

CESS-Consulta: Interfaz de consulta a un banco de datos sintáctico-semántico

Manuel Bertran

Universitat Politècnica de Catalunya,
Grup de Processament del Llenguatge
Natural (GPLN)
Jordi Girona 1-3, 08034
mbertran@lsi.upc.edu

M. Antònia Martí

Universitat de Barcelona
Centre de Llenguatge i Computació (CLiC)
Gran Via, 585, 08007-Barcelona,
amarti@ub.edu

Alicia Ageno

Universitat Politècnica de Catalunya,
Grup de Processament del Llenguatge
Natural (GPLN)
Jordi Girona 1-3, 08034
ageno@lsi.upc.edu

Mariona Taulé

Universitat de Barcelona
Centre de Llenguatge i Computació (CLiC)
Gran Via, 585, 08007-Barcelona,
mtaule@ub.edu

Lluís Márquez

Universitat Politècnica de Catalunya,
Grup de Processament del Llenguatge
Natural (GPLN)
Jordi Girona 1-3, 08034
lluism@lsi.upc.edu

CESS-consulta es una interfaz desarrollada en el marco del proyecto CESS-ECE¹, que permite hacer consultas a la información contenida en los corpus etiquetados morfosintáctica y semánticamente de dicho proyecto. Los corpus contienen información sobre la estructura de constituyentes, las funciones sintácticas de los mismos y los papeles temáticos asociados. Actualmente, los corpus disponibles contienen 100.000 palabras, CESS-CAT para el catalán, CESS-ESP para el castellano y CESS-EUS para el euskera. Cuando finalice el proyecto los corpus tendrán una dimensión de 500.000 palabras para cada lengua y la interfaz será accesible desde Internet. Las funcionalidades de la interfaz permiten realizar los siguientes tipos de búsquedas:

1. Frases que contengan una determinada palabra.
P.e.: Buscar todas las frases que contengan la palabra: **interfaz**
2. Frases que contengan un determinado lema.
P.e.: Buscar todas las frases que contengan el lema: **cerrar**
3. ¿Qué constituyentes realizan una determinada función sintáctica?
P.e.: Buscar los **SP** que realizan la función de **C.REG** (complemento del régimen)?
4. Los esquemas sintácticos de un predicado.
P.e.: Buscar la estructura de constituyentes del verbo: **cerrar**
5. ¿Qué oraciones contienen una determinada función sintáctica (o la combinación de dos)?

¹ CESS-ECE (Corpus Etiquetados sintáctica y semánticamente para el Español, Catalán y Euskera) (HUM 2004-21127-E).

P.e.: Buscar las frases que contengan la función: **SUJ**

Buscar las frases que contengan la función: **SUJ** y **CD**

6. ¿Qué funciones sintácticas realizan un determinado papel temático?

P.e.: Buscar los papeles temáticos asociados a la función: **CC**

7. ¿Qué papeles temáticos realizan una determinada función?

P.e.: Buscar las funciones que tienen asociado el papel temático: **PACIENTE**

8. ¿Qué frases contienen un determinado papel temático?

P.e.: Buscar las frases que contengan el papel temático: **CAUSA**

Buscar las frases que contengan el papel temático: **CAUSA** y **TEMA**

Para cada uno de los resultados se puede visualizar el árbol de análisis en forma gráfica o en estructura jerárquica parentizada (búsquedas de 3 a 10). Estos árboles a su vez pueden guardarse.

A continuación, se presenta una imagen que recoge las distintas búsquedas posibles de la interfaz CESS-Consulta (Figura 1) y el resultado de una consulta al corpus etiquetado semánticamente (Figura 2).

