La construcción del diccionario

LUIS FERNANDO LARA
del Colegio de México

Hay un error de base en la llamada «traducción automática», que consiste en la creencia de que el sentido de un discurso o de un texto es el resultado de la adyunción de los significados de las palabras que lo constituyen, junto con el que le agregan las reglas sintácticas. Agregaría que en el fondo de ese error repta la misma ambición del doctor Frankenstein, de reproducir al ser humano cosiendo sus miembros dispersos o, dicho de manera más contemporánea, con la gran metáfora del siglo presente, convirtiendo al ser humano en una máquina modular, interconectada a base de interfaces pero que, de hecho, aísla a cada uno de sus órganos en sus propios límites. Creo que de esa metáfora depende buena parte de la discusión científica contemporánea tanto en las humanidades como en la medicina. En la lingüística contemporánea, la distinción tajante entre léxico y sintaxis, junto con la preocupación por su «interfaz», heredada del paradigma de Noam CHOMSKY, corresponde a la misma concepción.

A diferencia de la traducción automática, la traducción «humana», que sigue y seguirá siendo la determinante de los contactos entre pueblos y culturas de lenguas diferentes, sabe que el sentido de un discurso se forja en los sutiles ires y venires de las palabras y sus combinaciones; de las palabras y las variadas posibilidades de significación que ofrece la estructura de las oraciones; pero además en la manera en que palabras y discursos adquieren su espesor significativo en la urdimbre de las experiencias culturales de cada lengua.

Los diccionarios nacieron como instrumento de trabajo de los traductores, no como productos descriptivos de los lingüistas. Desde la antigüedad de Mesopotamia y de Egipto, las primeras listas de palabras que podríamos considerar como primitivos diccionarios, consistían en equivalencias de nombres de objetos en las diferentes lenguas que interesaban, sobre todo, al comercio. Productos de intercambio entre el egipcio, el acadio, el fenicio, el griego. Trigo, vino, aceite, lino. Esas listas sólo ofrecían correspondencias de palabras para una comunicación

mínima, que permitiera a sus usuarios ponerse de acuerdo en lo que comerciaban.

Las primeras listas de correspondencias léxicas entre lenguas europeas modernas, en cambio, parecen haber nacido sobre todo de la necesidad de comprender las lenguas sagradas en los monasterios y en las universidades. Pero en un principio, como se puede apreciar en las *Glosas emilianenses*, los glosadores no se limitaban a ofrecer correspondencias léxicas, sino que indicaban relaciones gramaticales en el texto fuente y ordenaban los elementos de la oración. Es decir, se daban cuenta de que no basta con hacer corresponder dos palabras de dos lenguas diferentes, sino que era necesario tomar en consideración la manera en que intervenían en la construcción del sentido.

La labor de glosa no dio lugar a una lexicografía propiamente dicha, es decir, a una disciplina metódica de construcción de diccionarios, sino que siempre se mantuvo como un instrumento muy útil pero secundario, atenido a las necesidades de cada obra; y una vez delimitado un interés primordial por las palabras, más que por la construcción del sentido, la consideración primitiva del orden de palabras en la oración y de sus funciones sintácticas tendió a desaparecer. La costumbre de reunir voces de varias lenguas para ofrecer a sus lectores listas de correspondencias entre ellas, en cambio, sí dio lugar a la lexicografía multilingüe, por lo que se puede afirmar que el origen de toda la lexicografía europea —y de su continuación americana— es la práctica del multilingüismo.

Un elemento novedoso y de enorme importancia de esa lexicografía multilingüe es la introducción en los diccionarios de las lenguas sagradas: hebreo, arameo, griego y latín, acompañadas por varias de las lenguas modernas europeas: español, francés, inglés, neerlandés, alemán, etc. Los diccionarios multilingües de la época del Humanismo abandonaron sus primitivos intereses comerciales para convertirse en obras dedicadas fundamentalmente a la comprensión y la traducción de textos de la Antigüedad, a la enseñanza de las lenguas sagradas en las universidades y en los monasterios y, en América hispánica, a la evangelización. De nuestra tradición lexicográfica tomemos como ejemplo primero de esos diccionarios el *Vocabulario latino-español* de Antonio DE NEBRIJA, publicado en 1492, hecho para acompañar el estudio del latín en la Universidad de Salamanca, que ofrecía un vocabulario que interesaba a los humanistas, junto con breves explicaciones de cultura latina, como por ejemplo los nombres de los dioses, para que los estudiantes fueran capaces de traducir obras latinas al español. Como segundo ejemplo, el *Vocabulario en lengua castellana y mexicana* de Fray Alonso DE MOLINA, publicado en 1555, en que el evangelizador se esforzaba por encontrar

formas lingüísticas nahuas para poder transmitir a los indios la doctrina cristiana. Desde el punto de vista del necesario contraste entre las lenguas, que ilumina el sentido de los textos, he de decir que el *Vocabulario* de MOLINA es mucho más interesante que el de NEBRIJA, pues al fin y al cabo el latín y el español son lenguas cercanas, de la misma cultura, en tanto que el náhuatl y el español son de troncos y culturas radicalmente diferentes, con lo que los esfuerzos por hacer comprender un concepto religioso en náhuatl eran mayores. Pero lo cierto es que el origen de la lexicografía que hacemos ahora está en aquella multilingüe dedicada al conocimiento de la Antigüedad clásica y la evangelización.

La invención de la imprenta tuvo una influencia determinante para la lexicografía, pues muy pronto se vio que los diccionarios, además de responder a necesidades de conocimiento de las lenguas, eran buen negocio. El diccionario que el monje italiano Ambrogio CALEPINO publicó en 1502 como resultado de un interés muy semejante al de NEBRIJA, de inmediato fue ampliándose del latín y el italiano a siete lenguas europeas hacia mediados del siglo XVI y llegó a incluir once a principios del XVII, ya convertido en un negocio editorial y, obviamente, tras la muerte de su autor originario; desde entonces la palabra «calepino» vino a entenderse como «diccionario multilingüe» y se convirtió en un modelo para la lexicografía multilingüe, al que debemos la mayor parte de los diccionarios dedicados al español y las lenguas amerindias de la época colonial.

El *Tesoro de la lengua castellana o española*, de Sebastián DE COBARRUVIAS, publicado en 1611, un año antes del famoso *Voccabolario degli accademici della Crusca*, vale en cambio como el primero monolingüe de los nuevos Estados nacionales europeos. En cuanto a su factura, constituye un objeto de análisis que todavía merece más estudio; la interpretación dominante considera que corresponde más a una *summa* medieval muy tardía, que a una obra del clasicismo humanista —como el de la Crusca— o de la Ilustración —como el de *Autoridades*, de la Academia Española—. Para darle esa interpretación se toma como elemento clave el hecho de que muchos de sus artículos comiencen ofreciendo sus correspondencias latinas, griegas o hebreas, seguidas por comentarios sobre el origen de la palabra, como si fuera un diccionario etimológico del español; pero probablemente podamos atribuir esas correspondencias más a la tradición de la lexicografía multilingüe que a un interés bien definido por la etimología en sentido moderno. Pues para COBARRUVIAS, la *etimología* tenía el mismo sentido que para san ISIDORO, como conocimiento verdadero de las cosas, no como origen de las palabras.

La otra característica notable de este diccionario es su explicación de los vocablos, de carácter enciclopédico, basado en las obras que tenía COBARRUVIAS a su disposición, ya fueran historias naturales romanas, la *Biblia* o el conocimiento acumulado en su época, sin que mediara ni una distinción entre el significado y los objetos a que refieren las palabras, ni una ponderación clara de sus fuentes. Es decir, parece que a COBARRUVIAS le interesaba más el conocimiento que se obtiene por las palabras, y no ante todo las palabras, la lengua y los autores reconocidos de esa lengua, a diferencia de los diccionarios de la Crusca, de la Academia Francesa y de la Española. Yo diría que en ese aparente esfuerzo anticuado por lograr una *summa*, como lo había hecho San ISIDORO de Sevilla casi mil años antes, estriba el interés central de la obra: que nos ofrece un documento —parcial, ciertamente— del estado de la cultura española en la época de la publicación del *Quijote*, cuando no se habían disipado del todo las oscuridades de la Edad Media; cuando PLINIO, HIPÓCRATES, GALENO o ARISTÓTELES todavía determinaban el conocimiento; cuando el magisterio de la Iglesia definía lo que se podía decir acerca del mundo y la naturaleza, pero también cuando el español había alcanzado la primera conciencia de sí mismo como lengua de un Estado poderoso. Ese reflejo cultural que ofrece el diccionario de COBARRUVIAS merece un estudio dilatado que no se ha hecho, hasta donde llega mi conocimiento.

El diccionario de COBARRUVIAS dejó de ser un diccionario multilingüe de una manera casi natural, aunque notable, negativamente, para el público que, como señala Manuel SECO, no se interesó por él: no se veía qué sentido tenía un diccionario del español solamente, en español. Es decir, su carácter monolingüe, desmarcado de la tradición multilingüe, fue tan novedoso que al parecer no se entendió, en su tiempo, para qué servía. Pues como la lengua materna es obvia y transparente para sus hablantes, y cada uno cree que conoce bien su vocabulario, un diccionario de la propia lengua resulta una redundancia inútil[1]. Sólo habría de llegar a entenderse el valor del diccionario monolingüe cuando éste adquirió un papel simbólico, de afirmación de la propia lengua en relación con otras, de estados con los que el propio competía. Pero tal

[1] En este punto cabe distinguir el papel del diccionario monolingüe en la cultura inglesa: se había desarrollado una lexicografía de las palabras difíciles o *inkhorn terms*, para ayudar a comprender la literatura que, durante el siglo XVI, había hecho gala de erudición y conocimiento clásico; y además, las dificultades de la ortografía inglesa motivaban muchos esfuerzos eruditos por fijarla y ponerle reglas, que se materializaban en los diccionarios. Quizá por eso la aparición del *Dictionary of the English Language* de Samuel JOHNSON (1755), fue tan bien recibida por su público, en comparación con la indiferencia con que el público hispanohablante recibió los diccionarios de COBARRUVIAS y de la Academia.

papel correspondió realmente, un siglo más tarde, al *Diccionario de la lengua castellana*, primera obra de la Real Academia Española (1726), que hoy conocemos como *Diccionario de autoridades*, pues éste ya era una celebración de la gloria que la lengua había alcanzado en sus autores, como lo fueron antes el diccionario de la Crusca y el de la Academia Francesa, a los que la Academia Española tomó por ejemplos y por estímulos. El *Diccionario de autoridades* no sólo es un diccionario claramente monolingüe, sino que a diferencia del de COBARRUVIAS, que podemos catalogar como «enciclopédico», es el primero de nuestros diccionarios «de lengua»[2], es decir, de los que se dedican a la explicación del significado de sus vocablos, con relativa independencia de sus referencias a la realidad.

El *Diccionario de Autoridades* inauguró la tradición hispánica de los diccionarios de lengua, basados en el uso literario y ejemplificados con fragmentos de los autores que los miembros de la Academia consideraban imitables. Pero en esa elección, que al siglo siguiente habría de dar lugar a los diccionarios filológicos, de los que el *Oxford English Dictionary* es su mejor representante, las creencias y las concepciones tradicionales que forman el subsuelo de la cultura y se manifiestan en el espesor significativo de las palabras, como se ve en tantos artículos de COBARRUVIAS, quedaron reducidas a lo que el contexto de los ejemplos literarios pudiera traslucir.

El interés de la Academia Española al finalizar el siglo XVIII por ofrecer a su público un diccionario en un solo tomo terminó por dar al traste con el valor de los ejemplos, con las «autoridades», e impedir que durante el siglo XIX y hasta la fecha se produjera un diccionario filológico del español, como el ya mencionado de Oxford o el de LITTRÉ y el *Trésor de la langue française* a finales del XX (el largamente esperado *Diccionario histórico de la Academia Española*, tras varios inicios, todavía no se materializa). La lexicografía española —pues todavía no había una lexicografía americana del español— quedó así reducida a un solo diccionario con autoridad, cuya nomenclatura depende de los juicios caprichosos de valor y de uso de los vocablos que hagan los miembros de la Academia —y ahora, de las Academias hispanoamericanas, que no es decir mucho—, con un sistema defectuoso de definición, sin ejemplos, rodeado por diccionarios comerciales que comenzaron a competir entre sí

[2] El «diccionario de lengua», dedicado a explicar el significado de las palabras, más que las características de las cosas que nombran, es una tradición románica, que la comunidad hispanohablante comparte con la francesa o la italiana.

por hacer crecer sus nomenclaturas a cualquier precio, sin comprometerse con un registro amplio y digno de confianza del vocabulario hispánico, que les dieran orden, representación y pertinencia.

Para la gran comunidad hispanohablante los diccionarios de la Academia Española se convirtieron, sí, en una autoridad, pero también en acervos reducidos, sesgados hacia una sola variedad de la lengua y atrasados en su registro de vocablos no sólo hispanoamericanos —que a la fecha no ocupan el lugar que les correspondería en la lexicografía española— sino incluso españoles, pues las diferentes regiones peninsulares y de las islas Canarias quedaron subrepresentadas en comparación con el vocabulario culto madrileño y castellano. Por su parte, los diccionarios de las grandes casas editoriales fueron, durante mucho tiempo, subsidiarios de los de la Academia, a los que copiaban, modificaban o deformaban, por más que lograran aumentar comparativamente sus nomenclaturas y lograr buenas ventas en los mercados hispánicos (naturalmente, hubo alguna excepción). En mi opinión, esa situación sólo viene a modificarse parcialmente a partir del primer *Diccionario de uso del español*, de María MOLINER, y ahora en mayor grado gracias al *Diccionario del español actual,* de Manuel SECO, y el *Diccionario de uso del español de América y España*, de la editorial Vox, en los que sus autores no sólo han modificado sus nomenclaturas, sino también la información contenida en los artículos acerca de acepciones y apreciaciones del uso social de los vocablos, así como han cambiado la organización interna del artículo, han basado sus registros en usos reales —el caso del *DEA*— y han mejorado la calidad de la definición —en el caso del *DUEAE*—.

Este rápido y esquemático recorrido por una historia de la lexicografía española sobre la que se ha escrito mucho, pero todavía no lo suficiente para entenderla bien, ha tenido por objeto nutrir con los antecedentes más relevantes tres temas centrales de la construcción de los diccionarios contemporáneos, que me parece debieran caracterizar el futuro de la lexicografía hispánica en el siglo XXI: las características deseables de los diccionarios monolingües y multilingües, la situación de la lexicografía multilingüe en comparación con la monolingüe, y la situación de la lexicografía española e hispánica.

He venido destacando tres elementos centrales de los diccionarios: la base de su información; el carácter y la calidad de la información que ofrecen acerca de las palabras, tanto en cuanto a su significado como en cuanto a su entorno sintáctico y de uso; y su relación con las culturas verbales a las que pertenecen. El desarrollo futuro de la lexicografía hispánica, monolingüe y multilingüe, además de lo que está haciendo por

ella la tecnología informática, a la que no me referiré, tendrá que orientarse a mejorar esos tres elementos.

Trataré esos tres aspectos entremezclados, para no dar a esta conferencia el aspecto de un tratado: desde que se construyeron los primeros diccionarios monolingües de las lenguas europeas sus nomenclaturas se componen de voces registradas, mayoritariamente, en los usos cultos de las lenguas, y en obras de carácter literario, científico y técnico (sobre todo en los diccionarios ingleses, especialmente los norteamericanos; en el caso de los diccionarios españoles, el vocabulario científico y técnico había sido siempre postergado hasta muy recientemente). Gracias al hecho de que hay una unidad cultural europea, es decir, una cultura compartida por toda Europa desde que el Imperio romano se extendió por el continente y después, desde que el Imperio carolingio y el dominio franco hicieron crecer sus fronteras hasta Rusia y Ucrania, los vocabularios de las lenguas de Europa son compatibles entre sí, comparten muchos elementos de significado, y se refieren a la misma civilización. Eso posibilita la elaboración de diccionarios multilingües sin que tenga que mediar un estudio contrastivo previo entre todas las lenguas consideradas, en que se ponga atención especial al espesor cultural de los significados, y facilita el tratamiento de las correspondencias entre lenguas a base de vocablos aislados, con pocas explicaciones gramaticales, y con pocos o ninguno ejemplos de uso. Pero porque, a pesar de la comunidad cultural europea, cada lengua no sólo es estructuralmente diferente de las otras, sino que es vehículo de culturas con su propia historia y sus propias peculiaridades, hay diferencias en los significados de las palabras de cada lengua, que suelen soslayar o incluso dejar de notar los diccionarios multilingües, en tanto que los entornos sintácticos de las palabras apenas en algunos grandes diccionarios de los últimos veinte años comienzan a considerarse con la profundidad necesaria. A eso se debe que la estructura de los artículos lexicográficos de esos diccionarios multilingües sólo ayude, en el mejor de los casos, a una comprensión general de los textos de las otras lenguas, dejando a cada lector el trabajo de precisar su sentido, mientras que dificulta mucho la producción de textos en otra lengua. Se puede afirmar que la lexicografía multilingüe hoy en día va a la zaga, en calidad, de la monolingüe. Para estar a la altura de los diccionarios monolingües de que se dispone ahora, la lexicografía multilingüe tendrá que replantear sus métodos sobre la base del estudio contrastivo de los vocabularios de las lenguas que le interesen, y dejar la práctica de las correspondencias aisladas entre lenguas, para dotarse de una estructura del artículo lexicográfico en que a la entrada de una lengua corresponda una compleja estructura de voces, significados, matices y

posibles correspondencias en la otra lengua, con un análisis semántico rico y preciso de cada uno, acompañados por ejemplos de uso, recciones verbales, los patrones oracionales que sean pertinentes y colocaciones.

En estos aspectos las bases multilingües de datos terminológicos de que disponemos ahora superan a la lexicografía multilingüe y dan un buen ejemplo de lo que debieran ser sus diccionarios. Pues lo que han comenzado a lograr estas bases de datos es precisamente representatividad, en cuanto a los vocablos que es necesario registrar, ricos registros de usos reales, suficientes observaciones del entorno sintáctico y las colocaciones de las palabras, correlaciones complejas entre las palabras de una lengua y la otra en cuanto a sus significados, e incluso medidas del grado de confianza que se puede tener en el uso de un vocablo o de una expresión en una comunidad lingüística determinada. Por el contrario, los datos de que parecen disponer las casas editoras de diccionarios multilingües siguen proviniendo de los métodos de registro tradicionales en lexicografía, dependientes de observaciones individuales, azarosas y sin suficientes medios de confirmación, y atenidos crucialmente a los conocimientos personales de sus lexicógrafos.

En este estado de la lexicografía multilingüe debe jugar un papel importante el hecho de que es una práctica de casas editoras, orientadas al negocio lexicográfico, que no están dispuestas a hacer inversiones de largo plazo para dotarse de los equipos de lingüistas y los instrumentos de información que les hacen falta y, por el contrario, las universidades y otras instituciones públicas no parecen interesarse mayormente en las necesidades de la comunicación con pueblos y naciones de lenguas diferentes.

La lexicografía monolingüe, en cambio, ha mejorado mucho sus bases de datos, desde que se comenzó, hace menos de cincuenta años —por ejemplo, en el *Trésor de la langue française*, en los trabajos del diccionario combinatorio del inglés y en el *Diccionario del español de México*—, a reunir corpus de datos informatizados y cuantitativamente ponderados. Aun cuando la Academia Española no parece tomar del todo en cuenta en sus diccionarios los datos de su *Corpus de referencia del español actual*, un corpus como ese es una rica contribución a la calidad de la lexicografía hispánica de este siglo. La conformación de los corpus de datos para la lexicografía sigue siendo un tema de reflexión y de discusión en la lingüística contemporánea, pero no está en duda el creciente papel que irán teniendo en la calidad de los diccionarios.

No es nada clara, en cambio, la relación entre cultura y diccionario. A pesar de que Alain REY publicó hace una veintena de años un famoso artículo acerca de lo que, desde entonces, entendemos como «diccionario

cultural», salvo en su propia obra, por ejemplo en el *Dictionnaire historique de la langue française* (1998), en la mía para el *Diccionario del español de México* y en una pequeña serie de diccionarios monolingües de lenguas mayas actuales en el sur de México, que tuve la suerte de poder asesorar, no se entiende que el diccionario es un producto de la cultura y un depósito de la manifestación verbal de la cultura. Así entendido, la definición lexicográfica se propone revelar los muchos sutiles matices de significado de las palabras, que transportan experiencias sociales de las comunidades lingüísticas, de la misma clase que las que leemos en el diccionario de COBARRUVIAS, aunque ahora tratadas con un aparato crítico y hermenéutico contemporáneo. Cuando uno es hablante de una variedad de la lengua española como la mexicana, históricamente considerada periférica de la lengua española metropolitana, se da mejor cuenta de la presencia de ese carácter cultural del significado, precisamente porque la tarea de contrastar una variedad de la lengua con las otras permite descubrirlo. Se descubre, por supuesto, en cada variedad hispánica, en las expresiones andaluzas, aragonesas, o murcianas, pero también en el español cotidiano de España. La lexicografía hispánica tiene delante de ella, en el estudio profundo de la semántica de sus vocablos, un riquísimo filón de estudio y de material para los diccionarios, que podría cambiar radicalmente la calidad de sus obras.

Pero hacer esto en la lexicografía hispánica implica un cambio de enfoques radical, en comparación con lo que han venido haciendo la Academia Española y las editoriales. Pues requiere buenos corpus de datos de cada país o región cultural hispánica (por ejemplo, la región centroamericana) y un trabajo de análisis semántico original, a cargo de equipos de lingüistas originarios de cada país o región. El resultado sería una colección de 17 a 22 diccionarios, que sólo después se podrían reunir, entretejiéndolos en una gran base informática, para construir un gran diccionario de la lengua española contemporánea. La Academia Española no parece haber tomado en cuenta esta posibilidad y tampoco las universidades hispánicas, para las que embarcarse en un proyecto de largo plazo es casi imposible, dadas las condiciones de financiamiento que les ha impuesto el neoliberalismo. La Academia hoy en día está más interesada en inundar de diccionarios cada cinco años el mercado hispánico, en seguir introduciendo voces en su diccionario sin criterios de trabajo claros, a partir de datos defectuosos recogidos por sus miembros y los de las academias hispanoamericanas, y en fomentar una distinción entre la lengua española de sus diccionarios y la que aparezca en su diccionario de americanismos, que no hace sino repetir la dicotomía metrópoli / periferia que fue superada en la realidad política y cultural del

mundo hispánico desde hace casi doscientos años. Las grandes casas editoriales, por su parte, prefieren seguir vendiendo diccionarios «del español de España y de América» a base de datos parciales, información defectuosa y sin interés por reflejar realmente la gran riqueza de la lengua española contemporánea.

Me parece que hay mucho trabajo por hacer en la lexicografía hispánica multilingüe y monolingüe. La lexicografía en general se ha convertido en los últimos años en una materia de gran interés en muchos países europeos y unos cuantos hispanoamericanos. La ayuda que ofrece la informática a la lexicografía abre unas posibilidades de construcción de nuevos diccionarios, de mejor calidad, que habría costado imaginar hace cincuenta años. El conocimiento detallado de los métodos y las técnicas lexicográficas de que disponemos ahora permite planear muchos diccionarios de acuerdo con muy diferentes funciones sociales. El efecto que podrían tener esas obras sobre la traducción consistiría sobre todo en facilitar las tareas del traductor en la búsqueda de expresiones no sólo precisas, sino adecuadas a las diferentes variedades de la lengua española. Sólo falta convencer a las universidades y a las casas editoriales para que inviertan en su capital humano; ese capital que ahora niega la ideología económica imperante. Hay lexicógrafos, faltan diccionarios.