THE CPDB: A LEARNING AND TEACHING CORPUS-BASED METHODOLOGICAL TOOL

Natalia Judith Laso, Elisabet Comelles, Isabel Verdaguer, Montserrat Forcadell, Maria Luz Celaya

The use of both corpus-based methodological tools and computer technologies (NTICs) has recently become a commonplace in the teaching and learning of second and foreign languages (Granger 2003, Sinclair 2004, Bernardini 2004, Conrad 2005; Laso & Giménez 2007, Granger & Meunier 2008, Ajnär 2009, Comelles et al. 2010, Bennet 2010, MacDonald et al. 2011), as they allow the language learner, as a language observer, to become aware of the many complexities of real language in use. Similarly, corpus linguistics has also contributed to underlining the close relationship between lexis and grammar, since clause patterns depend on the presence of specific lexical verbs.

CLAUSE PATTERN DATABASE (CPDB): A lexicogrammatical tool

The CPDB provides:
- a wide representative sample of the valency and clause patterns of a selection of 217 lexical verbs extracted from a self-compiled corpus on mystery novels (the Whodunnit Corpus)
- a tree diagram for each example. Tree diagrams have been created with the assistance of the Chamiak parser (Chamiak and Johnson 2005) and the PhpSyntax Tree generator (http://www.ironcreek.net/phpsyntaxtree/).

The output of the resulting tree diagrams has been manually associated with each of the entries in the database.

The CPDB consists of 714 corpus-based registers. Each register shows the following fields:

a) Corpus-based example
b) Clause pattern, which can be selected from a scroll-down menu
c) Lemma (i.e., lexical verb)
d) Phrasal categories of the dependencies of the lexical verb

Possible Queries in the CPDB

- by Lexeme: enables the user to select a lexeme and decide whether to look for the several patterns linked to it and/or look for its valence.

The output of this query may show:

- the lexeme "put" and its corresponding patterns and examples:
- the valence performed by a specific verb:

PEDAGOGICAL APPLICATIONS

The CPDB

- contributes to the automatization of teaching and learning activities related to the specific contents on the syllabus
- helps the teacher explain dependency structure much clearly
- allows students to discriminate between obligatory and non-obligatory constituents
- favours students understanding of both the theoretical and the analytical perspective of the topic of the subject
- promotes collaborative work, both between peers and students and teacher

FUTURE WORK

The Grup de Recerca en Lexicologia i Lingüística de Corpus (GREUSC) is currently working on the following teaching innovation practices, which will contribute to underlining the interconnection between syntax and semantics.

- Discussing what kind of semantic information can be included in the CPDB
- Exploring semantic parsers which will automatize the incorporation of semantic information in the CPDB