

JOSÉ VIDAL BENEYTO, dir., *Las industrias de la lengua*, equipo de traductores coordinado por Manuel Alvar Ezquerro (Biblioteca del Libro, V), Salamanca - Madrid, Fundación Germán Sánchez Ruipérez, Madrid, Pirámide, 1991, 496 págs.

El volumen ofrece un variado panorama de estudios, al tiempo que da la imagen de un campo todavía un tanto inestable. La propia variedad nace de la gran cantidad de colaboraciones contenidas en el libro. Muchas de ellas reflejan investigaciones en marcha, proyectos en diversas fases de realización, con las inconexiones, retrocesos y vueltas de hoja de un campo todavía móvil y en proceso de definición. Aunque la materia presentada bien puede pertenecer a varios campos diferentes (constitución de *corpora* de referencia, lexicografía automatizada, etc.), están unidos por unas necesidades técnicas comunes, además de "por su utilidad industrial", como reza el propósito del volumen.

Una deficiencia de la compilación es que no refleja el estado de la cuestión tan al día como el lector interesado pudiera desear. Esto no deja de ser un contrasentido en campo que tanto necesita de una veloz transmisión de la información, y es ejemplo de la flexibilidad aún insuficiente de nuestras vías tradicionales de transmisión académica.

Debe aclararse que la compilación intenta exponer el estado del arte en Europa, o más exactamente en la Europa económica, intentando abarcar las perspectivas de lingüistas, informáticos, editores, industriales y organismos oficiales competentes.

El volumen contiene ocho secciones, además de una introducción, ésta a cargo de JOSÉ VIDAL BENEYTO<sup>1</sup>. Estos ocho capítulos o secciones tienden a cubrir un espectro más que amplio de problemas: el tratamiento automático del lenguaje natural (págs. 37-92); la constitución de *corpora* de referencia. Escritos y orales (págs. 9-124); elaboración de diccionarios electrónicos y sus aplicaciones (págs. 125-240); los inventarios terminológicos (págs. 241-264); ejemplos de cooperación europea (págs. 265-309); la comunicación hablada: reconocimiento y síntesis del habla (págs. 311-407); la enseñanza de la lingüística informática (págs. 409-427); muestras de una industria (págs. 429-494).

Vidal Beneyto recuerda en la introducción la importancia económica de diversas actividades que implican tratamiento de lenguas –valga por caso la traducción– y los programas de investigación aplicados al desarrollo

---

<sup>1</sup> Introducción: *La industrialización de las lenguas* (págs. 11-35).

automatizado de esos tratamientos, centrándose en particular en los logros y en los problemas europeos. De hecho, la ventaja y la desventaja europea aparecen recordadas en varias de las colaboraciones del libro. Europa dispondría, a su favor, de una situación multilingüe y multicultural, que debería funcionar como estímulo para las instituciones y, obviamente, como un inmenso mercado; además, se contaría con equipos y personalidades de reconocida valía científica. Sin embargo, la propia dispersión internacional y lingüística de esos equipos y la compleja selva de instituciones puede haber estado frenando el desarrollo europeo en este terreno. Diríase, en este sentido, que lo que el tono de todo el volumen deja entrever es que son las exigencias del propio mercado las que dirimirán el desarrollo de proyectos y aplicaciones económicamente adecuadas, sea cual sea la nacionalidad del producto, europea, estadounidense o japonesa. Por otra parte, el desarrollo industrial de las lenguas necesita una investigación de base de medio y largo plazo, de forma que sólo los laboratorios financiados públicamente pueden afrontar los retos del sector (esta necesidad de años de labor frustró el optimismo surgido en los primeros tiempos de trabajo en automatización lingüística). Por lo demás, podemos pensar que esta actividad lingüística europea interesa al menos a buena parte del resto del mundo, dada la extensión planetaria de varias de las lenguas de origen europeo. Y esto nos lleva a la preocupación por el trabajo con y sobre variedades americanas del español, realmente mayoritarias desde el punto de vista demográfico.

*El tratamiento automático del lenguaje natural*, primera sección del volumen, presenta varios panoramas más o menos de conjunto de cierta utilidad<sup>2</sup>. Entre las posibles aplicaciones del campo se encuentran la traducción automática, la realización de interfaces en lengua natural, la ayuda a la preparación de documentos, la indexación automática de textos, las instrucciones a robots, la enseñanza asistida por computadora

---

<sup>2</sup> Forman parte de esta sección los trabajos de GÉRARD SABAH, *El tratamiento del lenguaje escrito* (págs. 39-54); HENRY THOMPSON, *Análisis del texto: situación actual* (págs. 55-63); JEAN SYLVAIN LIÉNARD, *Estado del arte en comunicación hablada* (págs. 64-77); MAURICE GROSS, *Un nuevo tratamiento del francés escrito* (págs. 78-85); MAURICE NIVAT y DOMINIQUE PERRIN, *Los métodos informáticos aplicados al tratamiento de las lenguas naturales* (págs. 86-92). Puede encontrarse un amplio panorama sobre lingüística automatizada en ISTVÁN S. BATORI, WINFRIED LENDERS y WOLFGANG PUTSCHKE, eds., *Computational Linguistics*, Berlin - New York, Walter de Gruyter, 1989. Véase también Jacques Anis, ed., *La génération de textes. Langages*, 106, 1992. Un trabajo reciente sobre el español es JOS HALLEBEEK, *A Formal Approach to Spanish Syntax*, Amsterdam - Atlanta, Rodopi, 1992.

(v. SABAH, págs. 39-41). Todas estas aplicaciones comparten, sin embargo, la necesidad de disponer de abundante y explícita información lingüística; en ese sentido, la investigación previa necesaria aporta una base común de conocimiento aplicable de muchas maneras. Entre las múltiples y colosales dificultades de esa tarea representacional lingüística, destaca el problema de la representación del conocimiento del mundo. Nada más complejo que modelar en algoritmos lo que un hablante sabe sobre la realidad, lingüística o extralingüística; esta actividad modelizadora es la meta última de la automatización del lenguaje natural.

Sección especialmente interesante es la dedicada a *La constitución de corpora de referencia. Escritos y orales*<sup>3</sup>. Dado el alto costo económico y de tiempo<sup>4</sup>, el ideal con un corpus es que pueda emplearse para muchos fines diferentes, y que de él puedan extraerse segmentos de datos comparables a los producidos a través de corpus de otras latitudes. Se dispone ya de varios *corpora* del español, todos ellos de extensión modesta, si se comparan con los cien millones de palabras previstas en el corpus de inglés británico en proceso de recolección<sup>5</sup>. Como en toda la compilación, auspicia también este capítulo la idea de que una lengua internacional necesita, para serlo, ser sometida a un proceso de "industrialización"; en este contexto, el primer paso para disponer de una lengua competitiva es preparar un amplio corpus de ella<sup>6</sup>.

*La elaboración de diccionarios electrónicos y sus aplicaciones*<sup>7</sup> nos introduce ya en el terreno de los productos. Se insiste en esta sección

---

<sup>3</sup> JOHN M. SINCLAIR, *Creación de corpus* (págs. 95-107); RENÉ CARRÉ, *Los bancos de sonidos* (págs. 108-118); ANTONIO ZAMPOLLI, *Corpora de referencia* (págs. 119-124). El trabajo de Sinclair es especialmente claro para hacerse una buena idea del problema.

<sup>4</sup> Las dificultades aumentan según las características del corpus deseado. Es obvio que un corpus de lengua hablada es mucho más trabajoso de obtener. Sobre corpus, véase Jan Aarts y Willem Meijs, eds., *Theory and Practice in Corpus Linguistics*, Amsterdam, Rodopi, 1990; S. ATKINS, J. CLEAR y N. OSTLER, *Corpus design criteria*, *Literary & Linguistic Computing*, 7, 1 (1992), págs. 1-16.

<sup>5</sup> El proyecto del British National Corpus comenzó en enero de 1991; tiene una previsión de tres años y aspira a recoger una muestra representativa de un amplio conjunto de estilos de inglés hablado y escrito.

<sup>6</sup> Véase, por ejemplo, SINCLAIR, pág. 104.

<sup>7</sup> ANTONIO ZAMPOLLI, *Los bancos de datos léxicos: bases multifuncionales de datos léxicos* (págs. 127-146); MAURICE GROSS, *La elaboración de los diccionarios electrónicos del francés* (págs. 147-184); ANTONIO ZAMPOLLI, *Hacia bases multifuncionales de datos léxicos* (págs. 185-202); LAURENCE DANLOS, *La generación automática de textos en lengua natural* (203-218); MANUEL ALVAR EZQUERRA, *Desarrollos actuales en lexicografía automatizada del español* (págs. 219-240).

en una peculiaridad obvia: no todo diccionario en soporte electrónico es propiamente un diccionario electrónico; puestos a buscar una diferencia sustancial, ésta es, seguramente, la explicitud. Un diccionario o una gramática tradicionales, en sentido amplio, dejan de decir muchas cosas, bien porque se presupongan, porque sean evidentes o, simplemente, porque numerosas cuestiones no son tratadas. Pero si se aspira a aplicar de diversos modos una léxico-gramática electrónica, la computadora ha de tener completa y explícita toda la información, por obvia que hubiera sido para un lector humano. Sin embargo, lo cierto es que entre ambos extremos—el tradicional y el electrónico—existen, en la práctica, muchos productos intermedios; sin duda, esto se debe a la necesidad de rentabilizar las inversiones y, en cierto sentido, a la necesidad de *hacer lo que se pueda allí donde se pueda*. Una buena perspectiva sobre la metodología de trabajo electrónica puede encontrarse en la aportación de M. Gross. La lexicografía automatizada del español es presentada por M. Alvar Ezquerro, describiendo el proceso de elaboración de la nueva edición del *Diccionario general ilustrado de la lengua española* Vox (DGILE), trabajo actualmente “convertido en el centro de una actividad mucho más amplia” (pág. 240).

Las secciones dedicadas a *Los inventarios terminológicos* y a *Ejemplos de cooperación europea* tienen la utilidad de proporcionarnos información sobre las instituciones europeas envueltas en el sector<sup>8</sup>. La actividad terminológica tiene un importante papel normalizador; cualquier disciplina enfrenta hoy día el crecimiento espectacular de sus tecnicismos, y la comunicación científica puede llegar a ser enormemente complicada. El problema, entre otros, es que el contenido de numerosos términos es precisamente cuestión de debate en la disciplina en que son tecnicismos (¿qué es una palabra?) La cooperación y la coordinación de esfuerzos en Europa está, precisamente, en el camino de la homologación de

---

\* Forman la sección dedicada a terminología los trabajos de CHRISTIAN GALINSKY, *Terminología y sociedad moderna: el papel de INFOTERM* (págs. 243-251); WOLFGANG NEDOBITY, *Terminología de las ramas e ingeniería del conocimiento* (págs. 252-256); WOLFGANG NEDOBITY, *Perspectivas para la elaboración de una metodología común en la descripción del lenguaje* (págs. 257-264). La cooperación europea se ilustra con las aportaciones de MAURICE GROSS, *Red de laboratorios europeos. Construcción coordinada de léxicos-gramáticas* (págs. 267-269); MAURICE GROSS, *Los trabajos de la red europea de las industrias de la lengua en el campo de los diccionarios y gramáticas electrónicos* (págs. 270-282); LOLL ROLLING, *La Comisión de las Comunidades Europeas y las industrias de la lengua* (págs. 283-287); JEAN-FRANÇOIS DÉGREMONT, *Problemas en la descripción del ámbito de las industrias de la lengua. Algunas propuestas metodológicas* (págs. 288-309).

metodologías. Sólo por esa vía puede ahorrarse tiempo y dinero, y aspirar a construir cualquier clase de banco multilingüe de datos.

Entenderse con las máquinas a través de la lengua hablada es el problema central de *La comunicación hablada: reconocimiento y síntesis del habla*<sup>9</sup>. Ampliar las vías de comunicación más allá de la entrada vía teclado y la salida por el monitor o gráfica, es una tarea enormemente complicada. Ni siquiera es seguro que la comunicación hablada con la máquina vaya a ser más 'natural' que la mecánica, sobre todo si se piensa en ir más allá de las órdenes simples y precisas. Los productos ya elaborados, comercializados o no, disponibles en el momento de redactar las aportaciones de esta sección, parecen bastante limitados. Sin embargo, algunos programas de investigación confiaban en realizar avances sustanciales en los años siguientes (v. el gráfico que ofrece Mariani, pág. 369).

La perspectiva abarcadora del volumen explica la inclusión de las dos últimas secciones dedicadas, respectivamente, a *La Enseñanza de la lingüística informática*<sup>10</sup> y a *Muestras de una industria*<sup>11</sup>. Respecto a lo primero, la necesidad de estas enseñanzas es evidente en cualquier latitud. No es fácil, desde luego, romper las barreras disciplinarias tradicionales, coordinar esfuerzos y disponer del personal docente adecuado. El papel de las aplicaciones industriales nos introduce en un mundo apasionante, como cuando se habla de la máquina de escribir de entrada vocal o de las mejoras en la verificación y corrección ortográfica. Desde un punto de

<sup>9</sup> GUNNAR FANT, BJÖRN GRANSTRÖM y ROLF CARLSON, *La síntesis del habla como componente de la tecnología del habla y de los sistemas de información* (págs. 313-325); JOSEPH MARIANI, *Las tecnologías del lenguaje* (págs. 326-369); RENÉ CARRÉ, *Perspectivas de investigaciones en comunicación hablada* (págs. 370-374); JOHN LAVER, *Nuevos horizontes de la tecnología europea del lenguaje* (págs. 375-407).

<sup>10</sup> BERNARD CASSEN, JEAN-FRANÇOIS DÉGREMONT, HELMUT SCHNELLE y ANTONIO ZAMPOLLI, *Formación del personal investigador y estudios de doctorado en lingüística computacional* (págs. 411-415); MAURICE GROSS, *La informática lingüística y su enseñanza* (págs. 416-421); ANTONIO ZAMPOLLI, *Hacia un doctorado europeo en lingüística computacional* (págs. 422-427).

<sup>11</sup> PIERRE LE LOARER y BERNARD NORMIER, *ERL y el lenguaje natural* (págs. 431-454); ERIK LAMBERT, *La máquina de escribir de entrada vocal* (págs. 455-461); JEAN FRANÇOIS DÉGREMONT, *Verificación y corrección ortográfica informatizada. Un desafío para las lenguas europeas* (págs. 462-469); CHRISTINE POITEVIN, *Las industrias de la lengua: problemas de armonización jurídica de las lenguas europeas* (págs. 470-477); ROBERT MAHL, *El proyecto nacional de traducción asistida por ordenador* (págs. 478-488); PHILIPPE AMIEL, *Bases de datos lexicográficos en Europa: Eurolexic, proyecto europeo de origen editorial* (págs. 489-494).

vista industrial, las actividades de una empresa como ERLI, la primera francesa interesada en las aplicaciones de la lingüística automática, creada en 1974<sup>12</sup>, que ofrece productos como ALEXIS (que puede utilizarse como sistema de gestión de bases de datos lexicográficas o terminológicas, como gestor de *thesaurus*, bases documentales y nomenclaturas<sup>13</sup>, o el proyecto EUROLEXIC, que agrupa a editores franceses, británicos, españoles e italianos, y a sociedades francesas de servicios, son claras muestras de una de las tesis básicas de la compilación: la rentabilidad de la automatización lingüística, la existencia de una industria interesada en producir, y de un mercado dispuesto a consumir esos productos. De alguna forma, se propone una ecuación de equivalencia entre lengua internacional, lengua automatizada e industrializada y lengua rentable.

En conjunto, puede decirse que el volumen es interesante, aunque haya exceso de repeticiones y sean muy diferentes el tono y el estilo de los artículos, que oscilan entre la redacción académica y el informe apresurado; queda también en el aire la *plena* actualidad de la información suministrada, por los inevitables desfases editoriales. En cualquier caso, servirá para que industriales y lingüistas aún no implicados en el sector se hagan una idea multiforme de las perspectivas del inevitable futuro próximo.

PEDRO MARTÍN BUTRAGUEÑO

CSIC y  
El Colegio de México.

MARÍA BEATRIZ FONTANELLA DE WEINBERG, MERCEDES BLANCO DE MARGO, YOLANDA HIPPERDINGER, ELIZABETH RIGATUSO, SILVIA SUARDÍAZ DE ANTOLLINI y ANA VIRKEL DE SANDLER, *Lengua e inmigración: mantenimiento y cambio de lenguas inmigratorias*, Bahía Blanca, Departamento de Humanidades, Universidad Nacional del Sur, 1991, 143 págs.

Este libro lo constituyen una serie de estudios sobre la evolución idiomática de comunidades inmigrantes a la Argentina (italianos, alemanes, rusos, daneses, judíos), sobre las condiciones que han favorecido la conservación de la lengua originaria o su desplazamiento por el español.

---

<sup>12</sup> V. LE LOARER y NORMIER, pág. 431.

<sup>13</sup> V. LE LOARER y NORMIER, pág. 434.