

# LA ESTADÍSTICA Y LA SELECCIÓN DIDÁCTICA DEL LÉXICO TÍPICO DE UNA LENGUA EXTRANJERA PARA FINES ESPECÍFICOS: EL CASO DEL DISCURSO ECONÓMICO EMPRESARIAL EN ELE

Lieve Vangehuchten  
Universidad de Amberes

## INTRODUCCIÓN

En este trabajo se presenta el proceso de elaboración de una metodología para la selección adecuada del léxico propio del discurso económico empresarial a partir de un corpus de lenguaje económico empresarial. Se tratarán las dos cuestiones fundamentales para poder concebir esta metodología, es decir cómo identificar este léxico y cómo hacer una selección científica y objetiva. En cuanto a la identificación, se sigue la teoría de la corriente lingüística de la Terminología que estimula la investigación semasiológica en corpus (Cabré 1999). Para el proceso de selección se recurre a unas técnicas léxico-estadísticas. Este método se aplica a un corpus modélico de 136.932 *tokens*. Primero se comentarán la composición del corpus modélico así como su norma de lematización. Después se describirán las distintas categorías léxicas presentes en el corpus y la manera de la que la estadística contribuye a su clasificación y selección didáctica.

## EL CORPUS MODÉLICO

Se trata de un corpus de discurso didáctico, o sea un texto medianamente especializado, de comunicación entre un especialista y aprendices. Es un manual de introducción a la economía empresarial de la mano de Eduardo Pérez Gorostegui, catedrático de economía de la empresa en la UNED (*Universidad Nacional de Educación a Distancia*). El manual se titula *Introducción a la administración de empresas* y la primera edición data de 1997.

Con la denominación 'modélico' se quiere hacer hincapié en el hecho de que el objetivo de este trabajo no consiste en la elaboración de una herramienta pedagógica lista para ser utilizada en una situación de enseñanza, bajo la forma de un glosario o diccionario didáctico que pretende contener el léxico del lenguaje económico empresarial en español. El corpus es modélico en el sentido de que sirve de modelo del discurso analizado, por un lado, y de muestra necesaria para el desarrollo del método propuesto, por otro.

El tamaño es modesto: 136.932 formas léxicas, lo que es poco en comparación con los grandes corpora de varios millones de ocurrencias que se componen hoy día para propósitos lexicológicos. No obstante, según *The Handbook of Terminology Management*, un manual de terminología, un corpus de 100 000 formas léxicas ya es suficiente para sacar conclusiones significativas, por lo menos en el marco de un análisis de un corpus de