

# ECPC: European Parliamentary Comparable and Parallel Corpora / Corpus Comparables y Paralelos de Discursos Parlamentarios Europeos

María Calzada Pérez, Noemí Marín Cucala, José Manuel Martínez  
Martínez

Universitat Jaume I  
Dpto. de Traducción y Comunicación  
Campus de Riu Sec  
12071 Castellón de la Plana  
Tel.: 964729741  
calzada@trad.uji.es

## 1 Presentación del proyecto y miembros

El proyecto ECPC (European Parliamentary Comparable and Parallel Corpora) se enmarca dentro de los estudios traductológicos de corpus. La investigación se realiza partir de corpus comparables y paralelos, compuestos por discursos del Parlamento Europeo (EP/PE) así como de los parlamentos nacionales de España, Reino Unido e Irlanda. El grupo recibe financiación del Ministerio de Educación y Ciencia (HUM2005-03756/FILO). La coordinadora del equipo de investigadores, que está integrado por estudiosos de los ámbitos de la traductología, informática y estadística, es María Calzada Pérez y cuenta con la participación de Rosa María Agost Canós y María Pilar Jara Jiménez además de la colaboración de dos becarios (Noemí Marín Cucala y José Manuel Martínez Martínez) todos ellos de la UJI así como de otros investigadores de reconocido prestigio radicados en Reino Unido (Mona Baker y Gabriela Saldanha) e Irlanda (Dorothy Kenny, Marion Winters y Saturnino Luz).

## 2 Antecedentes

Este proyecto sigue la estela del trabajo realizado desde la lingüística de corpus por pioneros como Nelson Francis, Henry Kucera (Brown Corpus) o Geoffrey Leech (Lancaster-Oslo/Bergen Corpus, Lancaster-IBM Spoken English Corpus y el BNC). ECPC recoge también la aportación de los CTS realizada por Mona Baker (TEC) —véanse por ejemplo Baker (2004)—, Dorothy Kenny (2001), Sara Laviosa (2002), Lynne

Bowker (2002), Jennifer Pearson (1998) y, más recientemente, Gabriela Saldanha (2004) y Marion Winters (2004). Entre los corpus de referencia para el grupo ECPC destacan el UN Parallel Text Corpus, el ITU/CRATER Corpus, el JOC-CES Multilingual Corpus, el MULTEXTS, el Hansard Canadian English-French Parallel Corpus y el Oslo ENPC.

## 3 Hipótesis de trabajo

La principal hipótesis que el trabajo del grupo desea confirmar o refutar es la siguiente:

El grado de unificación europea es tan estrecho que instrumentos comunitarios, como los discursos del PE/EP, poseen ciertas dosis de autonomía macrolingüística y/o microlingüística respecto de instrumentos equivalentes en los Estados Miembros (discursos emitidos ante las respectivas Cámaras nacionales).

## 4 Objetivos esenciales

### 4.1 Principal:

- Comprobación o refutación de la hipótesis de trabajo.

### 4.2 Parciales:

- Creación, de un Archivo electrónico compuesto por corpus diversos, que permita la combinación y la comparación de los mismos, al que se someterá a un etiquetado, alineamiento y anotación.
- Desarrollo de unos parámetros para realizar estudios contrastivos a partir de los corpus comparables y paralelos del Archivo.

- Difusión del conocimiento procedente de los estudios contrastivos realizados y de los resultados del proyecto en distintos formatos y en distintos foros académicos.
- Desarrollo de una herramienta de búsqueda terminológica para permitir la consulta del archivo.
- Establecimiento de bases sólidas para la didáctica de la traducción del género del discurso parlamentario entre profesionales de la traducción y aprendices de traducción.

## 5 Metodología

Las pautas y fases metodológicas del trabajo son las siguientes:

### 5.1 Fase 1:

- Recopilación de los corpus: textos del Parlamento Europeo (PE), de la Cámara Baja del Parlamento británico (House of Commons, HA), de la Cámara Baja del Parlamento español (Congreso de los Diputados, CD) y de la Cámara Baja del Parlamento irlandés (Dáil Éireann, DE) - Periodo recopilado: 2005-2006.
- Tratamiento y limpieza de textos recopilados.

### 5.2 Fase 2:

Etiquetado y alineamiento del archivo recopilado en la fase anterior.

### 5.3 Fase 3:

Desarrollo de parámetros (macrotextuales y microtextuales) para el análisis contrastivo de los corpus recopilados, etiquetados y alineados.

### 5.4 Fase 4:

Estudios contrastivos a partir de las herramientas y los resultados de las tres etapas anteriores.

### 5.5 Fase 5:

Publicación de los resultados a través de diferentes modalidades de difusión, diseño de una página web en la que se recoja el trabajo y resultados parciales y totales del proyecto, creación de una interfaz de consultas contrastivas siguiendo la estela de TransSearch.

## 6 Aplicaciones (didáctica e investigación)

Las dos aplicaciones más importantes del corpus son la didáctica de la traducción y la investigación traductológica. En cuanto a la didáctica, la herramienta de consulta terminológica (contextual y bilingüe) que se pretende desarrollar y poner a disposición de la comunidad científica por Internet resultará de gran ayuda para los centros de formación de traductores y también para los aprendices de traducción. Y por lo que respecta a las aplicaciones del corpus en el campo de la investigación, la comunidad académica y la industria de la ingeniería lingüística contarán con la posibilidad de acceder a un archivo flexible y completo que estará etiquetado, alineado y anotado que, al mismo tiempo, les permitirá profundizar en los conocimientos sobre el género del discurso parlamentario en general y del discurso del Parlamento Europeo en particular. En la presentación del proyecto se ejemplificarán aplicaciones tanto en el terreno didáctico como en el ámbito puramente investigador.

## 7 Bibliografía

- Baker, M. 2004. A Corpus-based View of Similarity and Difference in Translation. *International Journal of Corpus Linguistics*, 9(2): 167-193.
- Bowker, L. 2002. *Computer-aided translation technology: A practical introduction*. University of Ottawa Press, Ottawa.
- Kenny, D. 2001. *Lexis and Creativity in Translation: A Corpus-based Study*. St. Jerome, Manchester.
- Laviosa, S. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Rodopi, Amsterdam & New York.
- Pearson, J. 1998. *Terms in context*. J. Benjamins, Amsterdam, Philadelphia.
- Saldanha, G. 2004. The Translator's Presence in the Text: A Corpus-based Exploration. *1st IATIS conference: Translation and the Construction of Identity*, Seoul, South Korea.
- Winters, M. 2004. F. Scott Fitzgerald's Die Schönen und Verdammt - A Corpus-based Study of Translators' Style: Modal particles and their influence on the narrative point of view. *1st IATIS Conference*, Seoul, South Korea.