

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

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

njlaso@ub.edu
elicomelles@ub.edu
i.verdaguer@ub.edu
forcadell@ub.edu
mluzcelaya@ub.edu

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, Aijmer 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.

CONTEXT

As part of a teaching innovation project (2011/PID-UB37), the *Grup de Recerca en Lexicologia i Lingüística de Corpus* (GRLeIC) at the University of Barcelona has recently developed a database of English clause patterns: the **Clause Pattern Database (CPDB)**, which has been implemented in the undergraduate course *Descriptive Grammar of English II* (DGE II) in the degree of English Studies

CLAUSE PATTERN DATABASE (CPDB)_A lexicogrammatical tool

The **CPDB** provides:

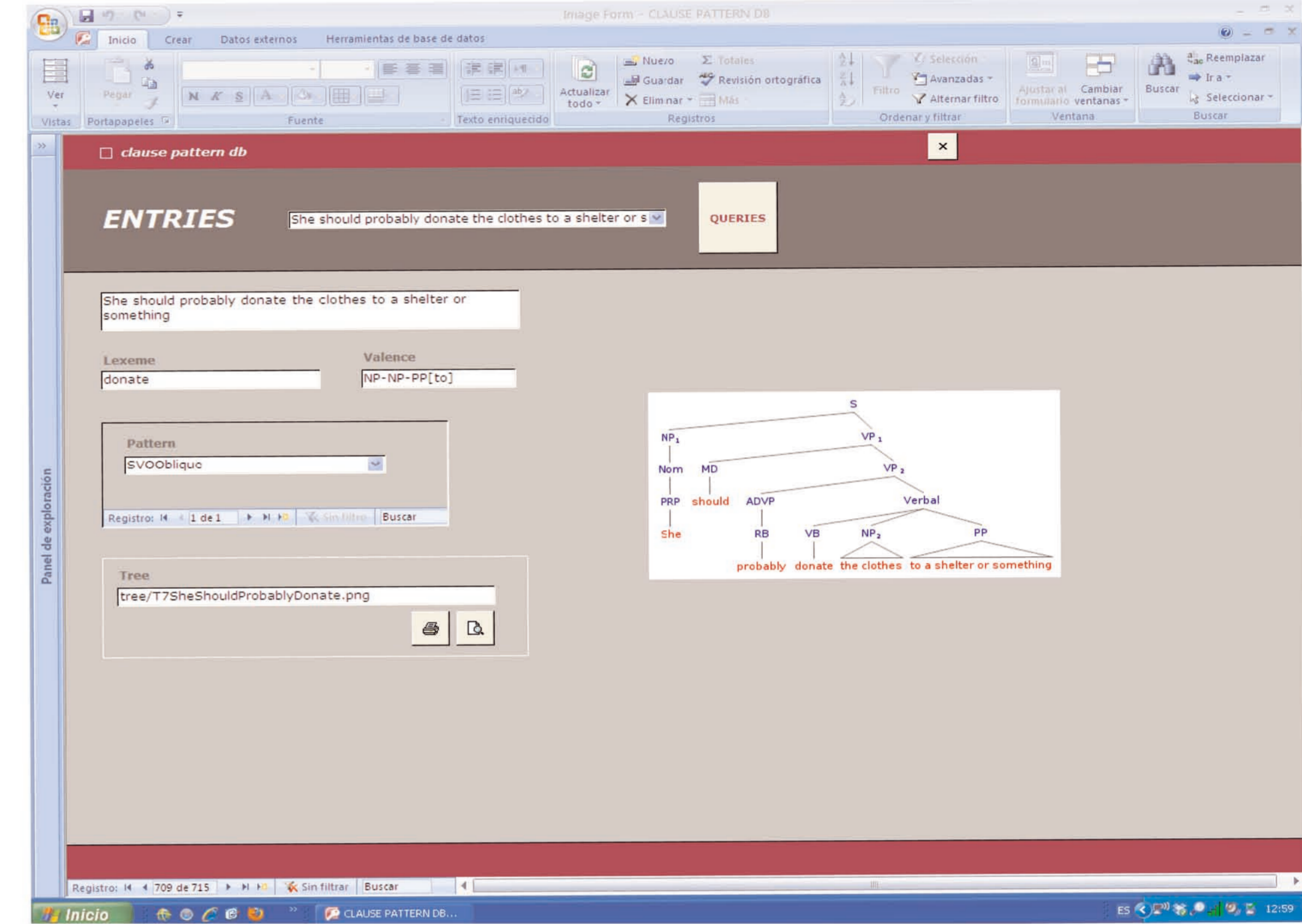
- a wide representative sample of the **valency and clause patterns** of a selection of 217 lexical verbs extracted from a self-compiled corpus on mystery novels (the **Whodunnit Corpus**)

- a **tree diagram** for each example. Tree diagrams have been created with the assistance of the *Charniak* parser (Charniak 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:

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

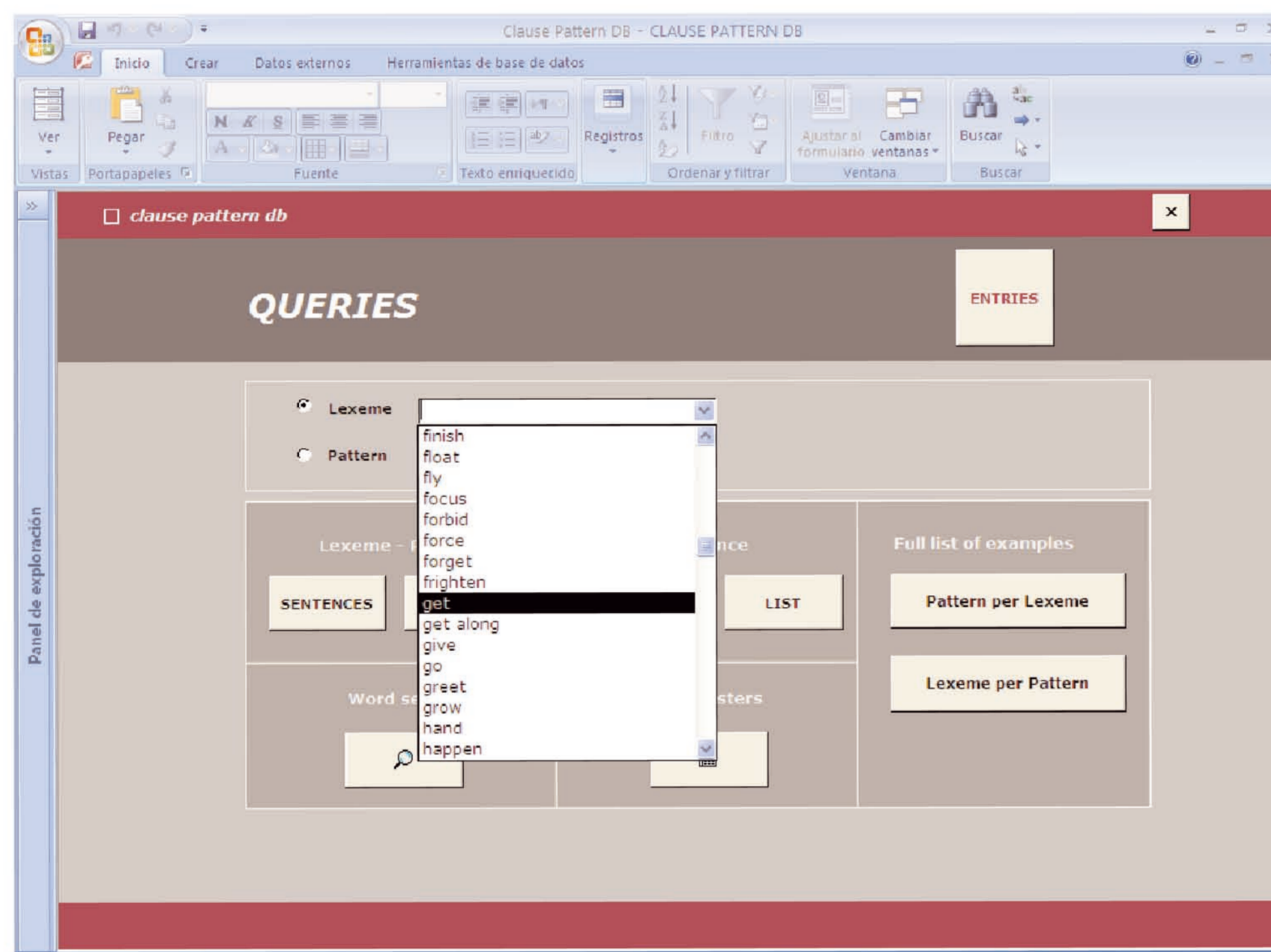
Full entry. Database entry corresponding to the example "She should probably donate the clothes to a shelter or something".



Possible Queries in the CPDB

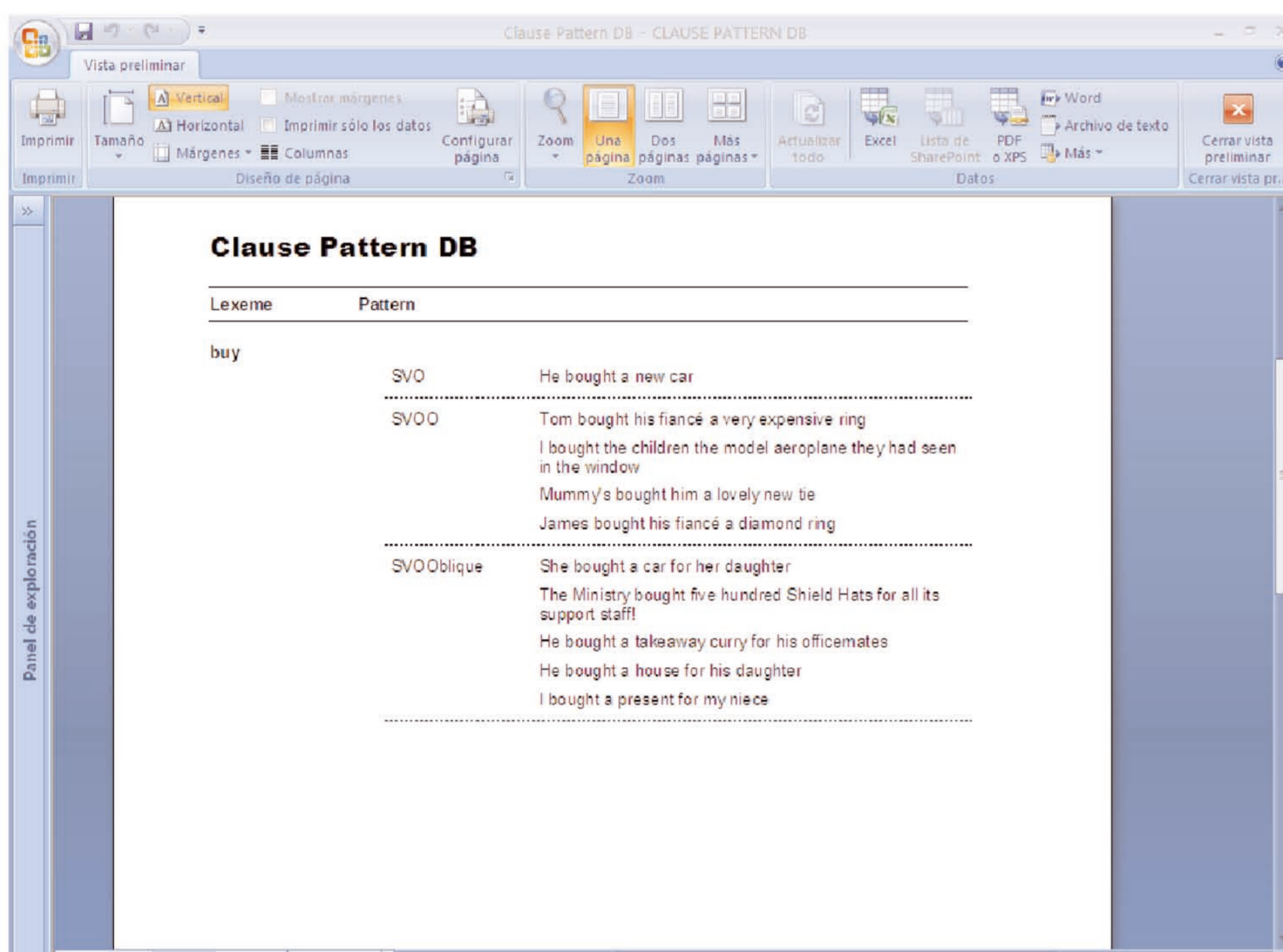
- by **Lexeme**: enables the user to **select a lexeme** and decide whether to **look for the several patterns** linked to it and/or look for its **valence**.

Screen displayed when clicking on **"Lexemes"**, which shows all the lexemes available in the database.



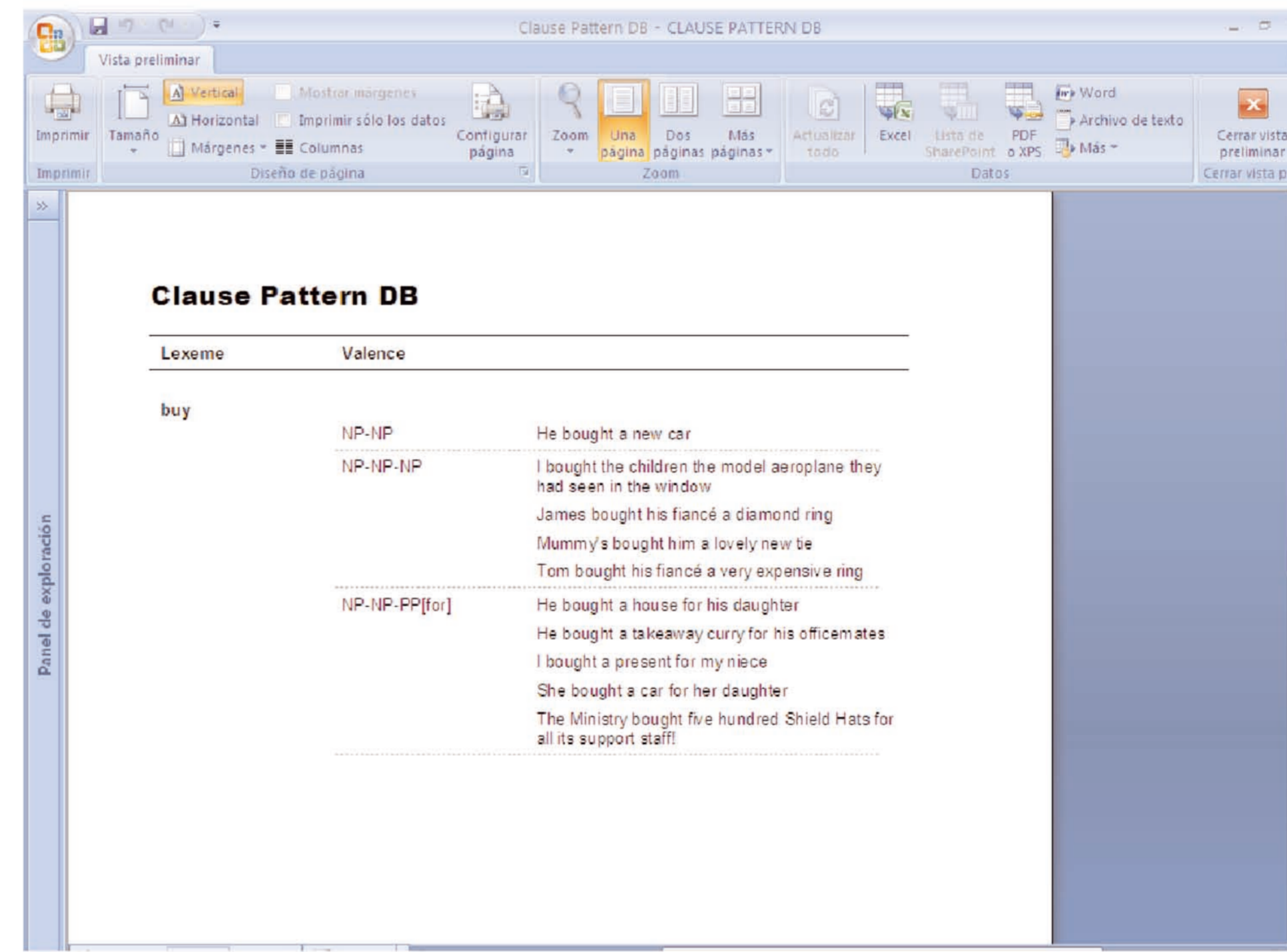
The **output** of this query may show:

- the **lexeme "get"** and its corresponding patterns and examples:



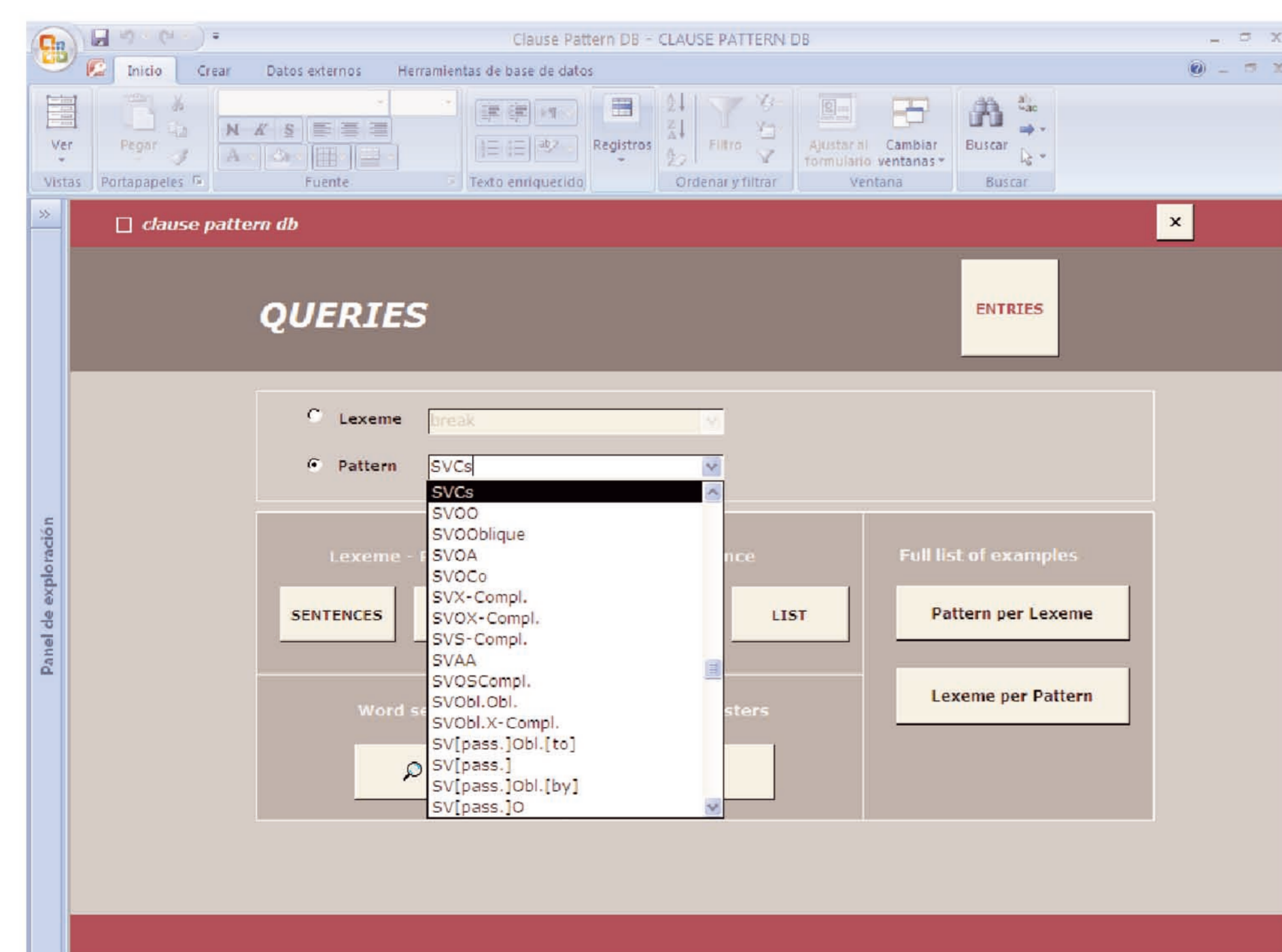
Output. List of patterns and their corresponding sentences displayed when searching for the information about the lexeme "buy".

- the **valence performed by a specific verb**:

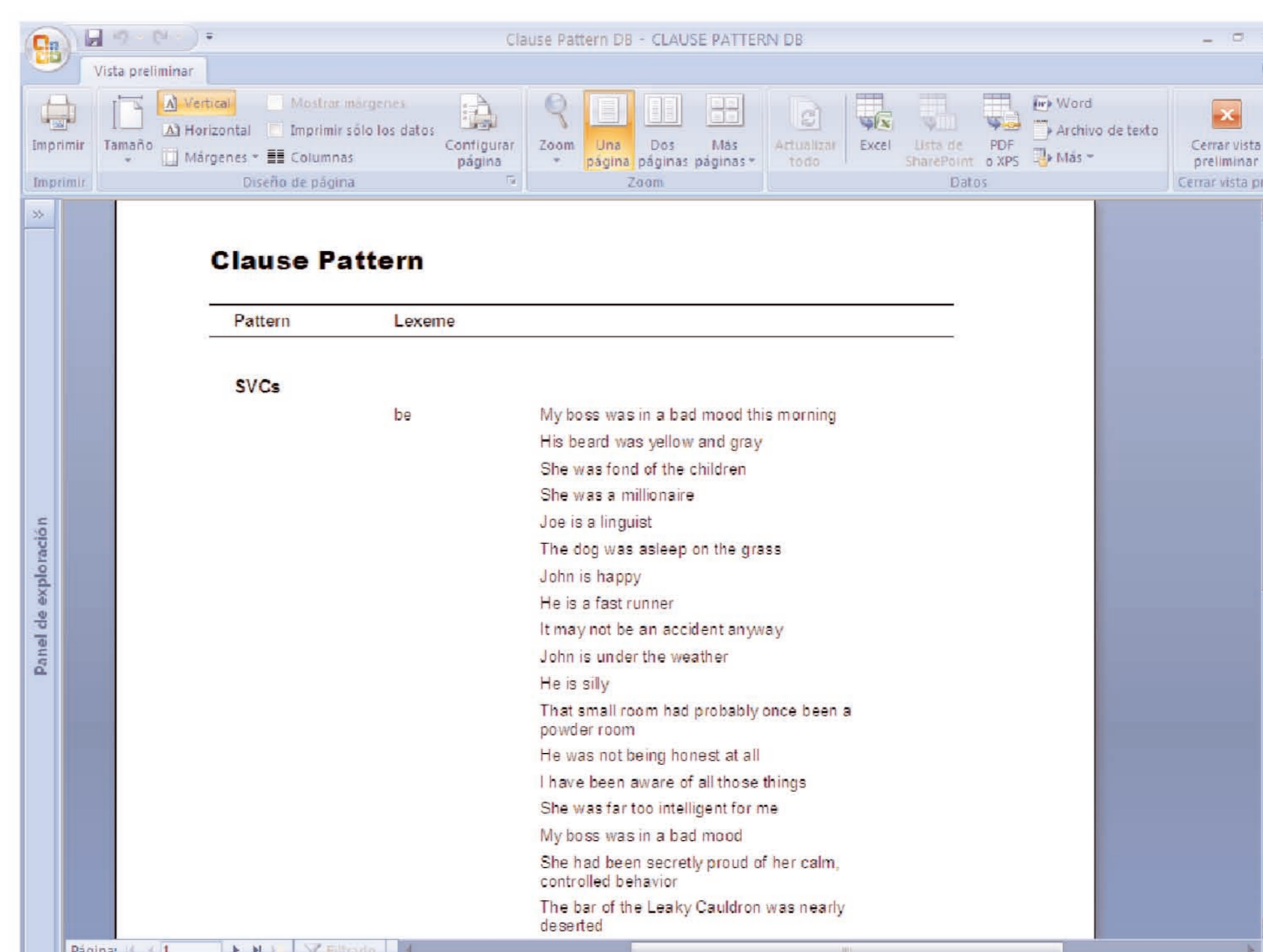


Output. List of patterns and their corresponding sentences displayed when searching for the information about the lexeme "get".

- by **Pattern**: when searching for the patterns performed by a specific lexeme, a list of the **patterns and their corresponding examples** are displayed.



Screen displayed when clicking on **"Patterns"**, which shows all the patterns available in the database.



Output. Pattern SVCs and its corresponding sentences

PEDAGOGICAL APPLICATIONS

The **CPDB**

- contributes to the **automatisation 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* (GRLeIC) 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 automatise the incorporation of semantic information in the **CPDB**

