Las bases del español

AGUSTÍN JIMÉNEZ MUÑOZ*

Servicio de Traducción, Comisión Europea, Bruselas
Tel. 0032-2953661. Agustin.Jimenez@cec.eu.int

Resumen

En un espacio informativo dominado por Internet, un texto sólo existe electrónicamente. Internet ha hecho nacer una metodología de trabajo particular y una apariencia de control de los datos. El efecto final de esa corriente han sido, desde los años noventa, los corpora. La base de datos es un caso especial de corpus. Hija del diccionario, ha perdido el prestigio intelectual que éste representaba. Contemporánea de Internet y de la multiplicación universal de enlaces, debe enfocar la terminología desde perspectivas nuevas pero retomando la tradición y la profesionalización lingüística perdidas. Ello exige un esfuerzo intelectual nuevo y un análisis computacional en el que español lleva retraso.

Palabras clave: base de datos, diccionario, texto, corpus, sinonimia, terminología técnica

1 Una convergencia sin método

Ya que hablamos de convergencia terminológica, lo más natural es debatir la función de las bases de datos, adonde convergen mecánicamente los términos desde que se inventó la posibilidad de almacenarlos electrónicamente y, mucho más, desde que se pudo distribuirlos a todo el mundo, especialistas o no, en la mayoría de los casos gratuitamente, por Internet. Internet lo hemos incorporado a nuestras vidas con una alegría tecnológica sin precedentes. Sin precedentes y sin memoria. Parece que siempre ha estado ahí, y hoy no concebiríamos el mundo sin él, y, sin embargo, es un invento jovencísimo. Netscape salió al mercado en 1994. Explorer, el imprescindible «browser de Microsoft» —¿qué vamos a

hacerle?, seguimos diciendo «browser»—, no se gestó hasta 1995. La única forma de atribuirle longevidad al asunto es reconocer, como ha sugerido un autor americano (Tom Paquin), la existencia de un «tiempo Internet» donde el ciclo de desarrollo se realiza a velocidad ultrarrápida y donde la duración real de un año se multiplica por siete como en la vida de los perros.

La magia de Internet, su combinación de almacenamiento ilimitado de datos y correo electrónico, nos ha hecho perder un poco la cabeza. Un mínimo de experiencia en gestión de bases nos obliga, no obstante, a recapacitar que ni siquiera el procesamiento de los datos es un asunto resuelto. Subsisten aún carencias notables en conversión de caracteres, en indización, en pasarelas entre aplicaciones. Los tiempos de respuesta se alargan sin remedio en proporción a la complejidad de los datos que se buscan —lo que dificulta la opción, tan en boga, de la integración de herramientas—, y esas carencias se amplifican cuando, siguiendo una moda optimista, todas las operaciones se efectúan en Internet. Una base de corte tradicional como Eurodicautom, aun después de iniciar su distribución en Internet, mantiene separadas una base de producción, receptáculo de datos en el que se lleva a cabo la alimentación, y una base de difusión, copia de la anterior, que es la que el usuario normal consulta en la Red. La futura base interinstitucional de la Unión Europea, que englobará todas las bases que ahora existen en las diferentes instituciones, se alimentará y se distribuirá directamente en la Red. En parte, una opción así se explica por otra ilusión, un poco simplista, de la que nos hemos enamorado en los últimos años: la interactividad, el derecho universal a que cualquier usuario alimente la base. La interactividad universal es una opción defendible en una serie de contextos, tiene consecuencias extraordinarias en la rapidez de recuperación y actualización de los datos y contribuye a la formación, y esto es esencial en los foros de usuarios, de una comunidad lingüística. Como modelo de gestión, el ejemplo de la base Logos nos hace soñar a todos los que trabajamos en bases o en normalización lingüística. Pero esta apertura de opciones hace más acuciante plantearse dos o tres problemas centrales de Internet y, en particular, el problema de la autoridad. Pues sobre la doctrina de la autoridad se ha fundado hasta ahora toda la tradición cultural. ¿Quién decide que un dato, que un término es correcto? ¿De dónde provienen una acepción o un giro que vemos documentado y sin referencia por primera vez?

Si nos referimos al español, el desarrollo de la lingüística computacional en español no ha ido probablemente parejo con la importancia y la difusión de nuestra lengua. Echamos en falta una tecnología lingüística propia que nos libere de los disparates, y de la

lentitud subsiguiente, que provocan sistemas de recuperación de información elaborados a partir del inglés. Los responsables de bases hemos incurrido en diversos pecados de papanatismo fiándonos del prestigio de diversos sistemas informáticos que, a partir de una lematización elaborada para el inglés, nos han vendido alegremente trucos para localizar desinencias gramaticales en español de la manera más pesada e inconsistente. Uno de esos famosos sistemas de indización está obligado a analizar más de cuatrocientas variaciones, ninguna pertinente, para dar con un plural español. A propósito de indización, convendría alertar de que, frente a las pretensiones de algunas empresas, las soluciones más simples son a menudo las más adecuadas. En la mayoría de los casos el problema de localizar una desinencia plural se ataja mucho más eficazmente con un simple truncamiento que barajando una ristra de variantes morfológicas que ocupan la memoria del servidor durante segundos interminables.

2 Gloria y desgracia de la información electrónica

Sea como sea, la rutina —y posiblemente el sentido común— impone que cualquier información se descargue en una base de datos. La rutina, que no el aprecio intelectual. Hasta la fecha, en información lingüística y probablemente en la mayoría de los campos del saber, a un receptáculo electrónico de información sólo se le concede valor intelectual si es una copia de una obra prestigiosa previamente ejecutada en papel. Es el caso de la *Encyclopaedia Britannica* o del *Diccionario* de la Real Academia. El prestigio intelectual de las obras realizadas directamente para soporte electrónico es, en la mayoría de los casos, limitado. En general su valor intelectual ni siquiera se plantea. De una base de datos nadie afirma que sea rigurosa, que obedezca a una metodología original ni que limpie, fije y dé esplendor. Se aplaude meramente, por este orden, su abundancia de entradas, su rapidez de acceso y que la presentación de la información no sea engorrosa. En lo que toca al número de entradas, es decisivo el factor psicológico del traductor técnico. Cada lengua cuenta con uno o varios diccionarios que, con escasas excepciones dictadas por la vetustez o el regionalismo de un término, ofrecen el abanico prácticamente completo de las palabras utilizables. Con un diccionario así, la comprensión de un texto general o la calidad de una traducción literaria dependerán casi exclusivamente del nivel de competencia lingüística del lector o del traductor. Pero, en mayor o menor medida, a quien lee textos técnicos,

sobre todo si no es un especialista, lo desasosiega sin parar la cantidad de cosas que no entiende y, mucho peor, que no sabe dónde encontrar. Si la situación es complicada para el simple lector que, al cabo, puede escurrir el bulto, tanto más lo es para el traductor, sometido por obligación profesional a ser un lector completo. Tal lector, tal traductor se sentirá aliviado en cuanto alguien le dé una pista. Y en ningún sitio encontrará tantas pistas y tan rápidamente como en Internet. Lo que, pese a sus posibles errores, explica el éxito de las grandes bases técnicas.

Una vez encontrado el término esquivo, el traductor no siempre entra en otras consideraciones. La base de datos es una más de las herramientas de traducir. Si el diccionario de D. Manuel Seco es universalmente aclamado como una proeza, Eurodicautom se conformaría con ser una buena herramienta. El término «herramienta», tan popular en las comunidades de traductores modernos, refleja una concepción de la traducción como industria, opción legítima pero que soslaya el aspecto intelectual de la operación. Y, si traducir es un proceso mecánico que cualquiera puede llevar a cabo si cuenta con una buena caja de herramientas, no es necesario en absoluto que una base de datos cuente con un Seco para supervisarla. Sin darnos cuenta, hemos llevado a cabo una rebaja, rebaja que hoy en día es bastante general. La lingüística y la metodología están muy bien —aducen muchos—, pero eso son cuestiones académicas y nosotros nos preocupamos por asuntos más prácticos. Confesadamente o no dividimos la lengua en dos grandes capítulos: la generalidad académica de los diccionarios, de Menéndez Pidal o de Manuel Seco, y la terminología técnica, que la mayoría llama simplemente «terminología». Y, en efecto, la mayoría de las bases de datos no se ocupan de cuestiones generales de lenguaje sino de «terminología». No tienen autoría, salvo, todo lo más, una colectiva cuando se escudan en una institución —pero institución administrativa y prestigio intelectual no siempre son sinónimos— y, cuando se pretende profesionalizar el trabajo, acuden a conceptos teóricos deslavazados leídos en un repertorio de normas recolectadas por ingenieros. Algunos piensan que el árbol de Porfirio es una norma ISO. Que el pobre Porfirio, filósofo y vegetariano, viviera en el siglo III y que sus diagramas, que él no llegó a realizar, puedan ya deducirse de las teorías de la división de Platón o de las categorías de Aristóteles, no tiene ninguna importancia.

3 Un lío conceptual

¿Qué es, pues, la terminología? Es esa disciplina lingüística que, contra lo que cabría pensar, no se ocupa de lengua ni, menos aun, de disciplina, sino de preparar, apoyada en una triste simplificación de la filosofía más burda, la redacción de textos más o menos administrativos. Objetivo principal de la terminología seria es la normalización. Nadie negará las ventajas de la normalización. Sin normalización, los aduaneros no sabrían cómo tasar las mercancías que pasan ante sus narices ni un aficionado a las manualidades seleccionaría con tino el número de roscas de la tuerca que precisan. A nivel nacional y, al menos en la intención, en el concierto internacional, es conveniente que los actos jurídicos o los títulos académicos se llamen siempre de la misma manera para facilitar los intercambios. Resta el pequeño detalle de decidir qué hacemos cuando la gente no respeta los nombres impuestos, lo que sucede a diario, o cuando interesadamente diferentes comunidades emplean nomenclaturas diversas. Las empresas de traducción saben que un mismo texto debe traducirse diferentemente según el cliente que lo haya solicitado.

Otra tendencia de la terminología técnica es considerar que las frases están compuestas exclusivamente de nombres sustantivos. Un terminólogo serio no tratará nunca los problemas que plantean frases como «Tarzán querer Jane» o «Chita mala», que carecen de relevancia lingüística, ni tampoco ocuparse en reseñar las frases hechas de la jurisprudencia, por no citar ocupaciones menos nobles. Para los teóricos de la «terminología técnica», los sustantivos presentan la ventaja indudable de que, detrás de cada uno se esconde un concepto. En las bases de datos multilingües, ajenas a todo relativismo, cuando se completa una ficha en una lengua, no se busca una correspondencia sino que «se traduce un concepto». Saussure simplificó la concepción de los signos en un enfoque dual que incluía un significante y un significado. Desde que Ogden y Richards sacaron al mercado su famoso triángulo, un vértice de éste se dedica a albergar el concepto, tan evidente como un dibujo o como el número exacto de roscas de un tornillo. En realidad la teoría del concepto es o mucho más sencilla o mucho más compleja. Kant postulaba ya un cierto mentalismo para que el signo fuera intersubjetivo. Está claro que, sin algún tipo de contenido mental común, sería imposible que dos interlocutores mantuvieran otra conversación que un diálogo de besugos. Pragmatistas como Searle han adoptado posiciones extremas en este campo. Pero, fuera de la pragmática, el concepto resulta ser un lastre para avanzar en la metodología. Wittgenstein quería una lógica sin conceptos. Umberto Eco ha criticado en diversos textos lo que él llama la «falacia referencial» y ha propuesto

superar el esquema dual de Saussure con la adición de un elemento —o una red de elementos— introducido por Peirce, el fundador de la semiótica. El «interpretante» de Peirce puede entenderse, en nuestro caso, como la red de relaciones lingüísticas explícitas que autorizan una correspondencia entre dos términos.

Íntimamente relacionada con el enfoque conceptual está la discusión del término. La mayoría de los diccionarios tradicionales priman la presentación por palabras aunque presentan igualmente, clasificadas como «expresiones» una serie de cadenas lexemáticas de las que las palabras son consideradas núcleo. Huelga decir que, si el núcleo no está claro, la «expresión» está ausente. Con notables excepciones —como el diccionario Longman de lengua inglesa—, estas obras no incluyen, ni siquiera como ejemplo, expresiones como «Lo que el viento se llevó» y mucho menos «la dama de hierro» o «la madre de todas las batallas», nomenclaturas o frases hechas, ristras de palabras que «no son conceptos». Pero un enfoque más formal, constructivista o puramente sintáctico, de la terminología debería considerar que un término es cualquier cadena fija de expresión. Según esto, cualquier entrada de un diccionario es un término, incluido un artículo o una conjunción y se abre la cuestión de saber cuándo una cadena de palabras lo es, aunque no su represente ningún concepto.

Los terminólogos aceptan de buen grado —de hecho, priman— los términos compuestos pero, en vez de justificarlos por su grado de lexematización, los incluyen o los excluyen según «representen o no un concepto». Así, en los repertorios terminológicos abundan entradas cuyo único mérito es «que son cosas muy importantes». Una lista de términos puede proponer «bandera de Francia» y, a continuación, «bandera de Escocia» y, después, «bandera de Castilla-La Mancha». Mucho más productivo que hacer una ficha con esas acepciones es investigar criterios objetivables que midan el grado de lexicalización de un grupo de palabras, del tipo de los propuestos por Berkema (flexibilidad) o por Gross (nominalización). Por supuesto, la lexicalización de un término no tiene por qué ser absoluta. La lingüística moderna se ocupa cada vez más de la «variación terminológica», pues con frecuencia la cadena del lexema se realiza en diversas distribuciones. Así, puede decirse «libertad de circulación» o «libre circulación», «derecho de voto» o «derecho a voto». Es interesante comprobar que el género de problemas que introduce la variación coincide con los que se plantean en la realización de los lexemas compuestos de la traducción automática. La integración en las bases de datos de sistemas de traducción automática permitiría un tratamiento común en áreas de interés. Un sistema de traducción automática

incorporado podría también ayudar a localizar fichas existentes en la base en una lengua diferente o analizar palabras clave en una definición para evitar dobles o relacionar sinónimos.

4 Del diccionario al diccionario pasando por la base de datos

Al principio fue el diccionario, que recogía esas palabras necesarias para entender una conversación y, preferentemente, un libro. El nacimiento del diccionario dio notoriedad al lexicógrafo que, con pocas excepciones —San Isidoro de Sevilla, a quien, por cierto, la Iglesia declaró patrón de Internet en 1999—, había sido un oficio anónimo. El diccionario fija las palabras —una tarea esencial para la Iglesia, que necesitaba conservar inalteradas las Escrituras—, cimenta los logros obtenidos por las armas —véase el prólogo al *Dictionnaire* de l'Académie française—, fomenta la autoridad —*Diccionario de Autoridades*—, funciona como estabilizador del gusto y, un poco, como resumen de todas las cosas. La llegada de las enciclopedias, que, decididamente, querían explicar el mundo —e incluso entenderlo: Novalis—, anuncia una cierta confusión sobre la definición de término. Las enciclopedias se presentan ordenadas en torno a entradas temáticas que, en muchos glosarios y bases de datos posteriores, serían consideradas términos. La enciclopedia produjo un cruce con el diccionario, el diccionario enciclopédico, y hoy hay terminólogos que, cuando redactan sus fichas, tienen dificultades en diferenciar la información lingüística de la puramente contextual. Los siglos XVIII y XIX son testigos de un acrecentamiento del léxico técnico que acompasa la progresiva industrialización de la vida cotidiana, y de abundantes gigantes de la lexicografía: un Samuel Johnson, un Littré son personajes impresionantes por su vocación, su preparación, su capacidad de trabajo. Sus biografías —a Samuel Johnson le dedicó una literaria su amigo, el escritor James Boswell— resaltan el carácter heroico del lexicógrafo. Samuel Johnson extrajo ese carácter de Henri Estienne, uno de los que, en el siglo XVI, dignificaron una labor que, hasta entonces, con contadas excepciones, como la de San Isidoro de Sevilla, había sido realizada anónimamente. Anonimato que regresó a la escena el siglo XX de manos de la terminología técnica. Pese a su torpeza metodológica, las escuelas de terminología del siglo XX tuvieron el mérito de focalizar la atención en la cuestión avasalladora de la terminología técnica. Para los diccionarios clásicos es una cuestión menor. En su estupenda *Introducción*

a la lexicografía, Julio Casares observa la inestabilidad de dicha terminología y reconoce el «descuido» en que se la ha tenido, aunque acaba recomendando que «procede... recoger tecnicismos a manos llenas».

Es lo que han tratado de hacer las bases de datos, que, no sólo se centraban en esta terminología, sino que tenían la capacidad de alimentarse rápidamente y de, liberadas del lento lastre del papel, reeditar de continuo la información: «El diccionario que no se renueva, acaba muriéndose«- recuerda Manuel Seco. Sin la informática y sin las bases de datos, una renovación continua es muy problemática. Pero esa alimentación a manos llenas no se ha hecho sin costes. O bien se hacía con metodología lingüística escasa o bien caía a su vez en el «descuido»de relegar una parte del vocabulario, despreciando el léxico general. Pero ¿cuál es el criterio de inclusión de un término?

5 Deducción, inducción, estadística y corpora

Como en cualquier actividad científica, el conocimiento léxico se acrecienta según las reglas de la deducción y de la experiencia. A la hora de emprender la redacción de un diccionario, los clásicos, como Casares, tomaban una enciclopedia que ordenaba el mundo de más a menos, de Dios al último representante del reino vegetal, e iban rellenando casillas de la ontología. Después abrían una lista de categorías gramaticales, otro modo entre filosófico y lógico de forzar una ontología, y realizaban la misma operación. Por último, realizaban una clasificación de estados semánticos y llegaban a otra ontología llamémosla semántica que, con criterios que oscilan entre las categorías de Aristóteles y la reciente teoría de prototipos, ensamblaban una retahíla de semas elementales. Es lo que hacía, en los años setenta, un Nida. La estadística vino a liberarnos de esas complejidades. Un anotador contable o, mejor aún, una máquina, recorría textos y seleccionaba las expresiones que aparecían más veces. En la práctica, la alimentación de las bases de datos procede muchas veces por inducción total. Se examinan documentos nuevos, seleccionados según criterios vagos de «importancia» o novedad, y se introducen en la base todos los términos no consignados (neologismos). Paralelamente ha quedado de manifiesto la insuficiencia de la categorización excesiva. Los documentalistas soviéticos intentaron construir sistemas que sirvieran tanto para la búsqueda documental como para la científica. Pero parece

demostrado que, cuanto más científico es un vocabulario, menos eficaz es su clasificación y su recuperación.

En los años noventa la lingüística se embala con la moda de los corpora. Un corpus es una simple colección de textos a los que, por formar un conjunto completo o por ser considerados representativos, se les da el valor de repositorios significativos de valores lingüísticos que se examinan según intereses distintos. Así, por ejemplo, el examen del vocabulario del corpus de historiales médicos ha permitido extraer conclusiones para mejorar la asistencia médica en Nueva York. El análisis del corpus permite un análisis circunstanciado de tipologías comparadas, imprescindible para la teoría de la traducción. Es lo que la lingüística ha practicado desde siempre. La afición a los corpora es el simple resultado de un enfoque formal de la textualidad que, desde los formalistas rusos de los años veinte, al constructivismo o a la narratología, atraviesa todo el siglo XX. Es más fácil reunir un corpus que analizarlo. Los corpora que más o menos se tienen por constituidos están a la espera de estudiosos que descifren sus reglas ocultas.

La copiosa abundancia de corpora disponibles no indica otra cosa que la existencia masiva de versiones electrónicas de los textos. Aunque coincidimos en aceptar que el sistema de corpora es, de momento, insustituible, su mera presencia no supone ninguna ventaja más allá del aumento desmedido de fuentes y referencias. Internet hace real el mundo posible, ese que antes sólo veía Dios. Pero sólo Dios tiene capacidad visual y eternidad para estudiarlo todo. Nosotros, pobres humanos con ojos y memoria más limitados, debemos movernos como siempre: seleccionando y evaluando. Nos queda el recurso a las máquinas que vean sin seleccionar. Pero alguien debe programar las máquinas.

Desde los años noventa se han multiplicado los trabajos de análisis de textos y el tratamiento formal de los mismos. Tales trabajos han permitido avances en campos limitados, como el alineamiento bilingüe de ciertos tipos de textos, pero han tenido poca influencia en la terminología. Y naturalmente deberíamos buscar vías de cooperación entre la investigación formal y académica y la práctica terminográfica que, como bien sabemos los que nos dedicamos a ella, sigue siendo casi al cien por cien artesanal. Esa vía de cooperación aún no la hemos encontrado. Es significativo comprobar que muchos de los proyectos de investigación al uso están apoyados por instituciones oficiales, pero que los resultados no se proyectan en absoluto en la metodología o en la mejora de la terminología oficial o de las bases de datos públicas.

6 Pero, ¿tienen sentido las bases de datos?

Son muchas las razones que abogan por un cambio radical en la orientación de las bases de datos, razones que vienen impuestas principalmente por la evolución de tres elementos:

- un tipo nuevo de usuario heterogéneo, polivalente y multilingüe,
- la «corporeización» prácticamente total del texto,
- la interconexión entre textos.

Los datos prueban que el traductor no es el único usuario de las bases de datos terminológicas. Hace medio siglo, Weinreich definió el bilingüismo como la situación, externa o psicológica, de dos «lenguas en contacto». Hoy en día, muy pocos habitantes de la tierra se libran de la influencia diaria de otra lengua. Esa lengua es mayoritariamente el inglés pero la influencia va, o debe ir, en todas las direcciones. Amin Maalouf ha sugerido que empleemos la globalización positivamente para que, en el inevitable espacio de experiencias común, pueda reconocerse cada identidad. No sólo la gente viaja cada vez y recurre afanosamente a listas de correspondencias entre idiomas sino que cada vez hay más personas confrontadas a textos en idiomas ajenos al suyo. A veces es preciso comprender un texto, a veces intriga sólo una frase corta (¿Qué significa *shut up?)* o una palabra determinada que viene de otras latitudes (*sherpa*, *sushi* o «movida», según los casos). Por eso, numerosos buscadores o portales informáticos incluyen en su página un traductor automático de idiomas. Finalmente, la sofisticación técnica media de los ciudadanos corrientes es tan alta en las sociedades desarrolladas que casi todos ellos son lectores de textos especializados, por necesidades perentorias de empleo de artefactos y por una normal curiosidad intelectual consagrada por la tecnificación cotidiana. Todos los grandes periódicos incluyen secciones de vulgarización, con frecuencia redactadas por especialistas, sobre asuntos abstrusos anteriormente muy reservados. Una base de datos técnica debería atender los requisitos no ya de un campo determinado sino de un usuario determinado. Ese usuario no es necesariamente un especialista ni es siempre un experto en el idioma en que surge la información. Por otra parte ni la separación entre las especialidades es casi nunca neta ni el usuario está siempre en condiciones de saber a qué lengua pertenece un término. El tratamiento de lenguas de especialidad es, con estas coordenadas, relativamente inútil, como lo es la separación entre vocabulario general y vocabulario especializado. A la hora de elaborar una base de datos lingüísticos, la pregunta sensata que debe hacerse es: ¿qué

información le es necesaria a un usuario de la misma para entender los textos a los que se enfrenta? Teniendo en cuenta que el usuario es plural.

En cuanto a la disponibilidad de textos, no podemos quejarnos. La corporeización de la lengua está suficientemente asentada. No es que, de la noche a la mañana, haya crecido monstruosamente el número de textos. Es que una de las características de los textos es hoy su naturaleza electrónica. Una vez colocados en Internet, el único problema es «leerlos». Debido a su número exagerado, «leer» no puede querer decir «comprender» (salvo para el ojo de Dios) pero, participando todos los textos de una naturaleza homogénea y operativa y contando nosotros con medios de relacionarlos mecánicos, lo normal es intentar aglutinarlos y buscar sentidos compartidos. La intuición religiosa primero, luego literaria (Borges) y después filosófica (Derrida) de que todo está relacionado con todo, es ya realidad cotidiana con Internet. El mundo se ha convertido en un solo texto y cualquier palabra puede enlazarse con otra. Semiológicamente estamos tentados a reducir la realidad a un red infinita que podría describir un sistema de grafos y que equivaldría a un modelo Quilliams generalizado. Quizás no se haya insistido suficientemente en que la operación semántica y la operación de traducción son de la misma naturaleza. Jakobson dio con la mejor formulación de la idea: «El sentido es lo que se traduce». En Greimas, ello autoriza «une conception de la forme sémiotique comme susceptible d'être soumise à différents transcodages, tant verticaux (métalinguistiques) qu'horizontaux». En su imprescindible *Estudios de lexicografía española*, Seco explica muy bien la importancia de la «ley de la sinonimia», que pone como condición para que las definiciones sean correctas. Una aplicación generalizada de la sinonimia es la única posibilidad de enlazar un número ilimitado de términos no sólo con sus definiciones sino con sus correspondencias en otros idiomas. Jean y Claude Dubois lo han enunciado claramente: *Les définitions des dictionnaires bilingues sont du même type que celles des glossaires: un terme est traduit dans l'autre langue par une série de synonymes.*

Este hecho plantea una serie de preguntas a las bases de datos terminológicas, que ya no pueden seguir poniendo puertas al campo. El léxico de una base de datos era inicialmente un sistema cerrado que se desarrollaba por adición. Ahora es un conjunto de puntos de encuentro que adquieren sentido por su relación con otros incluidos en un espacio superior que los incluye a todos. Empresas como Elsevier trabajan ya en desarrollar «webs semánticas» que relacionen los vocabularios existentes, y muchos traductores consideran ocioso consultar una base determinada y prefieren acudir directamente a Google antes que, por ejemplo, a

Eurodicautom. Está claro que, hoy por hoy, esa ilusión tiene sus limitaciones. Los buscadores no examinan al cien por cien el contenido interior de las bases de datos y, aunque haya quien crea lo contrario, las bases, como los corpora —pues las bases no dejan de ser un tipo particular de corpora—, no se alimentan solas. Si todos se dedicaran a mirar Internet y nadie lo hiciera crecer —«creced y multiplicaos», se nos dijo hace tiempo—, el mundo, tal como lo conocemos, se acabaría. Más seria es la duda en cuanto a la falta de autoridad comprobada de los datos encontrados al azar. En ciertos campos del saber tradicional, empezando por conglomerados académicos de filósofos, se intentan establecer listas de direcciones de Internet homologadas. Y mucho más radicales para el terminólogo profesional pueden llegar a ser las implicaciones metodológicas que plantea la nueva situación. ¿Cómo relacionar léxicos de formato y categorización heterogéneos? ¿Cómo, o para qué, relacionar términos controlados y analizados con otros que son una mera porción de un texto libre? ¿No pierde interés la propia ocupación terminográfica? Los diccionarios y otros léxicos se impusieron, por razones de economía intelectual, como compendios de puntos significativos del lenguaje. Hoy bastaría con aprender gramática y ver textos. La gramática fijaría las reglas y todo el resto serían palabras, por seguir el modelo de lenguaje presentado por Pinker. Aunque, en ese caso, a lo mejor lo sensato sería que las bases de datos ampliaran ellas mismas su parte de descripción gramatical. Mientras se hace la luz sobre estas peligrosas cuestiones, se sigue trabajando en desarrollar pasarelas entre bases. Una técnica afianzada es la de los *one-stop-shops*, especie de buscadores caseros que, al hacérseles una consulta, examinan sucesivamente el contenido de un número limitado y controlado de bases. Y lo que parece lógico, teniendo en cuenta la mezcla de globalización y de especialización en la producción de textos, sería que los alimentadores de una base no rehicieran el trabajo en léxicos primeramente controlados por otra base u otro colectivo. ¿Para qué va Eurodicautom a preparar una colección de terminología marítima que ya han preparado, y mejor sin duda, en la Organización Marítima Internacional de Londres? Naturalmente, en abstracto al menos, esta opción de especialización sugiere caminos que, para los creyentes en la libertad del lenguaje, puede tener visos de normalización aterradores. Imaginemos una situación en la que varias administraciones internacionales se reparten por retazos el lenguaje e imponen en cada caso lo que se ha de decir.

Por mucho que nos remuevan la cabeza todas estas prevenciones, no parece descabellado que, entretanto, nos organicemos para controlar el mayor universo terminográfico posible. En Almagro nos hemos reunido

atraídos por un título que todos hemos comprendido intuitivamente: «El español lengua de traducción». Un título que quiere decir, por lo menos, tres cosas: la evidencia de que el español es una lengua que constitutivamente está interrelacionada con otras, la extensión del español por el mundo, como prueba el origen variado de los que aquí nos encontramos, y la importancia digamos legislativa, es decir, normativa, del español por su presencia masiva en las principales organizaciones mundiales.

7 Propuesta de colaboración

Yo creo que no debíamos irnos de Almagro sin poner las bases de una colaboración continuada que vaya más allá del establecimiento de buenas relaciones personales. Ante todo, deberíamos poner al día y en común la lista de recursos lingüísticos y terminológicos del español, acompañada de un «Quién es quién» permanentemente actualizado. En segundo lugar, podríamos crear un foro fijo de intercambios u oficializar alguno de los ya existentes para constituir entre nosotros, usuarios privilegiados del español («privilegiados» en todos los sentidos, incluido el normalizador) una red continua de intercambios. ¿Por qué no crear, por ejemplo, un procedimiento para inventoriar y acordar neologismos en cualquier campo que se plantee? ¿Por qué no apoyarnos recíprocamente en la búsqueda de soluciones coherentes frente a las solicitaciones del inglés u otras lenguas? No debemos subestimar la importancia —o, si preferís, la responsabilidad— que tenemos en la constitución del español. Las instituciones internacionales son una inmensa fábrica de palabras. Los que tenemos experiencia de traducir en una institución internacional, sabemos con qué rapidez las palabras que hemos reunido en un momento del día más o menos lúcido se reproducen pocos días después en los periódicos dando nombre y sanción a una figura legislativa. Nuestros amigos los periodistas están en la misma guerra. Con muchísima frecuencia son los primeros en designar un hecho. Con harta frecuencia en campos de técnica, en política, en sociología, son los primeros traductores. En muchas ocasiones contradicen los términos que acuñan los traductores, apoyados por un uso nacional que, por aproximación física o por su naturaleza más marcadamente unilingüe, interpretan mejor que los traductores, bilingües por definición, en cuyas cabezas, para bien y para mal, las diversas lenguas siempre están en contacto.

De un modo más proactivo —no sé muy bien si la palabra, aparentemente latina, proviene del inglés—, no sería muy complicado establecer un «corpus» mínimo de textos corrientes del español que nos ayudaran a detectar neologismos. Hablando sólo de lo que conozco un poco, los textos producidos por la Comisión Europea y los teletipos de la Agencia EFE cubren un sinfín de parcelas del lenguaje. ¿Por qué no explotarlos en común, esperando que se sumen a la iniciativa un mínimo de actores relevantes? ¿No estarían dispuestas las universidades españolas o americanas a colaborar en el proyecto? El resultado podría alimentar una base de conocimientos accesible a todos los que se interesan por el español.

Referencias

BORGES, JORGE LUIS: *El Aleph*, 1949

BOSWELL, JAMES: *Life of Samuel Johnson*

CASARES, JULIO: *Introducción a la lexicografía moderna*, Consejo Superior de Investigaciones Científicas, Madrid, 1992

COYAUD, MAURICE: *Linguistique et Documentation*, Larousse, 19772

DERRIDA, JACQUES: *De la Grammatologie*, Minuit, 1967

DICCIONARIO DE AUTORIDADES, 1732

DICTIONNAIRE DE L'ACADEMIE FRANÇAISE, 1694

DUBOIS, JEAN et CLAUDE: *Introduction à la lexicographie: le dictionnaire*, Larousse, 1971

ECO, UMBERTO: *Tratado de Semiótica General*, Lumen, 1977

EURODICAUTOM: http://europa.eu.int/eurodicautom

GOZZI, RAYMOND: *A Brief History Of Internet Times*, ETC, Winter 2001-2

GREIMAS, A.J.: *Du Sens*, Seuil, 1970

GROSS, MAURICE: *Grammaire transformationnelle du français*, Cantilène, 1986

GROSS, MAURICE: «Lexicon-grammar, the representation of compound words». En *Proceedings, 11th International Conference on Computational Linguistics*. Bonn, 1986.

ISIDORO DE SEVILLA: *Etymologiae*

JACQUEMIN, CHRISTIAN: *Spotting and Discovering Terms through Natural Language Processing*, Massachusetts Institute of Technology, 2001

JAKOBSON, ROMAN: *Collected Writings*, Mouton, 1971

JOHANSSON, STIG: «Towards a multilingual corpus for contrastive analysis and translation studies», 1999, en *SPRIK reports* 1. October 2000; «Språk i kontrast»: http://www.hf.uio.no/german/sprik
KANT, IMMANUEL: *Kritik der reinen Vernunft*
LOGOS: http://www.logos.it
LONGMAN Dictionary of English Language and Culture, 1992
MAALOUF, AMIN: *Les identités meurtrières*, Grasset, 1998
NIDA, EUGENE A.: *Componential Analysis of Meaning*, Mouton, 1995
NOVALIS: «Die Enzyclopädie»
OGDEN, C.K.; RICHARDS, I.A.: *The Meaning Of Meaning*, 1923
PEIRCE, CHARLES: *Collected Papers*, 1931-35
PINKER, STEVEN: *Words and Rules*, Basic Books, 1999
PORFIRIO: *Isagoge*
QUILLIAM, ROSS M: «Semantic Memory» en: MINSKI, Marvin (ed.): *Semantic Information Processing*, 1968
SAGER, N.; FRIEDMAN, C; LYMAN, M.S. and members of the Linguistic String Project: *Medical Language Processing: Computer Management of Narrative Data*, 1987
SAUSSURE, FERDINAND DE: *Cours de linguistique générale*, 1916
SEARLE, JOHN: *The Social Construction of Reality*, Simon & Schuster, 1995
SECO, MANUEL: Entrevista en *Diario Sur*, 1.12.1999
SECO, MANUEL: *Estudios de lexicografía española*, Paraninfo, 1987
SECO, MANUEL; ANDRÉS, OLIMPIA; RAMOS, GABINO: *Diccionario del Español Actual*, Aguilar, 1999
WEINREICH, URIEL: *Languages in Contact*, 1953
WITTGENSTEIN, LUDWIG: *Philosophische Untersuchungen*, Blackwell, 1964

* Traductor y terminólogo; jefe de la base de datos terminológicos Eurodicautom. Ha trabajado en diversos medios de prensa escrita y audiovisual.