me answers to these questions, and so data generated by one researcher will lead to new tools for another. We need critical mass, however. Not a lot can happen with one dictionary, but if you suddenly have ten or twenty available that can be linked, then their value increases exponentially.

Will there be quality control for the data ingested in TextGrid?

Rapp: This is a delicate issue: To what degree should fellow

researchers be supervised and monitored? This has certainly not yet been fully worked out. If someone has been identified as a researcher and a member of a university or an academic institution, then in my opinion their research should be facilitated. If I think that an edition is bad, should it be kicked out of TextGrid? Bad editions are not removed from the library. The community itself must debate it and react to different quality levels – perhaps through user numbers. Of course, our license will contain a clause prohibiting copyright infringement or the dissemination of politically dangerous content, but there are researchers who do legitimate research on hateful or anti-Semitic texts. Their data needs to be made available as well. This is no different from printed dictionaries: some parts of the Grimms' dictionary, for example, were created during the Nazi era. You cannot just remove entries that reflect the perverted Zeitgeist of this period. Their original context, and how they should be evaluated, just needs to be pointed out very clearly.

Interview by Esther Lauer.

[Back to the press kit …](#)