Thesis proposal

A translation researcher's search tool

Suitable for a Bachelor's or Master's thesis (15 or 30 credits).

Background

Translation studies is the common name for the scientific study of translation. When the researcher's interest is on the translation product, or the strategies employed by the translator, the traditional methods are based on close reading of the texts – source text as well as translations. More recently, search tools and parallel concordance generators, have been developed to support translation research, the best known commercial tool probably being ParaConc (http://paraconc.com/). There are also some web-based systems that allow search in open corpora, such as Bilingwis (http://kitt.cl.uzh.ch/kitt/bilingwis/) and Paraquery (http://corpus.leeds.ac.uk/paraquery.html).

Project description

The goal of this project is to design and develop a system for use by a translation researcher who has a translation corpus at her disposal. The corpus is assumed to be XML-formatted according to some standard and aligned at the sentence level. It may or may not be word-aligned, tagged for parts-of-speech, or lemmatized. The most important functions that the tool should support are the following:

- selection of corpus sections based on metadata (including id:s)
- full text search in source and translations, and combined
- display of search results as parallel concordances
- display of frequency information for matches, including translation/source alternatives
- annotation of single and parallel sentences to be saved as metadata
- search based on the researcher's annotations, also in combination with in-built search mechanisms

Student profile

You should have taken a course in user-interface design/interaction design, preferably a course in language technology, and be able to program in Python, Java, PHP, or some other suitable language.

Contact

Lars Ahrenberg (lars.ahrenberg@liu.se)