

**PROCESAMIENTO DEL LENGUAJE NATURAL: FUENTES
BIBLIOGRAFICAS PARA SU ESTUDIO EN LENGUA ESPAÑOLA.**

Begoña Baeza Palacios
Lda. en Filología Española
3^{er} curso E.U. Biblioteconomía y Documentación
Universidad de Granada

INTRODUCCION

En 1976, Monserrat Meya publicaba una nota informativa

¹⁾ cuyo título era lo bastante explícito: "El hijo más joven de la Lingüística: la Lingüística Computacional".

Esta reseña presenta una breve historia de la nueva disciplina desde su nacimiento en 1946 en Estados Unidos, con los primeros intentos de traducción automática por computadora, hasta los años 70, en que consigue independizarse de la que había sido hasta entonces su tutora, la Inteligencia Artificial, y comienza a caminar por su propio pie, investigando los fenómenos del discurso y la comunicación, y desarrollando los campos de la semántica y la pragmática.

Su utilidad se fue reconociendo en distintas áreas como la medicina, la administración, etc., y se le auguraba un puesto de primera fila en los objetivos de las investigaciones científicas.

Sin embargo, otro autor español, Francisco Milla Lozano, en 1983⁽²⁾ todavía acusaba a la Lingüística de estar estancada en enfrentamientos sobre el generativismo y, según sus propias palabras, "no concede todavía a la Lingüística Computacional más atención que a las primeras 'cuentas' del primerizo escolar".

Aún quedaba mucho por hacer en nuestro país, aunque los primeros pasos estaban dados: del 21 al 23 de abril de 1975, se celebraba en la Universidad Complutense de Madrid un Simposio sobre la "Utilización de ordenadores en problemas de Lingüística". Esta iniciativa del doctor Torroja, entonces Vicerrector de Investigación Científica de la Universidad Complutense, responde al incremento producido en otros países de las investigaciones relacionadas con el tema: Francia, Italia, Bélgica, E.E.U.U., tenían desde hacía años centros dedicados especialmente a la aplicación de los ordenadores a la Lingüística y a los estudios de Humanidades en general.

En España, hasta ese momento, los esfuerzos en este campo se habían realizado de forma independiente, siendo preciso destacar las actividades llevadas a cabo por el Centro de Cálculo de la Universidad Complutense.

Sin embargo, y a pesar de que el Simposio pretendía establecer de forma oficial estos estudios dentro del ámbito nacional, es bastante significativo que en el VII Simposio de la Sociedad Española de Lingüística, celebrado en Sevilla en 1977, y versando sobre el tema "Relaciones de la Lingüística con otras ciencias", no se presentara ni una sola comunicación sobre Lingüística e Informática o Lingüística Computacional, reflejándose así el escaso interés que aún despertaba esta disciplina en el seno de la comunidad Lingüística tradicional.

Tres pasos van a ser fundamentales para que definitivamente se abra en España el

¹MEYA, M. "El hijo más joven de la lingüística: la Lingüística computacional". *Revista Española de Lingüística*. 1976, vol. 6, nº 2, p. 158-159.

²MILLA LOZANO, F. "Una formulación morfosintáctica: bases verbales codificables". *Revista Española de Lingüística*. En-jun. 1983, vol. 13, nº 1, p. 158-164.

campo de investigación sobre Lingüística Computacional: en 1977 se celebra en Madrid el Congreso "Informática y Lingüística"; en 1981, y por iniciativa del presidente de la sección de español de la Asociación de Literatura y Lingüística Computacionales (ALLC), Leopoldo Sáez Godoy, y gracias al departamento de español del Instituto de Lingüística Computacional de Pisa, se celebró en esta ciudad un encuentro de trabajo de "hispanistas computacionales", en el que se acordó, entre otras cosas, fomentar el intercambio de los distintos grupos de investigación y convertir las reuniones de hispanistas computacionales en una actividad regular; en 1982, el Departamento de Filología hispánica de la Universidad de Deusto, Bilbao, presenta el curso monográfico "Lingüística Computacional".

Así se sentaron las bases para que en abril de 1983 se celebraran en Sevilla las "I Jornadas sobre procesamiento del lenguaje natural". El grupo de asistentes acordó constituirse en asociación y publicar una revista como órgano de difusión.

La asociación que surgió de aquel encuentro es la SEPLN: Sociedad Española para el Procesamiento del Lenguaje Natural, y la revista es el boletín que la Sociedad comenzó a publicar en octubre de ese mismo año, del cual se siguen editando hasta nuestros días dos números anuales, con la finalidad de difundir todos los nuevos estudios sobre la materia.

Con el presente trabajo pretendemos ofrecer una panorámica de cómo han evolucionado en España los estudios referidos a la "Lingüística Computacional" o al "Procesamiento del lenguaje natural", dos nombres para una misma disciplina, y para ello nos ceñimos a lo que consideramos principal reflejo de las investigaciones más avanzadas en cada ciencia, esto es, los artículos de revistas y las tesis doctorales.

El período abarcado discurre desde 1976, hasta nuestros días, sentando así las bases para el análisis tanto cuantitativo como cualitativo de los trabajos publicados en lengua española en los últimos 16 años. A ello hemos querido también contribuir, al disponer una sección cronológica dentro de cada sección temática.

En cuanto a las fuentes bibliográficas utilizadas, se han extraído referencias de bibliografías especializadas, concretamente las bases de datos del Consejo Superior de Investigaciones Científicas (CSIC), y hemos de señalar que la información principal ha sido hallada en la base de Ciencias Sociales y Humanidades (ISOC). Se han vaciado manualmente publicaciones periódicas contenidas en la hemeroteca de la Facultad de Letras y de la Escuela de Biblioteconomía y Documentación de la Universidad de Granada, susceptibles de poseer información pertinente, y también ha sido consultada la edición impresa de TESEO, Tomo 2: Ciencias y Técnica, editado por el Ministerio de Educación y Ciencia, y que abarca tesis publicadas desde 1976/77 hasta 1988/89.

Las referencias bibliográficas siguen la norma ISO 690, siempre que en la fuente consultada se encontraran los datos suficientes, y se han incluido descriptores, sólo cuando se hallaban en dicha fuente.

A continuación se ofrece una síntesis del tipo de repertorio que presentamos:

1. Por el método de elaboración:

* Se trata de una bibliografía "de 2ª mano", pues las fuentes bibliográficas consultadas son "secundarias".

* Es "selectiva", ya que sólo recoge artículos de publicaciones periódicas y tesis doctorales o proyectos fin de carrera.

2. Por la naturaleza de los documentos reseñados:

a) Según el contenido es una bibliografía "especializada", pues se dedica a una materia concreta.

b) Según la lengua, es "nacional", ya que sólo recoge documentos en lengua española.

c) Según la fecha de publicación de los documentos reseñados, abarca, como se ha indicado, de 1976 a 1992. Podemos decir por tanto que se trata de una bibliografía actual.

3. Por la naturaleza de las referencias: es "signalética", ya que se indican estrictamente los elementos bibliográficos.

Acerca del procedimiento empleado en la ordenación de las referencias, se ha optado por estructurar el trabajo en 6 áreas temáticas, que aunque por supuesto se encuentren interconectadas entre sí, ayuden sin embargo a esquematizar la materia.

Estas áreas son las siguientes:

1. LINGUISTICA COMPUTACIONAL: GENERALIDADES.
2. ANALISIS FONETICO, MORFOLOGICO, SINTACTICO, SEMANTICO Y/O PRAGMATICO.
3. LEXICOGRAFIA COMPUTACIONAL.
4. TRADUCCION AUTOMATICA.
5. RECONOCIMIENTO AUTOMATICO DEL HABLA Y CONVERSION DE TEXTO A VOZ.
6. COMPRENSION Y ANALISIS DE TEXTOS.

Dentro de cada área se incluyen en primer lugar los artículos de revista, y en un segundo grupo las tesis doctorales o proyectos fin de carrera.

A su vez, los artículos y tesis han sido dispuestos en orden cronológico, de los más antiguos a los más actuales.

Para completar el trabajo se añaden finalmente 3 índices, uno de títulos de revistas

vaciadas, un segundo listado de autores en orden alfabético, y un tercero de las Universidades en que han sido realizadas las tesis que se recogen, esperando así proporcionar toda la información de interés sobre los artículos incluidos.