THE REPRESENTATION OF SEMANTIC INFORMATION IN GERMAN MACHINE READABLE DICTIONARIES

Winfried Lenders

Institut für Kommunikationsforschung,u. Phonetik, Univ. Bonn
Poppelsdorfer Allee 47, D-5300 Bonn, BRD

The basis of this paper is a comparative study on German machine readable dictionaries which has been conducted during the last two years in connection with a research project financed by the German minister for science and technology. The aim of this project was to compare and to describe about 12 dictionaries, used in computer systems, in more or less conventional language analysis and synthesis systems as well as in understanding systems. The comparison and description was oriented to the data basis, the theoretical background, the field of application, and especially the microstructure of these dictionaries. ¹n this paper we want to discuss a specific problem of lexical microstructure which has arisen in the context of this project, the problem of semantic information. This problem is most relevant since on the one side all of the existing machine readable dictionaries are relatively restricted in their semantic dimension, and on the other side more detailed information about the semantics of words is extremely necessary for language analysis and synthesis programs as well as for understanding systems.

The first point of my discussion is to explain how the meanings of German words are actually represented in the microstructure of the intended dictionaries. My examples are taken from the dictionary components of the systems SUSY, LIMAS, HAM-RPM, PLIDIS, BEAST/BACON, from the German word data

bank "Wahrig" and from several other more theoretical
approaches.

My second point is the question how far in these approach-
es the general linguistic problem of representation of word
meanings has been solved and which elements of the semantic
level are normally not to be found in these dictionaries. In
the third point I want to advocate the thesis that in a mach-
ine readable dictionary the meaning of a word can be repres-
ented by a set of semantic relations. In connection with this
thesis recent developments in word semantics especially in the
fields of frame theory and semantic network theory have to
be discussed.

The general assumption of this paper is that "dictionary"
- this concept is used synonymous with "lexicon" - designat-
es any collection of word elements of a specific language;
every element of such a collection should be connected with a
certain set of linguistic information. This set of linguistic
information is called the microstructure of the dictionary.
The microstructure should be considered as an n-tupel of
descriptive elements, each of which consists of an information
field. Examples are the field of morphological information,
the field of morphological information, the field of phonolog-
ical information and the field of semantic information. The
field of semantic information again may be regarded as a
n-tupel of information. The elements of this n-tupel may be
specific relations which combine the concerned lexical element
with other lexical elements of the intended dictionary. - The
object of this paper is not to discuss how the semantic di-
mension of an individuals dictionary or lexicon may be organ-
ised but to show in what way machine readable dictionaries
which are determined for use in different computer systems
could be constructed in their semantic level with respect to
the results of recent discussion in Computational Linguistics
and Artificial Intelligence Research.