

Data Model for a Lexical Resource Based on Lexical Functions

SOCORRO BERNARDOS

Universidad Politécnica de Madrid, Spain

MARÍA A. BARRIOS

Universidad Complutense de Madrid, Spain

ABSTRACT

The purpose of this paper is to describe a data model that has been used in the construction of a lexical database for Spanish. This lexical resource takes the Meaning-Text Theory (MTT) as the theoretical basis, especially lexical functions (LF), which associate a lexical unit (LU) to values that express a specific meaning related to that LF for that LU. We have developed a database, called BaDELE3000, for the 3,000 most frequently used nouns in Iberian Spanish. We followed a systematic process for the design, so that the lexical data are well structured and separated from the applications that might use them. This way, the features of the data model and the subsequent database make them useful for different purposes such as word sense disambiguation, machine translation and text generation.

1. INTRODUCTION

In 2002 the Laboratory of Computational Linguistics of Moscow (Apresjan et al. 2002), (Apresjan et al. 2003) *CALLEX*, a computer-assisted language learning tool for Russian, English and German which helps the learners of one of this languages to extend their lexical knowledge by means of five types of games. This tool is based on the general linguistic framework of the Meaning-Text Theory (MTT), proposed by Igor Mel'čuk (2003), especially on Lexical Functions (LFs), which are used to represent lexical relations: paradigmatic (synonyms, antonyms, hyperonyms, etc.) and syntagmatic (collocations).

The adaptation of that tool for Spanish, called *CALLEX-ESP*, is under development. This paper deals with the data model created for