Corpus - a tool in grammatical research

Rafał L. Górski (Polish Academy of Sciences, Poland)

A wide use of corpora is made in several subfields of linguistics, still it seems that in the study of grammar this use is less extensive than, e.g. in lexicology or stylistics. This picture gradually changes, there are more and more corpus based studies focusing on grammar, as well as an international conference (Grammar & Corpora). There is also an increasing number of publications devoted to theoretical and methodological issues of corpus based grammatical research.

The aim of the workshop is to show a possible use of corpora in grammatical research. As an introduction we shall discuss some theoretical problems. First of all, we shall pinpoint questions in which a corpus can supply information not accessible by an examination of linguistic competence. Also the difference between corpus based vs. corpus driven linguistics will be explained. These theoretical considerations will be then illustrated by four case studies:

1) An issue on (Polish) morphonology
2) Variation in (Polish) morphology
3) An analysis of an (English) particular construction
4) An alternation of two (English) synonymous constructions

The motivation of the choice of the topics is to show different theoretical approaches to the study of the language with a corpus, as well as topics in various subfields of grammar. We will analyse problems which can be solved with a scarce knowledge of the language (as in case 1 and partly 2), and tackle questions belonging to morph(ono)logy and syntax. We shall also focus on two important factors concerning the choice of a particular grammatical pattern, that is, lexis and information structure.

Participants of the workshop will be given hands-on experience in practical analysis of grammar using corpus evidence. They will have a chance to "discover" by themselves some rules underlying the variation between forms and structures in the language, even though (due to lack of time) we will work also on prepared concordances and data.

The case studies will be based on the BNC[1] and the National Corpus of Polish.

Note that we shall not cover two important topics: the query syntax and the statistics. The topics are broad and would require separate workshops.

---

[1] Participants are asked to register for an account on BNCweb http://bncweb.lancs.ac.uk/bncwebSignup/