

José Manuel Blecua et al., eds. (1999)

***Filología e Informática. Nuevas tecnologías en los estudios filológicos***

(pròl. de José Manuel Blecua), Barcelona, Seminario de Filología e Informática (UAB) i Editorial Milenio.

Salvador Climent

Aquest recull de vint treballs ofereix una visió panoràmica de la incidència de les tecnologies informàtiques en els estudis filològics a Espanya. Algun dels treballs tracta aspectes de Lingüística Computacional, però cal tenir en compte que el volum no es planteja com una presentació de l'estat de l'art en processament automàtic del llenguatge sinó que s'orienta clarament cap a la Lingüística Informàtica, és a dir, el camp de les tècniques i aplicacions informàtiques enteses com un ajut per al treball del filòleg.

Podem classificar els articles entre aquells que presenten de forma genèrica [eines o tècniques](#) i aquells de caire més específic que presenten les seves [aplicacions a treballs filològics concrets](#); apart dels que ja hem esmentat que enquadraríem dins la [Lingüística computacional](#). En tot cas, aquesta classificació s'ha de entendre de forma orientativa ja que, com és previsible, hi ha treballs que es podrien classificar de forma alternativa en més d'un d'aquests camps.

## Eines i tècniques

A *Análisis estadístico de textos*, **Mónica Bécue** (Universitat Politècnica de Catalunya) presenta mètodes d'anàlisi lexicomètrica, a partir de la repetició de paraules i les seves associacions, dins el marc de l'anàlisi de corpus. Aquests mètodes s'apliquen a l'extracció d'informació de respostes obertes a enquestes i de corpus diacrònics. En la mateixa línia de les anàlisis quantitatives, **Pilar García Mouton** (CSIC, Instituto de Filología) presenta a *Dialectometría* l'aplicació d'eines informàtiques al tractament i la gestió de dades per a l'estudi de la variació dialectal i la interpretació d'atles lingüístics. Explica l'aplicació de procediments objectius, estadístics i taxomètrics, per establir relacions de semblança o diferenciació dialectals, en un intent de sintetitzar els continguts dels atles lingüístics. Per la seva part, **María Morrás** (Universitat Pompeu Fabra), a *Informática y crítica textual: Realidades y deseos* fa una valoració de l'aportació dels mètodes informàtics a l'edició crítica de textos, expressant l'opinió que aquests mètodes es convertiran en una eina fonamental en aquest camp.

En l'apartat de terminologia, **Maria Teresa Cabré** (Universitat Pompeu Fabra), a *Informática y terminología* situa aquesta àrea dins el camp de les anomenades "indústries de la llengua" i presenta les aplicacions de la informàtica al treball terminològic (la anomenada *terminòtica*). Analitza dos camps principals: els bancs de dades terminològics i la utilització de la informàtica en els processos d'obtenció i constitució de dades terminològiques. **Juan Carlos Sager** (UMIST, Manchester) planteja a *Reflexiones sobre los equivalentes terminológicos en traducción* els problemes de la interpretació de variants terminològiques, en què l'anàlisi automàtica pot ser de gran valor a l'hora de tractar adequadament els equivalents terminològics de traducció.

L'editor digital **José Antonio Millán** presenta a *Estaciones filológicas* l'estació lexicogràfica amb què s'ha fet l'edició electrònica del Diccionario de la Real Academia Española, que controla la correcció del text, la creació d'enllaços hipertextuals i l'etiquetatge. A més a més reflexiona sobre com hauria de ser l'estació de treball filològic ideal, que hauria de contenir dades (textos electrònics i imatges), programes de treball textual i de treball amb imatges, obres de consulta, una plataforma de desenvolupament de noves obres (eines de processament de textos i d'altres) i una interfície.



En l'àrea de l'ensenyament, **Charles B. Faulhaber** (University of California), a *La enseñanza del catalán antiguo: educación a distancia y bibliotecas digitales, cara y cruz de una misma moneda*, presenta una experiència aïllada d'ensenyament a distància de català antic als EUA, realitzada el 1995, utilitzant les comunicacions per internet, materials en web i textos digitalitzats; i **Dolors Poch** (Universitat Autònoma de Barcelona) analitza a *Informática y enseñanza de lenguas a las puertas del tercer milenio* la influència de la informàtica en la fonètica aplicada a l'ensenyament de segones llengües. En fa una reflexió per concloure que les eines informàtiques només seran útils per a aquesta tasca quan permetin la interacció oral entre l'alumne i l'ordinador.

## Aplicacions a treballs filològics concrets

**Manuela Sassi** (Istituto di Lingüística Computazionale - Pisa) presenta a *Concordancias para filólogos: en pos de la simplicidad* l'aplicació de la tècnica de concordances utilitzant l'eina DBT per a l'anàlisi del corpus de l'obra de Santa Teresa de Jesús. A més, exposa el projecte de compilació electrònica d'un corpus format per les actuals constitucions iberoamericanes.

Quant a estudis diacrònics, **José Ramón Morala** (Universidad de León) explica a *Las bases de datos en la investigación diacrónica y dialectal* l'explotació de l'ús d'una base de dades relacional per a l'estudi diacrònic d'una llengua. En el cas que presenta, es tracta de documents notariais del lleonès del segle XIII. L'objectiu de l'estudi és analitzar el procés de castellanització i fixar els mecanismes emprats pels notaris lleonesos per constituir una norma gràfica del lleonès medieval a partir de les variants que s'hi parlaven. Per la seva part, **Carlos Sánchez** (Universitat Autònoma de Barcelona), a *Sintaxis histórica, informática y periodización del español* descriu una experiència d'ús d'un conjunt d'eines (OCR, base de dades relacional, programa de concordances i un lematitzador) per a l'anàlisi diacrònica de la sintaxi de l'espanyol.

Una aplicació als estudis dialectològics és la presentada a *Análisis cuantitativo de campos léxicos*. En aquest treball, **Francisco Moreno** (Universidad de Alcalá), descriu la metodologia i les tècniques emprades en l'estudi d'un camp semàntic a partir del corpus recollit a ALEA (Atlas Lingüístico y Etnográfico de Andalucía), cosa que li permet d'extreure conclusions sobre la distribució d'un cert nombre de realitzacions lèxiques dins les diferents àrees dialectals d'Andalusia.

Tres articles tracten d'aplicacions de la informàtica a la lexicografia. **Manuel Alvar** (Universidad Complutense de Madrid), a *La redacción lexicográfica asistida por ordenador: Dificultades y deseos* analitza el paper de la informatització en l'elaboració de diccionaris a partir de materials lexicogràfics preexistents, partint de la seva experiència en l'actualització dels diccionaris VOX. Assenyala una sèrie de necessitats inexcusables, com la substitució dels editors de textos per bases de dades, la sistematització dels continguts o l'estandardització dels sistemes de transferència de dades. **Mari Paz Battaner** (Universitat Pompeu Fabra) a *Estructura del vocabulario a través de las definiciones lexicográficas: Las entradas re- del DRAE 1992* explora la utilització dels diccionaris informatitzats com a corpus a partir del qual extreure una multiplicitat d'informacions lexicològiques: extracció de tesaures, taxonomies, trets semàntics, bases lèxiques i derivacions, o constants gramaticals. Ho exemplifica mitjançant l'anàlisi de les entrades que comencen pel prefix re- al DRAE-1992, a partir de les quals extreure patrons lèxico-semàntics. Finalment, el treball de **Gloria Clavería** (Universitat Autònoma de

Barcelona), *La documentación en el diccionario etimológico*, presenta l'experiència d'informatització del Diccionario Crítico Etimológico del Castellano Hispánico, de Corominas i Pascual, que s'està portant a terme a la Universitat Autònoma de Barcelona en col·laboració amb les universitats de Girona i León, i

discuteix els problemes sorgits i les possibilitats d'explotació de les dades per a estudis lexicològics diacrònics.

## Lingüística Computacional

**Gerardo Arrate** (Instituto Cervantes), a *Normas y estándares para la codificación de textos y para la ingeniería lingüística* tracta de la codificació estandarditzada de textos amb l'objectiu d'optimitzar l'intercanvi i la seva reutilització per a diferents programes informàtics i diferents tipus d'estudis i recerques. Presenta l'estàndard de marcatge SGML i els projectes TEI (de normativa per a la codificació digital de textos) i EAGLES (recomanacions per a la normalització de recursos lingüístics i sistemes de tecnologia i enginyeria lingüístiques: gramàtiques, corpus, lexicons...).

**Joan Torruella i Joaquim Llisterra** (Universitat Autònoma de Barcelona) presenten a *Diseño de corpus textuales y orales* les diferents tipologies de corpus, tant escrits com orals, i n'expliquen les seves principals aplicacions. A més, descriuen el procés que se segueix en la constitució d'un corpus i ressalten la incidència i importància que tenen per a la seva reutilització el disseny, la determinació dels objectius i l'adaptació als estàndards.

El treball d'**Horacio Rodríguez** (Universitat Politècnica de Catalunya), *Técnicas estadísticas en el tratamiento del lenguaje natural* és en realitat un tutorial breu però molt complet de l'aproximació al Processament del Llenguatge Natural que es basa en mètodes estadístics (per oposició a l'aproximació basada en coneixement lingüístic explícitament declarat). Aquests mètodes s'orienten al tractament massiu de text no restringit i parteixen de la informació disponible a corpus i recursos lèxics en format digital: tasques com l'extracció d'informació, la indexació, els resums automàtics, el filtratge d'informació o la traducció automàtica. Es presenten explícitament les següents àrees: etiquetatge i desambiguació morfosintàctica de textos, anàlisi sintàctica basada en gramàtiques probabilístiques, estudis de coocurrència i afinitat entre paraules (basada en el concepte d'informació mútua) per a l'extracció de restriccions selectives, la detecció d'expressions idiomàtiques, locucions i compostos nominals i, finalment, la desambiguació semàntica.

**Ignacio Moreno-Torres** (Universidad de Málaga), a *Técnicas informáticas para la traducción*, explica les aportacions de la informàtica a la traducció, des dels sistemes complementaris d'ajut al traductor, fins als sistemes de traducció automàtica, passant per l'accés a fonts d'informació (terminològica, enciclopèdica, estilística, corpus bilingües alineats...).

Finalment, **Joaquim Llisterra et al.** (Universitat Autònoma de Barcelona) descriuen a *Fonética y tecnologías del habla*, per un costat les principals eines informàtiques de les quals disposa la disciplina de la fonètica (anàlisi articulatòria, acústica i perceptiva), i per un altre, algunes de les aplicacions en el camp de les tecnologies de la parla, específicament en el camp de la síntesi o conversió de text a veu.

### Citació recomenada:

CLIMENT, Salvador (2000). "Filología e informática. Nuevas tecnologías en los estudios filológicos, de José Manuel Blecua et al.". *Digithum*, núm. 2 [artícle en línia].  
DOI: <http://dx.doi.org/10.7238/d.v0i2.614>