Report on my Internship at the LATL

Garance Paris

1st August 2002 – 13th September 2002

I started out looking for an internship in autumn 2001, when I was in my 7th semester. At first I had no idea where to look, but then I found the website of the Computational Linguistics Department at the University of Paris 7, on which the department's counsellor puts any ads which are sent to him. There were three or four ads on the page at the time I first checked it, so I wrote an application letter, and sent my CV along with the letter to the three or four email addresses mentioned on the page.

Not all of them replied, and some of them also wrote that they already had found someone for the job. The first to give me a positive answer was C. Walther-Green from the LATL, the "Laboratoire d'Analyse et de Technologie du Langage" (Laboratory for Language Analysis and Technology), which is part of the Department of Linguistics at the University of Geneva. They were looking for two interns for the summer of 2002, which were to participate in the FreeText project.

FreeText is a project to develop an advanced hypermedia CALL (computer assisted language learning) system featuring NLP tools for an intelligent treatment of authentic documents and free production exercises. It is funded under the "Information Society Technologies Programme" (IST), which is part of the European Commission's "5th Framework Programme", an EU-funded R&D programme. The IST programme aims at promoting a user-friendly information society, by favouring life-long, remote training, multilingualism, and hypermedia mastering in the European Community.

The software will incorporate authentic documents and various sentence processing tools. Thanks to an efficient and robust error diagnosis system that incorporates error typologies based on statistical French learner corpus analysis, FreeText will be able to handle open-ended answers to exercises, which will be automatically corrected by the system. The system will also include reference aids such as a linguistically motivated reference grammar and an evaluation tool. The complete system will be available both on CD-ROMs and via Internet/intranets.

Among other NLP tools, the system integrates a sentence analyser, from which have been derived a sentence structure viewer and a error diagnosis tool. The structure viewer offers two possible displays: a tree display, and a "colour grammar" display, which colours the words of a sentence according to their part-of-speech categories and also underlines the various functions of the constituents in a sentence.

My job was to test the sentence analyser, by looking at the output of both sentence structure viewers. This involved the following:

- building a test corpus from the authentic documents which will be used in the software;
- entering all words unknown to the analyser into the lexicon, complete with their syntactic and semantic features and their phonetic transcription;
- checking the output of the sentence structure viewers;
- whenever errors were apparent in the sentence structure viewers, finding out at what level the error had occurred (parsing error, or error in the XML output computed from a correct analysis, or incorrect display although the XML representation was correct) and writing a bug report.

Overall, I would not say that I have learned much during this internship, but it was very nice to meet the enthusiastic linguists and computational linguists of Geneva University's Linguistic Department, with whom I had some interesting conversations, so that in the end, I guess one thing makes up for another.

More information about the LATL and the FreeText project can be found on the following websites: `http://www.latl.unige.ch/` and `http://www.latl.unige.ch/freetext/index.html`.