

## SOME PROBLEMS OF LINGUISTIC DATA BASES

Jürgen Kunze

Zentralinstitut für Sprachwissenschaft, AdW DDR, Berlin, DDR

### 1. Data bases as parts of procedures

It is of some advantage to compose procedures for automatic language processing of two parts:

- (i) a logical part (representing the algorithm proper) as the procedural component;
- (ii) an information part (containing the specific items of information and depending on concrete languages and purposes) as the data base.

The logical part determines the structure of the items in the information part, preferences between them and their contribution to the final result. The information part is variable. By designing the logical part sufficiently general, one can use it even for various natural languages, if one combines it with suitable information parts. The variability of the information part admits easy corrections and completions of the whole procedure. This is illustrated by some examples taken from wordform analysis.

### 2. Modular structure of logical parts.

Another important aspect is the connection between the structures of the logical part and the information part (data base). If the logical part consists of some (rather independent) sub-procedures, the information part may be composed in a similar way: Either a separate information part for every

module or a correspondence between modules and informations of a certain type within one (or some) information part(s) for more than one module. Such principles allow an easy simplification of a full version and facilitates the extension of a simple version.

### 3. Data bases as tools for establishing procedures

In many cases the elaboration of procedures may be essentially facilitated by data bases: the base contains either a representative sample or the whole set of objects to be processed by the procedure together with the desired results. Combined with a certain (mostly rather simple) dialogue languages this type of data bases admits an automatic look up of specific types and combinations, which is very advantageous for the elaboration, if a manual look up is impossible (at least in a reasonable time). One special aspect is the compensation of lack of language competence. Furthermore the same data base allows an automatic test of the final version of the procedure. This is illustrated by works on automatic hyphenation.

### 4. Data bases as tools for linguistic research

The data bases contain again linguistic objects together with additional informations (e.g. structural descriptions, classifications). Using a (more sophisticated) dialogue language one can put a lot of questions to the data base and condense the results to frequencies, correlations, mutual exclusions and so on. This is illustrated by applications of data bases to research on syntax of natural languages.

### 5. Other applications

Here one may consider terminological data-banks, data bases as dictionaries or as a mean for the elaboration of (usual) dictionaries etc.