Descriptive Grammar – From Database

Zoltán Bódis, Judit Kleiber, Éva Szilágyi, Anita Viszket

Department of Linguistics, University of Pécs
`lile@btk.pte.hu`

Keywords: descriptive grammar, lexicon, relational database, argument lexicon

Our student research-team has been engaged in developing a linguistic database and coordinating the process of data recording. This database is a linguistic lexicon in which stems and suffixes are stored with their arguments and detailed features.

We have built up our lexicon in a relational database. This technology provides the opportunity to store lexical units (morphemes) as well as applying underlying representations, and also using the well-known features of unification morphology; in a dynamically expandable structure. Furthermore it is possible to define the meaning (intension) of a morpheme by linking optional number of morphemes from different languages. In our system not only the phonological, morphological, syntactic and semantic features of a morpheme are stored as records, but also the rules operating within or between the lexical items. Consequently, the set of rules can be dynamically expanded as well. The theory behind the definition-method of the rules is GASG (Generative/Generalized Argument Structure Grammar), a totally lexicalist grammar by Gábor Alberti.

As a development platform we have chosen Delphi, because it offers the ability to create a user-friendly interface; and we can program in assembler code in order to increase the efficiency of searching.

The database supports several purposes like developing computational linguistic applications at our department, supporting public and higher education with our lexicon as a teaching device (the methodology of transmitting grammatical knowledge needs to be improved especially in public education), and serves as a base for developing a future descriptive grammar which we do not consider as a pile of rules, but a set of unique phenomena. Compared to the well-known concept of "grammar" our program contains less generalization, but in case of generalization (procedures which are stored in separate records) the involved elements can be exactly defined, according to certain computational linguistic standards. The descriptive grammar itself can be statistically generated from the functioning of our rules: the procedures operating on larger sets of elements form the rules which should be denominated, and those operating on smaller sets of elements form the exceptions.

Collecting linguistic data in order to fill this lexicon is not corpus-, but competence-based; this approach raises the problem of linguistic register, as it seems to be inevitable in descriptive grammars and computational applications.

In our presentation we are going to introduce the structure of our database and a the morphological parser to demonstrate the applicability of our database.