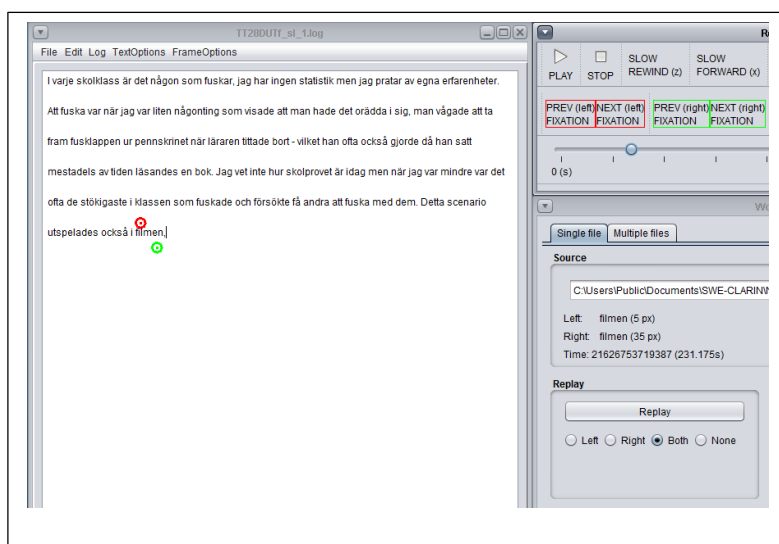Lund University Humanities Lab is a state-of-the-art research infrastructure and training facility financed by the Joint Faculties of Humanities and Theology (JFHT) at Lund University. The Lab has, among other things, equipment for eye-tracking, articulography, motion capture, virtual reality and EEG. The Lab provides access to and training in the use of various types of technology to the Humanities, among them language archiving technology and training in corpus and data management.

At the Humanities Lab, we develop ScriptLog, a tool for writing research. With ScriptLog it is possible to log the writing process while you write. This can then be played back and analyzed afterwards. We are also able to add eye-tracking to see which part of the text that the writer is looking at a particular time. The figure shows the program during a playback of a text plus eye movement measurement where the writer just fixated on the word 'filmen'.

ScriptLog is developed in connection with the Swedish department at Gothenburg University and the Department of Human Sciences at the University of Kristianstad. It is used in several projects within linguistics and speech therapy as well as in primary and secondary schools.

The Humanities Lab hosts a large number of projects that generate data in many different forms. The Lab provides and maintains a corpusserver, https://corpora.humlab.lu.se/, where different projects can store their data. An important part of describing the content and/or structure of a particular data collection is metadata. Efforts are ongoing to add material from different research groups and projects, and to contribute to the annotation and metadata description of this.

Metadata is done in the CMDI model, developed within CLARIN centrally. To illustrate the material on the server there is also a corpus browser, which enables search and display of metadata for different types of materials. The figure shows a portion of a corpus of the language Khmu, spoken in Laos, as well as how it is annotated with the tool Elan.

Swe-Clarin's first HSS workshop

Swe-Clarin arranged its first workshop for the Humanities and Social Sciences (HSS) on Friday April 17. The event was open to all but mainly targeted researchers from the HSS disciplines at University of Gothenburg. About 50 people attended the event. The workshop consisted of three sessions: the first concerned existing pilots and projects within the Swe-Clarin consortium, the second concerned material, existing and planned, and the third session was an open discussion.

During the first session Jonas Lindström, Uppsala University, discussed the uses and usefulness of language technology to the *Gender and Work* project in a presentation entitled "Peering into the darkness". The project uses a large historical text material, mostly handwritten court records, and searches for mentions of men's and women's activities. The project is a collaboration between historians and language technologists. Linn Sandberg, University of Oslo, presented a project in election and public opinion studies based on Twitter data, in a collaboration between political science and Språkbanken at University of Gothenburg, and computer science at Chalmers. The project aimed at using machine learning to investigate which topics were discussed on Twitter during televised election debates before the Swedish elections. After that, Leif-Jöran Olsson, University of Gothenburg, informed us of the Swedish Drama Web and its potential as a well-annotated material for research in humanities and also as training and test data for machine learning of professions and roles. Finally, Nina Tahmasebi, University of Gothenburg, presented three projects: on relation extraction, on Swedish pseudo-coordinations, and on a study of the concept of rhetoric over time.

During the second session, existing and planned material were presented and discussed. Lars Borin introduced the corpus material at Språkbanken and its potential for HSS research. Jonas Engman described the Nordiska museet ethnographic questionnaires and ceremoniously presented the material to Språkbanken. Anne Palmér and Annika Hillbom discussed a corpus of student texts under development at Uppsala University and asked for feedback regarding the design of such a dataset and potential research questions. Finally, Lars Björk discussed a future project at the Royal Library for digitization of a large collection of modern literature.

During the discussion session, topics like material, collaboration between with the libraries, and possible research questions were discussed. There was a request for a financial corpus, for example consisting of annual reports for social science and economic research. Swe-Clarin was also asked to provide basic vocabulary for language technology to aid HSS researchers for future discussions with the LT community.

The workshop ended with drinks and snacks and many participants stayed and carried on with discussions also after the scheduled sessions were over. All in all, it was a successful workshop and we look forward to the next event sometime during fall of 2015.

Calendar

8-10 June: Fonetik 2015, Lund,
http://konferens.ht.lu.se/fonetik-2015/
9-10 June: Nordic Clarin Network (NCN) Workshop.
5-6 October: NCN Workshop in conjunction with the Language Bank's autumn workshop.
11 November: SND's autumn workshop
19-20 November: Swe-Clarin general meeting in Stockholm, Swe-Clarin Workshop enclosed Friday afternoon. Everybody is welcome.

Partners

Swe-Clarin has nine partners from Lund, Gothenburg, Linköping, Stockholm and Uppsala, both in universities and public authorities. A list and description of all partners may be found here:
http://sweclarin.se/swe/centrum

News

We will not go on spamming you. Should you want more info on Swe-Clarin, please sign up for the news list here:
http://lists.sweclarin.se/mailman/listinfo/news_lists.sweclarin.se