

«TRANSCALC» DEL PROCESADOR DE TEXTOS A LA BASE DE DATOS: UN PROGRAMA AUTOMÁTICO

J. Torruella Casañas

En el *Seminario de Filología e Informática* de la Universidad Autónoma de Barcelona se está trabajando en la elaboración de un *Archivo Informatizado de Textos Catalanes Medievales* para su posterior conversión a un Banco de Palabras, bajo la dirección del Dr. José Manuel Blecua. El diseño y la dirección del proyecto está a cargo de Joan Torruella, y se cuenta con la colaboración de especialistas en lengua y literatura medievales como la Dra. Lola Badia, de la Universidad de Barcelona, o el Dr. Jeremy N. H. Lawrance, de la Universidad de Manchester.

La finalidad del tratamiento informático de documentos catalanes medievales es disponer de una base de datos con todas las palabras que aparezcan en ellos, acompañadas de otras informaciones complementarias, tanto históricas (primera documentación, zonas lingüísticas de sus usos, autores, géneros, etc.) como lexicográficas (grafía, lema, categoría gramatical, etimología, etc.).

Para esta finalidad ha sido necesario crear un programa informático que fuera capaz de agilizar todo el proceso de separación del texto en palabras y que posteriormente hiciera el análisis de éstas.

El programa que queremos presentar, está pensado para la gestión de corpus léxicos, considerando que éstos no se deben reducir solamente al almacenamiento de los documentos dentro de los ordenadores, sino que, además, tienen que proporcionarnos informaciones históricas y lexicográficas de las palabras.

Para ello el programa 'transcalc' pasa de forma automática el resultado de la concordancia del texto de un documento, hecha en OCP, a una base de datos. De este modo podemos organizar la información de un modo más flexible y la podemos completar con nuevos datos, a parte de poder ordenarla de formas muy diferentes. En esta base de datos, el programa crea una ficha para cada forma gráfica del documento que se analiza. En cada una de estas fichas aparecen, por un lado, los campos con las informaciones que provienen de la concordancia del documento (*grafía de la palabra-lemma, número de frecuencia, frase de contexto y localización de la frase en el documento*), y también los campos con las informaciones históricas que se desprenden del documento mismo, comunes a todas las palabras (*autor, género, sigla del manuscrito, fecha y zona lingüística*). Por otro lado, también aparecen en cada ficha

de nuestra base de datos unos campos lingüísticos que el programa llenará semi-automáticamente. Estos campos son: *lema*, *categoría gramatical*, *construcción*, *subcategoría*, *entrada*, *etimología* y *traducción al catalán moderno*. De este modo, por ejemplo, la forma gráfica *enprès*, que aparece en el manuscrito BARCELONA, Biblioteca de Catalunya Ms. 9, una vez incorporada a nuestra base de datos, formaría parte de una ficha con las informaciones siguientes:

Grafía:	[enprès]	Dialecto:	[]
Lema:	[emprès]			
Categoría:	[v,par.]			
Construcción:	[hageren enprès]			
Subcategoría:	[v,ind,ant,6.]			
Entrada:	[emprendre]			
Traducción:	[emprès]			
Etimología:	[imprehendere]	Frecuencia:	[2]
Procedencia:	[.llatí.]	Referencia:	[001-34-16 P.Q.]
Contexto:	[E savis molts hageren tant enprès si contr'Amor]
	valgés alguna ofensa				
Notas:	[*]				
Fecha:	[1499-01]	<u>DICCIONARIOS</u>		
Autor:	[Masdovelles, Joan Berenguer]	Moll:	[T]	
Género:	[A1.]	Corominas:	[T]	
Zona:	[Ce.]	Otros:	[]	
Sigla:	[BARCELONA, Biblioteca de Catalunya, Ms. 9.]	Situación:	[358 / 7923]

Este programa ha sido realizado en el *Seminario de Filología e Informática* de la Universidad Autónoma de Barcelona, por Ramon Capçada y Joan Torruella, en lenguaje de programación Clipper, de manera que su resultado puede ser fácilmente utilizado en la mayoría de bases de datos relacionales.