

DIFFERENT ISSUES IN THE CONSTRUCTION OF AN INTELLIGENT DICTIONARY HELP SYSTEM

E. Agirre, X. Arregi, X. Artola, A. Díaz de Ilarraza, K. Sarasola, A. Soroa

Informatika Fakultatea (Univ. of the Basque Country)
P. K. 649, 20080 DONOSTIA (Basque Country)
E-mail: jiparzux@si.ehu.es

ABSTRACT.

This paper shows different issues in the construction and knowledge representation of an intelligent dictionary help system. IDHS (Intelligent Dictionary Help System) is conceived as a monolingual (explanatory) dictionary system for human use (Artola & Evrard, 92). The fact that it is intended for people instead of automatic processing distinguishes it from other systems dealing with the acquisition of semantic knowledge from conventional dictionaries. The system provides various access possibilities to the data, allowing to deduce implicit knowledge from the explicit dictionary information. IDHS deals with reasoning mechanisms analogous to those used by humans when they consult a dictionary. User level functionality of the system has been specified and partially implemented (Agirre *et al.*, 94a).

A methodology for the extraction of semantic knowledge from a conventional dictionary is described. The method followed in the construction of the phrasal pattern hierarchies required by the parser (Alshawi, 89) is based on an empirical study carried out on the structure of definition sentences. The results of its application to a real dictionary has shown that the parsing method is particularly suited to the analysis of short definition sentences, as it was the case of the source dictionary. As a result of this process, the characterization of the different lexical-semantic relations between senses is established by means of semantic rules (attached to the patterns); these rules are used for the initial construction of the Dictionary Knowledge Base (DKB).

The representation schema proposed for the DKB (Agirre *et al.*, 94b) is basically a semantic network of frames representing word senses. After the construction of the initial DKB, several enrichment processes are performed on the DKB in order to add new facts to it; these processes are based on the exploitation of the properties of lexical-semantic relations, and also, on specially conceived deduction mechanisms. The result of the enrichment processes show the suitability of the representation schema chosen in order to deduce implicit knowledge. Erroneous deductions are mainly due to incorrect word sense disambiguation.