Year 1 report Multilingual FrameNet database

Hans C. Boas

The first year of the project was spent on compiling a list of off-the-shelf resource that can be used for extraction of data from corpora, annotating extracted data, and storing them, and to produce a proof of concept. This initial phase led us to download the English FrameNet database from the International Computer Science Institute in Berkeley, CA (http://framenet.icsi.berkeley.edu), to learn more about the structure of lexical entries and other technical aspects of the MySQL database.

Next, we decided to test the methodology of multilingual annotation of corpus data by focusing on the Personal_Relationship frame, which describes relationships between different people with nouns such as father, sister, and sugar daddy, or verbs such as to date, to love, to see, etc. Working with these data, Judith Atzler completed her dissertation in May 2011 which looked into using Frame Semantics as a tool to help improve vocabulary learning in second and third semester German instruction. Another frame was the Theft frame, which is evoked by nouns such as thief and stealer and verbs such as to pickpocket, to steal, and to shoplift. Working with these data, Ryan Dux finished his M.A. thesis (May 2011) on verbs of stealing in English and German, focusing specifically on how the semantics of the same underlying frame is realized differently at the syntactic level in the two languages.

During the spring 2011 semester we also set up the infrastructure to extract corpus data from different resources in order to annotate them using e-Comma, an in-house annotation tool developed at UT Austin. During the summer 2011, Ryan Dux spent a considerable amount of time extracting lexical entries from the Personal_Relationship frame (together with corpus data) (with the help of COERLL staff) from English FrameNet in order to help him find comparable German corpus data in a variety of resources. With the extracted German corpus data, Ryan annotated relevant example sentences that exemplify the use of the Personal_Relationship frame. COERLL staff programmed a small database in which the annotated examples sentences can now be viewed right next to the English examples, thereby serving as a tool to help students understand how the semantics of the frame is realized similarly and differently in English and German.

At the end of the summer we evaluated our results and were able to determine that although the information displayed by such a resource is useful it might not be of direct help for immediately improving a student’s learning of new vocabulary. To remedy this situation we decided to spend year 2 of the grant period on developing some more concrete deliverables. Specifically, we are planning to annotate one or two texts from Deutsch im Blick (the German textbook used at UT) with semantic frame information and to then devise fill-in-the-blank exercises that will help students learn new vocabulary more efficiently. We will spend the fall 2011 semester with
annotating the texts and developing exercises. In the spring 2012 semester we will use these new teaching materials in German classes at UT to conduct a study about their effectiveness.

Completed theses and dissertations:


Presentations based on the project so far:

Boas, Hans C. 2011. “Neuere Ansätze für eine Grammatik des Deutschen” (New approaches towards a grammar of German), Institut für Deutsche Sprache (Institute for the German Language), Mannheim, Germany, January 2011.

Ryan Dux and Hans C. Boas “On categorizing semantic roles and frame elements: What is the proper level of granularity?” Workshop on semantic roles, University of Zürich, Switzerland, April 2011.


Publications based on the project so far:

