

# “ESTUDIO LEXICOGRÁFICO PARA EL TRATAMIENTO AUTOMÁTICO DE LA LENGUA: DICCIONARIO BILINGÜE FRANCÉS-ESPAÑOL DE LOS SUSTANTIVOS PREDICATIVOS DE LAS <AYUDAS ECONÓMICAS>”.

## Section 5.

Elena DIEGO HERNÁNDEZ

Dos de las principales dificultades a las que se enfrenta el tratamiento automático de las lenguas naturales son el reconocimiento de las secuencias fijas (por ejemplo, dentro del marco de nuestro estudio, *paracaídas dorado*) y la interpretación de las palabras polisémicas u homónimas (*prima*).

El hombre es capaz de solventar estas dificultades gracias al contexto lingüístico en el que tales unidades léxicas aparecen. Las palabras no pueden ser aprehendidas como entidades aisladas, sino que es dentro del marco de la frase donde adquieren un significado específico. Así, concebimos la semántica como el resultado de la inscripción de los elementos léxicos dentro de una sintaxis (combinación de palabras) determinada. Atendiendo a las especificidades de los autómatas, consideramos que para que estos sean capaces de interpretar las unidades polisémicas y de identificar las secuencias fijas necesitan una descripción exhaustiva y sistemática de las características morfológicas, sintácticas y semánticas de cada unidad léxica.

Queremos presentar en esta comunicación el diccionario electrónico bilingüe de sustantivos predicativos de <ayudas económicas><sup>1</sup> en español y en francés que estamos elaborando. Un diccionario cuyo principal propósito es la inscripción del sustantivo dentro del marco de la frase, y poder así predecir de manera automática las colocaciones verbales.

Para llevar a cabo estos objetivos, hemos considerado que la teoría de las clases de objetos (Gaston Gross y el laboratorio *Lexiques, Dictionnaires, Informatique*, CNRS-Université Paris 13) se adapta a las especificidades de los autómatas.

Siendo la polisemia y la homonimia las dos grandes dificultades que las lenguas naturales plantean al tratamiento automático, hemos adoptado el postulado de considerar la frase como unidad mínima de análisis, ya que el contexto de la frase hace desaparecer la ambigüedad de interpretación. Así, en nuestra descripción de la lengua no se puede separar el léxico (las entradas del diccionario), la sintaxis (la combinación de las palabras) y la semántica (el resultado de los elementos lexicales organizados de una manera determinada dentro del marco de la frase). Estos tres niveles están agrupados bajo la noción de *empleo*. El diccionario elaborado es un diccionario de empleos y cada empleo constituye una entrada diferente.

---

<sup>1</sup> *Subsidio, patrocinio, beca, etc.*

Primeramente, hemos elaborado una tabla de análisis, para la que hemos retenido dos tipos de propiedades: las propiedades configuracionales (número de argumentos inferidos por cada sustantivo predicativo, su modo de estructuración en la frase y su naturaleza morfológica y semántica) y combinatorias (los verbos soportes, verbos aspectuales y verbos predicativos apropiados seleccionados por cada sustantivo predicativo).

Hemos concebido esta descripción como la elaboración de dos diccionarios monolingües, uno en francés y otro en español. Hemos aplicado de manera sistemática esta tabla a todas las palabras recolectadas, estudiando de manera independiente cada lengua.

Tal descripción nos ha permitido agrupar en clases de objeto las entradas-*empleos* que operan las mismas restricciones sintáctico-semánticas. De este modo, gracias a esta clasificación el ordenador puede predecir de manera automática cuál es el esquema argumental de un sustantivo predicativo determinado y sus colocaciones verbales.

En un segundo tiempo, después de haber establecido y descrito cada entrada del diccionario de <ayudas económicas> en español y del diccionario de <aides financières> en francés, cuando hemos establecido correspondencias entre los predicados de ambas lenguas. Este diccionario permite al autómeta reconocer cada colocación en la lengua de origen y generarla en la lengua de destino. No se trabaja pues con un diccionario bilingüe, sino con dos diccionarios monolingües coordinados.

## **BIBLIOGRAFÍA:**

BOSQUE. I. y MALDONADO C., 2004: *REDES. Diccionario combinatorio del español contemporáneo*, SM, Madrid.

GREZKA A. et BUVET P.-A., 2007 : « Élaboration d'outils méthodologiques pour décrire les prédicats du français », in *Linguisticae Investigationes* 30, n° 2, Philadelphia John Benjamins, Amsterdam , pp. 217-245.

GROSS G., 2008: « Les classes d'objets », en *LALIES* 28, Éditions Rue d'Ulm, Paris, pp. 111-165.

LE FUR D., 2007: *Dictionnaire des combinaisons des mots: les synonymes en contexte*, Collection Les Usuels du Robert, Dictionnaires LE ROBERT, Paris.

MEL'ČUK I., 1984, 1988, 1992, 1999: *Dictionnaire explicatif et combinatoire du français contemporain. Recherche lexico-sémantique* I, II, III et IV, Les Presses Universitaires de Montréal, Montreal.