

*Luis Fernando Lara*

---

# DIMENSIONES DE LA LEXICOGRAFÍA

---

A propósito del Diccionario  
del español de México



**abc**

ICE  
308  
J88  
no. 116

---

**jornadas**

---

**116**

---

EL COLEGIO DE MÉXICO

**CE**

324564

308/J88/no.116

Lara Ramos, Luis...  
Dimensiones de la...



JORNADAS 116

Fecha de vencimiento

---

---

|

EL COLEGIO DE MÉXICO



**EL COLEGIO DE MEXICO**

308/188/no. 116/CE



\*3 905 0334499 -\*

**DICCIONARIO DEL ESPAÑOL DE MÉXICO**

*Biblioteca Daniel Calvo Villegas*  
EL COLEGIO DE MÉXICO, A. C.

*Luis Fernando Lara Ramos*

# DIMENSIONES DE LA LEXICOGRAFÍA

A propósito del *Diccionario  
del español de México*



X

JORNADAS 116

308

128

Mr

EL COLEGIO DE MÉXICO

**324564**

308  
J 88  
no. 116  
ej. 2

*Open access edition funded by the National Endowment for the Humanities/Andrew W. Mellon Foundation Humanities Open Book Program.*



*The text of this book is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License: <https://creativecommons.org/licenses/by-nc-nd/4.0/>*

**Portada de Mónica Diez Martínez**

**Primera edición, 1990**

**© D.R. El Colegio de México  
Camino al Ajusco 20  
Pedregal de Santa Teresa  
10740 México, D.F.**

**ISBN 968-12-0454-9**

**Impreso en México / Printed in Mexico**

## ÍNDICE

Prefacio	21
El objeto diccionario	39
El diccionario, instrumento de la etnicidad	51
La cuantificación en el <i>Diccionario del español de México</i>	85
Caracterización metódica del <i>corpus</i> del <i>Diccionario del español de México</i>	107
Sociolingüística del <i>Diccionario del español de México</i>	133
El diccionario entre la tradición y la realidad	157
La cuestión de la norma en el <i>Diccionario del español de México</i>	195
Problemas y métodos del significado estereotípico	213
Diccionario de lengua, enciclopedia y diccionario enciclopédico: el sentido de sus distinciones	233
Lagunas y debilidades de la lexicografía hispánica	239
Bibliografía	239



## PREFACIO

Hace dieciséis años iniciamos en El Colegio de México la preparación de los trabajos para redactar el *Diccionario del español de México*. Entonces, como todavía ahora, se trataba de una empresa extraña en el ámbito de la lexicografía de la lengua española, pues por primera vez un equipo de lingüistas hispanoamericanos pretendía dar un tratamiento integral a su lengua materna, en vez de continuar con la tradición diferencial de los *diccionarios de regionalismos* que, desde finales del siglo XIX hasta la fecha, han definido todo el horizonte de la lexicografía del español americano. Antes de comenzar nuestro trabajo se consideraba peregrino, un tanto ocioso y, a veces, hasta subversivo para la idealizada unidad de la lengua que algún hispanoamericano planteara la posibilidad —ya no digamos la necesidad— de dar a su dialecto regional o nacional un tratamiento lexicográfico completo, congruente con los valores de su propia cultura y libre de los prejuicios elaborados por cientos de años de purismo y academicismo. Se nos exigía —como todavía se nos exige— que nuestra lexicografía sólo tratara el vocabulario de nuestra marginalidad: aquel que, por provenir de las lenguas aborígenes de América y por circunscribirse a las regiones limitadas en donde esas lenguas vivían, sólo nosotros usamos y sólo nosotros seríamos capaces de comprender; o aquel que, llegado

con los primeros colonizadores y arraigado en una América aislada y aldeana, resultaba pintoresco arcaísmo para la idea de la lengua que evolucionaba, con su propio paso, en la metrópoli; o incluso aquel que, descuidado por la lengua literaria, desarrollaba “vicios” perseguidos por la prescripción académica, aunque se reconociera que formaba parte de las costumbres y las lealtades de esas remotas regiones de la América hispana. La larga bibliografía existente de la lexicografía del español en América, tanto en México como en la Argentina, en Honduras como en el Paraguay es un testimonio entrañable y doloroso de la *conciencia del desvío* con que sus autores tratan las variedades del español en América en relación con la lengua de la antigua metrópoli.

La exigencia impuesta por esa tradición a Hispanoamérica dio lugar a una concomitante peculiar en España, pues al asignar a la vieja madre patria, implícitamente, el derecho exclusivo de redactar diccionarios generales, basados en la esencial centralidad de nuestra lengua en la Península, depositó también sobre ella una carga demasiado pesada y difícil de llevar: la responsabilidad de tratar toda la lengua, de documentar y analizar los cientos de miles de voces que forman su vocabulario, y de determinar, de un lado, lo que habría de ser general a todas las comarcas hispanohablantes; del otro, lo que debiera considerarse marginal o “hispanoamericano y provincial” como dicen los prólogos del *Diccionario* de la Academia Española. Los efectos de esa doble tradición los tenemos a la vista: el *Diccionario* de la Academia ha sido incapaz de registrar la verdadera riqueza léxica de la lengua española y está condenado, por eso mismo, a

padecer todos los ataques habidos y por haber, por no haber logrado cumplir con las exigencias, en especial con las hispanoamericanas, que no perdonan la falta de voces queridas en cada región hispánica (como *estufa*, *chino* 'rizado' o *morralla* en México), ni un solo error o una falsa lectura (como la de *cacahuete* en vez de *cacahuate*), ni una falsa etimología o una falsa asignación de uso regional (como sucede con *chapotote*, sólo reconocido como *chapapote* y atribuido a las Antillas). Los diccionarios de regionalismos, a su vez, se ven siempre sometidos a la tensión que se crea entre la necesidad de asegurarse de que un vocablo es exclusivo de la región que tratan y la necesidad de encontrar cada vez más voces que nutran y justifiquen su exotismo. La lengua común, en cambio, aquella en la que nos entendemos todos en cada región hispánica, no siempre sancionada por la literatura, muchas veces resultado nuevo de las necesidades de expresión que va creando la existencia de cada comunidad lingüística particular en un momento de la historia, se ha ido quedando al margen de esos diccionarios; y las necesidades de información del público hispanohablante se han venido viendo traicionadas por unos diccionarios quizá demasiado cultos, quizá demasiado eruditos, quizá realmente lejanos de la lengua que la gente efectivamente utiliza. De ahí la afirmación, hace años, del académico colombiano Germán Arciniegas, de que el diccionario más útil para los hispanohablantes es el *Webster*.

Pero si bien nuestro trabajo para el *Diccionario del español de México* encuentra su justificación en el deseo de romper esa vieja dicotomía entre la metrópoli y su periferia hispanoamericana, entre el diccionario

general de la Academia y los diccionarios de mexicanismos, su planteamiento bajo otro punto de vista y su realización con métodos y técnicas lexicográficas novedosas estaban —y están— lejos de ser obvios para los especialistas y para el público interesado en los diccionarios del español. De ahí la necesidad, sentida desde el primer momento de nuestro trabajo, de desarrollar los fundamentos en los que se basa la importancia y la validez de nuestro diccionario ante el diccionario general y ante los de regionalismos.

Por esa causa el inicio del *Diccionario del español de México* fue una investigación lingüística y no simplemente un acopio de documentos lexicográficos que hubiéramos de componer con tijeras y pegamento en las manos, y también por esa causa nuestra selección de vocabulario no se limitó a listar las voces que, desde nuestra competencia individual de la lengua, nos parecieran necesarias y representativas del español mexicano. Por lo contrario, optamos por estudiar el español de México poniendo entre él y nosotros, sus autores, la distancia que todo método de objetivación científica requiere entre el investigador y su materia, y decidimos detenernos cuanto tiempo fuera necesario en una documentación imparcial y variada de nuestro vocabulario, como para que, al final, pudiéramos al menos asegurar que lo encontrado corresponde a la realidad, y no a nuestra voluntad ni a nuestro gusto particulares.

Esa investigación preparatoria quedó concluida cinco años después de haber iniciado nuestro trabajo, y nos produjo una enorme cantidad de resultados de primera mano, que son los que utilizamos como materia prima del posterior análisis lingüístico y de la redacción definitiva de la obra. Los resultados fueron:

cerca de cuarenta mil vocablos documentados en nuestro *Corpus del español mexicano contemporáneo (1921-1974)* —el primero, y hasta hoy el único de la lengua española contemporánea—; un rico diccionario estadístico, que nos provee de datos acerca del uso y la extensión geográfica y social del vocabulario; un sistema de análisis gramatical automático —el *Analizador del DEM*— que nos permitió procesar cerca de dos millones de ocurrencias de palabras de nuestro *Corpus* (también el primero creado para la lengua española), y una creciente documentación lingüística y filológica del vocabulario español, con la que completamos nuestras necesidades de información para redactar el diccionario.

A principios de 1981, convencidos de la necesidad de hacer explícitos nuestros puntos de vista y nuestros métodos, ofrecimos al público interesado, Isabel García Hidalgo, nuestra analista de cómputo, Roberto Ham Chande, nuestro estadígrafo, y yo un volumen titulado *Investigaciones lingüísticas en lexicografía* (Col. Jornadas 89, El Colegio de México) en el que exponíamos los criterios que utilizamos para armar la muestra de textos del español mexicano contemporáneo que es el *Corpus*, así como los procedimientos cuantitativos seguidos para determinar lo que fuera general al español de México, la estructura del sistema del Analizador computacional, y las primeras bases del análisis semántico que habríamos de seguir en la redacción de la obra. Ese libro ha gozado del permanente interés de nuestros colegas en México y en el extranjero y a él deseo referir al lector interesado en éste cuando requiera la información previa pertinente a los estudios aquí reunidos. Éste que ahora presento

al lector es una suerte de continuación de aquél, motivado por la misma necesidad, pero diferente en cuanto ya es capaz de ofrecer algunas respuestas acerca de nuestros postulados teóricos y de nuestros métodos de trabajo.

La lexicografía es sin duda un “arte”, como la definen muchos diccionarios antiguos y modernos; pues la delimitación precisa de las varias acepciones de un vocablo y su reducción al pequeño tamaño de un artículo lexicográfico no son acontecimientos mecánicos ni se rigen por un conjunto único y bien definido de reglas y operaciones. La capacidad del instrumento de elaboración de datos que ha dado el siglo XX a los lexicógrafos, la computadora, no sirve a esa parte del arte ni tiene comparación con el fino hilado, la mezcla de sutileza y realidad que, diariamente y por años, tiene que ir haciendo el redactor de un diccionario para cada una de las palabras que habrán de componerlo. En cambio, la cantidad de conocimientos lingüísticos que pone en juego el manejo de los datos léxicos durante el análisis previo a la redacción hace de la lexicografía algo más que el arte: la pone en el centro de la ciencia del lenguaje y la obliga a definirse en todas sus dimensiones.

Por años la lingüística ha despreciado a la lexicografía, como ha despreciado la realidad de los hablantes de las lenguas para poder llegar al nivel de abstracción requerido por sus necesidades de sistematicidad y rigor científicos. Cuando no la desprecia, la lingüística la agrupa con las disciplinas que forman la *lingüística aplicada*, a la que entiende generalmente como una puesta en práctica subordinada y defectuosa de los conocimientos producidos en su teoría y no como lo

que es realmente la lingüística aplicada: una prueba rigurosa del valor de la lingüística para la sociedad. Quizá no habría yo llegado a este último convencimiento, si no me hubiera visto precisado a comprender la lexicografía sin renunciar a mi derecho de seguir siendo lingüista, y si mi formación profesional no me exigiera, por ella misma, construir los puentes que relacionan, sin contradicción ni separación, el conocimiento científico del lenguaje con la práctica sistemática de la lexicografía. Lo que descubrí con sorpresa y entusiasmo, en estos años de dedicación al diccionario, es que la lexicografía pone radicalmente en cuestión muchos de los sobrentendidos y de las seguridades de la lingüística, tanto como para que sea posible una crítica lexicográfica de la lingüística y una superación de la lexicografía desde el campo práctico de las técnicas y los métodos de trabajo hasta el teórico de los fundamentos mismos de la lengua. La lexicografía viene entonces a situarse frente a la lingüística como una crítica, cuyo sustento proviene tanto del campo estrictamente científico, como del de su legitimación social.

La lexicografía adquiere así múltiples dimensiones, que lingüistas y lexicógrafos no solemos ver. Se nutre de la vieja filología para situar las palabras en su contexto histórico y evolutivo; de la hermenéutica para formular sus esquemas interpretativos de los sentidos de un vocablo; de la gramática para determinar funcionamientos y posiciones de las palabras, así como para decidir el tratamiento que habrán de recibir entidades formalmente mayores que ellas y compuestas por varias de ellas, pero semánticamente unitarias; de la dialectología para reconocer las zonas geográfi-

cas y etnográficas en las que se registran ciertos usos; de la sociolingüística para delimitar los campos sociales de uso y de aceptación de los vocablos; de la lingüística contrastiva, para mejor discernir —con la comparación con otras lenguas— sentidos y funciones de los vocablos de la lengua en estudio; de la estadística lingüística para procurarse indicadores objetivos e imparciales de la extensión de los usos de las palabras, así como de varias de sus características de escritura, de patrones de derivación, y de fijación en el movimiento histórico; de la computación electrónica, tanto para poder manejar los grandes volúmenes de datos que requiere su trabajo, como para probar la regularidad de ciertos acontecimientos léxicos; de la lingüística teórica para encontrar su razón de ser y el sentido de sus relaciones con la sociedad; y de todos los campos de la cultura humana (incluidas en ella, como debe ser, las ciencias en general) para interpretar los sentidos de las palabras, esas maravillosas entidades sonoras o gráficas que transfieren a nuestro intelecto la realidad del mundo.

El trabajo del lexicógrafo oscila siempre entre la práctica del análisis y la redacción y la reflexión sobre su práctica. Simultáneamente con el trabajo cotidiano el lexicógrafo observa y razona varios de los elementos que intervienen en su labor. En nuestro caso la reflexión sobre la lexicografía tiene un papel tanto más importante, cuanto muchas de las cuestiones que aparecen al tratar al español de México como lo hacemos no han sido tocadas por los que nos han antecedido en nuestra disciplina. Y si con ello no bastara, las preguntas que se hace el público mexicano acerca de *su* diccionario, tan heterodoxo en comparación con las ideas

recibidas acerca de lo que “debiera” ser un diccionario mexicano, requieren siempre respuestas razonadas, que logren modificar los pétreos sobrentendidos que ha creado nuestra cultura de la lengua española y que, a menudo, tantas dificultades nos causan a los autores del diccionario. Los trabajos que he reunido en este tomo han surgido de ahí. Responden, con argumentos, muchas de las preguntas que nos hicimos al iniciar el trabajo y que nos siguen haciendo muchos colegas al respecto. Tratan de dar inteligibilidad a nuestra tarea y de contribuir a los conocimientos generales de la lexicografía. Tratan también, por qué no, de abrir y ventilar cuestiones que la lexicografía solía ocultar, temerosa del juicio de la lingüística a lo que aparentaban ser procedimientos más prácticos que razonables o teóricamente justificables.

De las varias dimensiones que tiene la lexicografía he reunido en este libro solamente tres y no por azar, sino porque forman parte de las cuestiones de principio de todo trabajo lexicográfico, y en particular del nuestro: he tratado de explorar lo que será una verdadera *teoría del diccionario*, preguntándome qué clase de *objeto* es (en el sentido aristotélico) y sobre qué elementos de la sociedad y la cultura se instituye (los artículos “El objeto diccionario”, “El diccionario, instrumento de la etnicidad” y “Diccionario de lengua, enciclopedia y diccionario enciclopédico: el sentido de sus distinciones”); después he sometido a análisis las ventajas y las limitaciones de nuestra exploración cuantitativa del español de México, tratando así de responder, con hechos, la pregunta que permanentemente se hace mucha gente: ¿hasta qué punto sirve un corpus de datos? y, a la vez, tratando

de desenmascarar el prejuicio que hace a tantos afirmar que lo encontrado en una investigación empírica “no corresponde a la realidad”, como si hubiera otra manera de conocer la realidad que no fuera, precisamente, mediante una investigación empírica (los artículos “Caracterización metódica del *Corpus del español mexicano contemporáneo*” y “La cuantificación en el *Diccionario del español de México*”); con el mismo método cuantitativo y sobre la base de nuestros datos, he investigado algunos fenómenos interesantes del léxico, que tienen repercusiones sobre nuestra redacción (en “El diccionario entre la tradición y la realidad”, “La sociolingüística del *Diccionario del español de México*” y “La cuestión de la norma en el *Diccionario del español de México*”); por último, presento uno de los elementos del sistema de análisis semántico que efectuamos con cada vocablo del diccionario (“Problemas y métodos del significado estereotípico”) y una pequeña pieza polémica (“Lagunas y debilidades de la lexicografía hispánica”) con la que debía yo promover una discusión en la mesa redonda de lexicografía del XVIII Congreso Internacional de Lingüística y Filología Románicas.

La mayoría de estos trabajos ha sido presentada y publicada en el exterior, pues allí es donde se dan las discusiones de nuestra disciplina y no en México, donde por desgracia somos todavía muy pocos —y con la miseria que nos hacen padecer a los universitarios pronto seremos menos— los que profesamos esta actividad. Los artículos del primer bloque son resultado parcial de mi trabajo como investigador visitante del Seminario de Romanística de la Universidad de Heidelberg entre 1983 y 1984, realizado con el generoso

apoyo de la Fundación Alexander von Humboldt, de la República Federal de Alemania y con el interés de mi querido maestro Kurt Baldinger; “El objeto diccionario” lo leí en el Simposio de lexicografía de la Asociación Mexicana de Lingüística Aplicada, realizado en el Instituto de Filología de la UNAM en 1987; “El diccionario, instrumento de la etnicidad” en el Encuentro sobre políticas del lenguaje, que organizó la Dirección General de Culturas Populares en Oaxaca, en 1984; “Diccionario de lengua, enciclopedia, etc.” me fue encargado por Franz Josef Hausmann, Herbert E. Wiegand y Ladislav Zgusta, editores de la obra *Dictionaries, an International Handbook of Lexicography*, que pronto aparecerá en la editorial W. De Gruyter, en Berlín; su original está en francés. La “Caracterización metódica del CEMC” fue mi ponencia al 1er. Coloquio sobre lexicografía del español de América, que organizó Günther Haensch, de la Universidad de Augsburg, en el Instituto Caro y Cuervo de Bogotá, con ayuda generosa del Instituto Goethe, en 1988; “La cuantificación en el DEM” es mi contribución al homenaje al gran lexicógrafo francés Bernard Quemada, que prepara Antonio Zampolli en el Laboratorio di Linguistica Computazionale de Pisa; “El diccionario entre la tradición y la realidad” fue una ponencia al Simposio sobre lenguaje y tradición en México, que organizó Herón Pérez Martínez en El Colegio de Michoacán, en Zamora, en 1988 (el título lo modifiqué para este libro; antes era: “Entre la realidad y el diccionario”); “La sociolingüística del DEM” corresponde a una invitación de Yolanda Lastra y Joshua Fishman para participar en un volumen del *International Journal of the Sociology of Language*

dedicado a la sociolingüística en México; “La cuestión de la norma en el DEM” apareció, en francés, en la *Revue québécoise de linguistique*, 17,2 (1988) como parte de un volumen dedicado a “Problemas de la lexicografía en América”, y gracias al interés de Claude Poirier, quien se ocupa de lexicografía quebequense, un tema con problemas muy similares a los nuestros. Los “Problemas y métodos del significado estereotípico”, también parte de mi investigación sobre la semántica del diccionario en Heidelberg, en 1983-1984, los presenté en el XVIII Congreso Internacional de Lingüística y Filología Románicas, en la Universidad de Treveris, en 1987, y aparecieron, junto con “Lagunas y debilidades de la lexicografía hispánica” en el tomo IV de las *Actas* correspondientes, publicadas por Dieter Krämer en la editorial Max Niemeyer de Tübingen, en 1989. De todos ellos he recibido autorización para reproducir estos artículos y, en su caso, para presentar las versiones correspondientes en español, lo cual les agradezco sinceramente.

El lector notará que las explicaciones iniciales sobre nuestro trabajo, y algunas ideas centrales, se repiten artículo tras artículo. Deseo pedirle que disculpe tal redundancia, causada por el hecho de que todos ellos estuvieron destinados a públicos, en principio, distintos, que no podrían conocer los antecedentes de nuestra labor y que echarían de menos esa clase de información.

Agradezco a mis compañeros de trabajo: a Gilberto Anguiano, Luz Fernández, Isabel García Hidalgo, Roberto Ham, Natalia Mata, Francisco Segovia, Laura Sosa, Carmen Delia Valadez y Carlos Villanueva,

su colaboración en la preparación de estos estudios. A Josefina Camacho, su paciente organización y transcripción de todos ellos.

Octubre de 1989



## EL OBJETO DICCIONARIO\*

**0.** Durante cientos de años la lexicografía ha sido una práctica lingüística. Sus métodos, sus técnicas y sus doctrinas, que han producido tantos y tan buenos diccionarios, se han creado y perfeccionado en independencia de la lingüística (pues de hecho la anteceden en la historia) y por lo general no han gozado de la estima de esta ciencia, para la que la lexicografía ha sido siempre una práctica arbitraria —como el orden alfabético y la noción de “palabra”—, comercial, y lingüísticamente injustificada.

Sin embargo, desde hace aproximadamente quince años la lexicografía ha venido despertando el interés de la lingüística hasta convertirse, hoy en día, en el tema central de múltiples reuniones y congresos internacionales. No es fácil discernir los motivos por los que la lingüística contemporánea se ha venido a interesar en la lexicografía. Posiblemente uno de ellos sea la ansiedad que se percibe en el gremio lingüístico por reencontrar la práctica y el empirismo, como una reacción en contra del exceso de especulación que la ha caracterizado durante los últimos veinticinco años. Y posiblemente otro motivo lo constituya la propia evolución de uno de los paradigmas dominantes de la

\*Ponencia presentada en el VII Simposio de la Asociación Mexicana de Lingüística Aplicada, dedicado a la lexicografía, en septiembre de 1987.

lingüística moderna, la gramática transformacional, que se ha interesado notablemente por estudiar el papel que toca jugar al léxico en el esquema general de su teoría de la lengua.

La lexicografía, por su cuenta, también ha contribuido a alimentar el interés contemporáneo en ella misma, desde el momento en que varios lingüistas, quizá por azares del destino, se vieron envueltos en proyectos lexicográficos de importancia y decidieron comprometerse en la investigación de las características de esta disciplina, para poder delimitarla entre las ciencias del lenguaje y constituir la cabalmente en una de ellas. Prueba de ello es la coincidencia afortunada de tres importantes obras de la lexicografía contemporánea editadas el mismo año de 1971 y provenientes de destacados lingüistas: el *Etude linguistique et sémiotique des dictionnaires français contemporains*, de Jossette Rey-Debove, del equipo de trabajo del diccionario *Robert*; el *Manual of Lexicography*, de Ladislav Zgusta, famoso indoeuropeísta; y la *Introduction à la lexicographie: le dictionnaire*, de Jean y Claude Dubois, lingüistas de la casa Larousse. Estas tres obras, seguidas posteriormente por *Le lexique, images et modèles* (1977) de Alain Rey, editor en jefe del *Robert* y por varios artículos sustanciales de Bernard Quemada,<sup>1</sup> Robert Martin, Herbert Ernst Wiegand, Franz Joseph Hausmann, R.K.K. Hartman, el mismo Alain Rey y otros más, han sido los que no solamente hoy por hoy marcan la pauta en lexicografía, introduciendo mejoras importantes en la técnica lexicográfica y

<sup>1</sup>Además de su obra de carácter histórico sobre la lexicografía francesa: *Les dictionnaires du français moderne* (1969).

una racionalidad teórica que se echaba de menos en muchos trabajos anteriores, sino que además han inaugurado una cuestión crucial para poder situar la lexicografía dentro del campo de las ciencias del lenguaje: la de la existencia posible o no de una “teoría de la lexicografía”.<sup>2</sup> Tal cuestión, sin duda, se da en la medida en que la experiencia práctica de la elaboración de diccionarios ha podido reunir, a lo largo de los siglos, suficiente reflexión como para que hoy madure la necesidad de construir un marco general de las características de la lexicografía, en cuanto disciplina que toca profundamente el lenguaje, hasta el grado de llegar a influir sobre la misma idea de la lengua que tienen los miembros de una sociedad.

**0.1.** Aunque también es cierto que por “teoría de la lexicografía” todavía pueden entenderse dos cosas diferentes: por un lado, se puede llamar “teoría” el conjunto de los aspectos metódicos y técnicos que intervienen en la elaboración de los diccionarios;<sup>3</sup> por el otro, se puede llamar “teoría” al desarrollo de un

<sup>2</sup>Algunos autores prefieren nombrar a esta teoría *metalexigrafía*; en un artículo de próxima publicación, titulado “Une critique du concept de métalangage”, me he colocado entre los partidarios de un significado restringido de ese concepto (así como del prefijo *meta-*), en consonancia con su sentido originario, dado por David Hilbert, su creador. De ahí mi consciente omisión de este termino y la aparición de algunos otros efectos terminológicos que, a lo largo de este trabajo, se harán evidentes.

<sup>3</sup>Un ejemplo, ya anticuado, de esta clase de concepción de la lexicografía lo constituye el libro de Julio Casares, *Introducción a la lexicografía moderna* (1950); otro, la parte del libro *La lexicografía* (1983) redactada por Gunther Haensch (independientemente del hecho de que él dirigió toda la composición de la obra).

conjunto de ideas referentes a la naturaleza de los diccionarios y de su papel en la vida social de una comunidad lingüística.

**0.1.1.** La primera clase de “teorías” es sin duda muy útil para los lexicógrafos y conduce al desarrollo de diccionarios cada vez mejor documentados y más consecuentes con la realidad de la lengua a la que se refieren y las necesidades de sus usuarios; sin embargo, limitadas esas “teorías” a los aspectos de método lexicográfico, no son capaces de convertirse en una verdadera teoría, en el sentido científico del término, y más bien tocan ciertos planteamientos lingüísticos fundamentales con tanto temor y tan verdadera insuficiencia, que fácilmente se vuelven blanco de las críticas de la lingüística.

Hay problemas de la elaboración de los diccionarios que no puede resolver una metodología por ella misma. Son precisamente todos aquellos que atañen a la naturaleza de la lengua, o del léxico, y que se deben tratar con algo más que métodos y técnicas, considerando en profundidad las cuestiones de principio que se plantean legítimamente en la teoría de la lengua y no en el ámbito reducido de la lexicografía, entendida como método de la elaboración de diccionarios. Así por ejemplo, la cuestión de cómo se delimitan los signos que pueden aparecer como entradas en el artículo lexicográfico (afijos, raíces, palabras, locuciones) no se soluciona con un concepto trivial de “palabra” como el muy usual de “expresión sonora o gráfica delimitada por dos pausas o por dos espacios en blanco”, sino atendiendo a la naturaleza de la palabra desde la construcción teórica que se tiene que elaborar para concebir la naturaleza de la lengua. Lo

mismo se requiere cuando se trata de establecer los límites entre varios posibles significados de una palabra, que habrán de traducirse en acepciones linealmente ordenadas en el artículo lexicográfico, pues no basta con la intuición del lexicógrafo y una receta para separarlos y numerarlos, sino que hay que tener previamente elaborado un concepto de la polisemia, considerada como fenómeno propio de la lengua. Estos dos problemas, con muchos más a los que no me referiré para no hacer este artículo excesivamente prolijo, no se resuelven con métodos y técnicas solamente, sino que requieren una solución del orden teórico, que habrá de encontrarse en la teoría de la lengua y en la teoría de la lingüística que tiene por objeto a aquélla.

**0.1.2.** Por el otro lado, afirmar lo anterior no implica una dependencia total de la lexicografía en relación con la lingüística, pues es posible proponerse otra manera de concebir lo que es la teoría de la lexicografía, a partir del convencimiento de que la lexicografía ha venido desarrollando durante centurias conocimientos del orden práctico —en el sentido usual de la palabra, pero también en sentido kantiano— que no se reducen a recetas técnicas, sino que refieren a cuestiones nucleares de la teoría de la lengua, como lo es su percepción del sentido y el significado, que da lugar al hecho asombroso de que nada menos que sociedades completas *crean* que, en efecto, el significado de las palabras de su lengua y el sentido de sus discursos se puede recomponer a partir de las definiciones asentadas en un diccionario. La adecuada ponderación de este fenómeno, que forma parte de la práctica cotidiana de un lexicógrafo, es ya materia suficiente como

para sostener que la lexicografía puede desarrollar su propio pensamiento teórico y contribuir al conocimiento del lenguaje de una manera desconocida para la lingüística.

Ha sido precisamente esa clase de experiencias de la lexicografía la que parece estar en el origen del florecimiento contemporáneo del interés por esta disciplina. Entre ellas es en donde deseo situar la mía, de lingüista y de lexicógrafo, que también azares del destino me han llevado a tener durante los catorce años que ya dura la elaboración del *Diccionario del español de México*.<sup>4</sup> A partir de esa experiencia, lo que pretendo es construir una teoría completa de lo que es un diccionario monolingüe, desde el punto de vista semántico. En el breve espacio del que dispongo ahora, me limitaré a un par de cuestiones iniciales de esa teoría.

1. Toda teoría requiere, como punto de partida, una definición precisa de lo que constituye su objeto. En este caso, lo que requiere es una definición de lo que es el diccionario, de manera que la objetivación que produzca establezca las pautas que puede seguir y los límites que puede encontrar a lo largo de su desarrollo.

1.1. A diferencia de lo que pasa con otros objetos de la lingüística, como por ejemplo el fonema, el morfema o

<sup>4</sup>En la elaboración teórica y práctica del *DEM* está la base de este artículo, que constituye un breve resumen del primer capítulo de un trabajo que, con el título de *La complejidad semántica del diccionario*, he venido preparando desde 1983, cuando la generosidad de la Fundación Alexander von Humboldt me permitió gozar de un año como investigador visitante de la Universidad de Heidelberg.

la oración que, no obstante su íntima constitución social, revelan también un carácter natural, el diccionario no es un objeto de la misma clase, que la lingüística conciba espontáneamente como algo tan natural como aquéllos elementos de la lengua;<sup>5</sup> todo lo contrario, como el número de comunidades lingüísticas del planeta que tienen diccionarios es, sin duda, mucho menor que el de las comunidades que no los conocen, para la lingüística parece evidente que los diccionarios no tienen carácter necesario, ni parecen corresponder a ningún tipo de comportamiento lingüístico de carácter universal. A la vez, las convenciones de presentación de los diccionarios (y aquí me referiré, sólo para abreviar, a diccionarios de lenguas del tipo de la nuestra), como el orden alfabético; el arreglo de la macroestructura sobre la base de formas canónicas relativamente arbitrarias, como lo son las formas singular y masculina —cuando es el caso, por supuesto— de los sustantivos, y las formas infinitivas de los verbos; la eliminación, muchas veces, de las flexiones y las conjugaciones de esas clases

<sup>5</sup>Aquí me refiero, por supuesto, a la sensación que tiene todo hablante, de que su lengua materna “ya está allí” desde siempre, naturalmente dispuesta a servirle de medio de expresión. Que esta idea es la común entre los hablantes de cualquier lengua lo prueban dos hechos: el primero, y socialmente más importante, es la dificultad que tiene cualquier persona que intente convencer a un hablante común de que su lengua materna se puede *educar* y por lo tanto se puede modificar el uso que hace de ella (cuestión tan difícil, que no dudo en situarla como el primero de los problemas que tiene la institución escolar para justificar ante sus alumnos la necesidad de una materia llamada, en nuestra circunstancia, “español” y para que los maestros encuentren el sentido y la dirección de lo que deben enseñar en esa materia); el segundo, la dificultad que tenemos los lingüistas para explicar cuál es nuestro papel en la división social del trabajo.

de vocablos cuando se consideran regulares; la coerción ejercida por la linealidad del texto sobre el orden de las acepciones, basado además en criterios de jerarquía de los significados que muchas veces parecen caprichosos; la autolimitación a nociones de la palabra que fácilmente se contradicen (como la famosa discusión, en la lexicografía francesa, sobre la *pomme-de-terre*); etc., parecen suficiente motivo para que la lingüística se desentienda de los diccionarios y los considere extrañas construcciones comerciales, que no tienen nada que ver con la realidad de las lenguas.

Desde esa clase de puntos de vista, todavía dominantes en lingüística, si lo que constituye el objeto del diccionario es el léxico, y el léxico es parte de un sistema lingüístico, la función del diccionario se reduce a la de simple catálogo de elementos de ese léxico; es decir, el diccionario, como documento descriptivo, merece, a lo más, una consideración del orden metódico o técnico, que lo vuelva útil para las finalidades descriptivas generales de la lingüística. Aun más, si la lingüística es rigurosamente post-saussureana, el diccionario, que sólo contiene elementos léxicos ya realizados y no puede contener realizaciones posibles determinadas por la virtualidad del sistema lingüístico, no es parte de la *lingua* —y por definición tiene que quedar fuera de los objetivos de la lingüística— sino del *habla*, de la “suma de hablas”, que establece mi maestro Klaus Heger. Evidentemente, si tal fuera el caso, que los diccionarios fueran meros catálogos descriptivos del *habla*, la lexicografía no sería, en efecto, ni siquiera una lingüística aplicada, sino apenas una parte de la metodología de la descripción.

Es esa manera de concebir el diccionario lo que nos lleva, en lexicografía, a considerar urgente una defini-

ción del diccionario como objeto, en que se gane para la teoría una especificidad del diccionario que, hasta donde llegan mis conocimientos, nunca ha sido tomada en cuenta.

**1.2.** Para ello, consideremos el diccionario en cuanto libro, pues fue así como apareció en la historia de muchas lenguas y me atrevo a afirmar que tal hecho no fue fortuito, como veremos en seguida. En cuanto libro, el diccionario presupone la existencia, por un lado, de un sistema de escritura que permita plasmarlo de una manera perceptible; por el otro, de una cultura en la cual la fijación y la conservación de un discurso tenga una función específica para sus instituciones sociales. El interés de fijar y conservar un discurso tampoco es gratuito; corresponde al valor que conceden muchas sociedades a la memoria de sus experiencias, tanto con finalidades prácticas inmediatas, como podría suceder con los cálculos calendáricos que requiere una sociedad agrícola para asegurar sus cosechas, como con finalidades históricas y míticas, en las que encuentra explicaciones de los acontecimientos pasados y predicciones de los futuros. Escritura y libro han sido, consecuentemente, dos extraordinarias invenciones sociales, que han ampliado enormemente la capacidad de memoria de las sociedades que los utilizan. Si ahora consideramos qué clase de discurso es el que conserva el libro-diccionario, veremos que se trata nada menos que de aquella parte de las lenguas que entra en contacto inmediato con la realidad; es decir, el léxico, con el cual se manifiesta la experiencia y se nombra el mundo que nos rodea. El diccionario es, entonces, fundamentalmente un depósito de la experiencia social manifiesta en palabras, que

deriva su valor de su capacidad de superar la memoria de cada uno de los individuos que componen la sociedad. En la medida en que una sociedad crece en número de miembros, en complejidad de su organización y en fragmentación de sus conocimientos, el diccionario adquiere un papel más preponderante como acervo social al que concurre la totalidad de las experiencias individuales y al que habrá de referirse cada individuo cuando desee comunicarse con los demás acerca de temas que no domina.

El diccionario revela en esa forma su carácter radicalmente social. En efecto, como piensa la lingüística, se ha dado como una invención que no tiene nada que ver con la naturaleza de la lengua a la que se refiere; sin embargo, la propia lengua, en cuanto es el vehículo privilegiado de los seres humanos para representar y conocer el mundo, al dar lugar a la formación de un léxico teóricamente ilimitado, que crece con el paso del tiempo y supera fácilmente la memoria de cada hablante, hace del diccionario una necesaria memoria social que de inmediato trasciende su carácter utilitario para venir a representar nada menos que el propio carácter social de la lengua.

Que el diccionario sea un libro escrito le confiere además, una cualidad esencial de la escritura, cuyo efecto es de suma importancia para la lengua: pues no solamente representa la lengua y la fija, sino que, al formalizarla en los trazos de las letras y la configuración de las palabras, la hace patente a la reflexión de sus hablantes —lo que algunos amigos míos, de los que discrepo, llaman la “conciencia lingüística” o hasta “metalingüística”—, operando así un determinante extrañamiento del individuo frente a su lengua, por el

cual ésta se vuelve inteligible en cierta medida y abandona, de una vez para siempre, la inmediatez natural con que se le presenta al hablante no educado. Cuando las palabras se escriben, el diccionario pasa a desempeñar dos funciones: primero, como acervo de la lengua social (permítaseme el pleonasma); después, como representador de la lengua en cuanto ella misma. Lo que encuentra el hablante en el diccionario es, consecuentemente, un léxico que se le presenta como hecho social, como algo en lo cual la sociedad a la que pertenece, está de acuerdo; y al mismo tiempo, ese hecho, en cuanto manifestación de una lengua que por ese medio se objetiva, adquiere para el hablante un carácter simbólico: la lengua que habla no solamente ocurre en el fugaz instante de su discurso, sino que tiene unas fronteras y un carácter que el diccionario le devela.

Siendo así, el diccionario es más que un mero catálogo descriptivo del léxico. Es la presentación de una realidad social y de una memoria colectiva, que constituye el punto de referencia de todo hablar que quiera tener sentido para la sociedad.

Es en ese carácter social, en su calidad de consenso operado en la comunicación de las experiencias a las que la sociedad confiere un sentido, en donde se encuentra el primer elemento de la constitución de objeto del diccionario.<sup>6</sup> El diccionario, entonces, tiene un

<sup>6</sup>Hay que aclarar la enorme diferencia que hay entre la idea, desgraciadamente originada en Saussure y otros fundadores de la lingüística moderna, de *convención* en el origen ontogenético de la lengua, y la de *consenso* que estoy defendiendo: lo que quiere destacar la primera es el hecho de que la lengua no tiene una motivación en la propia naturaleza de las cosas. Pero de este hecho

sentido en relación con la sociedad a la que pertenece; y para los hablantes la selección de vocablos que contiene, la presencia o ausencia de transcripciones fonológicas de la pronunciación de una palabra, la clase de información etimológica que ofrezca, las marcas de uso social o regional, el orden de sus acepciones y el texto que las define, los ejemplos de uso que introduce, etc. adquieren, por el sólo hecho de estar ahí, un valor que se reúne y se entremezcla con el resto de los valores sociales: el hablante cree que lo que indica el diccionario es lo que vale para la sociedad en su conjunto. Se hace así evidente la causa para que todo diccionario, aun sin que se lo proponga, tenga carácter normativo y pase a formar parte de las instituciones sociales, junto con las constituciones políticas y, en algunas sociedades, con la Biblia.

---

a la creencia de que realmente existe un “contrato social” de la comunidad lingüística acerca de su lengua hay mucho trecho, en particular, si sobre la base de la convencionalidad se hacen propuestas de reforma inmediata, racional y planeada de algún aspecto de una lengua. No hay tal convencionalidad en ninguna lengua; lo que las mantiene unidas para que sus hablantes se puedan comunicar entre sí, se da en el permanente proceso de significación dialógica que conlleva todo discurso, por el simple hecho de darse con la intención de comunicar algo a otra persona. De ese proceso de significación, que siempre va a recalar en las posibilidades de comprensión que ofrece la colección de discursos ya experimentados que forman la urdimbre de la tradición, se forma un *consenso* social acerca de lo que tiene o no significado en cada discurso nuevo. Es entonces la memoria colectiva de la sociedad, manifiesta en su tradición lingüística y en su lengua real, concreta —no en su virtualidad combinatoria—, la que origina el consenso en el seno de la sociedad y antes de la aparición o no de las instituciones sociales, como los diccionarios, los usos cortesanos, o las academias.

2. Hay un segundo elemento con que se puede definir el carácter de objeto del diccionario. Como recordarán, en su *Teoría del lenguaje* Karl Bühler propone que toda expresión verbal, en cualquier situación, constituye una *acción verbal*. Así, un grito de dolor, la llamada de atención de un niño a su madre o los soliloquios a que son tan dados los niños pequeños, son acciones verbales. Entre todas las acciones de este tipo, la comunidad lingüística destaca varias, que por ser desencadenadoras o portadoras de un sentido que supera la pura necesidad individual de expresión, y se refieren a experiencias que interesan a la sociedad en cuanto tal, adquieren un valor específico entre todas las acciones posibles: se vuelven *actos verbales*. Los actos verbales corresponden, entonces, a acciones portadoras de un sentido social, localizadas dentro de una red de actos socialmente instituidos. Como más tarde señalará John Searle, y aparentemente sin haber tomado en consideración la obra de Bühler, los actos verbales son instituciones sociales.

2.1. Hay cierto tipo de actos verbales que ocurren en la mayor parte de las sociedades, si no es que en todas, dado su carácter.<sup>7</sup> Pues quizá en todas ellas existe la

<sup>7</sup>Como lingüista, me siento más proclive a considerar que los actos verbales tienen una especificidad antropológica, determinada por las características de cada sociedad. Su investigación forma parte, en consecuencia, de lo que Dell Hymes llama "etnografía del habla", aun cuando su correcta interpretación está más cerca de una antropología filosófica. Sin embargo, tratándose de actos verbales como el de la pregunta acerca del significado y su respuesta, es posible que se trate de actos de carácter universal y que, en consecuencia, formen parte de la pragmática trascendental propuesta por Jürgen Habermas y Karl Otto Apel.

posibilidad, reconocida por la sociedad, de preguntar qué significa una palabra y de responder a esa pregunta. Lo hacen desde el niño pequeño hasta el individuo más viejo y más sabio de una comunidad. El acto de la pregunta acerca del significado de una palabra, y el acto de respuesta a esa pregunta son actos típicos entre todos los que tienen sentido para una sociedad. Pues bien, la existencia de un diccionario depende de la existencia de esa clase de actos verbales; si no los hay, el diccionario no tiene sentido; si los hay, el diccionario deriva de ellos su función.

Las preguntas y las respuestas acerca del significado de una palabra son actos verbales socialmente instituidos; es decir, la sociedad reconoce su intención y les confiere un lugar determinado entre todos los actos con sentido que se pueden llevar a cabo. Cuando quien responde una pregunta de esa clase es un diccionario, se revela en él su carácter de catálogo codificado de actos verbales de respuesta.<sup>8</sup>

**2.2.** De acuerdo con esta interpretación del diccionario, el artículo lexicográfico es un acto verbal escrito. Y en cuanto acto verbal, tenemos que situarlo bajo la especie

<sup>8</sup>De ese carácter deriva la disciplina escolar, todavía importante en muchas comunidades lingüísticas, que consiste en enseñar a usar el diccionario. Como, de nuevo, con el diccionario no se trata de un fenómeno “natural” del habla, es claro que la sociedad necesita socializar a sus miembros en su manejo (y es lamentable que, en México, muchos maestros de escuela crean que la única finalidad de la enseñanza del uso del diccionario sea el aprendizaje del orden alfabético, para lo cual convierten el ejercicio de consulta del diccionario en un simple acertijo de búsqueda de palabras muy raras, que por lo general tienen que resolver los padres de familia y que siempre borra el sentido del diccionario).

de los *generis dicendi* (lo que muchos lectores de Searle traducen como “actos performativos”). En el artículo lexicográfico, la palabra que constituye la entrada, el *lema*, es una *mención* del signo (cf. mi trabajo, antes citado, sobre el metalenguaje). Es decir, el acto verbal comienza por la *cita* de un vocablo, acerca del cual se da una respuesta. El resto de las proposiciones que siguen al lema, como las transcripciones fonológicas, las marcas gramaticales y de uso y, sobre todo, las definiciones, que reconstruyen el significado del vocablo mencionado por medio de una perífrasis, constituyen el contenido proposicional del acto, ligado al lema mediante lo que podríamos llamar un “conectivo de lengua” (no en balde la mayoría de los lexicógrafos contemporáneos hablan de una elisión del verbo *significar* en el espacio en blanco entre la entrada y la microestructura del artículo lexicográfico). Este conectivo, que en el caso de la relación entre lema y definición es un conectivo sinónimo (pues define la relación como una sinonimia entre lema y definición), tiene el mismo carácter que los llamados “verbos de lengua” o *verbum dicendi* de la gramática antigua (los “verbos performativos” de las traducciones de Searle); carácter que consiste en afirmar que la relación entre ambos es verdadera. Tal afirmación no viene, entonces, del arbitrio del lexicógrafo o de una simple convención social y tipográfica, sino que se deriva de la naturaleza del acto verbal de respuesta, que es de carácter social. Sólo un proceso de socialización, que a partir del reconocimiento del valor de una memoria colectiva de la lengua, da sentido al juego de actos verbales de pregunta y de respuesta acerca del significado de una palabra, es el que puede definir la verdad de la relación entre lema y definición.

En comunidades lingüísticas como la nuestra, ese proceso es el que determina la creencia del lector de un diccionario, de que el significado de una palabra contenida en el diccionario *es* realmente su verdadero significado; y lo que eso quiere decir, es que ese significado se considera socialmente válido.

**2.3.** El diccionario, por lo tanto, en cuanto objeto, tiene una naturaleza fundamentalmente social, basada en la existencia de una memoria colectiva de la lengua y en la institucionalización de la posibilidad que tienen los miembros de la sociedad de preguntar y recibir respuestas acerca del significado de los vocablos incluidos en esa memoria. Sin duda el diccionario es un producto cultural que, en nuestro caso, ha aparecido como efecto de dos invenciones importantes: la escritura y el libro. Pero este carácter delimitado por la cultura no lo convierte en un objeto desdeñable por particular y por secundario en relación con la lengua, sino que por el contrario lo hace tema de una elaboración que la lingüística debe tomar en cuenta, si lo que quiere es verdaderamente conocer todas las dimensiones del lenguaje, que es lo que corresponde a una verdadera lingüística pragmática, capaz de trascender las limitaciones del estudio restringido al sistema lingüístico y a los “componentes” sociales, prácticos, etnológicos, etc. que caracterizan a varias “interdisciplinas” de la lingüística contemporánea.

**3.** Las consecuencias que tiene todo lo anterior para la lexicografía son enormemente importantes, pues por un lado, nos podemos dar cuenta de que la “teoría del diccionario”, que debe formar el núcleo de una

posterior “teoría de la lexicografía”, tiene elementos hasta la fecha intactos tanto por las teorías lingüísticas como por las teorías de la lengua. Lo que se revela de la lengua en una consideración radicalmente lexicográfica como la presente, es más que lo que ha podido notar la lingüística; y de allí, la importancia y la necesidad de que la lexicografía se dedique a elaborar una teoría a partir de la reflexión que le produce su propia experiencia práctica. Como se ve, la lexicografía es lingüística aplicada en el sentido más noble del término: es lingüística dedicada a establecer sus lazos con la sociedad; no es, en cambio, simple “aplicación” de teorías lingüísticas a la elaboración del diccionario.<sup>9</sup> Si lo fuera, no sería posible hacer más teoría de la lexicografía que lo que supone la primera de las interpretaciones a las que me referí al inicio de este artículo: una verdadera metodología.

Por el otro lado, el hecho de que el lexicógrafo, por las características indicadas del acto verbal de respuesta sobre el significado de una palabra, se vea obligado a presuponer que todo lo que él diga en la definición de un vocablo será tomado como cierto por su lector, introduce, en la propia teoría del diccionario, una dimensión ética nunca considerada por la metodología lexicográfica ni por la lingüística.

Pues el acto verbal de respuesta a la pregunta acerca del significado de una palabra tiene una condición de verdad, según la cual el lector del diccionario cree que lo que le responde el lexicógrafo es cierto. El lexicogra-

<sup>9</sup>Para una discusión inicial de estas dos maneras de comprender la lexicografía, cf. mi reseña a Haensch *et al.*, *La lexicografía*, en *Anuario de letras*, 23 (1985), pp. 313-323.

fo, en consecuencia, no puede soslayar su presencia en el diccionario como verdadero sujeto del acto verbal. El diccionario, por lo tanto, tiene siempre un autor, no en el sentido —socialmente menos importante— literario (como sucede con los llamados “diccionarios de autor”, como el *Diccionario secreto* de Camilo José Cela, o el *Diccionario del diablo*, de Ambrose Bierce, que nunca son juzgados por el lector a partir de la condición de verdad mencionada), sino en el sentido ético que involucra al lexicógrafo como sujeto responsable de transmitir el conocimiento del significado socialmente importante. Visto lo anterior desde una consideración científica, la lexicografía viene a establecer el punto de encuentro entre la “razón pura” de la ciencia del lenguaje y la “razón práctica” de la moral (bajo la suposición, que no concesión, de que la lingüística sea solamente “razón pura”).

En este sentido, el lexicógrafo es un mediador entre la sociedad y su lengua; su trabajo no solamente consiste en estudiar y analizar cuidadosamente los documentos de que dispone para comprender el significado de un vocablo; ni sus criterios de trabajo se quedan dentro de los famosos tres principios científicos de la coherencia interna, la exhaustividad y la sencillez. Su trabajo consiste, ante todo, en su capacidad de reconstruir, mediante criterios científicos, el significado que tienen los vocablos socialmente. Este tema, desde hace mucho tiempo reconocido y tratado por la hermenéutica, confiere a la teoría lexicográfica una segunda característica que hoy no puedo ya desarrollar.

## EL DICCIONARIO, INSTRUMENTO DE LA ETNICIDAD\*

Que para Claude Lévy-Strauss, según relata George Steiner, el diccionario sea el mejor documento de una cultura, no implica que el diccionario sea un objeto natural, algo que aparece inevitablemente en un momento del desarrollo intelectual de una comunidad lingüística (¿una etnia?). Me explico: en contra de la opinión común, de que un diccionario es una lista de palabras cuyo principal interés es la consulta ortográfica o la legitimación del uso de un vocablo, el diccionario es fundamentalmente un depósito de contenidos culturales, objetivados en la unidad privilegiada que es la palabra o su equivalente morfológico en lenguas aglutinantes o polisintéticas. El diccionario recopila y presenta en un orden que se quiere siempre pertinente —sea etimológico, histórico o estereotípico— el acervo léxico de una sociedad, es decir, el conjunto de vocablos que se utilizan en el hablar de una comunidad y mediante los cuales se establece la inteligibilidad de cada discurso individual. El diccionario no contiene lo que es posible decir, sino solamente aquello que hace posible comprender algo dicho. El diccionario es así una memoria de la sociedad. Si, de acuerdo con las

\*Ponencia presentada en el Coloquio sobre Política del Lenguaje en México, Oaxaca, 28 de noviembre de 1984.

teorías modernas de la biología, una característica privilegiada de la especie humana es su capacidad de memoria, habría que agregar que la existencia de la sociedad contribuye igualmente a la formación de una memoria que da sentido, histórica e individualmente, a cada acto verbal y que establece el telón en donde se objetiva una identidad colectiva y una personalidad individual. Para esta memoria, el diccionario es su instrumento, pero además, a partir de él, es su principio de organización, su base de reflexión.

Sin embargo, los diccionarios no aparecen, como decía antes, como productos “naturales” o “necesarios” en el desarrollo de una sociedad, sino que son construcciones intelectuales que dependen de otras anteriores. Históricamente, para dar algunos ejemplos, sabemos que en el antiguo Egipto aparecieron como glosarios, o diccionarios multilingües que facilitaban el comercio entre Egipto y otros pueblos del Medio Oriente; en la India, interesaban más como registros detalladísimos de la fonética del sánscrito, que permitían la repetición ritual de los textos sagrados y garantizaban su efecto sobrenatural; en China, fueron catálogos de la infinita escritura logográfica, que facilitaban el reconocimiento y la generación de ideogramas. Sólo en la edad moderna, y marcado por el ascenso de Occidente, el diccionario comienza a entenderse como memoria social y obra socializadora (¿pedagógica?). Así se puede entender por qué no en todas las lenguas hay diccionarios y por qué el diccionario no tiene carácter necesario.

Esta discusión previa de lo que es el diccionario me parece importante en el contexto de una discusión sobre etnicidad porque no quiero dar lugar a una

aceptación instantánea y poco reflexionada de mis proposiciones lexicográficas, ni quiero pretender que la lexicografía es la parte crucial o de principio de una política lingüística. Pero esto último tampoco quiere decir que la lexicografía no forme parte importante de tal política.

El diccionario tiene ciertos requisitos previos: primero, en cuanto obra escrita (pues el diccionario hablado es una posibilidad hasta hoy no realizada), implica que la comunidad sabe leer y escribir; aun más, implica que para la comunidad tiene sentido saber leer y escribir (algo que muchos alfabetizadores parecen nunca haber tomado en cuenta); segundo, implica que para la comunidad tiene sentido proveerse de una obra de consulta acerca de ella misma, lo cual presupone que saber acerca de sí mismos es un interés compartido por los miembros de la comunidad; tercero, implica que la comunidad reconoce, aunque sea intuitiva o inconscientemente, que tiene límites y que hay otras comunidades diferentes a ella, por lo que la comparación o el contraste con otras comunidades, sea para legitimarse a sí misma o por puro interés universalista, es un valor de la socialización y la educación comunitarias; y cuarto, implica que existe una actividad verbal mayor que la simple comunicación cotidiana de necesidades, sea religiosa, laboral, escolar, científica, técnica, periodística, deportiva o política. Entonces, y sólo entonces, es pertinente hablar de lexicografía.

En la metodología lexicográfica existen dos clases de diccionarios y dos objetivos de diccionarios: las clases son el diccionario monolingüe —como el *Diccionario del español de México*— o el multilingüe, sea de sólo dos lenguas, como el *Diccionario maya Cordemex* (Barrera

Vázquez, 1980), o sea de varias. Los objetivos son el diccionario lingüístico, como el *Great Tzotzil Dictionary of San Lorenzo Zinacantán* (Laughlin, 1975), y el diccionario de lengua, como el que dirijo, o como el *Larousse* o el de la *Academia Española*.

Las clases de diccionarios son fácilmente comprensibles, por lo que no me detendré en ellas; los objetivos de los diccionarios requieren una explicación: un diccionario lingüístico es una herramienta de trabajo científico; lo compone un lingüista como parte de su labor descriptiva y tiene por objetivo central proveer a la descripción global de una lengua de su correspondiente parte léxica, así como recopilar y presentar datos morfológicos, sintácticos, y de derivación y formación de palabras (o los sintagmas equivalentes). Un buen ejemplo es el antes citado diccionario del tzotzil de Zinacantán: la entrada ofrece el sintagma autónomo similar a nuestra *palabra* española, pero en la entrada misma hay una transcripción fonológica (que no necesariamente sería ortográfica), seguida de una serie de marcas de funciones gramaticales. El campo de la microestructura asociada a la entrada lo ocupa una equivalencia aproximada al inglés, que no puede considerarse definición, sino sólo una ayuda heurística para el usuario de la obra. Dadas las características estructurales del tzotzil, estas equivalencias ofrecen varias formas de interpretación del sintagma entrada que, aparentemente, no llegan a ser acepciones, tal como las entendemos en los diccionarios de lenguas flexionales como el español.

Lo que incluye este diccionario, su “vocabulario”, es, según relata Laughlin en la introducción, el resultado de un dilatado levantamiento de datos en el campo, acom-

pañado de los resultados que obtuvo en encuestas directas, interrogatorios a sus informantes, durante los cuales el lingüista operaba con largas combinatorias morfológicas, para investigar cuáles de ellas tenían sentido; es decir, “creaba” palabras artificialmente y pedía a su informante que decidiera cuáles de ellas eran palabras reales, cuáles eran posibles y cuáles no.

Casos interesantes de este diccionario los constituyen las entradas de signos que refieren a objetos de la naturaleza, como los animales y las plantas. Para estos, Laughlin reunió muestras o especímenes de cada objeto y los entregó a especialistas norteamericanos en zoología y botánica. Hecho el reconocimiento científico de esos objetos, la entrada correspondiente ofrece sus nombres en las taxonomías científicas occidentales. Un ejemplo es la serie de nombres de los hongos que, como para muchos otros pueblos amerindios, implica una larga tradición religiosa y médica. El micólogo que llevó a cabo el estudio de los diez o quince especímenes recolectados por Laughlin en el campo no sólo identificó todos —como era de esperarse— sino que descubrió contradicciones entre la clasificación occidental y la aborígen; para los tzotziles hay hongos llamados del mismo modo, pero que corresponden a especies diferentes de la micología. Desgraciadamente, Laughlin no fue más allá, a buscar el sentido de la clasificación tzotzil, sino que se contentó con exponer el problema.

Y bien, un diccionario lingüístico, elaborado para las finalidades estructurales de la descripción, suele contentarse con esa clase de trabajo: una descripción de sintagmas autónomos, de funciones gramaticales básicas, y una aclaración de carácter heurístico sobre el significado

de cada sintagma. Los signos referidos al mundo sensible se clasifican desde el reconocimiento que se presupone objetivo en las ciencias de Occidente.

No sólo por la factura del diccionario lingüístico, sino también por otros hechos, como el que la microestructura aparezca en inglés, o el todavía más claro, que es su publicación por el Instituto Smithsonian, su venta en Washington y su precio en dólares, es evidente que no se trata de una obra hecha para la comunidad de Zinacantán.

Una variante del diccionario lingüístico es la que se puede ejemplificar con el *Diccionario maya CorJemex* de Alfredo Barrera Vázquez. En esta obra, dedicada a la recolección léxica del maya desde el siglo XVI hasta el presente, el objetivo principal es la elaboración de un instrumento de consulta para la filología maya. Se reúnen en ella todos los datos encontrados sobre léxico maya y sobre su interpretación desde los primeros cronistas españoles del pasado maya, y se espera poder alumbrar así el conomiento histórico de esta lengua. La parte microestructural está en español, mediante vocablos o expresiones relativamente equivalentes; hay acepciones e ilustraciones de sintagmas mayores, tomados de la bibliografía maya anterior a esta obra. Tiene una parte español-maya que difícilmente podría ser usada por un hispanohablante para traducir al maya; sino que más bien sirve como apoyo para la investigación de un filólogo hispanohablante. Cabe señalar aquí que la parte inglés-tzotzil del diccionario de Zinacantán tiene la misma función.

Por el contrario, los diccionarios de lengua se orientan al uso de la comunidad. Pretenden servir al hablante en su vida diaria. Aunque se hagan con criterios

lingüísticos —lo que es el caso mayoritario en los diccionarios de lenguas amerindias— pretenden ser útiles para una comunidad mayor y de intereses más diversificados. Tratándose de lenguas indígenas americanas, suelen ser bilingües, y tratándose de los que ha hecho el Instituto Lingüístico de Verano, se anuncian con el objetivo de fijar la lengua de acuerdo con una fonología y una gramática previamente establecidas, por una parte y, por la otra, servir de instrumentos de la llamada “castellanización”. Basta comparar el diccionario de Laughlin antes mencionado con el *Diccionario tzotzil de San Andrés* (Hurley/Ruiz Sánchez, 1978) para encontrar importantes diferencias, que sólo un especialista y no yo, podría definir como discrepancias: para comenzar, la grafía de ambos diccionarios es distinta; si la de Laughlin es fonológica, la del ILV implica una toma de posición ortográfica, lo cual a su vez implica una política de la lengua tzotzil; el arreglo alfabético del diccionario varía cuando el segundo diccionario toma en consideración lo que aparentemente son morfemas prefijos obligatorios; el artículo lexicográfico en general parece ser más rico en el primero, tanto en el registro de expresiones como en la profundidad del significado. Las equivalencias en español, aunque del mismo tipo en el segundo que las de Laughlin en inglés, tienen en aquél un objetivo más que heurístico y se orientan a la traducción entre ambas lenguas. Sin embargo, ninguno de ellos deja de dar la impresión de aleatoriedad en el registro de sus datos y de capricho en la redacción de sus vocablos.

Una variante del diccionario de lengua es el diccionario monolingüe y en este caso referiré al mío, al *Diccionario del español de México*, del cual han apa-

recido una versión mínima, llamada *fundamental* y una mayor, llamada *básica*. El diccionario de lengua monolingüe se centra en su comunidad lingüística e implica, como decía antes, una traducción interna, cuyo principal objetivo es la explicación detallada del significado de cada vocablo y la educación de la lengua del usuario de la obra. Esta clase de diccionarios profundiza en el significado, en la historia y en las tradiciones de la comunidad; ofrece a su usuario la posibilidad de una reflexión sobre la lengua, orientada hacia la cultura y hacia el aumento de precisión en la expresión verbal. Es obra pedagógica más que descriptiva e implica una política de la lengua que trasciende al campo de lo social y de los valores de la comunidad. Que yo sepa, no hay diccionarios de esta clase para las lenguas amerindias de México.

Para el diccionario de lengua y monolingüe —lo que también valdría para el bilingüe, aunque por desgracia casi no se hace— la exploración del significado busca verdaderas perífrasis del vocablo e intenta destacar los elementos formativos del sentido. Si se trata de verbos, por ejemplo, se interesa en todos sus detalles aspectuales —si son incoativos, iterativos, etc.—, en la relación que se establece entre sus actantes, en las causas o finalidades del proceso designado, y en cada uno de los elementos que hacen que haya tantos verbos distintos que, para decirlo rápido, justifican que no toda la lengua se reduzca a *hacer* más un modificador. Si se trata de signos referidos a objetos de la naturaleza, lo que importa es el cómo de sus significados, la tradición de sentido que está en su base; así, que los zorros sean listos o astutos —lo que ningún zoólogo estaría dispuesto a afirmar—, que haya tecolotes que, al cantar,

predicen la muerte —tradición hispánica ajena a la ornitología—, que la salamandra viva en el fuego, etc. pueden ser parte del significado social y tradicional de estas palabras, que según el caso y la lengua debe quedar incluida, en alguna forma, en el artículo correspondiente al vocablo en cuestión. En el caso de los hongos de los tzotziles, el objeto de un diccionario de lengua sería la exposición de la lógica interna de la clasificación tzotzil —lo que algunos antropólogos segregacionistas llaman etnosemántica.

Qué decir de los vocablos referidos a instituciones sociales, a conceptos de la cultura y de la civilización. *Libertad, pueblo, amparo, trabajo, dios*, no son equivalencias de *freedom o liberty, people, work, god, liberté, peuple*, etc., sino que requieren su exploración semántica en el interior del sentido de la comunidad. El *tequio*, la *mayordomía*, el *pozol* ¿no merecen un estudio cuidadoso de sus relaciones con la vida comunal, la repartición del trabajo, el sistema de poder y la función alimentaria de los pueblos que los designan?

Con esto paso a la parte propositiva que pide la convocatoria a esta reunión:

Hasta la fecha, la mayor parte de los diccionarios de lenguas amerindias mexicanas es de tipo lingüístico o filológico, que no reporta ninguna ventaja para las comunidades aborígenes existentes. Los diccionarios bilingües que se preparan como parte de programas de castellanización adolecen de defectos similares a los de los diccionarios lingüísticos; defectos, por supuesto, vistos con la lente de los valores semánticos que he expuesto al hablar del diccionario monolingüe de lengua. Salvo la opinión más autorizada de los propios miembros de las comunidades indígenas, supongo que

estas obras son de poca utilidad tanto para la comunidad en sí misma como para su castellanización.

Una política de la lengua para una comunidad particular implica una importante y grave toma de decisiones técnicas a partir de lo que se considere útil y necesario para la comunidad. Como Leopoldo Valiñas (1983) sostiene en su artículo “Alfabetización y la experiencia Mixe” tal política debiera proceder de la comunidad misma, que fijara el papel de la lingüística en relación con el sentido de la comunicación verbal en su interior y con sus expectativas de mejor vida.

Un diccionario de lengua puede ser parte importante de tal política, en la medida en que difunde y fija un sistema ortográfico, una gramática sistematizada y una visión de la cultura y la lengua. El diccionario puede llegar a ser el lugar de encuentro, por excelencia, de la imagen y los valores étnicos de una comunidad lingüística.

Si una política de la lengua llega a determinar la necesidad de elaborar un diccionario ¿Qué clase de diccionario conviene poner en obra?

He aquí algunas respuestas, a las que por supuesto no deseo darles el carácter de dogmas o de condiciones *sine quae non*:

1. Interesa más un diccionario de lengua que uno lingüístico.

2. Sea que el diccionario sea monolingüe o bilingüe (pues para muchas comunidades puede tener más sentido práctico el contraste y la operación orientada hacia la lengua española), interesa que la definición profundice en el análisis del significado cultural de los vocablos y que la equivalencia —cuando se trata de

diccionarios bilingües— implique un contraste previo de los elementos semánticos formativos (lo que se puede alcanzar organizando el vocabulario, para el trabajo exclusivamente, en microcosmos analizables con técnicas de campo semántico), con objeto de que sea verdaderamente útil, tanto para la cabal comprensión de un discurso, como para la producción de discursos.

3. La técnica de recolección de datos léxicos no debiera ser la de interrogatorio a informantes confrontados con combinatorias morfológicas, o a partir de listas procedentes de los estudios de genealogías lingüísticas, sino que podría basarse, sea en un levantamiento estadístico a partir de grabaciones libres entre distintos componentes de la comunidad, sea en un acopio de textos de toda clase —cuando los hay—, manejados también con métodos estadísticos, o también mediante un sistema de módulos de conversación dirigida, a la manera de William Labov, aplicados a una muestra de hablantes.

Naturalmente un diccionario como el que les propongo no es trabajo sencillo ni rápido; implica, además de técnicas lexicográficas cuyo rigor nunca dejaremos de subrayar suficientemente, un conocimiento profundo de la lengua en cuestión por parte del lingüista. Cuando se trabaja con significados no es posible manejar la lengua objeto con la misma exterioridad y el mismo desconocimiento con que solemos manejar sus aspectos formales, como la fonología, la morfología y la sintaxis. Trabajar la semántica de una lengua implica una sumersión profunda en su sentido y esto solamente se logra con años de dedicación exclusiva a ella y con una competencia más que suficiente en su

práctica. Por eso no es exagerado sostener que el destino de un lingüista queda determinado por la lengua a la que decide dedicarse y que tal elección no se hace sin una profunda crisis existencial, que altera por completo la vida de una persona. Un diccionario es un trabajo de muchos años y de ello debieran darse cuenta, tanto los que en un momento determinado ven la lexicografía como posible especialización, como las instituciones que sostienen el trabajo. En otros lugares del mundo es ya una verdad por todos sabida que la lexicografía requiere muchos años; en México, sería ya tiempo de que los funcionarios responsables de tantas instituciones y agencias gubernamentales entiendan lo que ello implica y dejen de proponerse alegres planes de docenas de diccionarios por año o por sexenio.

Un diccionario de lengua por comunidad lingüística mexicana sería el coronamiento de una comprensión total de cada uno de nuestros pueblos amerindios y el elemento indicador de que una política de la lengua está en marcha. Después de terminado el diccionario, los programas de periodismo, de radio, de enseñanza escolar, de creación cultural y de defensa propia de las comunidades pueden convertirse en sólidas realidades. A Levy-Strauss le interesa el documento; quizá también le interesaría, como a nosotros, la realidad vital de nuestros pueblos indígenas.

## LA CUANTIFICACIÓN EN EL DICCIONARIO DEL ESPAÑOL DE MÉXICO

0. Poco a poco la estadística lingüística se ha venido imponiendo como un instrumento útil para el trabajo lexicográfico; en ese proceso el papel que ha jugado la obra difusora de Bernard Quemada<sup>1</sup> ha sido determinante para muchos de los que nos dedicamos a esta área de la lingüística. Este artículo está dedicado a exponer las ventajas y las limitaciones que ha tenido la estadística lingüística y su utilización con la ayuda de una computadora electrónica en la elaboración del *Diccionario del español de México*.<sup>2</sup>

1. El objetivo que persigue nuestro trabajo, iniciado en 1973, es redactar una obra lexicográfica en la que se vea registrado el léxico del español de México, por primera vez, como una variante nacional de la lengua española internacional, y con finalidades tanto sociales y educativas prácticas, como lingüísticas descripti-

<sup>1</sup>En particular desde la publicación de las *Actes du colloque international sur la mécanisation des recherches lexicologiques*, que él organizó en Besançon en 1961 y publicó en *Cahiers de lexicologie*, 3 (1962).

<sup>2</sup>Una exposición de conjunto sobre los planteamientos estadísticos y el *Analizador automático* del español, que elaboramos para realizar la investigación previa a la redacción del diccionario —que no hemos terminado todavía— se encuentra en Lara/Ham, 1974, Lara 1976a y Lara/García Hidalgo/Ham, 1981.

vas. Como se trata, desde 1973 hasta ahora, del primer diccionario de la lengua española que se documenta y se elabora en una variedad lingüística diferente de la de Castilla y la Península Ibérica, la cuestión más importante entre todas las que se nos planteaban al inicio de nuestro trabajo era precisamente la de cómo lograr deslindar y reconocer las características del léxico del español de México, de manera que pudiéramos elaborar un diccionario en que efectivamente quedaran contenidos, tanto el núcleo del vocabulario usado en México, como una gran cantidad de vocablos pertenecientes a nuestra variedad lingüística, independientemente de su mayor o menor difusión en otras áreas hispanohablantes. Sólo una descripción objetiva y lo más amplia posible, basada en criterios y métodos lingüísticos, nos podría ofrecer los datos necesarios para llevar a cabo tal tarea y, sobre todo, para librar-nos de caer en el “intuicionismo” característico de muchos de los que se han ocupado de variantes regionales — como la mexicana— de una lengua, que históricamente han preferido la minucia pintoresca de una palabra sobre el uso real de otros vocablos, menos característicos, ciertamente, pero más importantes para poder comprender lo que es una lengua como la española.

Esta dificultad inicial debe destacarse, pues lo que en realidad tiene que confrontar muchas veces la lexicografía no es un conjunto de intereses y de argumentos teóricos y metódicos que orienten y sometan a crítica su trabajo, sino la imposición ideológica y política de un conjunto de creencias generalizadas acerca de lo que “debe ser” el vocabulario de una lengua. La lingüística sabe que la idea del léxico que tiene un

hablante común no se forma sobre bases objetivas, sino que depende fundamentalmente de la educación que ha recibido, que lo provee de una versión, social y tradicionalmente elaborada, de su propia lengua, a partir de la cual juzga lo que escucha y define las características de lo que puede reconocer en ella; sin embargo, lo cierto es que ese conocimiento suele soslayarse cuando los resultados de una investigación objetiva difieren en alguna medida de la idea que tienen los miembros de la sociedad a propósito de su lengua. Así por ejemplo, para la mayor parte de los mexicanos, el español de México se caracteriza sobre todo por los vocablos de origen amerindio (náhuatl, en especial) que se han introducido al idioma desde el siglo XVI y por ciertos usos populares de larga tradición. Son muy pocos, por lo contrario, los que están dispuestos a reconocer, a partir de ciertos resultados de investigación, que los vocablos de origen amerindio en el español mexicano contemporáneo —en su nivel estándar— no lo pueden caracterizar, así como tampoco son ciertos giros populares tradicionales los que dan el perfil de un idioma que sirve de vehículo de comunicación, de conocimientos y de expresión a una cultura que ha evolucionado notablemente más allá de sus raíces y se esfuerza por manifestar el mundo de hoy en día.<sup>3</sup> En esa situación, es imprescindible para la lexicografía en general, y para un trabajo como el

<sup>3</sup> Acerca de la cantidad de palabras de origen amerindio que se encuentran en el español mexicano se ha venido dando una polémica que todavía no termina, pues es un hecho que, en el nivel estándar de la lengua, como lo demuestra Lope Blanch 1969, son poquísimos (0.477% de su *corpus*, pero descontando los topónimos, que en verdad abundan por toda la República, se reducen a 0.073%. Cf.

nuestro en particular, encontrar un método de registro y acopio de datos lo más adecuado y lo menos controvertible posible, que nos permita descubrir los contornos y los elementos principales de un léxico como el del español de México, sobre los cuales pueda construirse el diccionario y en los que no intervengan los prejuicios que suele tener la sociedad —y nosotros mismos, al fin y al cabo hablantes también y miembros de ella— acerca de nuestra propia lengua.

Precisamente la posibilidad abierta por la utilización de la computadora electrónica, de manejar grandes cantidades de datos con criterios bien delimitados e independientes de juicios personales que pueden variar a lo largo de una investigación tan compleja como la nuestra, y la experiencia que nos ofrecen importantes estudios previos realizados con métodos cuantitativos (como los de Pierre Guiraud, Gustav Herdan y Charles Muller, y los de Alphonse Juilland, Henry Kucera y Nelson Francis),<sup>4</sup> nos permitían considerar, en 1973, la idea de construir un gran *corpus* de textos,

---

también Lara, 1983:594-5). Diversas variedades regionales mexicanas del español, en particular las que se dan en territorios que todavía tienen población india, tienen cantidades mayores de esas palabras. Pero para la capa intelectual nacionalista de México, este hecho es inaceptable, por cuanto contradice las creencias ligadas al esfuerzo por definir una identidad nacional en términos exclusivamente indios. Sin duda, tomando como objeto de estudio las variedades regionales del español mexicano y buscando y rebuscando en textos antiguos y modernos, dedicados a cuestiones prehispánicas, se puede aumentar notablemente el número de vocablos de procedencia amerindia.

<sup>4</sup>Cf. *Problèmes et méthodes de la statistique linguistique*, París, 1960; *The advanced theory of language as choice and chance*, Berlín, 1966; *Initiation à la statistique linguistique*, París, 1968; la serie de diccionarios estadísticos de las lenguas romance, dirigida por Jui-

hablados y escritos, recolectados por toda la República Mexicana, procedentes de todas las capas sociales y de todos los tipos de discurso, que nos sirviera como registro amplio y representativo de la realidad del español mexicano, y al que pudiéramos aplicar un análisis estadístico. Fue así como elaboramos el *Corpus del español mexicano contemporáneo (1921-1974)*, al que en adelante me referiré con la abreviatura *CEMC*.

1.1. Este *corpus* está constituido por 996 textos,<sup>5</sup> de ca. 2 000 palabras gráficas cada uno, seleccionados de obras escritas y de grabaciones magnetofónicas procedentes de toda la República Mexicana, y agrupados en catorce “géneros” diferentes (*poblaciones*, en términos estadísticos), que a su vez corresponden a la lengua culta, la lengua estándar coloquial —llamada “subcultura” en Lara/Ham, 1974— y la lengua “no-estándar”, bajo la cual se engloban las variedades regionales del español de México.

La estructura que dimos a nuestro *corpus* ya constituye un tema de discusión, que resulta particularmente interesante ahora, después de que hemos obtenido resultados y podemos juzgarlo a partir de ellos; para llevar a cabo tal discusión, hay que partir de las siguientes consideraciones:

---

lland, de la cual Juilland/Chang Rodríguez 1964 es el primero; *A Computational Analysis of Present-day American English*, Providence, 1961, respectivamente.

<sup>5</sup>El plan original era reunir mil textos, pero dificultades de último momento en la selección de textos y la presión del tiempo, nos obligaron a cerrar el *corpus* cuando teníamos 996 textos. Este hecho tiene consecuencias insignificantes.

La lengua española ha sido objeto de varios estudios estadísticos basados en *corpora* elaborados con diferentes criterios; así por ejemplo, mientras que M.A. Buchanan<sup>6</sup> y L. Rodríguez Bou<sup>7</sup> estructuraron sus *corpora* en diversas poblaciones de tamaño heterogéneo, V. García Hoz<sup>8</sup> y A. Juilland y E. Chang Rodríguez<sup>9</sup> compusieron sus poblaciones en tamaños homogéneos. Y si los dos primeros mezclaron entre sus poblaciones textos hispanoamericanos, los dos siguientes no tomaron en cuenta el español de esta región, sino que se centraron exclusivamente en el español de la Península Ibérica (Juilland y Chang Rodríguez en el español estándar y culto de España). Por último, si Buchanan y Rodríguez Bou incorporaron textos de épocas muy distintas entre sí, García Hoz y Juilland/Chang Rodríguez prefirieron concentrarse en lapsos históricos más pequeños, con el objeto de asegurar la sincronía de los datos obtenidos.

**1.1.1.** La homogeneidad de las poblaciones estadísticas tiene conocidas ventajas para el cálculo de las diversas medidas que se propone un estudio estadístico (Cf. Juilland/Chang Rodríguez 1964 y Bortolini/Tagliavini/Zampolli 1971), en cuanto que las comparaciones entre diversas poblaciones se facilitan. En cambio, siempre que se construya un *corpus* heterogéneo —como en nuestro caso—, los diversos tamaños de las poblaciones que lo componen obligan a elaborar me-

<sup>6</sup>*A Graded Spanish Word Book*, Toronto, 1927.

<sup>7</sup>*Recuento de vocabulario español*, Puerto Rico, 1952.

<sup>8</sup>*Vocabulario usual, vocabulario común y vocabulario fundamental*, Madrid, 1953.

<sup>9</sup>*Frequency Dictionary of Spanish Words*, La Haya, 1964.

didadas estadísticas que permitan establecer las comparaciones necesarias entre ellas a pesar de su heterogeneidad. En este artículo no habré de referirme a las medidas adoptadas para analizar nuestro *corpus*, para no hacerlo demasiado largo y porque ya fueron expuestas en trabajos anteriores (cf. Lara/Ham 1974 y Lara/García Hidalgo/Ham 1981); pero cabe señalar que tales medidas nos permiten manejar un *corpus* heterogéneo sin distorsionar los resultados cuantitativos que se obtengan.

Sin embargo, la cuestión de la heterogeneidad del *corpus* tiene un aspecto que va más allá del de las medidas que lo manejen, y es el que se refiere a los motivos para establecer poblaciones de diferentes tamaños y al juicio implícito en ellos acerca de su “valor” dentro del total del *corpus*. Pues, si bien es verdad que, como señalan Juilland/Chang Rodríguez, “unequal worlds are not necessarily arbitrary or accidental [since] the varying size of the subsamples can be assumed to reflect the varying importance assigned to the socio-cultural area they manifest” (1964:XXV), a la vez, tal desigualdad de tamaños puede ser efecto de una ponderación “largely subjective (...); errors of educated guessing could introduce serious distortions and gravely endanger the representative character of the universe as a whole” (1964:XXVI). En nuestro caso ese riesgo era manifiesto desde un principio y, si bien intentamos fundamentar adecuadamente nuestra ponderación del tamaño de los géneros de nuestro *corpus* acudiendo a una “hipótesis sociolingüística” (Lara/Ham, 1974: 3.2) de carácter intersubjetivo, no ha dejado de constituir una de las principales interrogantes en el estudio de nuestros resultados. En este

artículo no podré discutir con la prolijidad necesaria este aspecto de nuestro análisis,<sup>10</sup> por más que, considerando sus resultados, tengamos suficientes razones para creer que la heterogeneidad así introducida no ha sesgado de manera importante los datos cuantitativos y cualitativos recolectados.

**1.1.2.** En cuanto a la cuestión de la homogeneidad del “estado de lengua” que debe definirse previamente a la suposición de la existencia de una homogeneidad sincrónica de los datos, el hecho de que nuestro *corpus* cubra 53 años parece contradecir ese requerimiento. La naturaleza de nuestros resultados, al contrario del caso anterior, no nos permite juzgar si los datos recolectados de veras pertenecen a un “estado de lengua” homogéneo, por más que nos hayamos guiado por el concepto de “sincronía práctica” que defiende J. Rey-Debove (1971). Las razones que nos llevaron a definir nuestro “estado de lengua” en esa forma parecen estar justificadas por la necesidad de delimitar el español mexicano contemporáneo de acuerdo con un conjunto de criterios válidos para los mexicanos.<sup>11</sup> Solamente

<sup>10</sup>Al estructurar el *corpus* dejamos abierta la posibilidad de modificar la composición de los géneros en el momento en que los datos obtenidos nos señalaran errores de conceptualización en la correspondencia entre la hipótesis sociolingüística y el arreglo de los textos en géneros, algo que habremos de hacer en el futuro, pues en especial varios elementos del género coloquial deben agruparse en la lengua culta (sobre todo la “novela rosa”, que es un género romántico-trivial, muy dependiente del léxico literario más formal y solemne). Cf. aquí mismo “La sociolingüística del *DEM*”.

<sup>11</sup>Tales criterios son de los órdenes histórico y cultural: el México contemporáneo es resultado del movimiento revolucionario que se inició en 1910 y terminó en 1921, cuando la creación de nuestras

cuando contemos con nuevos *corpora* que, por un lado, permitan observar posibles diferencias entre el español mexicano prerrevolucionario (por ejemplo, a partir de 1876, un año importante en la historia política, social y económica de México) y el que hemos determinado a partir de 1921 como “mexicano contemporáneo”; y por el otro, nos permitan encontrar diferencias importantes al interior del lapso histórico definido por nosotros (incluyendo en el *corpus* textos periodísticos y científicos de esos años, lo que no hicimos ahora, por el interés de considerar solamente los vocabularios modernos de las ciencias y por aligerarnos el trabajo de recopilación hemerográfica), estaremos en la posibilidad de concluir en la adecuación o la inadecuación de nuestro “estado de lengua” para basar en él un verdadero estudio sincrónico.

**1.1.3.** La cuestión de la homogeneidad dialectal de nuestro *CEMC* puede verse de dos maneras: por un lado, no plantea problemas desde el punto de vista de que el español de México es la lengua de una macro-región lingüística, dentro de la cual se han desarrollado tanto una variedad histórica de la lengua española, que la caracteriza unitariamente en relación con el español

---

actuales instituciones políticas devolvió la paz a la población y detuvo su movilidad, efecto de la huída permanente de la guerra y de las caprichosas condiciones que imponían los distintos caudillos revolucionarios sobre su vida cotidiana. Desde el punto de vista cultural, la Revolución originó profundas renovaciones, tanto en la educación como en las artes y las ciencias. Simbólicamente, en 1921 se publicó la novela que se toma como estandarte de esa renovación cultural: *Los de abajo*, de Mariano Azuela. Por último, de entonces data la creación del importante medio de difusión nacional que es el radio.

de América en su totalidad (y, en particular, en relación con las unidades dialectales vecinas: el español centroamericano, y los dialectos originariamente mexicanos de California, Arizona, Colorado, Nuevo México y Texas, que quedaron relativamente aislados de México después de su anexión a Estados Unidos de América, pero que, sin embargo, forman parte de la unidad dialectal histórica de México), como una “lengua culta” válida para todo el país, difundida por sus medios de comunicación y de valor normativo, que refuerza la existencia de aquella unidad macro-regional. Por el otro, sí los plantea, si se considera que, por ejemplo, el español contemporáneo de la Alta California o de Texas han restablecido una continuidad con el español de México, y no por cierto solamente con los dialectos de los estados mexicanos limítrofes de Baja California, Sonora, Coahuila, Chihuahua, Nuevo León y Tamaulipas, sino con dialectos del centro del país (de Jalisco, Durango, Zacatecas, Michoacán, etc.); zonas todas ellas de donde han partido las corrientes migratorias de trabajadores que hoy caracterizan a la mayoría de la población llamada “hispana” en esas regiones de los EE.UU. Es decir, el “español mexicano” puede considerarse, social y geográficamente hablando, más amplio que el que delimita la frontera nacional en el norte.

Igualmente, el dialecto del estado mexicano sureño de Chiapas, en la frontera con Guatemala, podría considerarse, o bien una continuación del español centroamericano, o bien precisamente el dialecto que establece la transición entre el español de México y el de Centroamérica. Solamente la hipótesis, sostenida con criterios sociolingüísticos basados en la teoría de la

lengua estándar (cf. *supra* 1.1.2), de que sobre su constitución geográfica dialectal se ha venido imponiendo en México —precisamente a partir del final de la Revolución mexicana— una unificación lingüística, es la que nos permite suponer que el español mexicano es una verdadera unidad, de carácter cultural, social y político. *Corpora* futuros, dedicados a Centroamérica y al suroeste de los Estados Unidos de América, habrán de confirmar o invalidar nuestra hipótesis.

La heterogeneidad dialectal al interior de la República Mexicana queda incluida, por lo demás, dentro de nuestro *CEMC*. Igualmente, la heterogeneidad de registros y sociolectos del español de México está considerada en nuestro *corpus*, pues forma parte de los objetivos centrales de nuestro trabajo.

1.2. La estructura formal del *corpus* constituye un aspecto igualmente importante para poder evaluar las ventajas de la estadística lingüística y su manejo con la computadora electrónica. En nuestro caso, decidimos que el *CEMC* constara de aproximadamente dos millones de palabras gráficas, reunidas en mil textos formados por tantos párrafos como fueran necesarios para alcanzar esa suma de ocurrencias léxicas.<sup>12</sup> Seleccionamos el párrafo como unidad de muestreo de nuestras fuentes porque, dada la necesidad lexicográfica de elaborar un análisis semántico de los vocablos del *corpus*, nos era necesario partir de unidades lingüísticas con un sentido completo, que facilitarían el análisis. Seguramente un

<sup>12</sup>Como explicamos en Lara/Ham, 1974. En ese trabajo se dan los argumentos que utilizamos para optar por una recolección de dos millones de palabras gráficas.

muestreo por palabras aisladas habría tenido por resultado una mayor riqueza de unidades léxicas —entendida como cantidad de vocablos diferentes— que la que nosotros obtuvimos, pero nos habría impedido, o al menos dificultado, el estudio semántico (cf. Lara/Ham, 1974: 3); a la vez, la selección por párrafos hace que el muestreo sea más rico que cuando, como en el caso de Juilland/Chang Rodríguez, se selecciona un texto seguido e ininterrumpido hasta completar el número de palabras gráficas deseado (sin soslayar tampoco el hecho de que un texto seguido se presta más a la distorsión que causa el estilo del autor del texto y el tema que se está tratando, como lo han señalado los especialistas en estadística lingüística).

Los textos de nuestro *CEMC* conservaron todos los signos de puntuación del original, fechas, cifras y nombres propios, también con el objeto de que el sentido del párrafo no se dañara. En el momento de los recuentos estadísticos, sin embargo, se eliminaron todos ellos, para que los resultados cuantitativos sólo se refirieran a palabras comunes (cf. *infra* 2.2.2).

**1.2.1.** Los textos del *CEMC* se registraron en la computadora electrónica con el menor esfuerzo posible de pre-edición: solamente se incluyeron diacríticos para señalar fechas, nombres propios y topónimos, mientras que no se pre-codificó dato alguno del orden gramatical. En Lara, 1976a, se explican las razones para actuar así y se dan las características generales de nuestro trabajo en gramática computacional; Isabel García Hidalgo ha hecho un relato pormenorizado de este trabajo en 1981 y en 1986. El trabajo central de análisis gramatical, de reunión de las ocurrencias de palabras en *tipos estadís-*

*ticos* (en la terminología de Charles Muller, que adoptamos desde un principio), de cálculo de las frecuencias de palabras, de dispersión y de uso (específicamente, de “frecuencia corregida” y del “índice normalizado de dispersión”, cf. Lara/Ham, 1974: 4.4.) correspondió a nuestro *Analizador gramatical automático*, que es un *parser* morfosintáctico, dotado de un componente estadístico y de un productor de concordancias (cf. García Hidalgo, *loc. cit.*).

**1.2.2.** Los resultados obtenidos se han arreglado de varias maneras: listas de tipos léxicos, en orden alfabético y por orden decreciente de frecuencias, y concordancias de tipos léxicos, en las que el tipo en cuestión aparece al centro de un contexto de tres líneas. Las listas se ordenan, primero, por resultados obtenidos para los tres grandes “géneros”: lengua culta, lengua estándar coloquial y lengua no-estándar; después, por los resultados obtenidos en catorce géneros, que es la subdivisión que dimos a los tres grandes géneros anteriores.<sup>13</sup> Hemos sacado resultados también para las subdivisiones de los géneros de ciencia (físico-matemáticas, químico-biológicas, etc.) y de técnica (por ejemplo, las diferentes ingenierías). De hecho, es posible también subdividir

<sup>13</sup>He modificado algunas etiquetas de los nombres de los géneros, como aclararé a continuación. Los sub-géneros son: literatura, periodismo, ciencia, técnica, discurso político, discurso religioso (o religión), “habla culta” de la ciudad de México (tal como la define Lope Blanch, 1968), literatura popular, “habla media” de la ciudad de México (cf. Lope Blanch, *ibidem*), lírica popular mexicana, textos regionales (provenientes de todas las zonas estudiadas en el *Atlas lingüístico de México*, dirigido también por Lope Blanch y de próxima publicación), documentos antropológicos, jergas (o argot) y “habla popular” de la ciudad de México (cf. Lope Blanch *ibidem*.)

otros géneros, cuando hay interés por establecer comparaciones en su interior; por ejemplo, se puede subdividir el género del discurso político en los de cada presidente de la República y los de algunos otros políticos. Los textos regionales se pueden subdividir de acuerdo con las zonas dialectales establecidas por el *Atlas* mencionado en la nota 13, etcétera.<sup>14</sup>

En todo lo anterior se ha venido señalando que nuestros resultados se dan para tipos léxicos y no para vocablos. Esto es así porque el *Analizador automático* no contiene algoritmos suficientes para lematizar los tipos en todos los casos (por ejemplo, para llevar los tipos de sustantivos con diferentes flexiones a una forma canónica única), debido a un compromiso que el tiempo nos obligó a establecer entre rapidez del análisis y capacidad del programa para realizarlo. Esta limitación se puede superar en el futuro, naturalmente. En este artículo, sin embargo, los resultados que se muestran corresponden a vocablos y no a tipos.

**2. Una vez dadas todas estas explicaciones, he aquí un recuento de varios resultados cuantitativos obtenidos en nuestro estudio, mediante los cuales trataré de delimitar las principales ventajas y algunas limitaciones de la cuantificación en lexicografía.**

**2.1. Descontadas las ocurrencias de fechas, cifras, nombres propios y topónimos de nuestro CEMC, el número**

<sup>14</sup>Mi colega Raúl Ávila acaba de realizar un estudio estadístico en que combina textos de los géneros hablados del CEMC (habla culta, media, popular y regionales), cuyos resultados cuantitativos demuestran también la representatividad de nuestros materiales. Cf. Ávila 1987.

total de ocurrencias que lo componen es de 1 891 045. Estas ocurrencias se reagrupan en 64 183 tipos.<sup>15</sup> Por géneros se reparten como se muestra en la tabla 1:

**Tabla 1**

<i>Género</i>	<i>Número ocurrencias</i>	<i>Porcentaje ocurrencias</i>	<i>Número tipos</i>	<i>Porcentaje de tipos</i>
L. culta	1 241 313	65.6416	53 714	83.6888
L.est.col.	232 175	12.2776	16 982	26.4587
L.no-est.	417 557	22.0807	18 250	28.4334
Literatura	269 788	14.2666	24 483	38.1456
Periodismo	299 775	15.8523	21 072	32.8311
Ciencias	346 313	18.3133	26 487	41.2679
Técnicas	202 716	10.7198	17 836	27.7893
Disc. pol.	31 971	1.6907	5 516	8.5942
Disc. rel.	21 277	1.1251	3 979	6.1995
Habla culta	69 473	3.6758	6 291	9.8017
Lit. popular	127 459	6.7401	12 876	20.0614
Habla media	59 567	3.1500	5 870	9.1457
Lírica pop.	45 149	2.3875	4 172	6.5002
Txt. regs.	259 881	13.7427	13 095	20.4026
Docs. antrp.	68 376	3.6158	6 082	9.4703
Argot	34 839	1.8423	4 710	7.3384
Habla pop.	54 461	2.8799	4 548	7.0860

**2.1.1.** Roberto Ham (1981:76) mostró ya cómo la relación entre el tamaño de los géneros y el número de tipos léxicos obtenidos se mantuvo directamente proporcional (dentro de la tendencia asintótica que carac-

<sup>15</sup>Según se dijo antes, nuestros resultados no están lematizados, aunque, tratándose de las formas verbales, sí nos fue posible crear algoritmos que las redujeran a la forma canónica del infinitivo, por lo que estos resultados son híbridos.

teriza la curva de crecimiento del vocabulario en relación con el número de sus ocurrencias, naturalmente) a excepción de los casos del género de periodismo (segundo en el *CEMC*, pero tercero en cuanto al número de tipos producidos), del de literatura (tercero en el *CEMC*, pero segundo en tipos producidos), y de los pequeños géneros del habla popular, lírica popular, argot y discursos políticos, que pasaron de los lugares 10 al 13 del *CEMC* a los lugares 12, 13, 11 y 10 respectivamente en los resultados obtenidos.

Se puede entonces formular una hipótesis en el sentido de que el tamaño previamente definido de cada género determina la cantidad de tipos léxicos encontrados en ellos; y a la vez, que las variaciones en esta relación entre tamaño del género y total de tipos obtenidos por género, deben ser señales de alguna característica cualitativa de los textos incluidos en ellos. En Ham 1981:69-72 se ofrecen algunos ejemplos de ello; es el caso del pronombre átono de primera persona en complemento directo o indirecto, que varía notablemente entre algunos géneros de textos escritos y otros de textos hablados:

**Tabla 2**

<i>Género</i>	<i>Pronombre me Tamaño relativo en el corpus (%)</i>	<i>Porcentaje entre géneros</i>
Literatura	14.27	14.64
Periodismo	15.85	1.40
Ciencias	18.31	1.03
Lit. popular	6.74	13.50
Textos regs.	13.74	18.31
Docs. antropos.	3.62	15.36

Si se toma el género de literatura como representante del comportamiento normal de la relación tamaño del género/número de tipos léxicos encontrados, las diferencias existentes en los otros géneros considerados muestran claramente la variación cualitativa de éstos. El caso particular del pronombre *me* parece deberse a un fenómeno característico de los textos hablados de la literatura popular —fotonovelas, telenovelas, “novelas rosa”— y de las grabaciones realizadas por lingüistas y antropólogos en diversas regiones del país, en los que predominan el diálogo orientado al hablante y la localización precisa de los participantes en el acto verbal, situaciones que no ocurren en textos científicos y que pueden servir para caracterizar, en negativo, en este caso, los textos periodísticos. En otras palabras, aun cuando el método de selección de párrafos en cada texto aspira a obstaculizar la influencia de temas y de estilos, vocablos como los pronombres (y la mayor parte de los déicticos), que se ligan a actos verbales localizados en la situación de habla, se ven influidos en su repartición o dispersión entre todos los géneros considerados. Tendríamos que estudiar detalladamente los textos que se incluyeron en estos géneros y dieron resultados como los anteriores, para poder afirmar definitivamente que la variación cualitativa es la causa de las diferencias cuantitativas.<sup>16</sup>

<sup>16</sup>Hay otros elementos que se deben considerar en este aspecto: desde que transcribimos las grabaciones que tomamos en cuenta en estos géneros notamos que, en la mayor parte de ellas, las circunstancias de la encuesta (dialectológica) se imponían claramente sobre el tipo de discurso que podía desarrollar el informante. La presencia de un encuestador llegado de la Universidad de México (la más

Lo anterior podría considerarse una prueba de las desventajas que tiene el uso de poblaciones heterogéneas en la creación de un *corpus*, pues tanto el hecho de que haya una relación directamente proporcional entre el tamaño de la población y los resultados obtenidos (en la medida en que la estadística hace comparaciones entre poblaciones de esa clase), como el que cada una de las poblaciones refleje muy diversas prácticas verbales (la literaria, la científica, la periodística, y la coloquial, la argótica o la rural), se convierten en factores de importancia para la interpretación comparativa de los resultados estadísticos y pueden conducir a versiones distorsionadas del carácter cuantitativo de la lengua.

**2.1.2.** Para intentar contradecir esta última proposición, consideremos ahora algunos de los resultados obtenidos en el análisis del CEMC en comparación con los que obtuvieron A. Juilland y E. Chang Rodríguez.

Estos autores, interesados en obtener datos estadísticos sobre el léxico de la lengua española que pudieran tener validez estructural,<sup>17</sup> optaron por tomar en con-

---

importante del país y de la capital de la República) obligaba al informante a optar por su registro de habla más elevado y a tratar temas de un orden diferente del que supondría su localización etnológica. Tales hechos seguramente tuvieron consecuencias en el tipo de vocabulario encontrado y en el tipo de las estructuras gramaticales que aparecieron en esos textos.

<sup>17</sup>Los criterios de Juilland sobre las características estructurales de una lengua son más claros en su interesante —y extrañamente poco citado— *Outline of a General Theory of Structural Relations*, La Haya, 1961, en donde defiende la idea de que las características estadísticas de los signos lingüísticos deben formar parte del conjunto de esas características estructurales.

sideración para su diccionario estadístico vocablos cuya frecuencia total mínima fuera  $F \geq 5$ .

De esa manera, seleccionaron como buenos representantes de la estructura cuantitativa de la lengua española 5 024 vocablos, que son los publicados en su *Frequency Dictionary of Spanish Words*. Con una frecuencia igual, en nuestro CEMC aparecieron 17,903 tipos léxicos.<sup>18</sup> Eso significa que, a mayor tamaño del *corpus*, mayor cantidad de vocablos obtenidos, si bien esa relación no llega a ser directamente proporcional con la diferencia de tamaños de ambos *corpora*<sup>19</sup> y en particular, si se considera que, una vez lematizados esos tipos, se reducirá el número de vocablos obtenidos. Sea como sea, es claro que un *corpus* de mayor tamaño resulta más útil, al menos para una investigación lexicológica como la nuestra.

**2.2.2.** Pero es quizá más interesante comparar los resultados de Juilland/Chang Rodríguez con los nuestros desde el punto de vista de los elementos que constituyen el vocabulario encontrado; es decir, viendo si el vocabulario que se obtiene en ambos *corpora* es semejante (bajo la suposición de que entre el español

<sup>18</sup>Bortolini/Tagliavini/Zampolli, 1971 redujeron la frecuencia total admitida en su *Lessico* a  $F \geq 4$ ; en nuestro CEMC encontramos, entonces, 20 781 tipos.

<sup>19</sup>Queda abierta la pregunta acerca de cuál podrá ser el tamaño máximo conveniente de un *corpus* de esta clase, pues, como se anota en Lara/Ham 1974, el del *Trésor de la langue française* parece haber resultado excesivamente grande, en comparación con los resultados que obtuvo, en tanto que el de Juilland/Chang Rodríguez resulta quizá muy pequeño. A mayor tamaño del *corpus*, menor proporcionalidad de vocablos encontrados y, consecuentemente, menor rendimiento.

de México y el peninsular no hay grandes diferencias, como pensamos todos), pues este hecho sería un buen elemento de comprobación del efecto que puede tener una estructura heterogénea y muy diferente del *corpus* en cuestión. Para ello consideremos las acumulaciones por clases de frecuencia de 500 en 500 vocablos que establecieron esos autores, en comparación con las nuestras correspondientes:

**Tabla 3**

<i>Clase</i>	<i>Juillard/Chang</i>		<i>CEMC*</i>	
	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Primeras 500	364 518	79.10	1 270 602	67.19
Segundas 500	30 672	6.66	124 872	6.60
Terceras 500	18 528	4.02	73 291	3.94
Cuartas 500	12 860	2.79	54 064	2.78
Quintas 500	9 487	2.05	38 695	2.04
Sextas 500	7 369	1.60	33 654	1.78
Séptimas 500	5 873	1.28	24 519	1.30
Octavas 500	4 726	1.03	22 720	1.20
Novenas 500	3 790	0.82	19 153	1.01
Décimas 500	2 990	0.65	14 218	0.75

\*en tipos

Es posible que la diferencia de porcentajes entre las primeras quinientas de Juillard/Chang Rodríguez y las nuestras se reduzca si, en vez de tipos léxicos, lematizamos nuestros resultados para lograr una comparación exacta. Es precisamente lo que hicimos para determinar el “vocabulario fundamental del español de México”, constituido por todos los vocablos cuya frecuencia acumulada diera el 75% de las ocurrencias

del *corpus* (cf. Ham 1981:81); ese vocabulario, para todo el *CEMC*, queda formado por 1 131 vocablos.<sup>20</sup> En consecuencia, sigue existiendo una diferencia entre los resultados de Juilland/Chang Rodríguez y los nuestros, pues la acumulación correspondiente del *corpus* de aquéllos (79.10%) a la del nuestro (75%), les produce sólo 500 vocablos, en tanto que a nosotros, 1 131; es decir, nuestros resultados duplican los suyos.

Cualitativamente, si comparamos la lista de esas primeras quinientas palabras de Juilland/Chang Rodríguez con la nuestra de 1 131, encontraremos que nuestros resultados contienen todos los vocablos de esa clase de su diccionario, menos once (véase tabla 4) que, sin embargo, aparecen documentados en nuestro *corpus*, sólo que no entre los “fundamentales”; por lo contrario, nuestros datos cuantitativos los definen como verdaderas “palabras temáticas”. Considérese la tabla 4.

Nuestros resultados de frecuencia corregida ponderan la frecuencia, la dispersión y el tamaño heterogéneo de los géneros del *CEMC*; los del índice de dispersión los refieren a un valor entre cero y uno, según el cual mientras más cerca se encuentra un vocablo del uno, está mejor repartido entre todos los géneros (cf. Lara/Ham, 1974: 4.4). Como se ve en la tabla, los índices de dispersión de esos vocablos no son muy elevados, por lo que se trata de palabras que aparecen en el *CEMC* muy disparejamente repartidas. En efecto, *científico*, *descripción*, *geografía* e *histórico* predomi-

<sup>20</sup>Pudimos definir “vocabularios fundamentales” para cada uno de los géneros de nuestro *CEMC*; para el de “lengua culta”, su vocabulario fundamental resultó de 1 451 vocablos; para el de “lengua coloquial estándar”, de 560, y para el de “lengua no-estándar”, de 298. El de las ciencias es el más cercano al de la lengua culta: 1 382 vocablos.

Tabla 4

<i>Vocablo</i>	<i>Frecuencia corregida</i>	<i>Índice de dispersión</i>
científico	28.67	0.67
descripción	38.73	0.61
geografía	12.76	0.79
histórico	57.87	0.61
literario	14.65	0.49
marchar	45.09	0.70
ministro	27.06	0.54
península	9.23	0.53
provincia	41.22	0.66
república	12.19	0.71
valle	43.26	0.76

nan notablemente sólo en el género de ciencias; *ministro*, *península*, *provincia*, *república* y *valle* en periodismo; *literario* en literatura, y *marchar* en literatura popular. Cualitativamente, se pueden dar algunas explicaciones acerca de estos vocablos en México: *marchar* se usa generalmente con los significados de ‘andar con ritmo y paso continuo, como los militares’ y ‘funcionar una máquina’, por lo que no es tan usual como en España, en donde significa ‘ir(se)’ y ‘caminar’. Que aparezca mayoritariamente en el género de literatura popular, se debe al hecho de que en él incluimos “novelas rosa” leídas por una gran cantidad de personas en México, pero escritas por una escritora española: Corín Tellado. *Península* es evidente que tiene más uso precisamente en la Península Ibérica, mientras que en México es sobre todo un término geográfico. En cuanto a *ministro* y *provincia*, hay que señalar que en México no son vocablos de uso jurídico que, por esa razón, tengan mucho uso (como república federal,

México tiene *estados* y no provincias; como régimen presidencialista, tiene *secretarios* encargados del despacho de los asuntos del poder ejecutivo y no ministros); de ahí que sólo aparezcan en textos periodísticos, designando realidades extranjeras a México (si bien *provincia* es la designación popular para el interior de la república). Resulta extraño, en cambio, que siendo México precisamente una *república*, este vocablo no tenga el rango estadístico que intuitivamente le correspondería. Es posible que aquí haya habido un efecto inadvertido por nosotros en el momento oportuno: como todos los nombres propios en el *corpus* se marcaron para eliminarlos de la estadística, los capturistas de datos deben haber considerado que *República Mexicana* (y *República*, que en México se tiende a escribir con mayúscula) era uno de ellos, por lo que su frecuencia se vio artificialmente rebajada.

Lo importante es que tanto Juilland/Chang Rodríguez como nuestro *CEMC* coinciden en sus resultados en una gran medida, lo que no deja de constituir una agradable comprobación, pues da lugar a una conclusión contraria a lo que temíamos arriba: que la diferencia de tamaños de los *corpora* y la heterogeneidad de sus poblaciones no solamente no impide el reconocimiento del núcleo del español de México sobre bases objetivas y no sobre intuiciones de hablantes, sino que, además, nos ha permitido reconocer, mediante la comparación de nuestros resultados con los de Juilland/Chang Rodríguez, una parte central de la constitución cuantitativa de la lengua española, considerada como diasistema; la que podríamos llamar “núcleo cuantitativo del diasistema hispánico”.

Pero al mismo tiempo, el hecho de que nuestros resultados de la acumulación de 75% de las ocurrencias del *CEMC* sobrepasen por 631 vocablos los resultados estadísticamente correspondientes de Juilland/Chang Rodríguez, nos fuerza a preguntarnos a qué parte corresponden de la constitución cuantitativa del léxico, tanto del español mexicano como, ahora, en vista de nuestro inesperado descubrimiento de un “núcleo cuantitativo del diasistema hispánico”, del de ese diasistema.

**2.2.3.** Podemos entonces revisar qué vocablos son los que contiene nuestro “vocabulario fundamental” y que, o bien aparecen en el diccionario estadístico de Juilland/Chang Rodríguez, pero más allá de la frontera de la clase de las primeras 500 palabras (acumulación de ocurrencias correspondiente a 79.10% de su *corpus*), o bien no aparecen del todo. Así encontramos que 498 de aquellos vocablos sí están incluidos en la obra de Juilland/Chang Rodríguez y que faltan en ella por completo 132 vocablos. Es claramente más interesante estudiar estos últimos, en cuanto son ellos los que podrían dar algún síntoma de interés para las cuestiones que nos hemos planteado.

Habrá que asumir el riesgo de caracterizar equivocadamente las causas de la ausencia de estas palabras en el diccionario de Juilland/Chang Rodríguez; sin embargo, tal caracterización se hace necesaria en favor de la cuestión general que ha dado origen a este artículo. Los vocablos faltantes pertenecen a varios campos de la realidad. Así, encontramos entre ellos palabras referentes a alimentos, como *camarón*, *cebolla*, *harina*, *jugo*, *naranja*, *pimienta*, *queso*, y otros que pueden ser

propriadamente mexicanismos, como: *chile, frijol, manteca, tomate y tortilla*.<sup>21</sup> Hay una importante cantidad de vocablos del lenguaje intelectual y científico, entre los cuales se encuentran algunas medidas, como *centímetro, gramo, hectárea, kilo*, etc., y otros más generales pero igualmente cultos, como *circuito, comportamiento, dosis, estimular, incremento o química*. Del vocabulario común —al menos para México—, encontramos, por ejemplo: *calentar, campeonato, ceremonia, cocer, entusiasmo, equipo, fumar, futbol, gallo, hervir, moler, papá, rama, san, vaca y voltear*. Un posible mexicanismo de significado entre éstos es *falla*, pues en España prefieren, para el mismo significado, *fallo*.<sup>22</sup> Se puede discernir también un campo léxico de las finanzas, con vocablos como: *contabilidad, costo, eficiencia, financiero, gerente, inflación y tasa*. Quizá se podría también caracterizar como novedades que no podían aparecer en el *corpus* de Juilland/Chang Rodríguez, dado el lapso histórico que cubre, *película, televisión y reactor*. Posibles mexicanismos, que tampoco tendrían por qué aparecer en esa obra son: *compadre, control, controlar, llorona, nomás, platicar, preparatoria y rancho* (al menos por su uso, pues *platicar* es también andalucismo; *preparatoria* se distingue por designar el bachillerato en México; todos los demás deben tener uso en España). Por último, causa extrañe-

<sup>21</sup> *Chile, frijol, manteca y tortilla* no son palabras desconocidas en España y en otros territorios hispanohablantes, pero los dos últimos vocablos designan en México objetos diferentes: *manteca* es el unto del cerdo; y *tortilla* es una lámina redonda de harina de maíz cocida al fuego, que forma parte esencial de la dieta mexicana.

<sup>22</sup> *Falla* en España sólo tiene su significado geológico y también designa las conocidas fiestas típicas gallegas.

za no encontrar entre esos resultados una buena cantidad de pronombres, como *le, me, se, ti, nos, nosotros, conmigo* y *contigo*, así como los vocablos *muy, (no) obstante* y *gran*. Como en su mayor parte no parecen ser vocablos extraños para la Península Ibérica, no es conveniente tomarlos como verdaderos caracterizadores del español de México. Por lo cual las explicaciones probables para estas ausencias se deben buscar en las características de las poblaciones de textos que analizaron Juillard/Chang Rodríguez; quizá sus textos no recogieron vocabularios de los campos léxicos mencionados y quizá la mayor parte de ellos no contengan situaciones de acto verbal en las que ocurran los pronombres (cf. *supra* 2.1.1).

Independientemente de las causas que expliquen estas discrepancias, parece ser evidente que lo que encontramos en el “vocabulario fundamental” de nuestro *CEMC* es otra parte del núcleo del español de México, y en buena parte del diasistema hispánico (con las pequeñas dudas que nos dejan 132 vocablos entre 1 131). Es decir, nuestro *corpus* realmente delimita al español mexicano en relación con el español peninsular.

Con estos resultados, la finalidad principal del *CEMC*, que era definir objetivamente lo que es el léxico del español de México, se cumple con suficiente seguridad para el núcleo cuantitativo de la lengua (lo cual nos permitió redactar dos pequeños diccionarios: el *Diccionario fundamental del español de México* y el *Diccionario básico del español de México*,<sup>23</sup>

<sup>23</sup> Ambos diccionarios se basan, en realidad, en el vocabulario fundamental de la lengua culta (cf. *supra*, n. 20) y en el vocabulario disponible que contienen los libros de texto oficiales, obligatorios y gratuitos para la enseñanza elemental.

destinados a la educación elemental en nuestro país) que, hay que recalcarlo, corresponde a 75% de las ocurrencias de vocablos en el habla y la escritura mexicanas. Pero un diccionario destinado al público no puede quedarse en el nivel del núcleo del vocabulario de una lengua, en particular cuando se trata de una lengua como la española, que tiene una tradición milenaria de reflexión y de cultura y una lexicografía que, sin llegar a ser todo lo buena que desearíamos, nos ha deparado una importante cantidad de diccionarios. En particular, el público mexicano espera, como se había señalado antes, una obra en cuyas páginas encuentre palabras que no solamente conforman un vehículo de su cultura, sino que además caracterizan su idioma.

3. En nuestro diccionario estadístico quedan todavía 16 772 tipos con  $F \geq 5$ , 12 551 tipos con  $4 \geq F > 1$  y 29 729 tipos con  $F = 1$ . Es de suponerse que, si ya en el núcleo cuantitativo del español mexicano habíamos encontrado vocablos que parecen caracterizar al español mexicano en relación con el núcleo hispánico, conforme bajemos en la tabla de frecuencias y dispersión este hecho se vuelva más pronunciado y a la vez dependa cada vez más de las características específicas de los textos que se utilizaron como muestra. Para poder ilustrarlo, tomemos algunos ejemplos que sirvan, de paso, para mostrar la utilidad de las fórmulas estadísticas aplicadas.

3.1. El motivo principal para componer el *CEMC* a base de varias poblaciones heterogéneas era, por un lado, siguiendo las recomendaciones de Charles Mu-

ller y René Moreau,<sup>24</sup> diversificar las fuentes de manera que los cálculos de dispersión resultaran más informativos al tener más campo de comparación; por el otro, captar un mayor número de vocablos diferentes, procedentes de la mayor parte de las situaciones de habla en México, que nos diera la riqueza necesaria para elaborar un diccionario de tamaño medio (aproximadamente 30 mil artículos lexicográficos), a partir del cual nos sintiéramos relativamente seguros de haber tomado en cuenta lo más importante de nuestro léxico en las dimensiones social y geográfica de nuestro país. La estratificación del *corpus* en catorce géneros nos ha dado una cantidad importante de resultados en esa dirección. Si la lematización de los tipos con  $F \geq 2$  (29 323) no reduce demasiado el número de vocablos obtenidos en esos rangos de frecuencia, casi podríamos elaborar todo el diccionario a base de la documentación que nos ofrece el *CEMC*. Queda, sin embargo, pendiente la pregunta: ¿serán *representativos* esos vocablos del español de México?

**3.1.1.** Para tratar de responderla, aunque no con seguridad definitiva, consideremos algunos ejemplos de vocablos que tienen baja frecuencia en el *CEMC* (tabla 5).

En todos estos casos la frecuencia corregida y el índice de dispersión nos indican que se trata de vocablos poco repartidos en el *CEMC*. Ya el hecho de que

<sup>24</sup>Cf. Ch. Muller, "Un dictionnaire de fréquence de l'espagnol moderne", *ZRPh*, 81 (1965), pp. 476-483 y R. Moreau, "Au sujet de l'utilisation de la notion de fréquence en linguistique", *Cah. Lex.*, 3 (1962), pp. 140-159.

Tabla 5

<i>Vocablo</i>	<i>F. total</i>	<i>F. correg.</i>	<i>Índice de dispersión</i>	<i>Géneros predominantes</i>
taco	15	5.1435	0.3354	Regs.lit.pop.
teológico	15	7.1710	0.4721	Religión
tórax	15	6.6859	0.4394	Ciencias
agostadero	14	4.5885	0.3201	Técnicas
barbacoa	14	5.3466	0.3749	Regionales
mamar	14	6.0246	0.4238	Argot
petate	14	7.6701	0.5427	Regionales
bato	13	0.4361	0.0225	Argot
gacho	13	0.9858	0.0653	Argot
leucemia	6	2.1522	0.3514	Ciencias, Lit.
mamey	6	2.7000	0.4438	Téc., Regs.
miche	6	0.1105	0.0073	Argot
nistamal	6	0.1728	0.0177	Habla pop.
otate	6	1.1182	0.1771	Lírica pop.
súchel	3	0.0716	0.0127	Lírica pop.
tarisbel	3	0.0553	0.0073	Argot
teosófico	3	0.1102	0.0258	Lit. pop.
tupaya	3	0.1102	0.0258	Lit., cienc.

nuestras medidas estadísticas nos sirvan por sí solas como indicadores de diferentes reparticiones de las palabras dentro del *corpus* es una enorme ventaja para el lexicógrafo, que a partir de ellas sabe que debe buscar en qué géneros ocurren predominantemente y, una vez caracterizadas en ellos, decide el tratamiento que debe dar tanto a la definición lexicográfica como, por ejemplo, a las marcas de uso dentro del diccionario. Así, *teológico*, *tórax* y *leucemia* se distinguen inmediatamente como vocablos marcados en religión y en ciencias; *mamar*, *bato*, *gacho*, *miche* y *tarisbel* se reconocen entre los textos del argot del

hampa;<sup>25</sup> etc. Sin duda que todas esas palabras son del español de México; entre ellas se encuentran varias que la intuición de los hablantes consideraría representativas de nuestra variedad regional, como *petate*, *barbacoa*, *súchel*, *mamey*, *otate* y *nistamal*.<sup>26</sup> Es decir, nuestros datos, tanto por lo que respecta a la selección de vocablos, como en cuanto a su caracterización cuantitativa, son útiles para los objetivos del diccionario.

**3.1.2.** Lo mismo resultan útiles nuestros datos cuantitativos para decidir qué tratamiento deben llevar los vocablos que presentan sinonimia entre variantes dialectales o sociolectales. Considérense los ejemplos de la tabla 6:

<sup>25</sup> *Mamar*, tanto en un sector de la lengua popular, como en el *caló* (que es el nombre mexicano de la *germanía* o argot del hampa) significa 'molestar'; *bato* designa a cualquier persona, sea desconocida para el hablante o sea considerada solamente tonta o ingenua (el vocablo procede del *caló* gitano y de las representaciones teatrales pastoriles españolas desde el siglo XVI, en donde es el nombre de un pastor tonto); *gacho* es cualquier cosa de mala calidad o echada a perder, o cualquier comportamiento deshonesto (y la persona que así se comporta); *miche* significa 'cincuenta'; *tarisbel* es una variante festiva de *estaripele*, también de procedencia gitana en la *germanía*.

<sup>26</sup> *Petate* es una estera de paja tejida, que se pone en el suelo para dormir en ella; forma parte de la cultura doméstica prehispánica. *Barbacoa* es un platillo tradicional mexicano, hecho a base de carne de oveja; *súchel* es una variante de *súchil* y ésta de *cempasúchil*, nombre náhuatl de una flor con que se honra a los muertos el dos de noviembre de cada año. *Mamey* es un fruto comestible; *otate* es una caña con que se hacen bastones y otros instrumentos; *nistamal* es la masa de harina de maíz ya preparada para hacer las *tortillas*.

Tabla 6

<i>Vocablo</i>	<i>F. total</i>	<i>F. correg.</i>	<i>Índice de dispersión</i>	<i>Géneros predominantes</i>
romper	171	166.2927	0.9722	General
quebrar	58	26.2977	0.4472	Regional gen.
cinto	8	2.7915	0.3415	Lit., regional
cinturón	15	9.6949	0.6423	General
demorar	11	6.1883	0.4826	Regional
dilatar	19	9.9789	0.5198	Regional gen.
búho	3	0.9256	0.3006	Culto
tecolote	3	0.3409	0.1035	Habla y lit.pop.

Entre los verbos *romper* y *quebrar* hay sinonimia entre variantes dialectales; como se ve, el índice de dispersión de *romper* lo sitúa como de uso general en México, y si uno busca su aparición entre los géneros, comprueba su existencia en todos ellos; en cambio, el índice de dispersión de *quebrar* (0.4472) lo caracteriza como un vocablo poco repartido, lo que lleva al lexicógrafo a buscar en qué géneros predomina y le permite descubrir que, si bien general en todas las zonas dialectales de la República, no aparece bien documentado en la lengua culta ni en la estándar coloquial. De ahí concluye que se trata de un vocablo que habrá de llevar marca de *rural* en México. Es el mismo caso de la pareja *cinto/cinturón*, de la cual el segundo vocablo resulta general en el español de México, mientras que *cinto* es de uso regional. *Demorar* y *dilatar*, con el sentido de 'tardar' son un interesante caso de concurrencia entre formas de uso regional; entre ellas, *dilatar* es más general (aparentemente la forma *demorar* es característica del dialecto veracruzano —una región del centro de la costa del golfo de México—). Por último, entre *buho* y *tecolote* hay una clara diferencia

de niveles de lengua; el primero es exclusivamente de uso culto (ocurre en la literatura y en las ciencias), mientras que el segundo es de uso popular.

Si bien en los ejemplos anteriores los resultados parecen concordar con la intuición de los hablantes (nuestro grupo de lexicógrafos, que ha desarrollado una notable perspicacia para distinguir variantes), es posible que no en todos los casos los datos cuantitativos manifiesten la realidad del léxico en México. En especial, datos que se sustentan en tan bajos resultados de frecuencia y dispersión pueden dar lugar a equívocos. La lexicografía no puede depender exclusivamente de ellos cuando trata de componer un diccionario de uso social y práctico.

No se diga las dificultades para tratar el enorme número de vocablos documentados una sola vez ( $F = 1$ ), entre los cuales hay desde simples errores de captura hasta verdaderos *hapax* que se introdujeron al *corpus* de modo tan aleatorio como el mismo método de recolección de los textos que lo constituyen. Pero ¿no sucede lo mismo a la lexicografía tradicional? Se podría afirmar que, una vez reconocido el núcleo del léxico de una lengua y elaborado en las páginas de un diccionario, la aleatoriedad de aparición de los demás vocablos ya no se puede controlar ni con métodos cuantitativos ni con los métodos tradicionales de captación y acopio de palabras que siguen los equipos de lexicografía de las grandes empresas editoras de diccionarios. Es decir, los vocablos con baja frecuencia de aparición en un *corpus* ya no son “representativos” de una lengua como lo son los del núcleo, sino que corresponden a una parte de la realidad, precisamente aquella que ha quedado plasmada, al azar, en los textos considerados. En ese sentido, lo que aporta

el *CEMC* es, sin duda, vocabulario del español de México, pero no todo el vocabulario ni el más “necesario” para la intuición del hablante. A la vez, lo mismo sucede a la lexicografía tradicional, que tiene que espigar los vocablos en los textos de manera particular y muchas veces a base de apariciones únicas, a sabiendas de que más de un hablante los considerará caprichosos.

Se puede pensar que, a pesar de lo anterior, el aporte de los métodos cuantitativos basados en *corpora* es mucho menor en términos de cantidad de vocablos reconocidos, que el de la lexicografía tradicional. Y bien, cuando se trata de lenguas con una larga tradición lexicográfica, en la que cada diccionario refunde los datos proporcionados por sus antecesores y agrega nuevas palabras, recolectadas precisamente por su extrañeza en comparación con el vocabulario usual, indudablemente los métodos estadísticos resultan pobres y limitados. El hablante, además, tendrá menos oportunidades de darse cuenta de la aleatoriedad y el sesgo de la selección de palabras porque ya los diccionarios son lo suficientemente grandes como para superar su intuición de las “palabras faltantes”. Los métodos cuantitativos no son útiles, en consecuencia para esa clase de lexicografía, para la lexicografía de los “diccionarios-tesoro”, como alguna vez los llamó Ramón Menéndez Pidal.

La lengua española caería dentro del campo de las que tienen una larga tradición lexicográfica, pero ¿qué lengua española?, pues lo cierto es que no se trata, realmente hablando, de la de la Academia Española, ni de la de Madrid, ni de la de Castilla, ni de la de la Península Ibérica, que son las entidades sociales que suelen verse suficientemente tratadas por la lexicogra-

fía española. “Lengua española” es la que se habla en 23 países, separados entre sí por cientos de kilómetros y distinguidos unos de otros por diversas evoluciones sociales y culturales. La lengua española de estos países es la que verdaderamente se desconoce, pues no consiste en la simple adyunción del léxico que contienen los diccionarios peninsulares con el que contienen los diccionarios de regionalismos de cada país. El vocabulario, como la sintaxis, también forma un todo, en que los valores semánticos de cada uno de sus elementos varía de acuerdo con los valores de los demás. Y esa unidad es la que constituye el objetivo de una lexicografía orientada por el rigor de la lingüística. En consecuencia, los métodos que se requieren no son los del “diccionario-tesoro”, sino aquellos que permitan reconocer la lengua en su totalidad, a sabiendas de que se descubrirá, como lo hemos hecho, la existencia de un núcleo común para todas las variedades regionales. Sólo cuando se haga esa clase de reconocimiento, objetivo y verificable, de todas y cada una de las variedades regionales y nacionales del español, se podrá hablar con sentido de un léxico de toda la lengua española y, correspondientemente, de un gran diccionario de la lengua española (quizá compuesto por una colección de 23 diccionarios como el nuestro). Para alcanzar ese objetivo, los métodos cuantitativos como los que he ilustrado en este artículo son la única garantía de objetividad para nuestra moderna lexicografía general y monolingüe. Conocer sus ventajas es tan importante como especificar sus limitaciones y éstas, como hemos visto, no son diferentes de las que ofrece cualquier otro método de recopilación de datos en lexicografía.

CARACTERIZACIÓN METÓDICA DEL  
CORPUS DEL *DICCIONARIO DEL ESPAÑOL*  
*DE MÉXICO*

0. Desde 1973, gracias al interés de don Antonio Carrillo Flores, director del Fondo de Cultura Económica en esa época, he venido elaborando, con varios colegas más, el *Diccionario del español de México (DEM)* en El Colegio de México. Aunque ya han aparecido varias noticias de nuestro trabajo tanto en libros y revistas, como en algunos encuentros de especialistas, la ocasión presente, en que, a mi entender, por primera vez nos reunimos lexicógrafos de la lengua española en un encuentro académico de esta altura y bajo los auspicios de tan prestigiosas instituciones como el Instituto Caro y Cuervo y el Instituto Goethe, me ofrece la oportunidad de presentar ante ustedes un panorama global de los efectos metódicos que ha tenido nuestro *Corpus del español mexicano contemporáneo (1921-1974)* en la confección de nuestro diccionario.

Si bien el DEM está todavía lejos de verse terminado —tenemos un poco más de la mitad del total de la obra relativamente redactada—, el habernos visto obligados<sup>1</sup> a publicar un pequeño *Diccionario básico del español de México* (El Colegio de México, 1986) nos

<sup>1</sup>*Obligados* por dos situaciones: por un lado, la fatalidad que siempre ha enfrentado la lexicografía, de tomarse mucho más tiem-

ha permitido elaborar una imagen de conjunto de la realidad de nuestros métodos y, en consecuencia, de sus virtudes y sus defectos. Siendo el interés principal del Profr. Haensch para esta reunión precisamente la realidad de la práctica lexicográfica, y no las teorías que la subyacen, me ha parecido apropiado abordar el tema del método lexicográfico a propósito de un corpus de datos lingüístico-estadísticos como una manera de motivar la discusión de tan selecto grupo de especialistas como el aquí reunido.

### 1. El objetivo del diccionario y su público

Para Carrillo Flores dos diccionarios norteamericanos del inglés, el afamado *Merriam-Webster* y el *American Heritage*, constituían sendos ejemplos de lo que un país americano debía ser capaz de hacer con la gran lengua de cultura que le tocó recibir y arraigar en la idea de su propia identidad nacional. Este pensamiento de don Antonio era consecuencia de una ya larga tradición nacionalista de México que, en ese aspecto, sólo venía a reconocer, en la nación de allende el río Bravo, un paralelismo de orígenes históricos y de esfuerzos de ambos países por definirse a sí mismos frente a sus respectivas y antiguas metrópolis; a la vez, no era sino la manera sintética de manifestar un sentimiento mexicano muy generalizado, que consiste en la

---

po para elaborar un diccionario que el que juzgan conveniente los patrocinadores de una empresa tal, en especial en un país como México, en donde los plazos largos de la filología y la lingüística son prácticamente desconocidos; por el otro, la solicitud del ministerio mexicano de educación pública, de que pusieramos a su disposición un pequeño diccionario escolar, que sirviera como auxiliar para la enseñanza elemental.

necesidad de reivindicar la libertad y el derecho que tenemos a sentir la lengua española tan arraigada en nuestro origen como en el de los propios españoles, primero ante España, que repetidas veces se nos muestra, todavía, como soberbia madre patria y reguladora de nuestra lengua; y después ante nosotros mismos, también todavía en ocasiones dispuestos a aceptar nuestra marginalidad colonial en cuanto al idioma o, lo que es peor, a negar nuestro español como si sólo fuera una imposición arbitraria de los conquistadores sobre nosotros, mexicanos ideológicamente indios, fatalmente desposeídos de nuestras lenguas precolombinas y condenados, por ello, a un hiato lingüístico de proporciones abrumadoramente metafísicas.<sup>2</sup>

Así que el reto que tuve la fortuna de recibir hace quince años por parte de Carrillo Flores y de El Colegio de México era iniciar el largo camino que nos habría de llevar a un “Webster mexicano”, en el que se manifestara cabalmente la lengua española que hablamos todos, con lo que de común tenemos con el resto de los países hispanohablantes, y con lo que nos distinguimos de ellos y entre nosotros mismos. Esta manifestación habría de preocuparse, ante todo, por hacer patente la riqueza expresiva y la precisión intelectual del español —lengua de cultura— en campos tradicionalmente descuidados por la lexicografía hispánica, como lo son los diversos vocabularios de las ciencias y las técnicas, así como los de ciertas áreas de

<sup>2</sup>La fractura de la idea que tienen los mexicanos sobre su propia lengua forma parte del problema más general de la manera en que ven su identidad nacional; al respecto, he hecho tres intentos de explicar esa fractura y sus consecuencias en Lara, 1983, Lara, 1987b y Lara, 1988. Cf. aquí mismo “La cuestión de la norma en el DEM”.

las humanidades, como la jurisprudencia y la antropología mexicanas.

El objetivo era, entonces, hacer un diccionario de la lengua española común, pero tal como la usamos en México y definida a partir de los valores de sentido que se han creado en nuestra propia cultura. Se podría decir que, desde el punto de vista lingüístico, la tarea no era más que el cumplimiento de una necesidad primaria de la ciencia del lenguaje: la descripción del léxico de nuestra variedad del español tal como es, haciendo caso omiso de sus evidentes concordancias con el resto de las variedades hispánicas y de la necesidad filológica de marcar las diferencias con ellas; es decir, había que redactar un diccionario *regional* del español y no un diccionario más de *regionalismos* mexicanos.<sup>3</sup>

Pero dado el objetivo social del diccionario, definido por la idea que lo originó, la justificación lingüística del DEM quedaba subsumida en aquél y habría de dar lugar a una elección de métodos de trabajo en los que algunas finalidades sociales entrarían en tensión con algunos métodos, acerca de los cuales la lingüística y la tradición lexicográfica de nuestra lengua parecían tener con ellas todas las certezas y ninguna duda. Este tema, que en realidad pertenece al campo de la teoría del diccionario —no soy de los que gustan llamarla “metalexigrafía”— merece una larga reflexión que no haré aquí.<sup>4</sup>

<sup>3</sup>Cf. también Haensch, 1989 acerca de la distinción entre ambos tipos de diccionarios.

<sup>4</sup>Lo he desarrollado en “El objeto diccionario”. Mis razones para no utilizar en muchos casos, como en éste, el prefijo *meta-* (y, en

## 2. El *Corpus del español mexicano contemporáneo (1921-1974)*

El primer problema de método que se nos planteó para hacer el diccionario era la manera en que podríamos reconocer el léxico real del español de México, habida cuenta de que los estudios y los catálogos lexicográficos dedicados a nuestra variedad lingüística suelen interesarse únicamente por lo que diferencia al español mexicano de las otras variedades de la lengua, fundamentalmente del vocabulario incluido en el *Diccionario de la lengua española*, de la Real Academia Española (*DRAE*), pero no permiten establecer una descripción completa y armónica de lo que es el español mexicano común, y menos nuestro español considerado como lengua del intelecto y la civilización contemporáneos. Así que había que encontrar la manera de recolectar una gran cantidad de datos léxicos de nuestra variedad lingüística y el modo de determinar entre ellos lo que constituiría el núcleo del vocabulario del español mexicano. Fue por eso por lo que optamos por construir un gran *corpus* de textos del español mexicano contemporáneo (*CEMC*), poniendo en práctica los métodos propuestos por los especialistas en estadística léxica. En otros lugares he explicado la conformación de esta muestra de datos lingüísticos, que constituye la documentación primaria del *DEM*.<sup>5</sup>

---

consecuencia, tampoco el término “metalexicografía”) las he expuesto en Lara, 1987a.

<sup>5</sup>Cf. Lara/Ham, 1974, Ham, 1981 y, aquí mismo, “La cuantificación en el *DEM*”, en donde se ofrece un primer balance de la utilidad de la estadística léxica para nuestro trabajo.

2.1. El *CEMC* consta de 996 textos, cada uno formado por tantos párrafos aleatoriamente seleccionados en las obras que constituyen nuestra "muestra", como fuera necesario para reunir aproximadamente 2 000 palabras gráficas por texto. Los textos provienen de todo tipo de obras escritas en México, así como de grabaciones magnetofónicas realizadas por toda la República. Estos materiales los sometimos a un análisis automático por medio de una computadora electrónica, y les aplicamos varios algoritmos estadísticos y un programa de producción de concordancias.<sup>6</sup> El resultado ha sido un enorme "diccionario estadístico" del español de México, acerca de cuyas características hablaré en breve y un conjunto igualmente grande de documentaciones del uso mexicano de los vocablos que contiene el *corpus*. En total obtuvimos de nuestro análisis 1 891 045 ocurrencias de palabras gráficas, que dieron lugar a 64 183 tipos. No hemos lematizado todavía el total de tipos como para saber cuántos vocablos encontramos.<sup>7</sup>

2.1.1. Después de haber sometido estos resultados a un análisis que ya dura cerca de diez años, creemos poder estar relativamente seguros de haber logrado la finalidad principal de su elaboración, que fue encontrar un núcleo de nuestro léxico mexicano a partir del cual pudiéramos iniciar la redacción del diccionario; tal

<sup>6</sup>Una exposición detallada del *Analizador gramatical automático* del DEM se encuentra en Lara/García Hidalgo/Ham, 1981.

<sup>7</sup>Esta tarea se realiza manualmente, dado el hecho de que no incluimos suficientes algoritmos en nuestro *Analizador gramatical automático* como para que la computadora hiciera por sí sola la lematización; al respecto, cf. García Hidalgo, 1981.

núcleo se podría definir de dos maneras: en sentido restringido, como “vocabulario fundamental” (1 451 vocablos del nivel culto, o 1 131 vocablos en todo el *CEMC*), mediante la acumulación convencional del 75% de las ocurrencias de palabras en el *corpus* (en tal caso, después de comparar este “vocabulario fundamental” con el equivalente en la obra *A Frequency Dictionary of Spanish Words*, de A. Juilland y E. Chang Rodríguez (1964), podríamos también afirmar que encontramos el núcleo común del español mexicano y el peninsular, al menos; Cf. “La cuantificación en el *DEM*’”); en sentido amplio, definiéndolo a partir de las cotas de frecuencia mínima  $F \geq 5$  o  $F \geq 4$ , que los mismos Juilland y Chang-Rodríguez (1964) o Bortolini/Tagliavini/Zampolli (1971) establecieron para que un vocablo cualquiera tenga valor en el estudio de la estructura léxica de una lengua, tenemos 17 903 tipos con  $F \geq 5$  o 20 781 tipos con  $F \geq 4$  que seguramente forman parte del español mexicano. Sólo que en este caso, si bien la documentación de cuatro o cinco ocurrencias de un vocablo parece suficiente para un estudio lexicológico, cuando se trata de los materiales que requiere la lexicografía, tal cantidad de ocurrencias de un vocablo resulta, muchas veces, insuficiente si desea uno atenerse estrictamente a esos datos; pues salvo los vocablos que encuentran una dispersión homogénea en todo el *corpus* y que, por esa razón, suelen corresponder a un uso generalizado y no marcado social o terminológicamente, todos los demás no alcanzan a definir claramente sus usos y se convierten en accidentes aleatorios más o menos cercanos al verdadero núcleo léxico que uno pretende reconocer, y en *síntomas* solamente de la extensión real de un vocablo en el uso

mexicano. Más adelante volveré a tratar este problema. De todos modos, hemos considerado bastante la documentación obtenida de nuestro *CEMC* como para basar en ella la macroestructura inicial de nuestro diccionario.

**2.1.2.** El análisis estadístico del *CEMC* nos ofrece las medidas siguientes: además de la frecuencia total de cada tipo de vocablo, sus frecuencias relativas por género (tenemos 14 géneros diferentes), sus frecuencias relativas entre géneros, una medida de dispersión, basada en la fórmula de Lanke<sup>8</sup> de *frecuencia corregida*, y una medida normalizada de dispersión, llamada “índice normalizado de dispersión C”. Estas medidas nos sirven para diferentes finalidades: mediante el índice normalizado de dispersión determinamos qué vocablos tienen la mayor frecuencia y la mayor dispersión en el *corpus*, con objeto de dedicarnos a estudiarlos y redactar sus artículos lexicográficos correspondientes, pues deben pertenecer al español mexicano común. Cuando este índice señala baja frecuencia o poca dispersión, se convierte en un síntoma ya sea de uso restringido a cierto tipo de discurso, ya sea a cierta disciplina o a cierta región del país. En ese caso, el redactor tiene que realizar una pequeña investigación del vocablo para determinar de qué fenómeno se trata; muchas veces, la consideración del índice da por resultado la asociación de una marca de uso al vocablo en cuestión dentro del diccionario. El manejo práctico del índice es muy sencillo, pues opera entre cero y uno: mientras mejor repartido y más frecuente es un voca-

<sup>8</sup>Cf. Ham, 1981:4.4.

blo, el índice se acerca a uno; en caso contrario, tiende a cero.

La medida de frecuencia corregida es el dato de partida del índice de dispersión y constituye el resultado estadístico de importancia; tiene la ventaja, sobre otras medidas formuladas anteriormente, de que toma en cuenta la heterogeneidad de la estructura de nuestro *corpus* y de esa manera no está sesgada hacia los géneros del *corpus* de mayor tamaño. La dificultad práctica de su uso reside en el hecho de que varía de acuerdo con el resultado de frecuencia total de cada vocablo y requiere una comparación dilatada antes de poder sacar conclusiones de ella. En realidad, no obstante su importancia estadística, la usamos poco en la práctica.

Las medidas de frecuencia relativa sirven ya sea para refinar la exploración de usos de los vocablos por género —y llevan, en tales casos, a establecer significados característicos de ciertos temas o de ciertas disciplinas—, o ya sea para darse una idea de la mayor o menor frecuencia de un vocablo en comparación con todos los que ocurren en un solo género (lo cual se presta, más que a un tratamiento lexicográfico inmediato, a la realización de estudios lexicológicos de interés más estrictamente científico).

En “La cuantificación en el *DEM*” (incluido en este libro) he desarrollado otros aspectos del valor metódico de nuestro *CEMC* para el diccionario, que ahora no tendría tiempo de explicar aquí. Por ahora, me basta con haber podido ofrecerles los elementos suficientes para considerar, en seguida, algunos ejemplos de las ventajas y las limitaciones que encontramos en el manejo lexicográfico del *corpus*.

**2.2.1.** Como decía antes, nuestro *CEMC* parece cumplir con los objetivos para los cuales fue creado; sin embargo, no todas las necesidades de la lexicografía se pueden resolver con su ayuda. Por principio de cuentas, como se sabe desde hace tiempo, ninguna “muestra estadística” es capaz de registrar todo el vocabulario que le interesaría reconocer a un hablante o a un lingüista, ni lo que registra corresponde a lo que uno cree que constituye el vocabulario más “necesario”, más “pertinente” o, simplemente, más “usual”. El *corpus* es rico, sin duda, y a veces hasta sorprende por la manera en que “percibió” ciertos fenómenos léxicos que uno intuía o que ni siquiera sospechaba. Pero otras veces el *corpus* muestra pobreza o sesgos en la información que contiene y uno no deja de molestarse con él. Ilustraré en seguida unos cuantos casos, no precisamente de falla del *corpus*, pero sí de insuficiencia.

Nuestro *CEMC* no está estructurado de manera homogénea, de tal modo que para cada tipo de discurso, cada región de México o cada campo del vocabulario corresponda el mismo número de textos y aproximadamente el mismo número de ocurrencias; por el contrario, privilegamos aquellos vocabularios que constituyen el objetivo central del diccionario: los de las ciencias, los del periodismo, los de la literatura, etc., sobre todos los demás. En consecuencia, la documentación de vocablos de cada uno de esos campos o géneros varía proporcionalmente con la ponderación que les asignamos. Así por ejemplo, mientras que para las ciencias documentamos 26 487 tipos (incluso los de  $F = 1$ ) en los 180 textos que forman esa muestra, para el vocabulario religioso reunimos solamente 3 979 tipos en los cuatro textos explorados; y si para el periodismo documenta-

mos 21 072 tipos en 176 textos, para las jergas profesionales y generacionales sólo encontramos 4 710 tipos en los 30 textos coleccionados. Podemos formular una medida sencilla de la mayor o menor capacidad de cada género del *corpus* para documentar aceptablemente los vocablos reunidos; si dividimos el número de tipos encontrados con  $F = 1$  entre el número de tipos con  $F \geq 5$  en cada género del *corpus*, el resultado nos indicará qué tan fragmentado quedó el léxico en ese género y, por lo tanto, qué tan bajas fueron las frecuencias de los vocablos documentados en él, sobre la base del promedio de fragmentación de todo el *CEMC*, que fue de 1.66. Para las ciencias, fue de 2.00; 1.82 para el periodismo y 2.58 para la literatura, mientras que subió a 3.69 en discurso religioso y 3.31 en jergas; esto quiere decir, por un lado, que los vocablos encontrados en los géneros de periodismo, ciencias y literatura, resultaron mejor documentados, en términos de frecuencia, que los que obtuvimos en discurso religioso y en jergas; por el otro, combinada esta medida con la interpretación que hace mi colaborador Roberto Ham de la "riqueza léxica" de cada género del *corpus* (Ham, 1981:5), lo que significa esa medida es que el número de vocablos diferentes por género (aunque sea con  $F = 1$ ) es comparativamente menor, por ejemplo, en el de periodismo que en el de literatura o el de jergas. Eso da por consecuencia que nuestro *CEMC*, de acuerdo con lo planeado, sirva más para reconocer el español mexicano común que para documentar singularidades, pintoresquismos, indigenismos, etc. El género del *corpus* dedicado a las grabaciones de conversaciones provenientes de todas las regiones de México —tomadas de la investigación del Prof. Juan M. Lope Blanch para el *Atlas lingüístico de México* (de inminente

publicación)— nos dio 13 095 tipos en 130 textos, con un índice de fragmentación de 2.01; tal resultado no es, numéricamente, malo, sólo que, como se trata precisamente de vocabulario que varía de región en región de México, no alcanza para poder establecer la clase de isoglosas léxicas que necesitaríamos para dar el tratamiento lexicográfico adecuado a nuestros regionalismos. He aquí algunos ejemplos de lo que digo, tomados de un trabajo al respecto de mi colaboradora Luz Fernández (1987): *birote*, que es el nombre de una presentación ordinaria del pan en buena parte de México —quizás todo el norte y parte del altiplano central— aparece cuatro veces en nuestro *CEMC*, pero las cuatro documentado solamente en La Paz, Baja California. Tal documentación contradice nuestro conocimiento de hablantes, que sabe, por experiencia directa en varias zonas de México, que ese vocablo tiene una extensión mayor. Como hablantes, también diríamos que, junto con *bolillo*, opera una distinción de dos grandes zonas dialectales mexicanas: *birote*, como he dicho, al norte, y *bolillo* al sur (incluida la ciudad de México). *Bolillo*, además, parecería ser el vocablo general del español de México, pero su poca dispersión en el *corpus* tampoco nos permite afirmarlo. *Guineo* se documenta sólo tres veces, en textos del estado mexicano de Chiapas, mientras que *plátano* tiene variación importante en su dispersión (su índice “C” es de 0.7395 cuando se considera la división en tres grandes géneros, y de 0.3227 cuando se considera la división en catorce géneros) y *banano* aparece con una bajísima dispersión de 0.1738 (*banana* no se registra); de estos datos sólo podemos concluir que *plátano* es la voz más usual en lengua culta —primer índice de dispersión—, pero no que esté generalizada

por todo México, así como que *guineo* parece tener vida en nuestro país, aunque no sabemos hasta dónde se extienda su uso; tampoco podemos afirmar que *banana* y *banano* no se usen en México. *Pilmama*, un nahuatlismo, ocurre dos veces en el estado de Durango, en tanto que su sinónimo *nodriza* tiene un primer índice "C" de 0.6083, por lo que, si bien esta última es la palabra usual de la lengua culta, no podemos concluir que sea ésta la forma general en español mexicano —segundo índice "C" = 0.1329—, ni que *pilmama* en efecto sea un regionalismo de Durango —nuestro conocimiento de hablantes nos hace suponer que esta voz está mucho más extendida en México—, por lo que nos vemos imposibilitados de asignar una marca de región de uso a las entradas correspondientes, basados exclusivamente en nuestro *CEMC*. Algo semejante sucede con *zurrapa*, sólo documentada en La Paz, Baja California, frente a *migaja*, cuyo primer índice de dispersión es de 0.8519, lo que sí logra mostrarla claramente como la forma culta general del español mexicano, pero tampoco permite concluir que sea la más extendida regionalmente frente a *zurrapa*, que también hemos oído en la ciudad de México. Estos ejemplos, y muchos más que han aparecido en nuestro trabajo, nos llevan a concluir que nuestro *corpus* no ofrece el material adecuado para dotar a nuestro diccionario de la requerida solidez en materia de regionalismos mexicanos.<sup>9</sup>

<sup>9</sup>Tenemos esperanzas de que la publicación del *Atlas* de Lope Blanch nos ofrezca documentación fidedigna, al menos, de los vocablos que representan isoglosas léxicas importantes. Desde hace tiempo nos hemos convencido de la necesidad de hacer nosotros mismos una dilatada encuesta por todo México; por desgracia, nos faltan los fondos necesarios.

2.2.2. Lo que cuantifica nuestro *CEMC* son *significantes* y no la compleja relación que hay entre éstos y los sememas que vehiculan en cuanto signos polisémicos, por lo cual los resultados de nuestro análisis, con ser de importancia para permitirnos determinar qué vocablos deben constituir el *DEM* (o, por lo menos, su mayor parte), no nos ofrecen una información equivalente cuando se trata de significados. Este aspecto de la estadística léxica ha quedado por lo general intocado, aun cuando haya razón para exigirle un tratamiento parejo de ambos planos del signo lingüístico. Prácticamente, lo que ocurre es que la determinación de significados es imposible hacerla mediante la misma clase de procedimientos automáticos basados en algoritmos formales, y queda por completo en manos del redactor del diccionario, que habrá de realizar un análisis semántico para definir los sememas encontrados en cada vocablo del *CEMC*; pero entonces se plantea el grave problema de cómo saber qué sememas se han encontrado efectivamente en nuestros datos y si son esos sememas, y sólo esos, los que habrán de quedar registrados en el diccionario.

Con el plano de los significados se repiten las mismas dificultades que nos presentaban los vocablos registrados con una baja frecuencia; salvo los significados de vocablos pertenecientes al núcleo “fundamental” del español mexicano, que en general quedaron suficientemente documentados en el *CEMC*, y los significados de muchos otros vocablos más que, sin llegar a  $F = 5$ , se registraron suficientemente —me es imposible definir, por ahora, en qué frecuencia se dio el límite más bajo de suficiencia—, todos los demás significados registrados sólo pueden considerarse como indicadores de ciertos usos mexica-

nos, pero de ninguna manera agotan la realidad de esos usos semánticos.

Además, dada la naturaleza propia del plano semántico de los vocablos, ni siquiera los significados registrados tienen una identidad unívoca que nos permita, por ejemplo, llegar a afirmar algo como que la palabra *cabeza* tenga 23 acepciones y sólo 23, o que *guardar* tenga exactamente las cinco acepciones que le dimos en el *Diccionario básico* y no sea posible elaborar otro análisis, igualmente válido, que las reduzca a solamente tres.<sup>10</sup> De allí la causa para que, contrariamente al optimismo de varios especialistas en inteligencia artificial, parezca imposible diseñar algoritmos formales que lleven a una elaboración automática del significado de los vocablos. Así es que la documentación que ofrece el *CEMC* es solamente materia prima para un trabajo que no puede someterse a la estadística.

Antes decía yo que una frecuencia de cuatro o cinco en el *corpus* establecía la frontera entre los vocablos adecuadamente documentados y los que no lo eran. También señalé que, sin embargo, cuando se trata de regionalismos mexicanos, aun una frecuen-

<sup>10</sup>En Lara, 1986a, a propósito de una somera comparación entre lo que recoge el *DRAE* y lo que asienta nuestro *DBEM*, traté de distinguir, al menos en principio, lo que puede ser una diferencia de interpretación semántica, de lo que sería la existencia de un verdadero significado diferente. Por ejemplo, en el artículo del verbo *ser*, el *DRAE* da en tres acepciones (8, 10 y 11) lo que nuestro *DBEM* interpreta en una sola (núm. 2); en cambio, en el artículo de *lanzar* del *DRAE* no parece haber posibilidad de interpretar como ya existentes las acepciones 2 a 6 de nuestro *DBEM*. La cuestión, como se podrá ver, es muy sutil y merecería una discusión más larga por parte de la teoría semántica que determina la delimitación de las acepciones en los diccionarios. Cf. también Lara, 1986b.

cia de cinco era poco si lo que se busca es llegar a tener una idea clara de las isoglosas léxicas que revelan las diferencias regionales del español de México y, en consecuencia, las zonas en las que se utiliza un vocablo en vez de otro. En realidad, ya tres ocurrencias de un vocablo, repartidas en más de un texto, son suficientes como para considerarlo “existente” en una lengua (es decir, que no se trate de un vocablo de uso estrictamente individual y que tampoco sea un *hapax*), por lo que, quizás con poco valor estructural, pero con validez documental, se puede confiar uno en los resultados que ofrece el *CEMC* con tan baja frecuencia. Si a todo esto se suma la imposibilidad de cuantificar el plano del significado de los vocablos en un *corpus* de esta o de cualquier otra clase, no queda más que establecer una distinción muy clara entre el valor del *corpus* para delimitar lo que es el núcleo léxico de una lengua —lo que en nuestro caso era un objetivo primordial— y su valor como esclarecedor de *pistas* o como importante conjunto de *síntomas* de la existencia de una multitud de vocablos que habrá que investigar con medios diferentes de los de la estadística. Consecuentemente con ello, también habrá que distinguir entre lo que hemos logrado hacer para determinar un núcleo importante del español mexicano común, y lo que deberemos aceptar como español mexicano, ciertamente, pero ya no tan común y de ninguna manera como la única versión representativa de nuestra variedad lingüística. Dicho de otro modo: tenemos suficientes argumentos como para poder afirmar que el *DEM* (y su versión mínima, el *DBEM* sobre todo, pues se ha nutrido especialmente del núcleo léxico bien definido estadísticamente en

nuestro *corpus*) contiene el vocabulario de más uso y más central del español mexicano y que, en esa medida, es una fidedigna representación de nuestra variedad nacional. En cambio, cuando se trata de los vocablos registrados en el *CEMC* con baja frecuencia, habrá que reconocer que lo que ofrece nuestro diccionario es solamente *una* (la primera y, si se quiere, históricamente importante) versión del vocabulario que se usa en México, que no puede pretender ser la única, ni permitir que así se la considere, so pena de perder toda aspiración de rigor científico y hasta quizás de ayudar a que en nuestra república sobrevivan el autoritarismo y el dogmatismo que emponzoñan nuestra vida social.

### 3. Las fuentes de consulta

Esa función de síntoma de ciertos datos, realmente tan natural para la filología, ha sido siempre la que ha desencadenado la investigación lexicológica y la que, en la lexicografía tradicional, ha definido buena parte de sus métodos de trabajo. Puesto que nuestro objetivo *social*, como decía al principio, es hacer un diccionario del español mexicano dirigido a todo el público de nuestro país, y no solamente a nuestros colegas lingüistas, que podrían sentirse hasta más satisfechos si nos atuviéramos a lo bien documentado y no “mezcláramos” métodos, nos vemos en la necesidad, precisamente en ese punto, de acudir a otros instrumentos de método, que nos permitan seguir aprovechado la documentación recolectada en el *CEMC*, que desde ese momento pasa a convertirse en documentación de síntomas, objeto de estudios filológicos de carácter tradicional.

**3.1.** Para hacerlo acudimos a dos procedimientos bien conocidos: ante todo, a la búsqueda de otros registros de la palabra insuficientemente documentada en el *CEMC*, en estudios filológicos y lingüísticos del español de México o que traten algún vocablo también documentado por nosotros en México, y en otros diccionarios, tanto de mexicanismos y de regionalismos mexicanos, como generales. Para ello hemos tratado de nutrir de la mejor manera posible nuestra biblioteca de referencia y hemos iniciado una revisión de las revistas especializadas que suelen incluir investigaciones sobre el léxico hispánico, como la *Revista de filología española*, el *Thesaurus* del Instituto Caro y Cuervo, nuestra *Nueva revista de filología hispánica*, el *Boletín de la Academia Española*, etc. Con esa revisión, llevamos un fichero que ya asciende a los sesenta mil lemas que pueden tener interés en la búsqueda de una palabra determinada. Nuestros encargados de documentación revisan cada uno de esos estudios y toman nota de las palabras que allí se tratan, con el objeto de incluirlas en un fichero especial, también en orden alfabético, pero de vocablos hispánicos en general. Las indicaciones que se hacen en este proceso se integran en la carpeta monográfica que recibe el redactor como material de trabajo para cada palabra. La información procedente de diccionarios, a su vez, también forma parte de esos datos. Cuando se trata de diccionarios generales y mexicanos, éstos se convierten en “diccionarios-testigos” de nuestro trabajo, contra los cuales revisamos nuestra propia producción. El *Diccionario de mejicanismos* de Santamaría, el de *Aztequismos* de Luis Cabrera, el *Vocabulario sonorense* de Sobarzo, o el *Diccionario del caló*, de Chabat,

por sólo nombrar algunos, son valiosas ayudas para nuestro trabajo diario.

No necesito explicar aquí un método de trabajo que conocemos todos. A partir de vocablos registrados en el *CEMC*, cada redactor busca los datos necesarios para poder incluir las voces en cuestión en el diccionario. Por lo general, los vocablos que han aparecido en el *corpus* con baja frecuencia y poca dispersión son regionales o pertenecientes a terminologías técnicas. Cuando se trata de voces regionales, hemos tratado de acopiar cuanto diccionario o pequeño estudio hemos encontrado en México, independientemente de su calidad formal. Así hemos podido comprobar cómo una obra tan poco considerada en México por su valor lexicológico como el *Tumbaburras de la picardía mexicana*, de Alberto Jiménez (por lo demás, es un éxito de librería), contiene datos fieles del habla popular y vulgar de la ciudad de México; cómo el *Diccionario del caló*, de Carlos Chabat, aunque posiblemente incompleto, refleja cuidadosamente la jerga del hampa mexicana; o cómo el *Diccionario rural de México*, de Leovigildo Islas Escárcega suele dar verídica cuenta del vocabulario campesino del centro de México, en especial. A base de esas obras, que se cruzan con los registros fragmentarios de nuestro *CEMC* hemos podido ir reconstruyendo parte del vocabulario mexicano, particularmente del popular y rural.

Una obra como el gran *Diccionario de mejicanismos* de don Francisco J. Santamaría no deja de plantearnos dificultades, pues a menudo registra vocablos que ya no podemos documentar en ninguna otra fuente, como si se tratara de voces antiguas y ya desaparecidas, o de palabras de uso muy restringido en regiones como Tabasco, de donde era originario don Francisco.

En estos casos, nos vemos obligados a guardar la información y esperar a que aparezcan nuevos datos, que nos confirmen la información de Santamaría.

Por último, cuando se trata de voces técnicas, solemos acudir a un cuerpo de asesores en cada disciplina, que nos aconseja y nos permite evaluar lo encontrado en el *corpus* y en los diccionarios terminológicos que consultamos. En este campo, constituye una dificultad especial el vocabulario de la flora mexicana, sobre la cual lo que se conoce, como el *Catálogo de nombres vulgares y científicos de plantas mexicanas* de Maximino Martínez, no es todavía suficiente para desentrañar la enorme polisemia de las palabras españolas con que se designa una naturaleza ricamente variada, así como para encontrar al menos las voces más extendidas, procedentes de nuestras lenguas amerindias, con que la mayor parte de esa naturaleza sigue siendo nombrada.

**3.2.** Nuestro segundo procedimiento depende totalmente del conocimiento reflexivo, crítico y científico que tienen o deben tener nuestros redactores de su lengua, de su léxico, de su cultura y de su país. De por sí, la interpretación de los datos cuantitativos de cada palabra y de las referencias bibliográficas que obtenemos de las otras fuentes de consulta, así como la elaboración del análisis semántico y gramatical del vocablo en cuestión son tareas que ponen en juego toda la capacidad intelectual del lexicógrafo. Así que no es extraño que, una vez agotados los materiales con que nos provee el *CEMC* y que obtenemos del resto de nuestras fuentes de consulta, haya que pasar a depender del conocimiento y del juicio del grupo de redactores del *DEM*.

Para ello, la experiencia nos ha mostrado qué cualidades individuales son las que nos conviene buscar en un redactor. Estas cualidades no son de las que se aprenden en un manual y durante el plazo de los estudios universitarios. Más bien dependen de la historia familiar del redactor, de la pluralidad de sus intereses y del oficio que, por una razón u otra, haya desarrollado en la escritura. Nuestros redactores son todos mexicanos —una condición de método, en la medida en que la sutileza del significado de un vocablo sólo puede percibirla adecuadamente quien se ha formado en medio de la sociedad mexicana—, todos tienen experiencia en la escucha de nuestras voces regionales y en la lectura de nuestros autores, todos manejan alguna lengua extranjera —que les sirve como instrumento de extrañamiento de la lengua materna, una condición esencial de la crítica— y todos gustan de la variedad del conocimiento humano. Sobre esas bases, las últimas decisiones acerca de un análisis semántico, de la división en acepciones de un artículo, de la existencia de una voz o de un significado específicos y mal documentados, las tomamos en grupo y mediante acuerdos consensuales. He aquí otro motivo más para sostener que el *DEM* es una versión del español de México y que, del mismo modo en que nuestro *CEMC* tiene fechas: 1921-1974, la interpretación del sentido que se manifiesta en la obra debe datarse en nuestra condición social e histórica. Nuestro diccionario, por ello, es más que la transcripción descriptiva de los datos que recolectamos del español mexicano y menos que la manifestación unívoca y atemporal del hablar de los mexicanos; nuestro *corpus* estadístico se resuelve, finalmente, en la interpretación, en la hermenéutica.

No me es posible abusar de ustedes y continuar extendiéndome sobre una buena cantidad de aspectos de nuestro método de trabajo que, todos sabemos, merecen comentario. Me basta con haber podido convencerlos, quizá, de las ventajas que ofrece un *corpus* de datos lingüísticos a la lexicografía, así como haber podido confirmarlos, también, en su sano escepticismo ante la estadística, que no habrá de ser otra cosa que el aprecio y la admiración del trabajo del ser humano, que es el lexicógrafo.

## SOCIOLINGÜÍSTICA DEL *DICCIONARIO* *DEL ESPAÑOL DE MÉXICO*

Una característica de los diccionarios rara vez considerada es su fundamento social. No suele tomarse la compra-venta de diccionarios, y el papel destacado que suelen tener en las bibliotecas caseras, como señal de algo más que trascienda el comercio y el consumo. En efecto, los diccionarios comienzan por ser mercancías. En cuanto objetos físicos —libros, diskettes, microfichas, discos compactos— son mercancías insertadas en un sistema económico, que derivan su valor del interés social por comprarlos y usarlos como obras de consulta. Pero ese interés de la sociedad por comprar diccionarios no es, como aparenta serlo, de la misma clase que sus intereses por otros satisfactores como la comida, el vestido o los automóviles. Su interés, cuando uno reflexiona más detenidamente sobre ellos, proviene de otros ámbitos de la vida social: el de la memoria colectiva y el de las instituciones simbólicas de la sociedad. Ámbitos, quizá, de extraña realidad para quien no haya salido del materialismo burdo en que a veces parece recluirse la ciencia, pero familiares para una visión antropológica o sociológica. Como memoria colectiva, el diccionario se une a las bibliotecas, los archivos, las bases de datos y las enciclopedias; como institución simbólica se agrega a los libros sagrados, las leyes y las artes, que definen y plasman la identidad

de una sociedad, como efecto de la reflexión colectiva sobre aquellas experiencias de sus individuos que resultan valiosas y trascendentes para ella; es decir, el diccionario está compuesto de tradición y de cultura.

De ahí que el diccionario sea un producto social en dos sentidos: en el usual, característico de la economía, y en el filosófico, en el que corresponde a la concreción de una obra intencional y dirigida hacia una finalidad específica.

Pero un diccionario es también un fenómeno lingüístico; es, como decía Karl Bühler, un *producto lingüístico*, resultado de una infinidad de actos verbales que, en la experiencia social, se han ido desligando de sus actores para pasar a formar parte de lo que cada uno de los miembros de la sociedad puede llegar a comprender y, sobre todo, de lo que ha sido dicho inteligiblemente en su seno.<sup>1</sup>

El diccionario, por lo tanto, en sus tres aspectos más importantes: el de producto mercantil, el de producto

<sup>1</sup>He desarrollado la idea del diccionario como colección de actos verbales antes, en "El objeto diccionario". La "socialidad" del léxico que se encuentra en un diccionario no se ha tomado en cuenta entre los teóricos ni de la lingüística ni de la lexicografía. El vocabulario registrado en un diccionario siempre es más y es menos que el que puede conocer cada uno de los hablantes de una lengua y, sobre todo, nunca corresponde a las ideas que se hacen de él sus hablantes. Es más porque está formado por el conocimiento de la lengua reunido en la memoria social, lo que es un hecho tradicional e histórico siempre anterior a cada hablante individual; es menos porque cada hablante puede conocer vocablos que todavía no hayan sido detectados por la lexicografía o que nunca lleguen a fijarse en la memoria de la sociedad. En ese sentido, un diccionario no es nunca exclusivamente una *descripción* del uso del vocabulario en un lugar, un tiempo o un grupo social dado.

cultural y el de producto lingüístico, se manifiesta como fenómeno social por excelencia.

En este artículo me propongo tomar el *Diccionario del español de México* (en lo sucesivo abreviado como DEM) como un caso que ilustre los aspectos sociales que lo proveen de justificación y de contexto, y que, como dije antes, son los verdaderos fundamentos de la lexicografía, así como los sociolingüísticos más interesantes de la investigación que lo precede y de su redacción, y que vienen a dar al diccionario sus características concretas últimas.

### *El contexto*

Es bien sabido que la comunidad lingüística hispanohablante, cercana ya a los 400 millones de individuos y extendida desde la Península Ibérica hasta buena parte del Continente Americano (sin olvidar algunas regiones de África —no solamente Ceuta y Melilla, sino sobre todo la República Árabe Saharaí y Guinea Ecuatorial— e, históricamente al menos, las Islas Filipinas), ha formado 23 países y un número todavía indeterminado de culturas regionales dentro de ellos. También es bien sabido que, a diferencia de lo ocurrido con otras lenguas internacionales, como el inglés, la conciencia lingüística hispánica —más la extrapeninsular que la de España misma— se ha mantenido dependiente de un ideal de lengua simbolizado y definido por la Academia Española, como representante institucional de un español metropolitano al que se le asigna carácter universalmente normativo.

Entre las lexicografías de lenguas internacionales, probablemente sea la española la que más tajantemente distingue entre el *diccionario general* y el *diccionario de*

*regionalismos*. El primero es el que pretende registrar todo el vocabulario de la lengua y el segundo se ocupa solamente de aquel vocabulario que se limita a un uso regional. En la realidad de la lexicografía española, el único diccionario general reconocido como tal es el *Diccionario de la lengua española* (en lo sucesivo, DRAE), elaborado por la Academia metropolitana de Madrid (todos los demás, generales, y editados tanto en España como en Francia e Hispanoamérica son subsidiarios suyos, compuestos a base de refundiciones de sus materiales); hay, en cambio, decenas de diccionarios de regionalismos, ya sea por países (diccionarios de mexicanismos, venezolanismos, chilenismos), ya por regiones históricas (diccionario de andalucismos), ya por zonas políticas interiores de cada país (diccionario de sonorensismos, diccionario de bogotanismos), y, por supuesto, diccionarios más amplios, como los de americanismos (Haensch 1987). Tal repartición del trabajo lexicográfico obedece, sin embargo, más al fenómeno normativo por el cual se conserva la idea de una definición metropolitana de la lengua —representada por la Academia madrileña—, opuesta a la marginalidad regional (colonial) hispanoamericana, que a una razonable división del trabajo en una comunidad lingüística de tan grandes proporciones (Lara, 1988, incluido aquí como “La cuestión de la norma en el *DEM*”).

Este fenómeno es objeto de discusión en Hispanoamérica desde hace muchos años; ha trascendido el campo académico y filológico y ha llegado a la opinión pública, en donde se manifiesta con defensas apasionadas del DRAE frente a ataques encarnizados en contra suya, y con imposiciones inapelables de su vocabulario frente a desestimaciones tajantes de su valor. Sea

cual sea la opinión de cada persona en Hispanoamérica, lo cierto es que hay un clamor general, que pide reconocimiento del léxico hispanoamericano. Este clamor da al DEM su raíz social, y tanto el diseño de la obra como la investigación realizada para llevarlo a cabo proceden de ella.

Lo que se pide al DEM, dicho de manera concisa y efectiva por uno de sus impulsores, es que al cabo de los años llegue a tener el papel de "un Webster mexicano"; es decir, que presente y represente el vocabulario de la lengua española sobre la base de su uso integral en México. Haensch (1989) lo ha definido, precisamente, como "diccionario integral" frente a los diccionarios "contrastivos" que son los de regionalismos. En otra parte lo he definido también como *diccionario regional* del español (Lara, 1978).

Resulta, entonces, que la base social del DEM ha sido definida no por una decisión arbitraria (todo lo adecuada que se quisiera) de un lingüista, sino por la propia opinión pública mexicana en cuanto tiene de interpretable y a partir de la discusión entre varios intelectuales mexicanos interesados en ella. Que eso suceda en el caso del DEM sólo revela una característica radicalmente social de los diccionarios en general, no claramente vista con anterioridad, sobre todo porque la historia de la lexicografía se ha hecho más en la práctica que en la reflexión acerca de ella misma, y de manera independiente de la ciencia del lenguaje, que suele soslayarla como ilegítima disciplina ancilar, y por lo tanto ignorarla.

El diseño general de la obra concreta y específica las finalidades antes explicadas, y es desde entonces cuando interviene un conjunto de postulados teóricos, de

recursos de método y de procedimientos de trabajo, que paso a exponer brevemente a continuación.

### *El planteamiento sociolingüístico*

Tratándose, entonces, de elaborar un diccionario integral y original (no una refundición más del *DRAE*) de la lengua española en su realidad mexicana, era necesario poder reconocer el vocabulario de uso general en México. Es decir, había que definir lo que realmente es “español general”, al menos en su variante mexicana. Y como no se conoce ninguna obra que muestre la composición de ese “español general” en México o en cualquier otro país hispanohablante,<sup>2</sup> había que establecer los criterios que nos permitieran encontrarlo y reconocerlo. Para ello optamos por construir un *Corpus del español mexicano contemporáneo* (en adelante *CEMC*) guiados por dos criterios principales: el primero, de carácter sociolingüístico, que nos determinaba el tipo de textos que debíamos coleccionar para com-

<sup>2</sup>Es creencia generalizada y ha sido práctica común de los autores de diccionarios de regionalismos considerar que el *DRAE* manifiesta “el español general”; de acuerdo con esa creencia, bastaría revisarlo para entresacar “lo general” eliminando todo aquello que incluya y no se use en México —regionalismos de otros países, por ejemplo— y rectificando algunos de sus artículos, que contuvieran errores acerca de usos mexicanos de la lengua. Pero el *DRAE* es una obra selectiva, a-histórica, metropolitana y literaria, y el vocabulario que contiene no corresponde ni a toda la lengua española en su historia (la Academia prepara, por eso, un verdadero diccionario histórico), ni al español contemporáneo de alguna región hispánica (no se puede considerar que sea, por ejemplo, un diccionario del español de la Península Ibérica), ni mucho menos a una lengua considerada en su verdadera amplitud geográfica y social. De ahí que no haya ninguna manera de conocer lo que es verdaderamente general en el español moderno.

poner una muestra adecuada a la gran variedad de usos sociales del español en México, a sus diferentes registros y estilos y a sus diferencias regionales internas; el segundo, de carácter estadístico, que nos definía los procedimientos cuantitativos mediante los cuales podríamos reconocer la extensión de uso de los vocablos en los diferentes niveles arriba mencionados (Lara/Ham, 1974). El criterio sociolingüístico se estableció tomando como punto de partida la teoría de la lengua estándar (culto o de cultura, según las traducciones que se consulten y según el autor correspondiente) formulada por la escuela de Praga (Hávránek, Mathesius, Vachek, Garvin). El estadístico, sobre la base de dos experiencias fundamentales: la de H. Kucera y W. Nelson Francis (1970) en su *Computational Analysis of Present Day American English* y la de I. Rosengren (1971) en Suecia.<sup>3</sup>

De acuerdo con la teoría de la *lengua estándar*, definimos como perteneciente a ella todo uso de la lengua que fuera general en México, de carácter urbano y de valor irradiador para el resto de las regiones de México. La lengua estándar se subdividió en *lengua culta* y *sub-culta*; la primera es aquella que tiene un vocabulario intelectualizado y rico, es más escrita que oral, tiene una sintaxis rica y vale como modelo de corrección para los mexicanos; la segunda, en cambio, tiene un vocabulario menos intelectualizado, usos sintácticos limitados y no sólo no vale como modelo de corrección, sino que llega a desviarse de él. Después consideramos como parte de la *lengua no-estándar* todo uso de carácter limitado, ya fuera geográfica (dialectos del

<sup>3</sup>Estos criterios están explicados en Lara/Ham, 1974.

español mexicano) o socialmente (sociolectos), y de poca fuerza irradiadora al resto del país. El vocabulario dialectal se definió como no intelectualizado pero rico en su diversidad, y se supuso que los dialectos existentes podrían tener algunas peculiaridades sintácticas propias, así como sus propios modelos de corrección. De manera paralela, el vocabulario sociolectal se definió como limitado (jergas y argots) y sujeto a modas, y se supuso que la sintaxis de estos sociolectos sería más pobre que en los niveles anteriores. La lengua no-estándar sería más hablada que escrita.

Justificados por la experiencia de la escuela de Praga, los rasgos característicos de los niveles de lengua establecidos no correspondían, de todas maneras, a una realidad que ya hubiera sido empíricamente comprobada en el español de México, por lo que se consideraron hipotéticos, sujetos a la comprobación, pero de todas maneras rectores de nuestra selección de materiales.

El cuadro de niveles de lengua establecido pasó a sustanciarse con una colección de textos reales, escritos y hablados, seleccionados a su vez a partir de una suposición acerca de los diferentes "géneros" de texto que conforman el uso cotidiano del español en México. La muestra de lengua culta quedó compuesta por textos literarios, científicos, periodísticos, técnicos, políticos, y religiosos, así como por transcripciones de grabaciones de habla culta de la ciudad de México;<sup>4</sup> la

<sup>4</sup>Cada uno de esos géneros recoge, a su vez, diferencias más precisas; por ejemplo, los textos literarios se dividieron en novelísticos, ensayísticos, etc.; los periodísticos se clasificaron como repor-

de lengua sub-culta se formó con muestras de literatura "popular" y de conversaciones grabadas en distintos puntos de la ciudad de México;<sup>5</sup> para reunir los textos dialectales mexicanos nos basamos en selecciones de grabaciones hechas en todo el país para el proyecto del *Atlas lingüístico de México* (Lope Blanch, en prensa); hicimos nosotros mismos algunas grabaciones de conversaciones jergales. La selección de textos de lengua culta se hizo con la opinión de especialistas en cada materia considerada.

### *Los resultados de la investigación*

Analizado el *CEMC* mediante un analizador automático del español que construimos también exprefeso (Lara/Ham/García Hidalgo, 1979), obtuvimos 1 891 045 ocurrencias de palabras, que se reducen a 64 183 tipos estadísticos. Para cada ocurrencia recuperamos la fuente de la que procede (con página y renglón en ella) y los índices estadísticos obtenidos nos proveen información acerca de su frecuencia total y en cada "género", relativa entre los géneros y en cada género, su dispersión en todo el *CEMC* y una medida normalizada de dispersión que sirve de base a la decisión de si cierto vocablo es o no es general en

---

tajes de política, de nota roja, de cultura, etc. El habla culta de la ciudad de México, que parece ser representativa de cualquier habla culta en el país, ya estaba documentada en un estudio anterior al nuestro. Cf. Lara/Ham, 1974.

<sup>5</sup>Por literatura popular entendimos aquella que tiene una función de entretenimiento y de evasión en las sociedades industriales contemporáneas, llámese "Groschen-roman" en la comunidad de lengua alemana, "literatura de cordel" en varios países hispanoamericanos, etcétera.

México.<sup>6</sup> Es decir, obtuvimos, en primera instancia, un conjunto de concordancias de palabras, que se prestan al estudio sociolingüístico y semántico, y un diccionario cuantitativo, útil también para el estudio sociolingüístico.

### *La sociolingüística del DEM*

Como el objetivo central de nuestra investigación es elaborar un diccionario, no hemos hecho los estudios de conjunto detallados que requiere la comprobación de nuestras hipótesis sociolingüísticas. En cambio, sí nos hemos podido dar cuenta de algunos resultados interesantes en este sentido.

### *Niveles de la lengua estándar*

En primer lugar, todo hace suponer que la distinción dentro de la lengua estándar entre culta y sub-culta no se sostiene a base de nuestros resultados cuantitativos. No es que los materiales que documentan la segunda hayan resultado más pobres en términos de rendimiento cuantitativo que los de la primera,<sup>7</sup> pero sí resultaron más pobres en términos de exclusividad de su uso. Dicho más sencillamente: encontramos menos vocablos típicamente “sub-cultos” de los que esperaba-

<sup>6</sup>Se ha tratado ampliamente el aspecto estadístico aquí mismo, en “La cuantificación en el diccionario” y en “Caracterización del corpus”.

<sup>7</sup>Todo lo contrario. Si se considera que la lengua culta forma 65.6% de todo nuestro CEMC y, en cambio la sub-culta sólo 12.3% de aquél, el rendimiento cuantitativo, considerado como la relación que hay entre el número de ocurrencias obtenidas en cada nivel sobre el número correspondiente de tipos en ellos hace que nuestro corpus de lengua sub-culta haya sido más rendidor que el de lengua culta (7.314 frente a 4.327 por ciento).

mos, como *corretear* ‘perseguir a alguien corriendo tras él, correr de un lado a otro’, que se documenta en los géneros del habla “media” de la ciudad de México y en varias regiones dialectales, *acatarrar(se)* ‘resfriarse, y molestar a alguien’, documentada en habla “media” de la ciudad de México, *pepenar* ‘recoger algún objeto, y alcanzar a atrapar algo’, documentada en el mismo género, aunque con igual proporción que en lengua culta, o *rajarse* ‘acobardarse, desdecirse’, cuyos resultados cuantitativos revelan con claridad tanto la extensión popular de su uso como su adopción en la literatura mexicana.<sup>8</sup> Por el contrario, la mayor parte del vocabulario encontrado en nuestras fuentes “sub-cultas” revela un léxico muy extendido entre ambos niveles: por un lado, se registra persistentemente un vocabulario muy literario entre los textos que considerábamos “literatura popular”. Así por ejemplo, vocablos como *adorar*<sup>9</sup> (e2 = 67.53, e1 = 24.67, e3 = 7.79), *amar* (e2 = 66.33, e1 = 32.16, e3 = 1.50), *ausente* (e2 = 60.00, e1 = 37.5, e3 = 2.50), *hermosa* (e2 = 55.91, e1 = 35.48, e3 = 8.60), y *prenda* (e2 = 66.67, e1 = 19.30, e3 = 14.03), son todos ellos cultos y literarios, pero su uso predomina en los textos sub-cultos. Por el otro lado, se hace notable la capacidad de la lengua culta, tanto escrita como hablada, para inte-

<sup>8</sup>Su dispersión va de la literatura (e = 54.54) a la literatura popular (e = 9.09), el habla popular de la ciudad de México (e = 9.09), los textos dialectales (e = 18.18) y las jergas (e = 9.09), cf. nota siguiente.

<sup>9</sup>Los datos cuantitativos siguientes se refieren a la medida de la frecuencia relativa entre géneros (e; el índice 1 indica lengua culta, 2 la sub-culta y 3 la no-estándar), que detecta mayores usos de vocablos en un género que en otros; es decir, que si una palabra se usa más en un género que en otro u otros, este índice lo revela.

grar vocabulario coloquial o popular (y dialectal) en sus textos y discursos, con lo cual la distinción entre dos niveles de la lengua estándar, el culto y el sub-culto, se ve cuestionada.

Esto puede significar que, o bien nuestras fuentes no estuvieron bien seleccionadas en relación con las características que dimos a la lengua sub-culta en nuestra hipótesis (y, entonces, habría que ir a buscarlas sobre todo en grabaciones de conversaciones familiares espontáneas, que procedieran de mejores fuentes que las de los estudios del habla de la ciudad de México, pero no en la literatura popular investigada), o bien que no es posible distinguir claramente un nivel “sub-culto” en el español de México. Por ahora hemos considerado que, en efecto, los textos recopilados para ese nivel deben reagruparse como pertenecientes a géneros de la misma lengua culta. Y a la vez, nos parece muy probable que lo que suceda sea precisamente que la lengua culta —literaria— sea más flexible, cercana al habla popular y expresiva de ella que la que utilizan textos supuestamente dirigidos al pueblo, como la literatura de cordel, las telenovelas, etc., lo cual no deja de haber sido señalado desde hace muchos años como una característica tradicional de la lengua española en su conjunto. A ello hay que agregar que todo parece indicar que el criterio que debía haber privado en la selección de textos para registrar esos dos niveles de la lengua estándar es el de *lengua hablada/lengua escrita* (que habíamos definido como secundario), pues el vocabulario que hemos identificado como general a todo el país, urbano, irradiador, pero divergente de los modelos de corrección se escribe poco y predomina en la expresión oral.

Así por ejemplo, palabras como *tatemar* ‘quemar’, que aparece con igual frecuencia en la literatura y en las jergas, *latirle (algo a uno)* ‘tener uno la intuición o la premonición de que algo habrá de suceder o de ser de cierta manera’, que se reparte por casi todos los géneros de lengua hablada, *sangrón* ‘pretencioso, molesto, malhumorado’, que ocurre en la literatura popular y en los textos dialectales (hablados), o *amolar* ‘causarle a alguien algún daño o contratiempo’, que predomina en todos los géneros hablados, se nos revelan en el análisis de sus concordancias, en donde podemos tomar en cuenta su procedencia específica y el contexto en que aparecen, como vocablos de amplio uso, sobre todo hablado, eludidos por el discurso o el texto culto —a menos de que se trate de novelas mexicanas— y a menudo censurados como “incorrectos”.

Si una designación como “sub-culto” para esa clase de vocabulario es poco afortunada,<sup>10</sup> su definición a partir de los criterios de la Escuela de Praga acerca de lo que es la lengua estándar podría obviar la designación y llevarnos a escoger otros términos que la sustituyeran, como “popular”, “coloquial”, “familiar” y hasta “vulgar” (no en su sentido peyorativo). Pero fuere cual fuere la etiqueta que escogiéramos, no dejaría de arrojar una sombra valorativa y dañina de carácter social sobre el vocabulario que calificara,<sup>11</sup>

<sup>10</sup>Desde poco después de haber terminado de componer el CEMC nos dimos cuenta de que la etiqueta de “sub-culto” para esa clase de géneros de textos era poco afortunada. Su única conveniencia era que quedaba en la misma serie que la etiqueta “culto”.

<sup>11</sup>Esa clase de calificaciones ha sido tradicional en lexicografía, como se puede ver, por ejemplo, en María Moliner, *Diccionario de uso del español*, 1970, p. xxiii: “se multiplican las indicaciones

con lo que había que esperar hasta que pudiéramos ampliar nuestra experiencia en el tratamiento de esa clase de vocablos para intentar definirlos mejor.

Tomando en cuenta eso, más el objetivo del *DEM*, de reivindicar para los mexicanos la legitimidad de sus propios usos lingüísticos, y, sobre todo, los resultados obtenidos, que no revelan en ese aspecto una estratificación social del vocabulario sino, por el contrario, el uso extendido, predominantemente oral, de vocablos que tradicionalmente se han oído entre el pueblo no educado, optamos por incluir en el diccionario, como vocablos de la lengua estándar (y, en consecuencia, sin marca de uso social) todos aquellos que en los resultados cuantitativos y en el análisis contextual particular muestren una dispersión amplia entre las fuentes específicas de la lengua culta —sobre todo la literaria— y las fuentes designadas como “sub-cultas” o “dialec-tales”. Con ello el *DEM* produce un efecto extraño —por nuevo— en sus lectores mexicanos: encontrarán entre sus páginas cientos de palabras, de giros, de sentidos característicos en sus medios sociales y fami-

---

relativas a la amplitud de uso de las palabras, a su valoración lingüística y social y a los matices intencionales o afectivos que las acompañan, tales como “brusco”, “científico”, o “del escrito”, “soez, grosero, inconveniente, popular, rural, coloquial, informal, refinado, solemne, humorístico, irónico, jocosos, jocosamente refinado, jocosamente culto, arcaísmo usado jocosamente, ñoño” y cualquier otra indicación semejante que se considere oportuna en cada caso”. Precisamente el abandono de esas apreciaciones sociales del vocabulario jugó un papel determinante en la ola de protestas que despertó la tercera edición del *American Dictionary of the English Language* (el famoso Webster), de la editorial Merriam-Webster en 1962. Cf. Sledd, J. y W.R. Ebbitt, *Dictionaries and that Dictionary*, Scott Foresman & Co., Illinois, 1962.

liares que, por primera vez, no aparecen censurados como incorrectos o divergentes del ideal normativo.<sup>12</sup>

No sucede lo mismo con todo el vocabulario proveniente de esas fuentes coloquiales, populares y habladas. Hay muchos vocablos que, si bien dispersos entre la lengua culta y la hablada, predominan en ésta y marcan estilos característicos del pueblo en general (no de zonas dialectales particulares) a pesar de que la literatura los asimile —pues en muchos casos se trata de reproducciones o de citas del habla popular— y para ellos hemos decidido utilizar dos marcas de estilo: la de *coloquial* y la de *popular* precisamente. Con ellas, lo que señala el DEM es la existencia de *estilos* o de *registros* de la lengua estándar, de los que disponen todos los mexicanos para ciertas necesidades expresivas, y que no marcan una diferencia de estrato social. Quien utiliza voces como *achichinle* ‘ayudante servil’, *cuete* ‘borracho, borrachera’, *jaguas!* ‘¡cuidado!, ¡peligro!’, *mano* ‘amigo, compañero’, *esculcar* ‘inspeccionar algún recipiente con objetos, buscando alguno de éstos en particular’, *magullar* ‘dañar algún objeto delicado, como la fruta’, *chivearse*

<sup>12</sup>Tal efecto se hace notar cada vez que un lector mexicano reflexiona acerca del diccionario. Sobre la primerísima versión reducida de nuestro trabajo, el *Diccionario fundamental del español de México* (1982), en el diario *Excelsior* (10 de febrero de 1983), S. Salinas Álvarez escribió: “Cuando un diccionario tiene la intención de que “el lector tenga siempre la sensación de que su lengua no le ha sido robada ni se le devuelve como un objeto ajeno e impuesto”, entonces uno comienza a interesarse por él como instrumento de consulta y como evidencia histórica.” Cita después algunos ejemplos en este sentido: “En la definición del verbo “andar”, por ejemplo, la tercera acepción registra “andar a gatas” [y continúa con:] cantar el gallo, ir de gallo, llevar gallo a la novia, en menos que canta un gallo, pelar gallo, dormírsele a uno el gallo, etc.”, y

‘intimidarse, sonrojarse’<sup>13</sup> no queda marcado como perteneciente a algún estrato social bajo (para hacerlo, por lo general hacen falta más señales: entonación, cierta fonética, aspecto físico, vestido, zona de la ciudad en donde vive, educación, etc.), sino que sólo está utilizando un estilo *coloquial* generalizado, suficientemente aceptado en cualquier estrato social, pero divergente de lo que suele considerarse como *culto* o *correcto* (no falta, por supuesto, quien juzgue altamente incorrecto y hasta vulgar —en sentido peyorativo— el uso de estos vocablos en círculos sociales amplios).

La marca de *popular* se diferencia de la anterior en cuanto su desviación de lo “correcto” es mayor y logra despertar reacciones de molestia o de enojo entre hablantes que han recibido una educación tradicionalmente burguesa (y en esa medida, en cuanto la movilidad social de México es acelerada y las diferencias de clase burguesas se borran, para sustituirse por diferencias de nivel económico, es frecuente la utilización de este estilo *popular* en medios sociales amplios). Es el caso de vocablos como *jeringar* ‘molestar’, documentado en el habla popular de la ciudad de México, el

---

concluye que lo más importante del diccionario es que sistematiza “la expresión lingüística elemental de la sociedad mexicana sin despreñar las palabras y las acepciones que usualmente se han despreñado por ‘ser de mal gusto, vulgares, o propias de gente corriente’, pero que en realidad son de uso general y patrimonio colectivo”.

<sup>13</sup> *Achichinle* y *magullar* sólo se documentaron en los textos científicos!; *cuete* se registra en casi todos los géneros hablados y en la literatura; *mano* está disperso por todo el CEMC, pero predomina en lengua hablada; *esculcar* apareció en textos populares, y lo mismo *chivearse* y *jaguas!*.

adverbio *harto* (salvo en la expresión *estar harto*), *muina* 'enojo', *pedo* 'borracho' y *puta*.<sup>14</sup>

No obstante lo anterior, con todo lo objetivos que tratan de ser nuestros procedimientos de definición de los estilos de habla referidos, la frontera entre lo *coloquial* y lo *popular* es a menudo indiscernible sobre la base de los datos obtenidos, pues como se ve, vocablos marcados con uno y otro calificativo suelen encontrarse en las mismas fuentes; de ahí nuestra impresión de que no hay géneros característicos de lo *coloquial* y lo *popular* y que, en cuanto posible género, lo que habría que definir es el campo de lo *hablado* frente a lo *escrito* en lengua estándar. En todos esos casos, siempre de acuerdo con los objetivos del DEM y con el hecho de que intersubjetivamente hay lugar para distinguir esos dos estilos, preferimos utilizar más la marca menos restrictiva, que es la de *coloquial* y consecuentemente reducimos el número de vocablos o acepciones marcados como *populares*.

Una tercera marca de estilo dentro de la lengua estándar es la de *grosero*, que se ha manifestado de características muy diferentes a las de las dos anteriores, pues aquéllas se refieren a verdaderos estilos del habla, que se expresan en toda la articulación del discurso, desde el vocabulario y su sentido hasta ciertos usos sintácticos y morfológicos;<sup>15</sup> ésta, en cambio,

<sup>14</sup>*Jeringar* sólo se documenta en habla popular de la ciudad de México; *harto* se extiende por todos los géneros de lengua hablada; *muina* se concentra en las jergas y el habla popular de la ciudad de México; *pedo* aparece en la literatura mexicana, la literatura popular, las jergas y el habla popular de la ciudad de México; *puta* se usa en la literatura mexicana, en las jergas y en el habla popular de la ciudad de México.

<sup>15</sup>Por ejemplo, el predominio de la terminación -s en las segundas

corresponde sólo al uso de vocablos aislados, que se insertan en el discurso para ofender o insultar a alguien o, en ciertos círculos sociales —por ejemplo, el habla de hombres y ahora, de muchas mujeres— para indicar la rudeza “característica de la hombría” y despertar así la solidaridad de los circundantes o para reivindicar el trato igualitario de ambos sexos.<sup>16</sup> Las palabras marcadas como *groseras* son muy generales en todo México y no reciben socialmente una sanción normativa lingüística; es decir, no se juzgan como “correctas” o “incorrectas” como las otras, sino que su sanción ocurre en otro ámbito de la vida social, que es el de la ofensa y el insulto. Así, voces como *chingar* y sus derivados, *cabrón* y *ojete*, por ejemplo, forman parte de la lengua estándar, se hablan más de lo que se escriben, se insertan tanto en estilos coloquiales como populares, pero siempre, aun expresadas en aislamiento, producen el mismo efecto sobre sus oyentes.<sup>17</sup>

### *La lengua no-estándar*

La parte del *CEMC* que documenta lengua no-estándar corresponde en su totalidad a transcripciones de dis-

---

personas del singular del pretérito de indicativo (*cantastes*), la frecuencia con que se usan los diminutivos, la partícula exhortativa *-le* (“¡pásele usted marchantita!”), la pluralización del pronombre de objeto directo al hacerlo concordar con el número del complemento indirecto (“se los dije” para “lo dije a ustedes”), etcétera.

<sup>16</sup>Naturalmente, también la entonación y la intensidad de la emisión producen el mismo efecto, sólo que éstos no son materia de un diccionario.

<sup>17</sup>*Chingar* y *cabrón* se documentan de manera predominante en textos hablados de carácter popular, pero también aparecen en la literatura mexicana; *ojete* es menos frecuente que los anteriores, pero tiene la misma dispersión.

cursos hablados. El género de mayor tamaño en esta parte fue el de textos dialectales, es decir, el que privilegia la variante geográfica sobre la sociolectal. A base de los datos reunidos, sobre 13 095 tipos estadísticos encontrados de palabras, lo que hemos podido comprobar es que la fragmentación que dimos a los textos que integraron nuestro *CEMC* produjo la misma fragmentación de resultados. Es decir, los 130 textos reunidos en este género se repartieron entre las 16 zonas dialectales de México<sup>18</sup> y, dadas sus características particulares, el resultado fue una muestra suficientemente rendidora cuantitativamente<sup>19</sup> pero, tan fragmentada, que a menudo impide reconocer isoglosas léxicas, pues los registros obtenidos suelen ser muy aleatorios. Así por ejemplo, la oposición dialectal entre *birote* y *bolillo* —cierto tipo de pan—, que muchos mexicanos experimentan al viajar por el interior del país no se puede confirmar con nuestros datos, pues *birote* sólo se documentó en La Paz, Baja California,<sup>20</sup> aun cuando sabemos que se extiende probablemente

<sup>18</sup>Las zonas dialectales habían sido definidas de manera provisional por Lope Blanch (1971): de acuerdo con su estudio preliminar de los materiales del *Atlas lingüístico de México*, se podían reconocer 16 zonas. De cada una de ellas elegimos transcripciones de grabaciones correspondientes a sus “centros” y sus “márgenes”, proporcionalmente repartidas entre hombres y mujeres, de tres generaciones, de tres niveles de educación y posición social, y de oficios diversos; así, a cada zona correspondieron aproximadamente ocho textos.

<sup>19</sup>El rendimiento cuantitativo del *CEMC* en su totalidad fue de 3.324%; es decir, por cada cien ocurrencias de palabras resultaron tres tipos léxicos distintos. El rendimiento de nuestro género de textos dialectales fue de 5.058%, lo cual indica su mayor “riqueza” relativa.

<sup>20</sup>Ha estudiado con más detenimiento esta parte de nuestro *CEMC* Luz Fernández (1987).

desde Aguascalientes hacia el norte; lo mismo sucede con la pareja *calentón/calentador*, de la que *calentón* se documentó en Durango, pero nuestra experiencia nos dice que corresponde a la zona que Lope Blanch (1971) define como “Noroeste” (es decir, se extiende desde Baja California y Sonora hasta Nayarit y Durango); *enfadar* contrapuesto a *aburrir* se ve ligeramente mejor registrado en Sonora y Nayarit, pero faltan datos sobre su verdadera extensión en el Noroeste. Acerca de los otros vocablos de estas parejas, baste señalar que se documentan suficientemente como propios de la lengua estándar. A base de algunos otros datos se ha podido suponer que ciertos vocablos reflejan la existencia de isoglosas, como en los casos de *cayuco* —una pequeña embarcación hecha de un tronco de árbol— documentado en Ciudad del Carmen, Campeche y por mí mismo en Veracruz; *candela* ‘fuego’, que aparece en Chiapas, Campeche, Tabasco y Veracruz, y *costurar* ‘coser’, registrado por nuestro CEMC en Oaxaca y Veracruz, pero documentado también en Chiapas. La polisemia contradictoria característica de *buscar* ‘buscar y encontrar’ en Yucatán se registró en textos yucatecos y campechanos. Con esos registros y varios vocablos más, creemos poder reconocer la existencia de las regiones dialectales que Lope Blanch (1971) denominó “veracruzano meridional”, “yucateco-campechano”, “chiapaneco” y “tabasqueño”, pero unidas por el carácter transicional de Tabasco en una región mayor, costera en gran parte, que se podría llamar, según se acostumbra en México, “Sureste”, si no fuera porque buena parte del “chiapaneco” es de rasgos centroamericanos, aunque Tabasco también forma la transición entre él y los dialec-

tos mencionados. Nuestro *CEMC* también nos ha permitido hacer algunas confirmaciones dialectológicas ya anotadas en investigaciones anteriores, como la fluctuación del género de palabras como *calor* y *mar*,<sup>21</sup> pero, de nuevo, no es posible reconocer su verdadera extensión geográfica.

Las observaciones anteriores nos han llevado a concluir que el *CEMC* sólo tiene valor de índice para el registro de voces regionales mexicanas, pero que no basta para delimitar con certeza sus extensiones geográficas. De ahí que, al marcar los artículos lexicográficos con designaciones de regiones lo hagamos de tres maneras diferentes: a) introduciendo en el propio texto de la definición los estados de la República en los que se registró el vocablo o la acepción en cuestión, con lo que sólo asentamos un hecho, pero no afirmamos que sólo en esos lugares tenga uso el vocablo; b) colocando antes de la definición los nombres de las zonas dialectales en que se recogió el vocablo o la acepción, con lo cual nos atrevemos a afirmar que se usan seguramente en ellas; c) colocando en ese mismo lugar las marcas: *norte, sureste, Veracruz y Altiplano*, que señalan una extensión regional mayor que las de las zonas dialectales definidas. Naturalmente, en posteriores revisiones del *DEM* que puedan basarse en mejores datos, habrán de hacerse las correcciones necesarias.

<sup>21</sup>La *calor* se ha documentado en cuatro puntos del centro del país, a diferencia de *el calor* que se ha convertido en parte de la lengua estándar; *la mar* parece ser característica de regiones con tradición marinera: desde Veracruz hasta Campeche, en el golfo de México, y en la costa de Sonora, en el mar de Cortés.

### *Los sociolectos*

Nuestro *CEMC* reunió solamente materiales de las jergas juveniles (llamadas, en distintas épocas de las que estudia nuestro trabajo, “pachuco”, “lenguaje preparatorio” y “lenguaje de la onda”) y del argot del hampa, llamado tradicionalmente *germanía* y en México *caló*;<sup>22</sup> no encontramos, cuando elaboramos nuestros materiales, registros de jergas profesionales, como la de los médicos y las enfermeras, la de los albañiles, la de los mecánicos, etc. y no disponíamos del tiempo ni de los medios para realizar nosotros mismos esas investigaciones.<sup>23</sup> El rendimiento cuantitativo de los materiales coleccionados fue notablemente superior al de otros géneros de nuestra muestra (13.520%), por lo cual creemos haber podido documentar suficientemente al menos el núcleo de las jergas coleccionadas. Así por ejemplo, *greña* ‘cabellera’, *tronar* ‘fumar’, *onda*,<sup>24</sup> *movida* —cualquier acción—, *alivianarse* ‘ponerse en acción, liberarse de algo’ y

<sup>22</sup>Acerca de este último, véase mi trabajo “El caló revisitado”, en el homenaje a J.M. Lope Blanch, que prepara el Instituto de Investigaciones Filológicas de la UNAM, México.

<sup>23</sup>En general, al construir nuestro *CEMC* partimos de la idea de aprovechar, cuando se trataba de lengua hablada, ante todo, materiales registrados por otros investigadores, pues el costo de las investigaciones de campo y el tiempo que lleva realizarlas son muy grandes. Sólo investigamos directamente el caló, por su interés para la conformación de la lengua popular. Deseamos, en el futuro, y si disponemos de los medios necesarios, llevar a cabo una amplia recolección de materiales de esta clase, para mejorar radicalmente nuestros resultados.

<sup>24</sup>*Onda* es el vocablo representativo de la jerga juvenil desde la década de 1960. Sus significados varían mucho, como puede verse en las siguientes expresiones: ¿*qué onda?* ‘¿qué pasa?’, ¿*qué buena*

*aplatanarse* ‘apaciguarse, caer en la pasividad, agüitarse (pachuco y rural)’ aparecieron claramente registradas en los textos respectivos.<sup>25</sup> Del vocabulario caló, más de cien palabras quedaron fidedignamente registradas. Las marcas lexicográficas para esta clase de vocablos son: *juvenil*, *pachuco*, *de la onda* y *caló*.

### *Una conclusión provisional*

Como se habrá visto, la relación entre el *CEMC* y la elaboración del *DEM* es casi consustancial. Nuestro trabajo lexicográfico depende crucialmente de los datos reunidos en la exploración de nuestra muestra del español mexicano contemporáneo. En otros lugares he discutido las ventajas y desventajas de esta dependencia elegida de un corpus de datos como el nuestro (aquí mismo, en “La cuantificación” y “La cuestión de la norma”); es claro que tal dependencia, considerando los objetivos de nuestro diccionario, es más ventajosa que desventajosa, pues nos ha permitido: a) establecer de manera segura lo que es el español mexicano estándar, con lo que tiene de general (“español pan-hispánico”) y mexicano; b) reconocer suficientemente el español mexicano culto y fundamental, a base del cual se da la educación en lengua materna en nuestro país;<sup>26</sup> c) disponer de los

---

*onda de maestro!* ‘¡qué comprensivo, honrado, activo, comprometido maestro!’; *¡qué mala onda!* ‘¡qué mal, qué lástima, qué feo!’; *agarrar la onda*, ‘participar en algo, ponerse a tono, juntarse con alguien, entender algo, darse cuenta de algo’, etcétera.

<sup>25</sup>*Movida* y *onda* ya se habían extendido al resto de la lengua hablada cuando reunimos el *CEMC*, como lo muestran nuestros datos.

<sup>26</sup>Para definir el vocabulario fundamental del español de M.

datos necesarios para definir el vocabulario del español mexicano de manera original, basados exclusivamente en nuestros usos; d) reconocer los niveles de lengua y los estilos sociolingüísticos que se dan en el español mexicano; e) legitimar, sobre esas bases, los usos mexicanos del español, a diferencia de lo que ocurre en la lexicografía hispánica tradicional, y f) plantear al público la variedad, las diferencias y aun los conflictos que definen la realidad contemporánea del español mexicano.

Las lagunas que hemos encontrado en nuestro *CEMC*; los faltantes en el reconocimiento de zonas dialectales mexicanas; la necesidad de formular un nuevo *corpus* más adecuado a nuestros objetivos y más apto para el estudio sociolingüístico del español mexicano —lo que planeamos hacer en 1993— relativan cualquier pretensión de validez absoluta de nuestro trabajo, pero también muestran que la lexicografía es como la tela de Penélope, que se teje y se desteje todos los días.

Por último, la respuesta social a lo ya publicado, demasiado compleja como para poderla tratar en el

---

nos basamos en los criterios expuestos por Gougenheim (1957) y en nuestro propio análisis cuantitativo (Ham, 1981). Sobre esa base publicamos en 1982 el *Diccionario fundamental del español de México*, a cuya nomenclatura agregamos el vocabulario disponible de los libros oficiales de enseñanza elemental en México. Más tarde ampliamos ese diccionario con vocabulario de menor frecuencia que el fundamental, pero estándar, con vocabulario disponible de la enseñanza media y de diversas terminologías científicas para publicar en 1986 el *Diccionario básico del español de México* (en adelante DBEM) Cf. “La cuantificación” en donde hay una comparación de nuestra capacidad para reconocer el vocabulario general “pan-hispánico” con lo que se encuentra en otros estudios cuantitativos.

espacio limitado de este trabajo, indica que, por un lado, hemos tocado las expectativas del público mexicano y algunas de ellas han sido aceptablemente satisfechas; pero por el otro, que también hemos tocado la llaga de un largo conflicto normativo e ideológico en la idea colectiva de lo que es y debe ser el español mexicano. Es decir, el *DEM* no se despega de su raíz social.



## EL DICCIONARIO ENTRE LA TRADICIÓN Y LA REALIDAD\*

0. ¿Cuánto de la tradición es lengua y cuánto de la lengua es tradición? o, para decirlo sin el retruécano, ¿en qué medida la transmisión histórica de valores culturales, de costumbres, de experiencias, de modos de concebir y representar la realidad se logra sin lengua y sin relatos ni narraciones? y, a la vez, ¿en qué medida tiene existencia una lengua si no es precisamente por el hecho de que debe su sustancia significativa y su capacidad de significar a un dilatado proceso social e histórico de construcción de formas y de registro de experiencias valiosas para una comunidad, en el cual la tradición es la que decanta y establece lo que producen la capacidad biológica de hablar y de tener memoria, y la necesidad social de conservar inteligibles para el conjunto las experiencias parciales y pasadas de sus individuos?

La primera reflexión de todo ser humano sobre la lengua es del orden de la tradición: su lengua materna se le presenta siempre, antes que nada, como herencia del pasado, como transmisión de un hablar que lo ha precedido y que le llega ya hecho, establecido con la

\*Ponencia presentada con el título “Entre la realidad y el diccionario” en el X Coloquio de Antropología e Historia sobre “Lenguaje y tradición en México”, en El Colegio de Michoacán. Zamora, octubre de 1988

obligatoriedad que le ha impuesto la sociedad en la que hubo de nacer, y que no le da, desde el principio, casi ninguna posibilidad de elección. Para todo ser humano, su lengua materna constituye un umbral irrebasa-ble: nada puede concebirse fuera de ella, como no sea otra lengua, pero esa otra lengua, a su vez, volverá a manifestarse como la primera: un modo de hablar fijado y establecido en una tradición, que viene repitiéndose de una generación a otra, en una sucesión cuyo principio queda en el misterio del origen de la humanidad y de las sociedades.

Esa presencia casi inmediata de la lengua como umbral y como tradición para todo ser humano logra manifestarse, sin duda, gracias a la exterioridad con que se presentan a la reflexión las emisiones verbales; y esa exterioridad, o mejor digamos esa *objetividad* de la lengua, se debe a la excepcional distintividad de la materia sonora de la que está hecha la comunicación verbal. El sonido de un vocablo se hace claro y distinto desde tan temprano en la vida de un ser humano, que facilita su objetivación y con ello la reflexión a propósito de él. Pero tanto la objetivación como la reflexión acerca de la lengua no son, en la realidad, sino resultado de un proceso de socialización que comienza con el nacimiento de cada ser humano y no se detiene nunca, durante el cual la lengua se enseña (y se aprende) primero, y después se educa o se cultiva. Eso y no otra cosa quiere decir la expresión “saber hablar”, que tan felizmente introdujo en la teoría del lenguaje el lingüista rumano Eugenio Coseriu. *Saber hablar* es, en consecuencia, una técnica que se aprende —*aprender a hablar*— y una educación de esa técnica, normada por la sociedad y fijada por la tradición.

Así, emitir cadenas sonoras con un sentido inteligible para la sociedad, hablar y saber hablar, son dos aspectos de la lengua indisociables para cualquier persona, cuyas repercusiones en la vida social y en el sentido de la lengua para la sociedad deben, sin duda, considerarse más dilatadamente de como se ha hecho en la teoría del lenguaje y, en general, en la teoría antropológica. Si bien no es ese el objeto de este trabajo, su importancia para el tema que he de tratar es tan grande, que es necesario referirme a él con alguna prolijidad. Para lo uno y lo otro, la tradición es el componente de partida: se aprende a hablar tal como hablan los demás y como han hablado nuestros padres; se reflexiona acerca de la lengua tal como lo hace nuestra sociedad y sobre la base de los criterios que ella misma ha elaborado a lo largo de la historia y de su desarrollo cultural.

La tradición parece ser entonces una transmisión en dos canales: por un lado, la transmisión de la técnica para hablar una lengua específica: la de nuestra comunidad lingüística; por el otro la transmisión de una cultura elaborada sobre esa técnica. La tradición se muestra, en consecuencia, como verdadera matriz del saber hablar, pero también, en cuanto resulta de la reflexión y de la cultura, como sustentadora de ciertos valores sociales que, a lo largo de la historia, han ido determinando la idea que tiene la sociedad acerca de su propia lengua, lo cual es un hecho necesariamente segundo en relación con el hablar, por más que no pueda disociarse de él, pues para la lengua, como en general para toda acción humana, no hay prácticas espontáneas, sino siempre prácticas cultivadas, socialmente establecidas en mayor o en menor medida. Sin

embargo, ese carácter segundo de la idea de la lengua que la tradición sustenta arroja sobre ella la sombra de la ideología; es decir, la sombra de la posibilidad de que la cultura de la lengua que se transmite por la tradición sea víctima de una articulación global de interpretaciones que, en un momento dado, resulten falsas o encubridoras en cuanto al verdadero sentido de la tradición o en cuanto a otros sentidos dignos de ser reivindicados o defendidos en un momento determinado de la vida de la comunidad lingüística. Es así como llega a suceder que, al destacar el papel de la tradición en la existencia de una lengua, lo que se esté destacando sea, más bien, una manera de comprender esa tradición que, además, pretenda determinar todas las perspectivas y la propia realidad de la lengua (cuando la tradición se subraya en este sentido, es cuando se cae en el tradicionalismo).

Probablemente sea la lexicografía, entre las disciplinas de la lingüística, de las que más se presten para reconocer el papel de la tradición en la sustancia real de la lengua y de su correspondiente saber hablar. Pues, por un lado, la materia de trabajo de la lexicografía: el léxico, es, de los elementos que componen una lengua, uno de los más fáciles de objetivar, debido no solamente a la distintividad sonora de la palabra—reforzada por la escritura— sino también al hecho de que es en la palabra en donde se concretan las relaciones de la lengua con el mundo, con la realidad, con el sentido, que es al fin y al cabo lo que interesa al ser humano común cuando piensa en la lengua; por el otro lado, precisamente porque el producto de la lexicografía: el diccionario, materializa el léxico y por eso mismo se convierte en un vehículo privilegiado de la

reflexión y de la interpretación de la manera en que se articula la lengua con la sociedad, con su historia y con su tradición.

La lexicografía se ha debatido siempre en esos dos campos del saber hablar, y de ese debate provienen su valor para el establecimiento de una lengua como técnica del hablar y como institución social, y sus dificultades cada vez que muestra alguna discrepancia entre la idea de la lengua que tiene la sociedad y la realidad de esa lengua. Pues por lo general se da por sentado que los diccionarios son catálogos que definen la manera *correcta* de escribir las palabras, de pronunciarlas —en el caso de los diccionarios del español esto es menos importante—, de utilizarlas en cierto medio social y con cierto estilo, y de comprenderlas. Es decir, se da por sentado que los diccionarios estipulan la técnica del habla; pero como esa técnica se ve interpretada por la imagen social de la lengua, toda estipulación lexicográfica contraria a la imagen compartida por los miembros de la sociedad tiene por efecto una reacción social en contra del diccionario.

Una teoría del diccionario, que todavía no ha sido objeto de elaboración por los lexicógrafos ni por los lingüistas,<sup>1</sup> tiene que comenzar por plantearse el fenómeno de la representación de la lengua que se da en los diccionarios monolingües y tiene que asumir como

<sup>1</sup>He iniciado la construcción de una teoría del diccionario en ese sentido en un libro aun no terminado: *La complejidad semántica del diccionario*, pero ofrecí algunos elementos de ella en “El objeto diccionario”, ponencia presentada en septiembre de 1987 durante el Simposio de Lexicografía de la Asociación Mexicana de Lingüística Aplicada celebrado en el Instituto de Investigaciones Filológicas de la UNAM.

cuestión de partida la aceptación social de la representación que ofrece el diccionario. Es probable que tal fenómeno ocurra gracias a la combinación de los factores a los que me he referido antes: el carácter privilegiado del vocablo (escrito) para desencadenar una objetivación de la lengua; el carácter de la transmisión social que opera el diccionario al ofrecer a cada generación una memoria del vocabulario de la lengua recibida de sus padres, y el propio carácter de memoria colectiva de las experiencias pasadas, que tiene el léxico a diferencia de los otros elementos de la lengua. A reserva de que en el futuro podamos los lexicógrafos ofrecer una explicación convincente y global de esos fenómenos, pretendo fundar en los rudimentos que tengo de ella una explicación de la relación que guarda el *Diccionario del español de México* con la realidad de nuestra lengua, tomando en cuenta el papel conflictivo que desempeña en ella la tradición.

1. Ha sido tradicional tanto para México como para el resto del mundo hispánico considerar que la lengua española tiene su centro en nuestra vieja metrópoli imperial: España y su capital Madrid. Todavía hoy, la idea de la lengua que comparte la mayoría de nuestros pueblos hispanohablantes es que sólo España puede ofrecer un punto de referencia valedero en la determinación de la corrección al hablar. Lo vemos a diario, por ejemplo, en la actividad de los traductores, siempre atentos a los usos madrileños y siempre cuidadosos de no caer en regionalismos (sobre todo entre los que traducen libros científicos y técnicos, que se apegan rígidamente y muchas veces sin criterio a lo estipulado por el *Diccionario de la Lengua*, de la Academia Espa-

ñola —en lo sucesivo, *DRAE*—). Lo vemos también en algunas decisiones extrañas que a veces toman los periódicos (como cuando *unomásuno* optó por comenzar a escribir *Sájara* en vez de *Sahara*) o las cadenas de televisión (como ahora que los locutores deportivos dicen a sus entrevistados: “*Le* vi meter un gran gol”, en vez de “*Lo* vi meter un gran gol”). Pero a pesar de esa idea la realidad del español en México, en América, y hasta en España misma, es diferente de lo que estipula el *DRAE* y de lo que cree mucha gente; y, no sólo eso, sino que también obedece claramente a tradiciones lingüísticas relativamente distintas, forjadas a lo largo de los siglos, por eso mismo socialmente instauradas entre nosotros y consecuentemente más válidas para nuestra comunidad que las madrileñas de la Academia.

Se suele reconocer que nuestra variedad de la lengua española tiene sus propias tradiciones; así, se reconoce la base andaluza de nuestro español, como se manifiesta, por ejemplo, en los vocablos *alcayata*, *alfajor*, *atarjea*, *cobija*, *chícharo*, *frazada*, o *panteón*.<sup>2</sup> Se reconoce también el supuesto “arcaísmo” del español americano,<sup>3</sup> manifiesto en la supervivencia de voces como *lindo* (en España ‘hermoso’), *liviano* ‘ligero’, *candela* ‘fuego, lum-

<sup>2</sup>Me restrinjo a dar ejemplos del léxico por ser ese el objetivo primordial de un diccionario; un panorama breve y bien documentado del andalucismo americano (y mexicano) se puede leer en la *Historia de la lengua española* de Rafael Lapesa (Madrid, ahora en Gredos; hay muchas ediciones; las últimas son más completas en bibliografía y comentarios útiles) y en el reciente libro de José Moreno de Alba, *El español en América*, FCE, México, 1988.

<sup>3</sup>Cuestión que Lope Blanch ha dejado, diría yo, definitivamente aclarada en su “El supuesto arcaísmo del español americano”,

bre', *esculcar* 'registrar', *recordar* 'despertar', *pararse* 'ponerse de pie', *prieto* 'moreno', etc. Se resalta el papel que juegan en el español mexicano las palabras procedentes de nuestras lenguas amerindias, hasta el grado de que hay filólogos y aficionados siempre dispuestos a demostrar etimologías prehispánicas en buena parte de nuestro vocabulario.<sup>4</sup> Tal reconocimiento se ha vertido, tradicionalmente, en un buen número de recolecciones de vocabulario y de diccionarios de regionalismos (o provincialismos). Todos ellos forman el tronco de la tradición lexicográfica mexicana, de manera paralela a lo que sucede en el resto de los países hispanoamericanos, que han seguido las mismas pautas; a ellos debemos valiosos registros de vocabulario regional, vigente todavía, ya desaparecido o en vías de desaparecer. Sin embargo, lo que determina la existencia de esa lexicografía es, al mismo tiempo, el deseo contradictorio de atajar la desviación que el uso de tal vocabulario implica, en relación con la idea de que la lengua de las dos Castillas es "prototipo y norma del castellano de todas las demás regiones" —como sostenía, por ejemplo, en 1909 don Manuel G. Revilla.<sup>5</sup> El efecto de ese doble sentido de la lexicografía mexicana tradicional (e hispanoamericana,

---

ahora reunido con otros estudios suyos en *Estudios sobre el español de México*, UNAM, México, 1983, pp.33-53.

<sup>4</sup>No doy ejemplos de indigenismos, pero sí de algunos intentos de proponer etimologías amerindias a voces españolas, como en *acotar*, que para Suárez (*El español en Yucatán*) proviene del maya *kotah* 'cerca, valla'; el interesante caso de *ate*, que por tantos motivos se presta a considerarlo nahuatlismo y es, por el contrario, un catalanismo existente ya en español hacia 1525. Cf. J.M.Lope Blanch, "Un falso nahuatlismo", NRFH, 27 (1978), pp. 296-298.

<sup>5</sup>En "Provincialismos de expresión en Méjico; cuáles son los aceptables y cuáles los viciosos", *Memorias de la Academia Mexi-*

naturalmente), ha sido, más que el verdadero enriquecimiento de nuestra idea de la lengua en México y de nuestra cultura, la constante afirmación de que nuestro modo de hablar se encuentra siempre al filo del barbarismo y del vicio.<sup>6</sup> De esa concepción marginada del español de México, así como del resto del español americano, se ha venido consolidando, en nuestra educación lingüística, la idea de la lengua imperante entre nosotros: nuestro español, así responda a nuestra tradición, así gocemos su variedad y su riqueza, no deja de ser una desviación del peninsular.

Como dije antes, el diccionario siempre tiene un efecto objetivador de la lengua y fijador de sus características, por lo que fácilmente se convierte en una manifestación de la memoria colectiva depositada en el vocabulario y fijada por el consenso social. Pero por eso mismo, porque el diccionario representa el consen-

---

*cana*, 6 (1910), pp. 353-367 [reimpresión facsimilar de 1975].

<sup>6</sup>Como lo demuestran varios títulos de colecciones de vocablos tradicionales mexicanos: Feliz Ramos i Duarte, *Diccionario de mejicanismos. Colección de locuciones i frases viciosas*, Imp. de E. Dublán, Méjico, 1895; Prudencio Patrón Peniche, *Léxico yucateco. Barbarismos, provincialismos y mayismos*, Tenoxtitlán, México, 1932; Aurelio de León, *Barbarismos comunes en México*, Porrúa, 1937 y varios más. Ramos i Duarte reúne indistintamente “barbarismos” con “regionalismos”: junto con *abajo* ‘debajo’ (cf. *DBEM* s.v.), *abarrotos* ‘tienda de víveres’, *abonarse* ‘suscribirse’ (cf. *DBEM* s.v.), *accidente* ‘escabrosidad del terreno’ (cf. *DBEM* s.v.), y unos cuantos posibles “barbarismos”, recoge cientos de regionalismos comunes. Patrón Peniche y De León separan claramente los “barbarismos” como *aceptación* ‘tipo de letra de cambio’, *afectar* ‘expropiar’ (cf. *DBEM* s.v.), *carátula* ‘portada y parte anterior del reloj’ (cf. *DBEM* s.v.), *imponerse* ‘acostumbrarse’, *lonche* ‘emparedado’, *adjuntar* ‘remitir adjunto’, *aliviado* ‘curado, sano’ (cf. *DBEM* *aliviar*), etc. de los regionalismos.

so de la sociedad, no es difícil que pase a caracterizar, junto con otras instituciones, como la religión manifiesta en la Biblia y el derecho codificado en la Constitución, la identidad de la sociedad. Tal función simbólica del diccionario es fácilmente comprobable a partir de la observación de la manera en que se concibe el diccionario en la comunidad hispánica. Baste recordar cuántas discusiones y cuántos debates se zanján (o se vician) recurriendo al diccionario y asumiendo como ciertos los datos que ofrece; cuántos tratados de derecho internacional, cuántos amparos, cuántos juicios se resuelven acudiendo a la autoridad implícitamente reconocida al diccionario. También podemos comprobar el valor del diccionario para una identidad social como la nuestra por su lado negativo, si consideramos la animadversión que despierta en toda Hispanoamérica el diccionario de la Academia Española. Cualquier error del *DRAE*, cualquier etimología equivocada (sobre todo si es de un indigenismo), cualquier marca de uso regional mal determinada, cualquier definición contradictoria con un significado familiar para nosotros da lugar, no solamente a airadas y agrias censuras al *DRAE*, sino a obras completas de “apostillas”, de “perlas”, de “disparates de la Academia”, de “correcciones y adiciones” al diccionario. La burla, la complementación bien intencionada y crítica, el odio o la pura desestimación del diccionario académico son síntomas del valor que tiene ese diccionario para el reconocimiento de nuestra identidad social. Todo ello indica el papel crucial del diccionario y la medida en que una obra compuesta en la Península Ibérica y por una Academia central y autoritaria parece traicionar el sentimiento de identidad que tenemos los mexi-

canos. Tampoco ese efecto es extraño en la historia de la lexicografía: dio lugar, por ejemplo, al esfuerzo de Noah Webster, a principios del siglo pasado, de ofrecer, al recién independizado pueblo de los Estados Unidos de América un “diccionario americano del inglés”,<sup>7</sup> un *American Dictionary of the English Language* (1828) en el que la lengua inglesa se manifestara también como “lengua nacional” de su país; ha dado lugar, posteriormente, a nuestro trabajo y al apenas iniciado diccionario del francés de Quebec.<sup>8</sup>

El papel del diccionario académico corresponde, por lo tanto, al doble sentido de nuestra tradición lexicográfica: por un lado es el diccionario de “nuestra lengua” (y el único, pues la lexicografía española, hasta la fecha, no ha creado diccionarios originales, que sean verdaderamente autónomos en relación con la Academia Española), concebida a la manera de Revilla, como “prototipo y norma” del saber hablar español. Por el otro es el odiado perpetuador de nuestra situación colonial, que traiciona nuestro senti-

<sup>7</sup>No conozco una historia amplia acerca de los motivos de Noah Webster para elaborar su diccionario; pero la semblanza de su personalidad con que se inicia la tercera edición del diccionario Merriam-Webster (1961) indica que de su “patriotism and nationalism...came the conviction that lusty young America needed its own school books, its own uniform language, and its own intellectual life”, así como que “his dictionaries were suggested partly by his resentment against the ignorance concerning American institutions shown in contemporary British dictionaries” (cf. también algunos comentarios en James Root Hulbert, *Dictionaries: British and American*, A. Deutsch, Londres, 1968, pp. 25-28).

<sup>8</sup>Cf. C. Poirier, “Les avenues de la lexicographie québécoise”, en L. Boivert et al., *La lexicographie québécoise. Bilan et perspectives*, Laval, 1986, pp.269-285.

miento de la lengua y nos impone un sentimiento etnocéntrico ajeno a nuestra propia identidad. Correspondientemente, nuestra tradición lexicográfica tiene también ambos sentidos: complementa o suplementa al diccionario académico, añadiendo nuestros regionalismos y corrigiendo en buena medida lo contenido en el *DRAE* que se refiere a nuestra región; a la vez, destaca el vocabulario que nos resulta querido para legitimar o fundamentar nuestra identidad nacional, aunque nunca deja de señalar, en el vocabulario que recolecta, nuestro desvío respecto del español pretendidamente correcto.

En ese contexto es en donde se explica y se justifica nuestro trabajo para el *Diccionario del español de México*. En relación con las tradiciones lexicográfica y de idea de la lengua el *DEM* es producto de su crítica y se muestra como un rompimiento con ellas, pues se plantea, por primera vez en el ámbito hispánico, la elaboración de un diccionario integral del español y no de un nuevo diccionario de mexicanismos, lo que equivale, en términos lexicográficos, a abandonar los métodos contrastivos de la lexicografía regionalista para pasar a los métodos de la lexicografía integral. En términos de la idea de la lengua, al tomar el español en su totalidad, centrado, por así decirlo, en México, reivindica el papel del español en la base de nuestra identidad social y contradice la idea generalizada de que la variedad lingüística de un país como México sólo puede ser una variante marginal, a lo más simpática y pintoresca, de una lengua metropolitana e imperial como el español de la Academia madrileña. En términos, por último, de la tradición de nuestra variedad lingüística, se propone como otra manera de

reconocerla, basada en el estudio de su realidad y críticamente deslindada de la idea predominante de la lengua.

2. Tal ruptura con la tradición lexicográfica mexicana e hispanoamericana y con la idea tradicional de nuestra lengua requiere, naturalmente, de un reconocimiento amplio y riguroso de la realidad léxica del español mexicano, que nos dé los elementos necesarios para construir un diccionario completo y no un acervo de vocabulario particular y muchas veces marginal, como se suele encontrar en los diccionarios de mexicanismos que se han hecho desde el siglo pasado. Pues si bien el vocabulario tradicionalmente registrado por esos diccionarios, que se refiere, por ejemplo, a la variedad de la fauna y la flora de México, a los platillos de la comida mexicana en cada región cultural del país, a las culturas rancheras y campesinas, al habla coloquial de cada zona dialectal, o al acervo heredado de los nahuas o de los mayas, sin olvidar el vocabulario grosero y picaresco, forma parte indudable de nuestra realidad lingüística, su definición dependiente de las normas académicas y el uso limitado que tiene en relación con la lengua estándar que utilizamos los mexicanos diariamente, lo caracteriza como un conjunto de voces marginales en la comunicación cotidiana, y marginadas por la idea de la lengua que predomina entre nosotros.

Construir un diccionario integral del español, tal como lo hablamos en México implica, desde el punto de vista del método lexicográfico, contar con un registro amplio y suficiente de la realidad de nuestro léxico, que nos permita reconocer su verdadero uso dentro de

los límites de nuestras fronteras o, mejor dicho, dentro de los límites de la comunidad de comunicación que da su cohesión a la sociedad mexicana. El método de trabajo de la lexicografía tradicional no resulta adecuado para ese objetivo, precisamente en cuanto es un método que busca peculiaridades, independientemente de la extensión de uso que tengan; es decir, es un método de “diccionario-tesoro”, como lo caracterizaba don Ramón Menéndez Pidal. Por lo contrario, lo que se necesita para poder escribir un diccionario integral es un registro de lo que más se usa en una lengua. De ahí que sólo hubiera dos procedimientos posibles: o bien, como alguien propuso al inicio de nuestro trabajo, tomábamos como base el *DRAE* y marcábamos en él, de acuerdo con nuestra propia experiencia del vocabulario, las palabras que considerásemos efectivamente usuales en México, para basar en ellas nuestro trabajo, o hacíamos uso de alguno de los métodos cuantitativos que la lexicología venía aplicando desde dos décadas atrás para ese efecto. El primer método, aunque parece expedito, nos enfrentaría de inmediato con la limitación de nuestro conocimiento del vocabulario —pues al fin y al cabo los lingüistas somos hablantes también— y con el efecto ideológico que necesariamente causa en nosotros la idea de la lengua en que hemos sido educados. Lo que podríamos producir como resultado sería una deforme imagen de nuestro léxico, que sólo correspondiera al juicio de nuestro grupo de trabajo y que, en consecuencia, se prestara a justificadas críticas por parte de quien tuviera una idea diferente de nuestro vocabulario, en todo comparable con la nuestra. El segundo, en cambio, nos permitiría recoger vocabulario en México sin

que interviniera nuestro juicio y nos referiría, con números, a una realidad objetiva de nuestra lengua. Por esa razón optamos por un método cuantitativo para recoger nuestros materiales, que dio origen al *Corpus del español mexicano contemporáneo (1921-1974)* —en lo sucesivo *CEMC*— y a los resultados contenidos en nuestro “diccionario estadístico”.<sup>9</sup>

Los resultados obtenidos del estudio cuantitativo de nuestro *CEMC*, compuesto por 996 textos de dos mil ocurrencias de palabras cada uno, provenientes de una amplia variedad de obras y de grabaciones (exposiciones detalladas de estos aspectos se encuentran en los trabajos citados en la n. 9) nos han permitido documentar 64 183 tipos estadísticos de palabras, que se reducen a cerca de cuarenta mil vocablos del español mexicano. Con esos datos podemos afirmar que el vocabulario fundamental de nuestra lengua nacional ha quedado objetivamente reconocido, así como que hemos podido documentar suficientemente cerca de quince mil vocablos usuales en México. Para otros diez mil, nuestros datos nos ofrecen suficientes pistas acerca de su uso y la extensión territorial en la que aparecen, en tanto que otros quince o veinte mil más nos ofrecen registros sin valor cuantitativo, pero importantes desde el punto de vista de los métodos tra-

<sup>9</sup>He publicado varios trabajos en que se explica la estructura del *CEMC*, las fórmulas estadísticas empleadas y las ventajas y limitaciones de los resultados: L.F. Lara, R. Ham e I. García Hidalgo, *Investigaciones lingüísticas en lexicografía*, El Colegio de México, 1981; L.F. Lara, “Características del *CEMC*”, en *Actas del Ier. Congreso internacional sobre el español de América*, San Juan, Puerto Rico, 1987; L.F. Lara, “La cuantificación en el *DEM*”, aquí mismo; L.F. Lara, “Caracterización metódica del *CEMC*”, también aquí mismo.

dicionales de “diccionarios-tesoro”.<sup>10</sup> Son estos resultados los que constituyen la base documental del *DEM*; con ellos nos hemos podido formar una idea clara y detallada de la realidad de nuestra lengua y, en consecuencia, hemos podido comenzar a reconocer nuestra tradición léxica con métodos objetivos y verificables, no deformados por la idea de la lengua en la que hemos sido educados. Naturalmente, todo estudio cuantitativo tiene sus limitaciones, que derivan del carácter aleatorio de la recolección de materiales, por más que se cuide que la muestra de textos de los que proceden sea representativa de los usos que se pretende reconocer. Por ello, si bien es posible afirmar que ningún corpus de datos lingüísticos es capaz de ofrecer un reflejo único de la realidad de una lengua, también lo es que, como método de trabajo, es el más objetivo y el menos condicionado por ideas recibidas a propósito de las características del vocabulario de una lengua. Dicho de otra manera, un corpus como el nuestro nos ofrece una imagen cierta de la realidad de nuestra lengua, aunque no sea la única imagen posible que pueda obtenerse y aunque la idea tradicional de la lengua se vea contradicha.

### 3. Para poder ilustrar ese reconocimiento de nuestra realidad lingüística actual en términos de la tradición

<sup>10</sup>Las características de nuestros resultados, basadas en el reconocimiento automático del vocabulario, no nos permitirán ofrecer resultados exactos *por vocablos* sino hasta que terminemos la redacción del *DEM*. Por ello doy cifras aproximadas, que corresponden a una estimación basada en los resultados exactos *por tipos de palabra*: 1 131 vocablos componen el vocabulario fundamental; 17 903 tipos en nuestro *CEMC* tienen frecuencia  $F \geq 5$ ; 12 551 tienen  $4 \geq F > 1$ ; 29 729 tipos tienen  $F = 1$ .

de la lengua y de nuestra tradición lexicográfica es necesario establecer una base de comparaciones adecuada. No existe ningún corpus de datos del español mexicano anterior y comparable con el nuestro, sobre el cual podamos basarnos; los diccionarios de regionalismos, desde el *Vocabulario de mexicanismos* de García Icazbalceta hasta el *Diccionario de mejicanismos* de Francisco J. Santamaría son, como dije antes, “diccionarios-tesoro” que documentan el vocabulario mexicano pero no dejan ver, con suficiente claridad, la vitalidad real de ese vocabulario ni la idea que tenía de él la sociedad mexicana desde mediados del siglo pasado hasta mediados del presente. Por ese motivo, he tomado como punto de referencia un interesante trabajo de Manuel G. Revilla (cf. n. 5), presentado ante la Academia Mexicana en 1909, con ocasión de la visita a México de un académico madrileño, en el cual se encuentra una larga lista de vocablos propios del español mexicano —a juicio de Revilla, por supuesto—, agrupados sobre la base de una idea de la lengua muy característica de nuestra tradición lexicográfica. No sé cuánta vitalidad tendrían en 1909, pero lo importante es que para Revilla manifestaban una peculiaridad de nuestra variedad lingüística, correspondiente a nuestra tradición cultural. Junto con esa lista de vocablos reunida por Revilla, he seleccionado algunos más de dos obras típicas de la lexicografía del regionalismo y el barbarismo: los diccionarios de Feliz Ramos i Duarte y de Prudencio Patrón Peniche. El efecto del contraste entre esa lista “calificada” de vocablos de uso en México y los resultados obtenidos en el estudio de nuestro *CEMC* deberá servir para confirmar, por un lado, la permanencia de nuestra tradición léxica real,

junto con el necesario cambio que se opera en la lengua al paso del tiempo, y por el otro, para mostrar hasta qué punto la idea de la lengua que tenían esos autores y que quizá compartimos nosotros suele abandonar rápidamente la realidad de la que parte para instaurarse como verdadera ideología de la lengua, que tiende a darle su propia homogeneidad, su propia “geometría” a la lengua, aunque contradiga su realidad.

Nuestro *CEMC*<sup>11</sup> nos confirma los usos mexicanos de los andalucismos *alcayata* (0.1329), *cobija* (0.4339) y *chícharo* (0.2350) frente a vocablos hispánicos de menos o de ningún uso en México, como los correspondientes *escarpia* (0.00), *frazada* (0.1489), y *guisante* (0.1738). De los “barbarismos” anotados por Ramos i Duarte y Patrón Peniche (cf. n.6), *abajo* como ‘debajo’, *abarrotos* (0.7940), *accidente* como ‘sinuosidad del terreno’, *aceptación*, *adjuntar* por ‘remitir adjunto’ y *aliviado* como ‘curado’ se muestran ya como parte común del uso mexicano y han perdido notoriedad como barbarismos; de los nahuatlismos anotados por Revilla (cf. n. 5) se conservan en uso, por ejemplo, *chiquihuite* (0.2435), aunque posiblemente en retirada frente a *cesto* (0.8519); *cajete* (0.6083), que es la única forma documentada para el *barreño* que ofrece Revilla

<sup>11</sup>Las cifras que siguen a cada vocablo citado corresponden al “índice normalizado de dispersión” con que analizamos cuantitativamente los resultados de nuestro *CEMC*. Ese índice relaciona la frecuencia de uso del vocablo con su dispersión en diferentes géneros de textos y con la ponderación que recibió cada género en la estructura del *CEMC*. Cuando el índice se acerca a 1, el vocablo en cuestión es más frecuente, tiene mejor dispersión y proviene de géneros de mayor importancia para el reconocimiento del español mexicano estándar; cuando tiende a cero, señala lo contrario. Explicaciones más precisas se encuentran en los trabajos citados en n. 9.

como peninsular; *tompiate*, que sobrevive ampliamente (0.6083) frente a *capacho* (que no documentamos en el CEMC); *olote* (0.5582) parece haber vencido a *carozo*; *milpa* (0.3570) se ha vuelto rural frente al culto *maizal* (0.6083); *malacate* (0.6083) predomina sobre *cabrestante* (0.0); *tlecuil* (0.1117) quizá está ya en vías de desaparición frente a *fogón* (0.6082); *achichincla* (0.6083) parece ya no tener rival en *paniaguado* (0.0); *chochocol*, *jehuite*, *tlalpiloya* y *calpixque*, en cambio, parecen haber desaparecido. Entre los vocablos que hace casi un siglo Revilla mostraba en competencia, *boleto* y *billete* se han especializado para designar diferentes tipos de comprobantes (ahora con el reto anglicista de *ticket*); lo mismo *luneta* y *butaca*; *chispar* (0.6277) se ha vuelto más familiar y rural que *zafar* (0.8966); *betabel* (0.8702) es la forma más usual en México frente a *remolacha* (0.6083) que aparece en los libros de texto gratuitos y en obras técnicas de la agricultura; se sigue manifestando nuestra preferencia regional por *recamarera* (0.2718) y *recámara* (0.6438) frente a *doncella* (0.2223) y *alcoba* (0.3623); *banqueta* (0.5160) sigue siendo más usual, desde hace siglos, que *acera* (0.4864), a pesar de la influencia del estilo culto en favor de ésta y no de aquella; *col* (0.3609) se usa más que *repollo* (0.1276); *amarrar* (0.7335) ya forma parte de nuestro vocabulario básico frente al culto y peninsular *atar* (0.6376); en cambio, el culto *aturdir* (0.3143) ha superado a nuestro coloquial *atarantar* (0.0805). *Fierro* (0.7567) es más familiar y designa más objetos de diversos metales, que *hierro* (0.5885), el nombre del elemento; *mercar* (0.2634) se ha reducido a los usos rurales frente a *comprar* (0.7826). Algunos usos mexicanos que nos son entrañables, siguen fuertes, como

*pararse* (0.6743) en vez de *ponerse de pie*, *pronunciamiento* (0.3337) frente a *alzamiento* (0.1489), *saco* (0.6434) y no *americana* (0.1117), *conectar* (0.4988) y no la perífrasis *ligar dos cosas*; *timbre* (0.6979) en vez de *estampilla* (0.1577); *fungir* (0.5405) como *hacer las veces*, y hasta el *susodicho* (0.1329) del lenguaje legalista, en vez del *antes citado*. Lo mismo en este grupo de vocablos, hay algunos que han desaparecido junto con la época y la moda a la que pertenecían: ya no documentamos *señorito*, sino sólo *joven* (0.8975); *flux* y *paletó* han desaparecido en nuestro CEMC; *calzada* (0.5775) ha cedido objetos que designar a *carretera* (0.7314), ya no digamos *camino real* que sólo se conserva como recuerdo de otros tiempos.

Como se habrá podido ver, aun dentro de la limitación del puñado de ejemplos ofrecido, la realidad de nuestra lengua es heterogénea: voces que hace un siglo parecían barbarismos forman parte hoy de nuestro español hasta el grado de que casi nadie podría calificarlas como lo hicieron esos tres iniciadores; los nahuatlismos, que son tan importantes para nuestra ideología nacionalista están desapareciendo poco a poco<sup>12</sup> y de ninguna manera caracterizan centralmente nuestro español; ciertos andalucismos y “arcaísmos”,

<sup>12</sup>Desde que J.M. Lope Blanch publicó su libro *El léxico indígena en el español de México*, El Colegio de México, 1969, basado en encuestas practicadas entre hablantes educados de la ciudad de México, el componente de indigenismos en nuestra lengua había bajado notablemente, hasta situarse en 0.07% del vocabulario de esos hablantes. No faltaron censuras a este resultado de Lope Blanch. Con métodos cuantitativos radicalmente diferentes, en nuestro CEMC hemos encontrado que el componente de indigenismos corresponde a 0.083% del total del corpus, lo que resulta sorpren-

así como muchas otras voces españolas que por diversas circunstancias arraigaron en México mejor que otras se conservan, aunque las hay que se especializaron para designar diferentes objetos, las hay que cedieron frente a vocablos más generales y más cultos, y otras han ido desapareciendo junto con la época de la que formaron parte. Se puede afirmar, con poco temor a equivocarse, que la tradición objetiva de una lengua es así de heterogénea y que suele ser el resultado de tendencias lingüísticas muchas veces contrapuestas; si por un lado pervive la tradición andalucista de nuestro vocabulario, por el otro hay una fuerza unificadora con el español académico, producto de la comunicación internacional entre los países hispanohablantes; si se han conservado en México voces que resultan arcaicas para España, también han aparecido entre nosotros neologismos de origen francés o inglés, traídos por la influencia cultural europea o norteamericana desde el siglo pasado; si nuestros indigenismos son tan valiosos para la determinación de nuestra identidad nacional, la lengua culta y urbana tiende a olvidarlos y éstos a conservarse en el mundo rural o en el habla coloquial. El diccionario de regionalismos se desentiende, por su naturaleza, de esa heterogeneidad, en cuanto priva en él la necesidad de documentar el

---

dente, dada la diferencia de los métodos empleados. Naturalmente, si uno busca cuidadosa y exhaustivamente más indigenismos, los va a encontrar, tanto en ciertos temas, como la botánica y la cocina tradicional, como en ciertas zonas, como todas aquéllas en donde sobreviven pueblos indígenas. Sin embargo, la aportación indigenista a nuestro español mexicano moderno es muy reducida, a pesar del deseo ideológico de que sea mayor. En este punto, el enfrentamiento entre la realidad y la idea predominante de la lengua es muy claro.

mayor número posible de vocablos diferentes de los que registra la Academia Española. Un diccionario integral, como el nuestro, tiene por objetivo, al contrario, reflejar la heterogeneidad y las contraposiciones que se dan en la realidad, sobre la base del vocabulario que verdaderamente se conserva en uso, independientemente de su procedencia o de su valor para la idea predominante de nuestra lengua. Diccionario integral y diccionario de regionalismos no se contraponen, sino que se complementan para documentar una cultura de la lengua amplia y rica; es más bien la idea de la lengua que se ha venido creando sobre la base de la lexicografía tradicional la que parece buscar una evaluación unilateral de nuestro léxico, que resulte acorde con determinados valores ideológicos, sea el indigenismo, sea el coloquialismo y el pintoresquismo de muchas formas de hablar, sea el anti-imperialismo, o sea el anti-anglicismo.

Un diccionario puede escribirse para apoyar una creencia de esa clase; como señalaba Morínigo hace muchos años, “los diccionarios de americanismos actuales rivalizan en incorporar a su léxico el mayor número de indigenismos, se usen o no se usen en el español de América, distorsionando de esta manera la realidad lingüística y confundiendo a los estudiosos”;<sup>13</sup> lo mismo, puede uno escribir sabrosos tratados sobre voces populares, picarescas o groseras, como el *Tumbaburros de la picardía mexicana* de Alberto Jiménez, o *El Chingolés* de José María de Usandizaga.

<sup>13</sup>M.A. Morínigo, “La penetración de los indigenismos en el español”, *Presente y futuro de la lengua española*, Madrid, 1963, t. II, p. 226, apud Lope Blanch, *El léxico indígena en el español de México*, El Colegio de México, 1969, pp.17-18.

Todos ellos —quizá a excepción del *Chingolés*— dan valiosa información sobre esos aspectos de nuestro vocabulario, pero no agotan la realidad de nuestra lengua, ni definen sus principales características. Son, más bien, refuerzos de la idea de la lengua tradicional, incapaz de centrar el español común y usual como verdadera realidad y como verdadera raíz de nuestra identidad.

Pero también hay que reconocer que el léxico de una lengua como la española es mayor que los treinta o cuarenta mil vocablos que hemos documentado para nuestro *DEM*; es aún mayor que los setenta u ochenta mil vocablos registrados por el *DRAE* y seguramente semejante a los cientos de miles que se encuentran, para su lengua, en diccionarios del inglés como el *New English Dictionary* de Oxford, el *American Dictionary of the English Language*, de Merriam-Webster, y el todavía en preparación *Dictionary of American Regional English* de la universidad de Wisconsin, aunque tampoco éstos podrán agotar la riqueza léxica del inglés. Es decir, un diccionario, por exhaustivo y rico que sea, siempre recoge menos vocabulario que el que existe en la realidad. Eso es lo que hace a la lexicografía una disciplina acumulativa y lo que no invalida trabajos anteriores, sino que los destaca en cuanto han aportado datos valiosos para la posteridad. En el caso de México, nuestra lexicografía, con ser relativamente joven, ya nos ha dejado más de una docena de obras de importancia, sobre las cuales descuella el *Diccionario de mejicanismos* de don Francisco J. Santamaría. Tratándose de la lexicografía hispánica en general, las obras que documentan variedades regionales superan a lo que nos ofrece el *DRAE*; sólo que en uno y otro caso

lo que nos falta es una lexicografía más moderna, capaz de volverse a acercar a la realidad de la lengua y de mirarla con rigor y con sentido crítico, no con la lente de la vieja relación entre la metrópoli y sus colonias, ni con la del gusto exclusivo por el pintoresquismo y el marginalismo. Lo que deseamos poder ofrecer con el *DEM* es una primera contribución en ese sentido: buscamos una definición *regional* y no regionalista de nuestro español, autónoma en relación con el español de los demás países hispanohablantes, pero no aislada de él; centrada en nuestro uso de la lengua y ligada a nuestra identidad nacional, que reivindique nuestras verdaderas tradiciones verbales y no que se someta a la idea de la lengua que, por diferentes circunstancias, se vino volviendo predominante en nuestra educación. Todo ello es, sin duda, una toma de posición frente a nuestras tradiciones: la de la realidad de la lengua, la de nuestra idea de ella y la de nuestra lexicografía.

## LA CUESTIÓN DE LA NORMA EN EL *DICCIONARIO DEL ESPAÑOL DE MÉXICO*

La lexicografía de la lengua española se caracteriza de manera determinante por el tratamiento que da a los vocablos originarios o usuales de las diversas regiones hispanohablantes del mundo, como México, Cuba, Chile o aun Andalucía, en contraposición con el que da a las palabras consideradas “generales”, es decir usuales en todas partes y relativamente originarias de la cepa castellana. Los diccionarios dedicados a las particularidades regionales del español forman la gran corriente lexicográfica de los llamados “diccionarios de regionalismos”,<sup>1</sup> en la cual el objetivo central es el acopio y la explicación de voces, sea procedentes de las lenguas aborígenes de un territorio particular —en México, por ejemplo, los “diccionarios de aztequismos”— o sea de base hispánica, pero claramente modificadas en su significante, en su significado o en su

<sup>1</sup>Hay diccionarios de esta clase en todos los países de lengua española y en muchas de sus regiones. Un panorama general de sus contenidos y sus métodos puede verse en Haensch y Werner 1978, Haensch 1989, Werner 1979, y Sala 1982. Es evidente la diferencia entre los diccionarios de regionalismos elaborados por lexicógrafos prácticos y aficionados, y los que han hecho lingüistas. La mayor parte de mis afirmaciones se refieren a los primeros, que constituyen la inmensa mayoría de estas obras.

uso por los hablantes de alguna región particular: los diccionarios de americanismos, de mexicanismos, de cubanismos, o también de zonas más reducidas, como los de andalucismos, o los de sonorensismos o de sinaloensismos en México, dedicados a vocablos originarios o usuales en los estados federales de Sonora y Sinaloa, respectivamente.

Los diccionarios dedicados, en cambio, al español “general”, suponen la existencia por lo menos de un nivel de lengua en el cual las particularidades regionales se superan y se alcanza un léxico válido y comprensible para ya casi 400 millones de hablantes de español en el mundo entero. Esta segunda corriente de la lexicografía española está orientada, en realidad, por una sola obra: el *Diccionario de la lengua española*, que publica la Real Academia Española de Madrid desde hace casi 250 años (en lo sucesivo abreviado como *DRAE*). El resto de los diccionarios “generales” del español que se encuentran en el mercado lo forman fundamentalmente refundiciones del *DRAE*, unas veces complementadas con éxito y otras más bien simplificadas, de acuerdo con los objetivos de venta de las editoriales.

Esta distinción entre lexicografía de regionalismos y lexicografía general del español sería muy razonable desde el punto de vista del método (contrastivo o diferencial, como señalan Haensch y Werner 1978) y de los objetivos (descripción exhaustiva de las variedades del español como muestra de su riqueza, frente a elaboración de un verdadero español general o “internacional” que garantizara la comunicación) si, por un lado, los diccionarios generales del español tuvieran una base documental en la que todas las regiones

hispanohablantes se vieran adecuadamente representadas y, a partir de ella, se pudiera afirmar, con relativa seguridad, que ciertos vocablos son realmente generales, desde la península ibérica y el suroeste de los Estados Unidos de América hasta el estrecho de Magallanes. Por el otro lado, si los diccionarios de regionalismos contaran también con bases documentales igualmente comparables y fidedignas, que permitieran asegurar el carácter estrictamente regional de otros vocablos.

La realidad es que ambas clases de diccionarios no cuentan con esas bases documentales. Los diccionarios de regionalismos se enfrentan permanentemente con una lamentable falta de datos, que a su vez es resultado del poco estudio que se ha dedicado a la lengua española en todos sus territorios. De ahí que el procedimiento de localizar el origen y el uso de un vocablo, sobre todo si se trata de palabras con poco colorido aborígen, se vuelva muchas veces imposible; y a la vez, el tanto más difícil proceso de confirmar el carácter regionalmente exclusivo de una palabra, se ve muchas veces condenado al error, como que no cuenta con la documentación contrastiva que haría falta. La consecuencia de ello es que la mayor parte de los lexicógrafos que se dedican a los regionalismos opten por registrar vocablos cuya procedencia es fácilmente reconocible, como muchos nahuatlismos o mayismos en México, y dejen de anotar las diferencias que encuentran en palabras de base hispánica, sea por temor a equivocarse, o sea porque al menos el significante parece español y en consecuencia difícilmente será exclusivo de una sola región. Para varios lexicógrafos este procedimiento ha llegado a reducirse al registro de

palabras de base no hispánica y no incluidas en el *DRAE*, como si todo vocablo faltante en este diccionario fuera necesariamente un regionalismo.<sup>2</sup>

Los diccionarios generales se encuentran en una situación relativamente más mala que la de los de regionalismos, pues su dependencia del *DRAE* no les permite encarar el problema de la falta de datos de una manera adecuada a sus finalidades. Y el *DRAE*, por su parte, aunque idealmente se elabore con la ayuda de todas las academias de la lengua correspondientes en América, tampoco parece tener ni las posibilidades ni los materiales necesarios para tan enorme tarea de redactar un diccionario realmente general.<sup>3</sup> El *DRAE* es una especie de “tesoro” de la lengua al que han venido ingresando materiales de diferentes procedencias, como lo revelan sus propios artículos lexicográficos: desde elementos tomados del diccionario de Cobarruvias (1611), y del trabajo central de la Academia originaria, el *Diccionario de autoridades* (1726-1739), hasta aportaciones tomadas de muchos diccionarios de regionalismos con diferentes grados de confiabilidad, y documentos aislados de académicos que no siempre se inclinan por una realidad más o menos general, sino

<sup>2</sup>Todavía a la fecha hay quienes utilizan como base de la comparación al *DRAE*, con lo que se desvirtúan los resultados. Cf. Werner 1985 en donde se critica este procedimiento.

<sup>3</sup>A ello me refería en Lara 1978. Una recopilación cuidadosa y amplia ya que no exhaustiva, del léxico total del español desde la península ibérica hasta la Patagonia excede los medios de cualquier equipo de lexicografía único, además de las dificultades de principio que tendrían sus métodos. Para la Academia Española, tan limitada por su propia tradición, tal tarea parece imposible. Por eso he propuesto la modalidad del “diccionario regional” como posible solución, como se verá más adelante.

por sus gustos literarios particulares. En ese sentido el *DRAE* no es del todo un diccionario histórico, ni un verdadero diccionario general del español, ni un verdadero diccionario del español peninsular, sino un agregado de materiales correspondientes a todos esos intereses.

Este panorama de la lexicografía hispánica corresponde al que se puede hacer con los valores descriptivos propios de la lingüística, pues ha sido esta ciencia la que ha venido estableciendo la importancia y la necesidad de bases documentales, la observación cuidadosa del léxico por regiones y la imparcialidad en relación con viejos criterios valorativos de las variedades o dialectos del español en relación con una lengua española ideal.

Para acabar de comprender la lexicografía hispánica falta considerar precisamente ese ingrediente valorativo y normativo que seguramente la antecede, pues obedece a actitudes y criterios sociales largamente elaborados y que, no por falsos, han dejado de determinarla.

Dicho de otra manera: la distinción entre diccionarios de regionalismos y diccionarios generales no es resultado de un planteamiento metódico de la lingüística (salvo en los diccionarios modernos de regionalismos, por supuesto), sino de una concepción normativa. De acuerdo con esa concepción, el español general sigue siendo determinado por la metrópoli castellana, en tanto que los españoles regionales solamente constituyen una periferia, colorida y pintoresca, importante para diversas lealtades nacionales, pero marginal al fin. Dígalo si no la explicación que ofrece todavía la última edición del *DRAE* (1984) acerca del orden de las

acepciones en sus artículos lexicográficos: “Dentro de cada artículo van colocadas por este orden las diversas acepciones de los vocablos: primero las de uso vulgar y corriente; después las anticuadas, las familiares, las figuradas, las *provinciales e hispanoamericanas*, y por último, las técnicas y de germanía” (el subrayado es mío). En ese orden no queda sino considerar que las acepciones “de uso vulgar y corriente” se refieren a alguna entidad geográfica diferente de las provincias españolas y de Hispanoamérica. A la vez, la serie “uso vulgar y corriente” —“anticuadas”— “familiares” —“figuradas”— “provinciales” —“hispanoamericanas”— ... supone un ordenamiento del centro hacia la periferia, impidiendo que voces generales de toda Hispanoamérica, por ejemplo, lleguen a ocupar el primer lugar en cuanto a sus acepciones, en la medida en que reflejan usos que superan, literaria y demográficamente, a otros que sean generales solamente en las dos Castillas o en Madrid. Voces “anticuadas” para la metrópoli pueden ser, a menudo, voces comunes en Hispanoamérica, cuyo supuesto “arcaísmo”, tan sostenido hasta hace poco tiempo por la filología peninsular, ha sido ya suficientemente contradicho.<sup>4</sup>

Por otra parte, el *DRAE*, a diferencia de lo que ocurre con los diccionarios ingleses, y más recientemente franceses, es un diccionario selectivo, cuyo objetivo no ha sido recoger la riqueza del vocabulario hispánico, sino solamente aquellos vocablos utilizados por la lengua literaria y, de ésta, sólo los utilizados por ciertos escritores y no por todos. Ello explica su menor tamaño en comparación con otros diccionarios de

<sup>4</sup>Sobre todo por el definitivo artículo de Lope Blanch 1983.

lenguas europeas; si el *New English Dictionary* de Oxford supera el medio millón, el *DRAE* apenas sobrepasa los 70 000 artículos.

Se puede proponer una interpretación del carácter normativo que subyace a la lexicografía hispánica. Para ello es necesario profundizar un poco en la manera en que la normatividad se relaciona con la realidad de una lengua, así como en el hecho de que toda normatividad tiene un carácter simbólico social que hace falta dilucidar para poder tomar una posición en relación con ella.

Como he sostenido en otros lugares (Lara, 1976b, 1983 y 1987b), hay una diferencia total entre la realidad de una lengua, comprobable con métodos objetivos, y la idea que tiene de ella la comunidad que la habla. En tanto que la realidad de una lengua es un hecho primario, que acontece como tal, la idea social de ella es un hecho secundario, que se da como resultado de una reflexión. Es decir, entre la percepción de un hecho verbal y la elaboración de una idea acerca de él, media en el hablante un proceso de reflexión que no consiste simplemente en una versión en espejo del hecho, sino en una construcción interpretativa. Al mismo tiempo, la lengua no se presenta al hablante como una novedad absoluta; el hablante no se enfrenta a ella como si no supiera nada de su existencia y como si no tuviera experiencias previas. El hablante no considera la lengua ni como un extraterrestre que la escucha por primera vez, ni como un lingüista, cuya razón de ser está precisamente en el dominio de los medios que le permiten extrañarse de su identidad de hablante e intentar objetivarla en cuanto ella misma. Para todo ser humano la lengua es un umbral irrebasa-ble: la totalidad de su existencia se da entre prácticas

verbales y se objetiva mediante prácticas verbales.<sup>5</sup> Precisamente esa radicación de la lengua en la práctica verbal la hace un hecho social: para el hablante individual la lengua es algo que aprende de sus padres y de su posición en la sociedad; es algo que hereda como hereda el resto de las experiencias propiamente humanas.

Así que la reflexión sobre la lengua no se da como un acontecimiento aislado e individual, sino como parte de un proceso social, determinado por la experiencia histórica de la comunidad hablante. En esa experiencia, que reúne de manera extremadamente compleja acontecimientos históricos, situaciones económicas, procesos políticos, tradiciones y situaciones de crisis, la reflexión sobre la lengua cifra sus puntos de referencia, encuentra sus valores y determina las explicaciones globales con que ha de continuar concibiéndose la propia lengua.

Se crea lo que se ha venido llamando una "ideología" de la lengua. Ideología que, en la medida en que ofrece una interpretación global y forma parte de los grandes medios de conservación de la identidad de una sociedad, llega a naturalizarse y a darse como versión definitiva y real de lo que es la lengua en sí misma.

La normatividad lingüística tiene su origen en esa ideología de la lengua. Ya que ésta ha definido sus valores sociales importantes depende de la norma para echar a andar sus mecanismos de control del uso de la

<sup>5</sup>Esta concepción corresponde a la corriente lingüística contemporánea conocida como "pragmática". Ha sido el campo del psicoanálisis el que ha desarrollado de manera más interesante esta idea, como en A. Lorenzer, *Zur Begründung einer materialistischen Sozialisationstheorie*, Frankfurt, 1981.

lengua, manifiestos siempre en un sentido de la corrección idiomática.

Pero como la ideología no es una versión especular de la realidad, tampoco puede ofrecer una idea de la lengua equivalente a toda ella, sino que selecciona aquellos elementos que más fácilmente se pueden prestar a una relación simbólica con los valores sociales que contiene. En esa forma la ideología no tiene explicación que ofrecer para todos los elementos de la lengua y lo que hace, por el contrario, es extrapolar a toda la lengua lo que ha elaborado para unos cuantos de sus elementos. Al final, la ideología se impone sobre la lengua y hace creer que aquello que ella define es *toda* la lengua. La norma, que es el medio por el que actúa la ideología, se convierte en la lengua y así se presenta a todos sus hablantes.

Bajo esta concepción, en cada caso particular, como lo es la lengua española y, en este artículo, la lengua española que se habla en México, hay que investigar qué elementos ideológicos son los que se ocultan bajo las características de una tradición como la antes señalada, pues sólo así se puede llegar a comprender la difícil cuestión de la norma en un diccionario que, como el *Diccionario del español de México*, precisamente pretende ofrecer una alternativa a la lexicografía bipolar de la lengua general y el regionalismo.

Se puede argumentar que hay al menos cuatro elementos constitutivos de la ideología dominante de la lengua española de los que nacen las características normativas de su lexicografía:

1. La historia de la lengua española ha estado siempre indisolublemente ligada a la historia de la expansión de Castilla: en la península ibérica primero, y

después en la conquista y la colonización de América, las islas Filipinas y algunas regiones africanas. No solamente fue el castellano de hace diez siglos el origen del español actual, sino la relación originaria de esa lengua con el reino de Castilla la que, desde entonces, se manifestó como un valor que se fue imponiendo sobre el resto de los dialectos románicos peninsulares durante la Reconquista y la constitución del imperio español, y la que se trajo a América con el descubrimiento. Una buena muestra de esta idea es, por ejemplo, el famoso prólogo de Antonio de Nebrija a su *Gramática de la lengua castellana* (1492), en el que, tras de la afirmación de que “siempre la lengua fue compañera del imperio”, propone esa relación constituyente entre el castellano y su expansión al añadir una observación que resultó profética en la víspera del descubrimiento de América: “que después que vuestra alteza metiese debajo de su yugo muchos pueblos bárbaros y naciones de peregrinas lenguas, y con el vencimiento aquéllos tendrían necesidad de recibir las leyes que el vencedor pone al vencido, y con ellas nuestra lengua”.<sup>6</sup>

2. La historia de la colonización española de América se caracteriza, en general, por la integración de las poblaciones aborígenes a la sociedad española y a su lengua, a diferencia de las otras colonizaciones europeas, que tendieron a segregar a los aborígenes y mantenerlos sin contacto con las lenguas de los conquistadores (sin que esto quiera decir que soslayo o minimizo

<sup>6</sup>He modificado la ortografía de este pasaje, para facilitar su reproducción. A. Quilis, autor del estudio preliminar a la edición citada, parece apoyar esta idea.

la crueldad con que una y otras colonizaciones se llevaron a cabo). El español se hizo pronto una lengua popular que, junto con el rápido proceso de mestizaje, que también caracteriza a las colonias españolas, se extendió por todos los territorios conquistados por España y terminó por convertirse en lengua materna de la mayoría de sus poblaciones.

3. Si la idea de una lengua unida era central para la concepción del imperio español —como también lo deja ver Nebrija en su prólogo—, posiblemente su transmisión a América se vio reforzada por la propia experiencia de las sociedades hispanoamericanas recién creadas, para las cuales el español se convirtió en uno de los principales fundamentos de su identidad. Y si ya Nebrija argumentaba en favor de una gramática que mantuviera unida la lengua para impedir su posterior fragmentación, como sucedió a las lenguas de los imperios de la antigüedad, no es raro que precisamente dos americanos: Andrés Bello y Rufino José Cuervo, al despuntar el siglo XIX e iniciarse los procesos de independencia de las colonias hispanoamericanas, volvieran a valorar la unidad de la lengua en toda Hispanoamérica y en España, para impedir que se repitiera el ciclo de fragmentación lingüística que sufrió el latín al desmembrarse el imperio romano (Ontañón 1967). Desde entonces la unidad del español es uno de los valores supremos de nuestra idea de la lengua.

4. Desde la antigüedad clásica existe un concepto de corrección idiomática, cuyo principal objetivo es asegurar la mutua comprensión y una calidad estética socialmente apreciada. Tal concepto tenía por norma la imitación de modelos socialmente sancionados, a los que se

consideraba “autoridades”. Sin embargo, también desde entonces, de los textos ejemplares que constituían las autoridades, se pasaba a los autores de aquellos textos, dando con ello un sentido diferente a la idea de “autoridad” lingüística: ya no el texto socialmente “autorizado” sino el escritor, el autor del texto, como autoridad activa. En tanto España —como el resto de las nuevas naciones europeas— no tuvo autores literarios en los cuales pudiera basar sus normas de corrección, el papel de las autoridades tocó a los usos cortesanos: de Toledo con Alfonso el Sabio, y después de las sucesivas capitales del imperio: Valladolid y Madrid. Los usos lingüísticos de los escritores españoles de los siglos de oro (siglos XVI-XVII) y la derrota ideológica de la nobleza por el pensamiento de la Revolución Francesa fueron determinantes para que reapareciera el concepto clásico de las autoridades, pero no alcanzaron a borrar la materialización de las autoridades en los escritores —todo lo contrario: el nacimiento de la burguesía hizo del escritor una personalidad autónoma, dotada de la absoluta libertad que ofrece la especialización del trabajo—<sup>7</sup> y en el prestigio de los usos ya no de las cortes, sino de las élites intelectuales de las capitales del imperio.

De esa manera la tradición ideológica que resumía en Castilla el origen y la fuente de legitimidad de la lengua española no solamente continuó determinando el sentido de la corrección y el establecimiento de normas lingüísticas en todo el ámbito hispanohablante, sino que se vio

<sup>7</sup>Me baso para esta afirmación en ideas de Sohn-Rethel 1973 y Habermas 1962, en que se liga la división kantiana entre razón pura y razón práctica con la instauración del ciudadano como individuo privado a partir de la Revolución francesa. Esto habrá dado al autor la identidad social que todavía hoy tiene.

reforzada por una necesidad de unidad y conservación de la mutua inteligibilidad de los hablantes de territorios muy lejanos entre sí. Las actitudes de Andrés Bello y Rufino José Cuervo, americanos eminentes, contribuyeron a extender en las nuevas repúblicas americanas el temor a la fragmentación lingüística y la idea de que el sentido de la corrección debía seguir siendo determinado por la vieja metrópoli colonial.

A la vez, la unión de ese reconocimiento histórico de Castilla como origen y fundamento de la lengua, con el sentido de la “autoridad” que se había venido gestando desde la antigüedad, vino a materializarse en una institución que, por su mismo nacimiento, reunía el poder de la corte con el nuevo papel del escritor como autoridad activa: la Real Academia Española, localizada en Madrid.

Hubo intentos en las repúblicas americanas de oponerse a la existencia de esa autoridad, que equívocamente mezclaba la autoridad de los textos —como lo proponía el *Diccionario de autoridades*—<sup>8</sup> con la autoridad de los escritores que la conformaban —no todos grandes escritores y probablemente muchos de ellos poco dignos de confianza en cuanto a su uso de la lengua— y con la autoridad real; las academias correspondientes en Hispanoamérica se fueron fundando a pesar de ello, paradójicamente en momentos de exaltación republicana, liberal e independentista.<sup>9</sup>

<sup>8</sup>Autoridades 1726:ii: “Como base y fundamento de este diccionario se ha puesto los autores que han parecido a la Academia han tratado la Lengua Española con la mayor propiedad y elegancia” (También modifiqué la ortografía).

<sup>9</sup>La oposición hispanoamericana a la Academia, hasta donde llega mi información, no se dio sino hasta que se constituyeron los

No resultará sorprendente que, en consecuencia, la idea de una lexicografía del español “general” no cruce por un planteamiento previo del orden metódico lingüístico, en el que se establezca la necesidad de una comparación entre usos del español en todos los territorios en los que se habla —una idea que quizá apenas esté despuntando en la Academia Española, gracias a la entrada de los lingüistas en su seno— sino que parta de una idea de lo “general” que se basa en el valor ideológico de la lengua de Castilla, ya no más localizada siquiera en Castilla la Vieja, sino en los usos de su capital imperial: la ciudad de Madrid. Y en la medida en que la Academia de Madrid se ha arrogado, por efecto de la misma ideología, la autoridad suficiente para ser ella la que sancione lo general y lo correcto, es el español que ella concibe el que se define como general, o “de uso vulgar y corriente”. Las academias hispanoamericanas correspondientes son, por efecto de esa misma ideología, fatalmente dependientes de la española, por más que exista formalmente una asocia-

---

estados independientes. Así por ejemplo, en contra de la primera academia hispanoamericana correspondiente a la española, la colombiana, un diputado local alegó, en 1875, que tal academia “era hostil a las instituciones patrias y basad(a) en el principio de autoridad, que bajo ningún aspecto social, ni aun en el del idioma, podía conciliarse con la verdadera república” (Romero 1972:xxx). En México, la Academia Mexicana se fundó como una institución conservadora precisamente en momentos de gran exaltación republicana y liberal. Hay una historia hispanoamericana de esfuerzos por “independizarse” lingüísticamente de España, que llega desde Argentina (Domingo F. Sarmiento) hasta México (Ignacio Ramírez e Ignacio M. Altamirano). Está por hacerse una investigación bien documentada y rigurosa de esos aspectos importantes de la constitución de las identidades nacionales hispanoamericanas.

ción de academias y unas y otras proclamen su igualdad. Basta considerar la realidad de las academias hispanoamericanas para comprobar que tal igualdad es falsa.

Como afirmaba antes, el diccionario académico es selectivo. Por un lado, dada la ideología que lo determina, sería imposible que adoptara una posición abierta, y respetuosa de realidades de la lengua que no pertenecen más a la cepa castellana castiza; por el otro, sigue determinando el valor de los vocablos en términos de autoridades, con lo que grandes cantidades de vocablos populares, rurales, científicos y técnicos, que por “mala” suerte no se ven plasmados en la obra de algún escritor considerado autoridad —también, hay que señalarlo, escritores hispanoamericanos— quedan al margen del interés académico.

La selectividad del *DRAE* y la posición ideológica que lo sustenta son las verdaderas determinantes normativas de la corriente lexicográfica hispánica del regionalismo. Pues una vez que la lengua “general” se concentra en un solo diccionario y éste se toma como único representante legítimo de los valores de la castellanidad y de la unidad de la lengua, la realidad de un vocabulario enormemente variado y rico, hecho de viejas palabras castizas —los “arcaísmos” hispanoamericanos o andaluces—, de arabismos, de multitud de voces de las lenguas aborígenes de los territorios conquistados por España, de galicismos incorporados durante siglos de influencia francesa y de anglicismos contemporáneos sólo puede concebirse como una marginalidad en relación con la lengua originaria de la metrópoli castellana; marginalidad que, además, atenta contra la unidad tan encarecida desde el siglo XV.

Los diccionarios de regionalismos obedecen por lo tanto, en la mayor parte de los casos, al interés por registrar la variedad léxica del español, a la que por lo general se aprecia por su carácter cultural, étnico o nacional; pero como su único papel ideológico consiste en el desvío respecto de lo castellano castizo y en el peligro que representan para la unidad de la lengua, tienen que aceptar su incapacidad de principio para competir con el *DRAE* y para postular valores lingüísticos independientes.

El regionalismo, en cuanto desvío de las normas académicas, se ve como un caso especial del “barbarismo” y adquiere, en relación con el ideal de la lengua general, el valor negativo que atestiguan muchas obras lexicográficas hispanoamericanas.<sup>10</sup> De ahí que no haga falta, para muchos lexicógrafos, una base documental que permita establecer el verdadero carácter regional de una voz, sino que baste con comprobar su ausencia en el *DRAE*. Sólo los lexicógrafos modernos,

<sup>10</sup>Lo que demuestran varios títulos de obras lexicográficas dedicadas al regionalismo, como: Edmundo Bolio, *Mayismos, barbarismos y provincialismos yucatecos* (1931); S. Cordero, *Barbarismos, galicismos y solecismos... con un apéndice de los principales modismos, refranes y provincialismos de empleo corriente en la República mexicana...* (1918); A de León, *Barbarismos comunes en México; solecismos, anglicismos, provincialismos...* (1936); P. Patrón Peniche, *Léxico yucateco. Barbarismos, provincialismos y mayismos* (1932); F. Ramos i Duarte, *Diccionario de mejicanismos. Colección de locuciones y frases viciosas...* (1895); lo mismo sucede en el resto de Hispanoamérica, como lo demuestran, en Honduras: E. Martínez López, *Provincialismos y barbarismos hondureños* (1905); en Costa Rica: C. Gagini, *Diccionario de barbarismos y provincialismos de Costa Rica* (1892); en Guatemala: A. Batres Jáuregui, *Vicios del lenguaje. Provincialismos de Guatemala* (1892), etc. Cf. Haensch 1987.

educados por la lingüística, han podido redefinir el valor de los diccionarios de regionalismos, pero hasta ahora solamente en el ámbito académico.

Hay regionalismos “aceptados” por la Academia. En estos casos su ingreso al *DRAE* ha sido por efecto del valor literario de los escritores que los utilizan. A veces esos regionalismos son muy restringidos en ciertas regiones y uno pensaría que esa sería razón suficiente para que la Academia no los tomara en cuenta. En realidad no han entrado al *DRAE* por su carácter regional, sino por su carácter literario.

Bajo esos criterios la Academia “acepta” y “rechaza” vocablos. Estos dos términos, perfectamente naturalizados en la lexicografía hispánica, manifiestan claramente cómo se concibe el diccionario y el papel de la Academia.

Como afirmaba antes, la ideología dominante de la lengua española tiene existencia en toda la comunidad hispanohablante y no solamente, como podría suponerse, en España, al fin y al cabo la “beneficiada” con ella. Esa ideología ha educado efectivamente a todas las poblaciones hispánicas, se ha vuelto “natural” para ellas y es la que determina toda la inteligibilidad de un diccionario, sea general o sea de regionalismos. De esa manera, especialmente los hispanoamericanos viven con una “conciencia del desvío” de su lengua y preocupados permanentemente por la aceptación o el rechazo, la “existencia” o la “inexistencia” de las palabras que cotidianamente usan.<sup>11</sup>

<sup>11</sup>Las calificaciones de “aceptado”, “existente” e “inexistente” para vocablos de la lengua española que estén o no incluidos en el *DRAE* son generales en todo el mundo hispánico. Llevan al público a

La idea de que si no se someten todos los hablantes al dictado de la Academia, guardiana de la corrección y de la unidad de la lengua, se desencadenará la temida fragmentación del español en varias lenguas incomprensibles entre sí, refuerza el papel del diccionario general y delimita la marginalidad de los diccionarios de regionalismos.

De esos efectos ideológicos derivan dos actitudes generalizadas: por un lado la del purismo, que se mira como un medio de defensa de la corrección y la unidad de la lengua; por el otro la de la crítica acerba a la Academia por cualquier faltante, cualquier defecto o cualquier error en su diccionario. Lo que podrían ser hechos comunes de todo trabajo de recopilación y de interpretación son faltas imperdonables de la Academia, a la que se llega a tachar de negligente. Purismo y condenación del trabajo de la Academia —que en muy pocos casos se convierte en un real anti-academicismo— llevan a muchos a posturas más intransigentes y a una búsqueda perpetua de los errores de la Academia, cuya debilidad y falta de seriedad intentan demostrar.

Otra actitud, menos generalizada, es la que, basada en el aprecio del uso de la lengua en una región determinada, se ocupa de demostrar constantemente que quienes se apartan de los ideales de la lengua española no son, por ejemplo, los hispanoamericanos, sino los españoles mismos y entonces aparecen creencias —de origen anónimo— como la muy conocida en Hispa-

---

menudo, a verdaderos conflictos cuando, para dirimir alguna discusión, uno de los adversarios descalifica a otro mostrándole que, uno de los vocablos con que manifiesta sus argumentos “no existe” en español (aunque lo haya pronunciado y tenga sentido).

noamérica, de que donde “mejor” se habla el español es en Colombia (?).

México no es una excepción en este panorama. Pero su larga tradición, que se remonta al siglo XVIII, de búsqueda de una identidad nacional (Lara, 1983 y 1987b) ha agregado elementos importantes a la ideología de la lengua. Hay una ideología nacional de México que opera de manera relativamente contradictoria con la ideología hispánica y que proviene del esfuerzo, logrado a lo largo de este siglo, de basar su legitimidad nacional no en su pasado hispánico, sino en el precolombino. Los mexicanos se consideran a sí mismos sucesores de los pueblos aborígenes que ocupaban su territorio antes, durante y después de la conquista española. Y en la medida en que el origen de los mexicanos no es España, la lengua española se concibe como una imposición ajena a las verdaderas raíces del país (Lara, 1987b).

Si el español se ha impuesto a los mexicanos, su realidad de lengua mayoritaria y de verdadero fundamento de la nacionalidad se tiene que concebir ya sea de manera contraria —con el insostenible argumento de que no proviene de Castilla, sino que es una especie de evolución independiente y aislada de la lengua de los conquistadores—<sup>12</sup> o al menos haciendo del desvío connatural a su regionalismo una diferencia trascendental: es decir, reivindicando el desvío como virtud

<sup>12</sup>Un maestro de escuela mexicano sostuvo, en una importante entrevista acerca del nombre de la lengua que se habla en México, que “el tronco idiomático hispano no basta para afirmar que la norma lingüística es la del español que, *supuestamente*, habla hoy el pueblo mexicano” (cursivas mías), en la revista semanal *Proceso*, 21.6.1982.

central. De ahí provienen algunos movimientos que se han venido dando sobre todo en el campo de la escritura y que intentan instaurar una "ortografía mexicana" del español, diferente de la académica, que, por ejemplo, elimine la distinción —gráfica para los mexicanos, que no distinguimos los fonemas /s/ y /s;/— entre las letras *c*, *z* y *s*.<sup>13</sup>

Esas actitudes no son mayoritarias. Lo común es el intento de valorar los aportes léxicos de las lenguas amerindias que se encuentran en México dentro del mismo marco de legitimidad y casticismo que tiene el castellano. Así aparece, dentro de la lexicografía del regionalismo, el interés por los diccionarios de nahualismos o aztequismos, de mayismos, etc., que rivalizan en mostrar la riqueza de los aportes de esas lenguas al caudal hispánico.<sup>14</sup>

Sin solucionar el conflicto ideológico entre la concepción predominante de la lengua española y la reivindicación del pasado precolombino, coexisten ambas y operan selectivamente, según la clase de cuestión que se propon-

<sup>13</sup>Intentos que pueden agruparse con otros, que surgen intermitentemente, de modificar la ortografía del español para hacerla todavía más cercana a su fonología. Naturalmente, aun cuando se modificara la ortografía para eliminar las letras *v* y *h*, como ésta no modificaría la distinción entre *s* y *c/z* (correspondientes a la fonología castellana), los mexicanos continuarían buscando una ortografía propia.

<sup>14</sup>La mayor o menor cantidad de voces nahuas o de las otras lenguas amerindias que se encuentran en territorio mexicano y que se han incorporado al español es motivo permanente de debate en México. La publicación de Lope Blanch 1969 motivó varias reacciones en su contra por demostrar que, entre hablantes educados, el conocimiento y uso de esos vocablos es muy reducido. Cf. también Lara 1983.

ga a la comunidad mexicana. Así por ejemplo, si se trata de defenderse del anglicismo —magnificado también por el purismo y por ciertos intereses políticos que intentan utilizarlo como sublimación de un conflicto mayor con los Estados Unidos de América (Lara, 1987b)— se apela al casticismo del español; si se trata de establecer las diferencias con España, se apela al carácter indio de México, y aun en ciertos casos, que no alcanzan dimensión pública por tratarse de actitudes de poder en pequeñas localidades, se apela al carácter nacional del español para impedir movimientos de reivindicación de su propia lengua de algunos pueblos indígenas mexicanos.

Fue dentro de este contexto ideológico en donde se propuso la elaboración de un *Diccionario del español de México* (abreviado en lo sucesivo como *DEM*) ya no como un diccionario de regionalismos, ni tampoco como un diccionario general del español, sino como un *diccionario regional* del español (que Haensch 1987 caracteriza como resultado de un “método integral”, diferente del “contrastivo” tradicional). Los objetivos centrales de este diccionario serían, en primer lugar, documentarlo en un estudio amplio del vocabulario del español hablado y escrito en México (Lara, 1972), a partir del cual se pudiera establecer, con el suficiente rigor teórico y metódico, el léxico efectivo de los mexicanos; en segundo lugar, considerar el español de México como un todo, que negara por principio la conciencia del desvío impuesta por la ideología dominante y la tradición lexicográfica del regionalismo; en tercero, lograr, plasmando la realidad del vocabulario mexicano del español, que la idea de la lengua de los mexicanos se aproximara a esa realidad y recuperara para ellos una legitimidad lingüística escamoteada por la ideología dominante.

Como se ve, los dos primeros objetivos corresponden a un tipo de planteamientos característico de la lingüística descriptiva: el valor de los datos y la noción de un sistema lingüístico autónomo por definición<sup>15</sup> resultan evidentes para cualquier lingüista, aun cuando no se preocupe por el aspecto normativo del problema, aspecto en realidad previamente recusado por el mismo origen de la lingüística descriptiva. El tercer objetivo corresponde, por el contrario, a una reacción en contra de la ideología dominante de la lengua española.

La documentación del *DEM* resulta ser un aspecto tan técnico para la opinión pública que no despierta ningún debate en el plano ideológico.<sup>16</sup> En cambio, ya la idea de que el español de México se puede tratar como un sistema autónomo despierta dudas y confusión.

En efecto, el concepto lingüístico del sistema supone su estructuración autónoma y no se contradice con el hecho histórico y real de que el español de México o de cualquier región forma parte de la lengua española. El concepto lingüístico del sistema corresponde a la axiomática de la descripción y no supone algún tipo de aislamiento o de fragmentación del español de México en contraposición con el español global. Sin embargo, dado el temor ideológico de la desmembración de la

<sup>15</sup>Ya sea porque se le concede existencia real, como en el estructuralismo, o porque se le da valor metódico, la noción de sistema como entidad autónoma no se pone en duda en el nivel descriptivo. Hay luego articulaciones con otros sistemas hacia arriba, que dan los llamados diasistemas (se dice que la lengua española es un diasistema), o hacia abajo, que dan sistemas dialectales, sociolectales, etc. hasta las llamadas "lenguas funcionales", como enseña cualquier manual de lingüística moderna.

<sup>16</sup>Se encuentran presentaciones de nuestro sistema de documentación en Lara/Ham/García Hidalgo, 1981.

lengua española a causa de las independencias hispanoamericanas, ha habido interpretaciones de este objetivo del *DEM* en el sentido de una última declaración de independencia y de un intento por separar el español mexicano del conjunto unido de la lengua española.<sup>17</sup>

Por otra parte, como la tradición lexicográfica sólo ha elaborado diccionarios de regionalismos, se intenta constantemente concebir el *DEM* como un nuevo “diccionario de mexicanismos”; a veces, como un útil catálogo de contrastes entre palabras mexicanas y de otras regiones hispánicas, en particular entre aquellas voces que pueden provocar incomprensión en otros países, como las que tienen significados sexuales, insultantes, etc. En cambio, la idea de un diccionario de toda la lengua española usada en México —que implica una enorme coincidencia real con los usos del español en otras partes del mundo— no resulta algo fácil de entender y hay quien se pregunta si tal esfuerzo podrá dar algún resultado interesante.

El tercero de los objetivos del *DEM*, que ya es del orden ideológico, ha invitado toda clase de aprobaciones, aun cuando muchas de ellas no correspondan a lo planteado en este diccionario. Desde los que vieron en él un anti-españolismo heredado de las malas versiones

<sup>17</sup> Así lo hizo, por ejemplo, el diario madrileño ABC en 1972 y lo llega a considerar como posible el académico español Valentín García Yebra: “Sería magnífico que hubiera un diccionario por cada país; pero sería lamentable que de cualquier manera se fomentase la disgregación del español como lengua de todos. ...si “el léxico del mexicano” quiere decir “el léxico de una lengua propia de México” y si esa expresión implica el deseo de que el español de México se aparte del español común hasta el punto de constituir otra lengua, creo que con ello se deseará el empobrecimiento cultural de México”. *El traductor*, 1,2 (1986), 12-14.

históricas del sentido de la independencia de México; o los que lo concibieron como la aparición de una anti-academia mexicana de la lengua, que viniera a suplantarse a la española dentro de sus mismos márgenes de autoridad; hasta los que esperaban que, consecuentemente con el valor ideológico de los vocablos procedentes de lenguas amerindias mexicanas, se llenaría el diccionario de vocablos de esta clase y se les impondría en contra de los de origen hispánico, y los que esperaban de él un sabroso catálogo de voces picarescas mexicanas, en las que el “genio del mexicano” se podría manifestar. Todas estas esperanzas han quedado defraudadas por lo que se ha dado a conocer del trabajo, el *Diccionario básico del español de México*, que es una versión de siete mil artículos, dedicados al vocabulario fundamental del español de México y al que aparece en los libros de texto oficiales de la escuela elemental mexicana.<sup>18</sup> Independientemente del valor de esas expectativas ideológicas, se trata de manifestaciones normativas que el *DEM* toma en cuenta para poder encontrar la comunicación que requiere con sus lectores.

En realidad el *DEM* tiene delante de sí varios conflictos normativos y provoca otros. En lo que sigue ilustraré un catálogo de esa clase de conflictos con las

<sup>18</sup>Esta versión requerida por la Secretaría de Educación Pública de México, se orienta a la educación elemental y toma en cuenta el vocabulario contenido en los libros oficiales que, como se sabe, son gratuitos y obligatorios para todo el sistema escolar mexicano. La composición léxica de este pequeño diccionario es, por ello, heterogénea, pues junto con las palabras más usuales de México aparecen muchos vocablos científicos que, cuantitativamente hablando, no son frecuentes ni mucho menos fundamentales.

soluciones que, en cada caso, hemos venido encontrando.

1. Descartamos cualquier pretensión de crear una “ortografía mexicana” del español, tanto por motivos prácticos como por razones culturales, históricas y tradicionales (Lara, 1987e). Sin embargo, hay cuestiones ortográficas dignas de consideración, por cuanto no están resueltas en el español general, como lo demuestra la vacilación de la propia Academia Española, e invitan a darles solución de una manera que se aproxime a los usos reales del público mexicano. Por ejemplo:

a) la letra *h* no corresponde a ningún fonema del español desde hace algunos siglos; se mantiene por razones etimológicas y, en ciertas zonas, entre las cuales se cuentan algunos dialectos mexicanos del español, como el veracruzano,<sup>19</sup> se pronuncia con aspiración. De la variedad de las pronunciaciones que suele recibir, algunas se han fosilizado en el español mexicano general, como en /xalár/ y en /regiléte/ cuya ortografía académica es, respectivamente *halar* y *rehilete*.<sup>20</sup> *El Corpus del español mexicano contemporáneo*

<sup>19</sup>En este dialecto los fonemas /x/ y /s/ en ciertas posiciones se aspiran. Cuando algún vocablo se pronunció históricamente con aspiración, como en el caso de *jalar* y muchos otros actualmente escritos con *h*, la posibilidad de aspirar fonemas ayuda a que se conserven las aspiraciones antiguas.

<sup>20</sup>Según Corominas (1954) *halar* es un germanismo que pasó al español a través del francés *haler*, que también se pronuncia con /x/ en Andalucía y Cuba y que tiene, en castellano, un sentido exclusivamente marineramente: “tirar de un cabo, de una lona o de un remo en el acto de bogar” (según *DRAE* 1984) y lo documenta por primera vez en las *Cartas* de Eugenio de Salazar —que para la lingüística hispanoamericana representan una primera documentación de la manera

(abreviado *CEMC*)<sup>21</sup> registra las escrituras *halar*, con un índice normalizado de dispersión de 0.82, y *jalar* (0.60), *jalas* (0.60), *jale* (0.62) —sin confundirlo con un sustantivo homógrafo— y *jalo* (0.11), lo que si bien muestra un conflicto normativo entre la ortografía dada por la Academia, mejor repartida en todo el Corpus, y la registrada en México, menos repartida, también indica que la ortografía mexicana tiene vitalidad, ya que permite conjugar el verbo, en tanto que la española académica está fosilizada en infinitivo solamente. El hecho de que en México se registren formas nominales como *jalón* (0.82), *jalones* (0.60), *jalonazo* (0.0011), y un nuevo verbo derivado de éstas: *jalonear* (0.87), con su respectivo participio nominalizado *jaloneada* (0.60), nos permite optar por la ortografía *jalar* (y *jalón*, etc.) como manifestación de un uso mexicano real, que puede dar lugar a un reconocimiento social de una norma latente, reprimida por la ortografía académica. En cuanto a *rehilete*, hemos percibido la tendencia del público a escribirlo como

---

en que se comenzaron a formar las variedades americanas del español—. En cuanto a *rehilete*, de origen incierto, Corominas documenta en Cuba el juguete designado por esta palabra, pues según el *DRAE* *rehilete* es una especie de dardo y también lo que en México se llama *gallito* del juego de badmington. Corominas reconoce la posibilidad de explicar el sonido /g/ sobre una base en parte fonética.

<sup>21</sup>Sobre este corpus y sus características, Cf. Lara/Ham 1981. Nuestro “índice normalizado de dispersión” es una relación estadística entre la frecuencia de un vocablo, su dispersión en todo el corpus y el tamaño heterogéneo de cada “género” de textos en el corpus. Varía entre 1 y 0. Mientras más se acerque a 1 está mejor repartido en el español de México. No cito los datos de frecuencia para facilitar esta exposición.

*reguilete*, aun cuando no esté documentado en nuestro *CEMC*. En este caso, como tampoco registramos la forma académica, tomamos el riesgo de actuar de la misma manera e instaurar *reguilete*, apoyados por el hecho de que Santamaría 1959 no registra la forma *rehilete* y sí *reguilete* con el comentario “forma vulgar de rehilete”.

b) Hay algunos conflictos normativos con los usos de las letras *b* y *v*, que no tienen distinción fonológica correspondiente en español y se pronuncian igual, pues son representaciones del fonema /b/ (sólo hacen la distinción algunas personas, que no solamente desconocen este hecho histórico, sino que además afectan su pronunciación de acuerdo con la ortografía y cierta influencia de otras lenguas). Es el caso del verbo *avalanzarse* o *abalanzarse*, cuya ortografía en el *DRAE* es la segunda, pero que en nuestro *CEMC* apareció escrito con la primera (0.60) solamente. Como apoyo al mantenimiento de la forma *avalanzarse* en nuestro diccionario hay una justificación del orden etimológico (Valadez, 1982:280-1), lo que nos ha llevado a decidimos a ir en contra de la norma académica.<sup>22</sup>

c) Hay conflicto en la ortografía del extrajerismo: de acuerdo con la Academia Española y muchos de sus seguidores, todo préstamo tomado de una lengua extranjera se debe hispanizar. Así *film* se vuelve *filme*, *nylon* se escribe *nailon* o *nilón*, *whisky* es *güisqui*. En México, si bien se registran usos como los de la Academia, predominan las escrituras que conservan la ortografía originaria. *Film* en nuestro *CEMC* tiene un

<sup>22</sup>Otros ejemplos de conflicto ortográfico sobre base etimológica se encuentran en Valadez (1982).

índice normalizado de dispersión de 0.85, en tanto que *filme* alcanza 0.60; *nilón* no lo registra el *CEMC*; *nailon* tiene 0.60 pero *nylon* alcanza 0.98. *Güisqui*, por último, no lo registra el *CEMC*, mientras que *whisky* recibe 0.99. Extranjerismos muy usuales, como *box*, *round*, *knockout*, *foul*, *smog*, *jet* y *jeep* se escriben de la misma manera en México, mientras que el *DRAE* no los ha incorporado. Por último, el respeto a la ortografía de la lengua originaria corresponde, en muchos casos, a un relativo respeto —limitado por lo fonología y la capacidad fonética de los hablantes— a su pronunciación. Así *pijama* (como la registra el *DRAE*, seguramente con pronunciación /pixáma/) se escribe del mismo modo en México, pero se pronuncia /piyáma/; *restaurante*, se escribe igual y se pronuncia /restorán/. Se puede decir que, en general, la norma mexicana consiste en la conservación de la ortografía y la pronunciación de la lengua originaria de los extranjerismos, en tanto que la académica es hispanizante.

d) Por último, el aprecio a los vocablos de origen indígena constituye un serio problema ortográfico, pues concurren varias grafías para el mismo vocablo y muchas de ellas suponen ser las más cercanas a una representación fonológica del vocablo originario. Así, mientras muchísimas voces se han hispanizado con el paso del tiempo, hasta el punto de que no es posible hallar escrituras alternativas de carácter fonológico, como en *gualote*, *ajolote*, *mole*, *escuincle* (procedentes, respectivamente, de *huey-xólotl*, *axólotl*, *mollī* e *itzcuintli*); en otros casos hay un esfuerzo normativo del español de México por conservar sus grafías cercanas a su origen, como en *ixtle* (0.60), *nahual* (0.78) o *achiote* (0.60), que predominan sobre las versiones *istle* (0.11), *nagual*

(0.64), y *achote* (0.0). Hay un sinnúmero de casos en que la solución no es tan sencilla, porque las formas predominantes ni se apegan a la hispanización, ni se acercan a sus grafías etimológicas, como en *chahuixtle* (0.67), que se impone sobre su supuesta forma etimológica *chahuiztle* (0.11), *chipotle* (0.86) sobre *chilpotle* (0.00) —sin contar la forma hispanizada *chipocle* (0.11)—.<sup>23</sup> La mayor parte de estas voces, naturalmente, no está registrada en el *DRAE*, pero muchas de ellas, cuando se encuentran en este diccionario, contradicen las formas mexicanas y se convierten en uno de los lugares comunes de crítica a la Academia, como sucede con los casos de *cacahuete* (*DRAE*) vs. *cacahuate*, *chapapote* (*DRAE*) vs. *chapopote*, *guacal* (*DRAE*) vs. *huacal*. Dado el valor simbólico de la escritura de los nahuatlismos en particular, hubimos de dedicar un estudio a esta clase de conflictos, que nos permitiera situar al *DEM* tanto en relación con el uso como con las diferentes normas que se aprecian en México (Maldonado, 1982).

2. Los aspectos morfológicos y sintácticos de una lengua en un diccionario suelen reducirse mucho como efecto del carácter esencialmente léxico de esta clase de

<sup>23</sup>Hay varias obras dedicadas a la etimología de voces amerindias, en particular provenientes del náhuatl, que con mucho son la mayoría de las que se han incorporado al español de México. Los diccionarios de Cabrera, 1974, Robelo, 1912, las obras de Dávila Garibi, 1939, R. del Castillo, 1919 y otros constituyen fuentes de consulta importantes. Sin embargo, para muchas voces falta documentación etimológica y muchas veces la que hay es de dudosa calidad. De ahí que en muchos casos no pueda uno estar seguro de la grafía etimológica de un vocablo y más bien, las contradicciones entre las explicaciones que dan esos autores, oscurezcan la posibilidad de encontrar una solución relativamente definitiva. Los ejemplos que aquí ofrezco podrían discutirse en ese sentido.

obras. A la vez en esos campos es en donde resulta más difícil a la sociedad asociar valores ideológicos a los fenómenos lingüísticos y por lo tanto definir normas que se les apliquen. Sin embargo, hemos encontrado algunos casos, que enumero a continuación:

a) Hay muchos sustantivos en español que pueden variar entre género masculino o femenino y marcarse consecuentemente con los artículos correspondientes. En varios casos se conserva la posibilidad —tan antigua como la propia lengua— de usar indistintamente artículos masculinos o femeninos, como lo demuestra la investigación dialectológica. Así, *el/la sartén*, *el/la lente*, *el/la mar*, *el/la pus* varían en todo el mundo hispánico y no son objeto de preferencias normativas en un plano general. En el español de México parece haber preferencia normativa por *el radio* (el medio de comunicación y el aparato), que registra nuestro *CEMC* 30 veces, sobre todo en textos literarios, frente a *la radio*, que es el uso académico y que registramos 17 veces en textos técnicos de radio.<sup>24</sup> De manera semejante, en México se prefiere *la pijama* frente al uso académico *el pijama*.

b) Hay diferencias semánticas ligadas a la selección de género, que también manifiestan normas mexicanas: *la bolsa* (0.97) se prefiere como designación de lo que el *DRAE* define como *el bolso* (0.74, en textos literarios); *la falla* (0.83) designa un defecto en el funcionamiento de algo a la vez que un fenómeno

<sup>24</sup>Los significados no se pueden cuantificar de la misma manera que los significantes o que los signos, cuando sólo tienen un significado, por eso doy solamente datos de frecuencia. En esos textos técnicos de radio, varios de ellos traducciones, en las que los traductores buscan conscientemente un español académico, 16 contextos se refieren al medio de comunicación y uno solo al aparato.

geológico, en tanto que *el fallo* designa exclusivamente una decisión judicial. Para el *DRAE*, un defecto de funcionamiento es también *un fallo*.

c) El español de México, como el resto del hispanoamericano, da a la forma antepresente *he cantado* valor durativo, mientras que utiliza la forma pretérito *canté* con valor perfectivo exclusivamente.<sup>25</sup> Se aparta claramente de los valores que la Academia da a ambas formas.

3. El campo léxico, por su flexibilidad sistemática y por su facilidad para verse objetivado se presta mucho más a la aparición de normas que pretendan controlarlo. Tales normas son de varios tipos: las que actúan seleccionando posibilidades de creación neológica (notables sobre todo en la formación de terminologías técnicas), las que se manifiestan como preferencias sociales por un vocablo en vez de otro, sin que lleguen a tener un efecto de obligación, y las que se dedican a negar y a desautorizar usos, tanto de vocablos como de significados. En este trabajo no hablaré de las primeras, porque el tratamiento de terminologías técnicas en el *DEM* es más conservador que en otros trabajos, ligados a él, en los cuales hemos ensayado una actividad neológica.<sup>26</sup> En cuanto a las segundas, he aquí una lista de preferencias encontradas en el español de México que el *DEM* asienta en su macroestructura:

<sup>25</sup>Este hecho, señalado por Lope Blanch 1982, hemos podido incluirlo en la parte del diccionario en que se explica, con fines pedagógicos, el uso de los tiempos verbales.

<sup>26</sup>Por ejemplo en el vocabulario técnico de la publicidad en México (Lara/Verdugo, 1982) y en un estudio de Natalia Mata, del equipo del *DEM*, sobre el léxico de la microcomputadora (todavía no publicado).

a) Vocablos que caracterizan al español de México frente al académico (no necesariamente correspondiente al de España): *estacionar* un automóvil (0.94) vs. *aparcar* un automóvil (0.60, en textos técnicos y en citas españolas); *plomero* (0.79) vs. *fontanero* (0.0); *papalote* (0.79) vs. *cometa* (0.00); *papa* (0.89) vs. *patata* (0.00).

b) Vocablos que señalan una preferencia mexicana general por ellos aunque otros vocablos, más generales, también se usen (el efecto de los primeros es que dan síntomas —en el sentido bühleriano— del habla mexicana): *banqueta* (0.95) vs. *acera* (0.60); *departamento* (0.96) vs. *apartamento* (0.83); *timbre* (0.87) vs. *estampilla* (0.74) y *sello* (0.00); *pasto* (0.84) vs. *césped* (0.60); *garage* (0.60) vs. *cochera* (0.00); *abarrotos* (0.79) vs. *ultramarcos* (0.11); *regadera* (0.60) vs. *ducha* (0.00); *bolero* (0.69) vs. *limpiabotas* (0.00).<sup>27</sup>

c) Vocablos que se usan en México pero se les persigue como “barbarismos” de distintas índoles, como ser extrajerismos, como tener significados discrepantes de los que tradicionalmente han tenido, etc. En estos casos se trata de normas negativas, dedicadas a señalar “lo que no se debe decir”, y los ejemplos que aquí ofrezco indican, no su aprecio normativo, como en los ejemplos anteriores, sino su existencia social, que en la mayor parte de los casos es abrumadora, si se considera que se trata de voces proscritas:<sup>28</sup> *contro-*

<sup>27</sup>Que haya vocablos con índice normalizado de dispersión 0.00 sólo señala que su aparición en el corpus es ínfima; sin embargo, sabemos que se conocen en México y que hay quien los usa.

<sup>28</sup>Varios de estos ejemplos se basan en una interesante lista de “disparates” publicada por un conocido purista mexicano: Gringoire, 1982.

*lar* (0.94), *chechar* (0.83), *revancha* (0.96), *sofisticado* (0.83), *lunch* (0.71; las formas hispanizadas *lonch* y *lonche*, 0.11 cada una); *clóset* (0.74), *kindergarten* (0.60), *sandwich* (0.94; la forma hispanizada *sangüich*, 0.11); *ponchar* (una llanta) 0.87, *chance* (0.76). Entre los “barbarismos” de significado destacan: *acceder* ‘tener acceso a algo’, que ocurre en 5 de 25 contextos; *ambición* (con sentido positivo), que ocurre en 28 de 78 contextos; y el uso de la preposición *hasta*, que se ha vuelto ambiguo en México (y en otras zonas americanas), pues indica tanto el límite final de una acción, como el inicial. Este uso está totalmente generalizado en toda clase de textos mexicanos, incluso los literarios.

d) Hay muchos vocablos de uso mexicano característico pero que, por estar marcados por la normatividad como muy regionales, quedan sometidos a otros, del español más general. En esos casos suele haber diferencias del orden de lengua escrita y lengua hablada, de lengua literaria y lengua coloquial, o de lengua técnica y científica y lengua común. Por ejemplo: *embrague* (0.60, en lengua escrita) vs. *clutch* o *cloch* (0.60, en lengua hablada); *camión* (0.76, en lengua coloquial, hablada) vs. *autobús* (0.86, en lengua cuidada, escrita); *cerillo* (0.60, en lengua hablada) vs. *fósforo* (0.60, en lengua escrita); *coche* (0.82) vs. *automóvil* (0.83, en lengua escrita), *auto* (0.88) y *carro* (0.75, en varias regiones de México). Entre estos cuatro vocablos parece haber preferencias generacionales: los mayores prefieren *coche* y los jóvenes *auto*. Las preferencias por *botica* (0.99) o *farmacia* (0.85) parecen corresponder a diferencias generacionales y comerciales: las *farmacias* se conciben como más moder-

nas que las *boticas*. También parece más moderno *taxi* (0.84) que *libre*.

e) Por último, parece haber multitud de innovaciones en el plano semántico del vocabulario (Lara 1986a y 1986b), pero por la naturaleza del significado, que depende fundamentalmente del instrumental teórico y metódico con que se lo trabaje, y por la falta de obras que permitan un verdadero contraste entre el español de México y el de otras regiones hispánicas, no es posible formarse una idea clara de hasta dónde llegan las diferencias y qué carácter normativo tienen. Es un hecho que, como se vio en c), un vocablo como *ambición* tiene en México un valor positivo totalmente generalizado; no sabemos si realmente esto es un fenómeno mexicano o es más general.

A pesar del reducido número de ejemplos con el que trato de demostrar mis afirmaciones en torno a la relación entre normatividad y lexicografía (reducción que me impone la necesidad de conservar este trabajo dentro de las dimensiones propias de un artículo), espero haber logrado mis objetivos: en primer lugar, sostener que la distinción entre lexicografía del español general y lexicografía del regionalismo ha dependido históricamente de un juicio normativo y no de una división razonable del trabajo. Tal división apenas comienza a darse en los diccionarios contemporáneos de regionalismos, hechos por lingüistas.

En segundo lugar, convencer de que dada la pobreza de la lexicografía del español general y las dificultades prácticas y económicas que conllevaría un proyecto de esta clase realizado conforme a criterios lingüísticos modernos, la tercera vía, la de una lexicografía *regional* del español puede contribuir a resolver,

tanto las necesidades documentales de la lexicografía del español general, como las correspondientes necesidades de la lexicografía del regionalismo, pues reduce el ámbito de la documentación a una sola región — en nuestro caso a un solo país— y permite un levantamiento de datos cuidadoso y suficientemente detallado como para que una variedad regional de una lengua como la española se registre adecuadamente y contribuya al reconocimiento real de lo “general” de la lengua y de lo verdaderamente “regional” en el territorio considerado. La suma de muchos diccionarios regionales del español, como el *DEM*, ofrecería como resultado una primera representación de la verdadera riqueza y la verdadera unidad de la lengua más allá de las limitaciones que hasta hoy impone sobre ambos conceptos la ideología dominante.

En cuanto al conflicto ideológico que subyace bajo la lexicografía hispánica, espero haber mostrado en qué medida la realidad lingüística y normativa de una región como México niega la existencia ideológica de la metrópoli castellana sin negar el valor de la unidad de la lengua, sino más bien confirmándolo, pero con una amplitud que la normatividad académica enturbia y soslaya. El vocabulario del español de México no es el vocabulario de una lengua desmembrada de la española, ni se ha visto hipostasiado en el *DEM* como una realidad aislada y autosuficiente, sino todo lo contrario: es el vocabulario de una lengua sorprendentemente unida y variada al mismo tiempo.

Naturalmente, el tercer objetivo del *DEM*, el ideológico, también se ha podido manifestar en los ejemplos ofrecidos: hay normas mexicanas de validez casi total para México, que reivindican una legitimidad mexica-

na del español en contra de la idea de la imposición colonialista y en contra de la idea del desvío en relación con una lengua “general”. La norma del respeto a la ortografía originaria de voces no hispánicas, como las provenientes de lenguas amerindias y de lenguas extranjeras, se revela como una novedad en el ámbito hispánico y, en su contradicción con la norma académica de la hispanización de estas voces, *puede* querer decir que, en el campo ideológico, lo que la sociedad mexicana y las sociedades hispánicas debieran reconocer es que la extranjería de algunos de sus vocablos —pues no son mayoritarios, como gusta creer el chauvinismo— no se contrapone con el sentimiento de identidad y de unidad, tan querido para nuestras culturas. Las normas que seleccionan unos vocablos en vez de otros y que manifiestan preferencias sociales pueden corresponder a necesidades profundas de reconocimiento de una identidad social, importante para el sentido total de la existencia de una sociedad como la mexicana. Oponer a estas normas dictados metropolitanos no solamente es iluso, pues no habrán de prosperar, sino que además revela un incalificable desdén por la cultura y la humanidad de una sociedad. La concurrencia de normas académicas y mexicanas, manifiesta tanto en el plano general como en el de los diferentes estilos de lengua (literaria, técnica, coloquial, etc.), señala en algunos casos un verdadero conflicto normativo; en otros, una aceptada subordinación de unas normas a otras, que contribuye y tiene por finalidad la conservación de un ideal de “español internacional”.

El *DEM* no habrá de solucionar todos esos conflictos, ni habrá de imponer una ideología. Articulado

como está a valores ideológicos mexicanos, más bien habrá de seguir siendo objeto de debate entre los mexicanos, pues esa articulación le da un sentido que no puede tener un diccionario elaborado fuera de sus fronteras, pero en la medida en que objetive los conflictos y proponga soluciones, esperamos que habrá contribuido a establecer una mejor representación de la realidad de la lengua y de la sociedad en México. Su contribución a la lexicografía de la lengua española, en el plano descriptivo, quedará fuera de toda duda.



## PROBLEMAS Y MÉTODOS DEL SIGNIFICADO ESTEREOTÍPICO\*

Es una vieja experiencia de la lexicografía la existencia de un “significado principal” a la cabeza de la microestructura de un artículo lexicográfico que, en particular en los diccionarios monolingües con ordenación llamada “lógica” de las acepciones del artículo, es el que guía y determina al resto de ellas. Tal “significado principal”, sin embargo, no ha podido definirse de una manera suficientemente coherente y detallada, que le permita configurarse como una noción semántica importante, capaz de encontrar fundamentación en la propia naturaleza de la lengua y, con ello, hacerse valer entre los temas que atraen el interés de la semántica teórica. En esta ponencia pretendo desarrollar los elementos que, a mi juicio, llevan a fundamentar el valor teórico del “significado principal”, reconociendo las dificultades que le resultan inherentes, y aclarando qué recursos de método lo hacen operar en lexicografía. Para ello, partiré de la hipótesis de que el “significado principal” que elaboran muchos diccionarios, tanto modernos como antiguos, es lo que el filósofo angloamericano Hillary Putnam llama “significado estereotípico” (Putnam 1970).

\*Ponencia presentada en el XVIII Congreso Internacional de Lingüística y Filología Románicas.

En efecto, si uno examina las primeras acepciones de los diccionarios que no profesan un ordenamiento radicalmente etimológico, histórico, o de uso, comprueba que, en todas ellas, se trata de pequeños textos, que definen las características “típicas” de un referente para la comunidad lingüística considerada; es decir, ofrecen como significado principal una articulación de características de cada objeto, que se consideran obligatorias para definirlo como tal y no otro; es esa articulación particular de características de un referente lo que Putnam llama “significado estereotípico”. Por ejemplo, el *Diccionario de Autoridades* (1726) de la Academia Española, define como *zorra* a un “animal astutísimo, del tamaño y casi la figura de un perro mediano, de color pardo obscuro...”, cualquier diccionario moderno define *tigre* como “mamífero, carnívoro, de la familia de los félidos, de pelaje amarillento a rayas negras, muy feroz y sanguinario”; la última edición del diccionario de la Academia Española define el adjetivo *azul* como “del color del cielo sin nubes”; nuestro *Diccionario básico del español de México* define *correr* como “ir una persona o un animal dando pasos rápidos y acelerados”. Estas primeras acepciones de los diccionarios corresponden a la primera idea que tiene una persona, un miembro de una comunidad lingüística, cuando escucha esas palabras en aislamiento; y aun cuando las oye en contexto, son las que le ayudan a determinar la mayor o menor lejanía de lo referido en relación con lo que resulta normal y común dentro de su comunidad. Cuando oye *tigre* no pensará en un jaguar —que es uno de los significados de la palabra en Hispanoamérica— y cuando oye *azul* no pensará en su posición dentro del

espectro de la luz; cuando piensa en *correr* no se le ocurrirá como típico el movimiento de los automóviles. Los “significados principales” de esos diccionarios corresponden a la noción de “significado estereotípico” que propuso Putnam a la discusión de la teoría semántica.

Los estereotipos tienen tres características centrales: la primera estriba en su carácter de productos del conocimiento; no son los significados de las palabras que ha teorizado la semántica estructural; no son haces de rasgos significativos determinados por un sistema, sino conjuntos de propiedades que se atribuyen a los referentes en cuanto tales, mediante un proceso cognoscitivo; la segunda consiste en que no constituyen definiciones intensionales —en sentido estricto, a la Carnap— de clases de objetos, sino articulaciones de propiedades asociadas a los objetos por un proceso de conocimiento que no postula una veracidad obligatoria, sino solamente una necesidad social; la tercera, que se concretan en el seno de la sociedad, como manifestaciones de un conocimiento de la realidad cuya inteligibilidad se determina en la memoria social del lenguaje.

Consideremos el caso de la definición de la *zorra*: el estereotipo no está formado por el conjunto de propiedades que constituyen la descripción zoológica de una zorra, sino que se refiere a la experiencia de los miembros de la comunidad considerada en relación con ella. Naturalmente, el estereotipo contiene elementos que proceden de un *relieve* del objeto mismo (*saliency* en las investigaciones de la psicóloga Eleanor Rosch 1977:6): su tamaño, su color. La definición lexicográfica de la *zorra* hoy en día, agrega elementos

de esta clase que, sin duda, provienen de la incorporación a la memoria, al saber social, del conocimiento zoológico. Sin embargo, también incluye, todavía hoy, la observación de que se trata de un animal muy astuto, observación que proviene, no de la zoología, tampoco directamente del comportamiento del animal, sino de la adjudicación de una valoración humana a ese comportamiento. Sólo sobre ese elemento estereotípico es comprensible que la lengua española, pero también por lo menos todas las lenguas occidentales, contenga locuciones, derivaciones y refranes que se basan en el carácter de astucia de la zorra (véase los artículos correspondientes en el apéndice de este trabajo). Algo semejante ocurre con el estereotipo de *tigre*: junto a la propiedad física de estos animales, de tener la piel a rayas negras, aparece una valoración humana que lo define como sanguinario y cruel. El estereotipo resulta ser, si ha de comprenderse en toda su profundidad, un producto de la memoria social, elaborado en la historia de la experiencia de la comunidad con el mundo que la rodea.

Esta noción del estereotipo es menos evidente cuando se trata de un verbo como *correr*, que no tiene, como dice Putnam, una "clase natural" por referente. En este caso, lo estereotípico es el movimiento que resulta ejemplar para la comunidad lingüística y, tratándose de un movimiento que ante todo se percibe en los humanos y muchos animales, sus propiedades se refieren a ellos y no a objetos que también corren, pero cuyo movimiento se viene a comprender así sólo aplicándoles un esquema conceptual humano anterior, como es el caso de los automóviles. (En lenguas que tienen "clasificadores semánticos" manifestados por

morfemas es más claro cómo procede la aplicación cognoscitiva del estereotipo a otros objetos; en mixteco —una lengua del sur de México— un clasificador cuyo significado es ‘madera’ se utiliza en palabras que designan utensilios hechos por el ser humano; la pala, el mazo o la bandeja, son, por ello, estereotípicos para la definición de utensilios; posteriormente, al aparecer otros artefactos, como los aviones, su nombre se compone con el mismo clasificador de ‘madera’, que viene a hacer inteligible su carácter de artefacto).

Eleanor Rosch ha venido ofreciendo pruebas empíricas de una parte importante del estereotipo: aquella que se refiere al carácter universal del proceso cognoscitivo que lo determina; ese carácter es lo que ella engloba en el concepto de *prototipo* que, desde hace algunos años, se cita en unión con los trabajos de Putnam; pero los diccionarios, mediante una lectura antropológica de su contenido, aportan otras pruebas más, en particular en el orden de los componentes tradicionales y culturales de los estereotipos. Un estereotipo, por lo tanto, es una compleja organización de elementos prototípicos, determinados por características inherentes a los objetos, y culturales, determinados por la historia particular de la formación del conocimiento en cada comunidad lingüística. De allí la importancia que parece venir ganando, en particular en lexicografía, la noción de significado estereotípico, y la razón de Putnam cuando afirma que “la cuestión de cómo entendemos una nueva palabra tiene más que ver con el fenómeno global de la definición y la escritura de los diccionarios. ...y es ese fenómeno, el fenómeno de que se escriban (y se necesiten) diccionarios,

lo que da lugar a la idea general de la teoría semántica” (Putnam 1970: 149-150; la traducción es mía).

La cuestión es ahora cómo ha podido la lexicografía desentrañar esos significados estereotípicos que luego ha convertido en sus “significados principales”. Esta cuestión resulta tanto más importante, en cuanto no solamente algunos lexicógrafos caracterizan ese proceso dentro del problema más general del establecimiento de un orden de acepciones, sosteniendo que proceden “de lo más general a lo más particular”, “de lo concreto a lo abstracto” o viceversa, oscureciendo, más que aclarando, su procedimiento con ese viejo y confuso lenguaje escolar, sino que además la teoría estructuralista imperante, tanto europea como angloamericana, no ofrece mejores posibilidades que las de la adición o sustracción de rasgos significativos en una serie de sememas (incluso bajo la forma de la preeminencia de los clasemas o marcadores semánticos).

Lo importante, teóricamente hablando, de la noción de estereotipo propuesta por Putnam es que se define como un hecho social que organiza la inteligibilidad de la comunicación entre los miembros de una comunidad lingüística. En términos saussureanos, se trata por ello de un hecho de lengua, por lo que los métodos que uno puede utilizar para desentrañarlo deben ser métodos del estudio de la lengua y no del habla. De aquí se desprende que la recolección cuantitativa de textos en los que se registran hablas o la aplicación de pruebas psicológicas de valor estadístico ofrecen datos necesarios e importantes para la determinación de un estereotipo pero son incapaces, por su naturaleza aislada, porque no reflejan los diferentes grados de cultura de la lengua que tienen los hablantes

y porque casi nunca contienen todos y solamente los elementos del estereotipo, de ofrecer una determinación suficiente del estereotipo. El estereotipo se concreta en el seno de la sociedad por un proceso de decantación de una plétora de significados individuales, que solamente ocurre en la memoria social y durante el tiempo histórico; para comprobar esto no hay un método “sincrónico”, como los arriba mencionados, que se pueda aplicar. El único método posible es histórico y éste, bien lo sabemos, implica un proceso hermenéutico en el cual la actividad del lexicógrafo, como sujeto de la interpretación, es un rasgo esencial. No es que el diccionario de un estado de lengua contemporáneo, o “sincrónico”, deba diluirse en un diccionario histórico, sino que un diccionario contemporáneo, como los que pide una sociedad determinada, no puede soslayar la sustancia histórica de los significados contemporáneos de las palabras, si lo que busca es precisamente aquel tipo de significado que tiene existencia legítima dentro de la sociedad. Esta paradoja, que Coseriu trata tan brillantemente en su famoso *Sincronía, diacronía e historia*, solamente tiene solución cuando se la repiensa a la luz de la experiencia hermenéutica.

En esa experiencia, hay que partir de dos condiciones iniciales: la primera, que el lexicógrafo se ocupa de la lengua bajo la asunción de que él también es hablante, que dispone de un “conocimiento originario” de ella en cuanto humano —lo que también ha señalado Coseriu— y de un conocimiento educado de la lengua que trata; la segunda, que le es posible desencadenar un proceso de distanciamiento de su lengua que le permite objetivarla, con lo que alcanza a situarse a una

distancia crítica de ella, a partir de la cual supera el círculo de su subjetividad —lo que tiene en común con los demás hablantes de la lengua— y logra una racionalización de sus características. Dicho brevemente: el lexicógrafo forma parte de la sociedad cuya lengua estudia y tiene un conocimiento privilegiado de ella, por lo que su método de determinación del estereotipo se fundamenta en su participación en la historia de esa lengua y en su experiencia del sentido legitimado socialmente.

Sobre esa base, el corpus de datos lingüísticos sobre el que trabaja el lexicógrafo constituye, más que un objeto completo del cual se limitará a hacer una descripción atendida a él, un conjunto de indicios del significado sobre los cuales basa su interpretación reflexiva y objetivante.

Y si en el sentido del conocimiento el trabajo del lexicógrafo se fundamenta en su participación como miembro de una comunidad lingüística, en el sentido de la formación de estereotipos en su sociedad, es parte integrante de aquellos que, en su actividad práctica, en el carácter específico de su participación en la división del trabajo, contribuyen a construir el estereotipo, pues al desentrañarlo y manifestarlo en una definición lo objetivan y lo presentan con caracteres definidos a la reflexión social sobre la lengua.

Así resulta que el lexicógrafo es a la vez observador de la lengua que estudia y mediador entre esta lengua y la sociedad a la que pertenece; el diccionario que produce resulta ser, en último y definitivo análisis, no un simple catálogo convencional de palabras y significados, sino un vehículo de la reflexión social sobre la lengua que llega a ocupar un lugar privilegiado en la

formación de la memoria social. De ahí deriva su carácter necesariamente normativo y su legitimidad.

El estereotipo que elabora el análisis del lexicógrafo adquiere entonces dos dimensiones: por un lado, en tanto producto de una interpretación basada en fuentes documentales que se guían por principios de objetividad comunes a la lingüística descriptiva, se propone como tesis interpretativa acerca de un fenómeno social de la significación; por el otro, en tanto elaboración racional y objetivante de la definición lexicográfica, se viene a considerar el *estereotipo mismo*, que entra en el ciclo de la socialización de los miembros de la comunidad lingüística y norma la inteligibilidad de la comunicación entre sus individuos.

En esa situación, la metodología del significado estereotípico no solamente debe verse desde el ángulo objetivante de la lingüística descriptiva, sino que además debe considerar la posición de *sujeto* del lexicógrafo, en la medida en que, de no hacerlo, induce una retroalimentación permanente y ciega de la interpretación lexicográfica sobre la reflexión social acerca de la lengua; en tal caso, el diccionario se vuelve un potente instrumento de la manipulación y la enajenación ideológica de los miembros de la sociedad.

No hay una receta metodológica que sirva para impedir estos efectos perjudiciales de un diccionario; lo único que puede hacerse es tomar en cuenta estas reflexiones, e incluirlas en el método general de determinación y definición del estereotipo como prevenciones para el lexicógrafo mismo.

Por otra parte, dado el carácter hermenéutico de la determinación del estereotipo, tampoco es posible llegar a proponer un estereotipo único para cada palabra,

sino que podrán elaborarse varios —siempre pocos— desde puntos de vista de un orden más general al del estereotipo mismo, como puede ser alguno que derive de una interpretación antropocéntrica, etnocéntrica, política, etc. de la realidad.

Y como el estereotipo constituye un nivel básico de la configuración de la memoria social de los objetos —algo que muestra con claridad Eleanor Rosch (1977:35ss.)— los otros significados que por lo general tiene una palabra, se arreglarán a partir de él en una red que la lexicografía traduce en un orden de acepciones; aun más, como lo demostró la semántica estructural, el estereotipo se deslinda gracias a la existencia de esa red, sólo que no se trata de una red binarista y cerrada, como la proponen los sistemas estructuralistas, sino de un tejido flexible, de fronteras poco definidas. La interpretación del estereotipo termina por ofrecer un plan general, un mapa, del conjunto de esas acepciones, en donde el estereotipo acaba de especificarse; de donde la inclusión del significado principal en la cuestión más amplia del orden de las acepciones en el artículo lexicográfico muestra su justificación y su sentido. Sólo que las explicaciones dadas para ese orden “lógico” —y para este mismo calificativo— no son suficientemente transparentes como para dejar ver sus características internas.

El estereotipo aparece como núcleo de la red de significados de una palabra y la configuración de esa red depende de la interpretación general de la realidad que uno proponga como criterio rector de la elaboración del diccionario.

En el caso del *Diccionario del español de México* hemos optado, por ejemplo, por interpretar el estereotipo (y la red de significados a la que pertenece) de las

palabras *cabeza* y *pie*, con una idea antropomórfica, aunque sería posible una interpretación abstracta, que propusiera como estereotipo algo como “parte superior, inicial o anterior, de un objeto que tiene varios extremos” para *cabeza* y algo semejante para *pie*; hemos interpretado el estereotipo de *poder* a partir de su localización en una idea de actor o agente, en vez de hacerlo a partir de su carácter de posibilidad de ocurrencia, que también se justificaría si se enfoca el uso auxiliar de este verbo. Las interpretaciones plasmadas en nuestro diccionario no son efecto arbitrario de un lexicógrafo, sino resultado de una interesante convergencia de análisis entre varios, que nos permite suponer su adecuación al significado de la sociedad a la cual pertenecemos. No obstante, de acuerdo con todo lo que he argumentado anteriormente, hay que reconocer y resaltar el hecho de que un diccionario como el nuestro no puede pretender un valor único en nuestra comunidad. ¿No es la hermenéutica una profesión de fe en la interpretabilidad de los signos y una invitación a la tolerancia?

- ZORRA.** f.f. Animal astutísimo del tamaño, y casi la figura de un perro mediano, de color pardo obscuro, y en algunas partes, según los temperamentos, es de otros colores. Es montaraz, y en su tanto no menos nociva, que el Lobo. Tiene la cola muy grande, y poblada, por lo que se llama también Raposa, quasi Rabosa; y Zorra dice Covarr. que se dice quasi Zorra; porque se pela toda en cierto tiempo del año. Lat. *Vulpes*. ALVAR. part. 1. lib. 1. cap. 3. A la bellaca de la Ventera, con el mucho calor, o que la zorra le mataffe la gallina, se quedaron (los huevos) empollados.
- ZORRA.** Llaman alusivamente à la mala muger, ò ramera. Lat. *Meretrix*. Lema. *Pellex*. ESPIN. Escud. Relac. 1. Desc. 13. Porque es oficio corriente por toda España, y en las Poblaciones de importancia tienen correspondencia, y avisos de las zorras comadres, para chupar la sangre à los corderos inocentes.
- ZORRA.** Llaman tambien al hombre astuto, y engañoso, que calladamente, y sin ruido buica su utilidad en lo que executa, y vs à lograr su intento: y para mayor energia suelen usar del diminutivo llamandole Zorrica. Lat. *Vulpes*. *Vulpecula*. MONTOR. Obs. Posth. tom. 1. pl. 291.
- Fueronse rabo entre piernas  
Palas, y Juno, à lo zortas,  
Con intencion de inventar  
El adagio: aqui fui Troya.*
- ZORRA.** Llaman asimismo en estilo familiar la borrachera: y así se dice: Dormir la zorra, desollar la zorra. Lat. *Ebrietas*, *otis*. *Crepula*.
- ZORRA.** Se toma tambien por lo mismo que Sorra.
- A la zorra candilazo.** Phrasé, con que se explica la habilidad de alguno, que engañó à otro alituo. Lat. *Sus Mirivam docet*.
- Caldo de zorra.** Apodo, que dan al disimulado, y que en lo exterior se muestra asfable, y modesto, para lograr allutamente su intencion: y así se suele decir: Es como caldo de zorra, que esta irio, y quema. Lat. *Blanditia fallax*. *Fraus mellita*. MORAT. Com. De fuera vendrà. Jorn. 1.
- ..... *Tontis bas masido:*  
*Sí: como caldo de zorra.*
- No es la primera zorra,** que ha desollado. Phrasé, con que se nota la costumbre, ò habilidad, que alguno tiene, y dá à entender por alguna accion. Lat. *Non semel dolbas fuit*.
- No hai zorra con dos rabos.** Phrasé, con que se explica la dificultad, que hai de encontrar dos cosas de un modo mismo semejantes, y en que concutran unas mismas circunstancias. Lat. *In eodem non conveniunt plures*.
- Mucho sabe la zorra;** pero mas quien la toma. Refr. que amonesta, que ninguno, por muy advertido que sea, debe confiar de su sagacidad; pues puede haber otro mas astuto, que lo engañe. Lat.
- Astu polles vulpis; verum qui vincere vulpam,  
Invenit astu, vincere teque, valet.*
- QUEV. Mus. 6. Canc. 3.
- Si el que toma la zorra, y la desuella,  
Canta el refrán que ha de saber mas que ella.*
- Tener zorra.** Phrasé, con que se explica la pesadéz de cabeza, ò melancolia, que alguno padece. Lat. *Veterno laborare*.

**RENARD** [ʀ(ə)naʀ]. *n. m.* (1240, *Renart*, *n. pr.*, du frq. *Reginhart*, donné à l'animal dans le Roman de *Renart*; a éliminé *Goupil*). ♦ 1° Mammifère carnivore (*Canidés*), aux oreilles droites, à la tête triangulaire assez effilée, à la queue touffue, au pelage fourni. *Renard commun*, à pelage jaune roux. *Renard argenté*. *Renard bleu*. V. *Isatis*. *Chasse au renard*. *Cri du renard*. V. *Glapisement*. — *Le renard a une réputation d'adresse et de ruse* (Cf. ci-dessous, 3°). *Le corbeau et le renard*, fable de La Fontaine. ♦ 2° Peau, fourrure du renard apprêtée. *Manteau à col de renard*. *Porter un renard argenté*. ♦ 3° *Par compar.* ou *fig.* (XIII<sup>e</sup>). *Personne fine et rusée, subtile*. *Un fin renard*. « *La souplesse et la ruse de ces vieux renards* » (MONTHERLANT). ♦ 4° *Loc. pop. et vieillie* (fin

*Le Petit Robert*

**fox** \ˈfɒks\ *n, pl foxes or fox often attrib* [ME, fr. OE; akin to OLF *vus* fox, OHG *fuhs* fox, *fōha* she-fox, ON *fōa* fox, Goth *fauho* fox, Skt *puccha* tail] **1**: any of various alert carnivorous mammals of the family Canidae related to the wolves but smaller, with shorter legs, more pointed muzzle, large erect ears, and long bushy tail and now placed in *Vulpes* and several other genera represented by one or more species in most parts of the world — see ARCTIC FOX, FENNEC, GRAY FOX, SILVER FOX; compare COLOR PHASE **2**: the fur of a fox **3**: a clever crafty man: a sly fellow (the ~es live by their wits and rely on fraud — J.H.Hallowell) **4** *archaic*

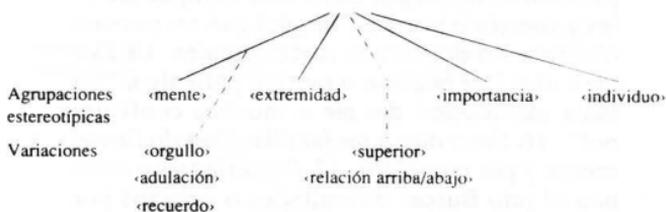


red fox

**cabeza** s f 1 Parte superior del cuerpo humano, y anterior y superior de los animales, donde se encuentran el encéfalo y los principales órganos de los sentidos 2 Parte superior del cráneo donde nace el pelo 3 Caja craneana 4 Mente, inteligencia, razón: *tener una buena cabeza, una cabeza brillante* 5 *De la cabeza a los pies, de pies a cabeza* De arriba abajo, por completo: *mojarse de pies a cabeza* 6 *Irse de cabeza* Caerse 7 *Sentar cabeza* Comenzar a llevar una vida ordenada y tranquila 8 *De mi (tu, su, etc) cabeza* De propia invención 9 *Hacer cabeza* Recordar algo 10 *Meterse de cabeza en algo* Dedicarse por completo a algo, concentrarse en ello 11 *Venirsele a uno algo a la cabeza* Recordar, pensar de pronto algo 12 *Metérsele a uno algo en la cabeza* Pensar en algo o querer hacer algo con terquedad, con obstinación 13 *Calentarle a uno la cabeza* Hacer que alguien crea en lo que le dice otra persona, generalmente con mala intención 14 *Tener la cabeza en los pies* Estar distraído o desorientado 15 *Subírsele a alguien algo a la cabeza* Enorgullecerse excesivamente de algo 16 *Echar de cabeza a alguien* Denunciar a alguien o descubrir sus actos 17 Parte superior de algo: *cabeza de un edificio* 18 Extremo abultado de un objeto: *cabeza de alfiler, cabeza de clavo, cabeza de hueso* 19 Extremo anterior de algo: *cabeza de una viga, cabeza de un puente* 20 Extremo inicial de algo: *cabeza del tren, cabeza de un desfile* 21 Parte más importante, principal, central o directora de algo: *cabeza de distrito, cabeza de un partido, cabeza de un gobierno* 22 Parte de una máquina en la que hay un instrumento movable para hacer algo: *cabeza de barrena* 23 Individuo o elemento de un conjunto: "Diez cabezas de ganado", "Tres dulces por cabeza".

Esquema estereotípico de *cabeza* según 15 informantes \*

Estereotipo: Parte superior del cuerpo humano, y anterior y superior de los animales, donde se encuentran el encéfalo y los principales órganos de los sentidos

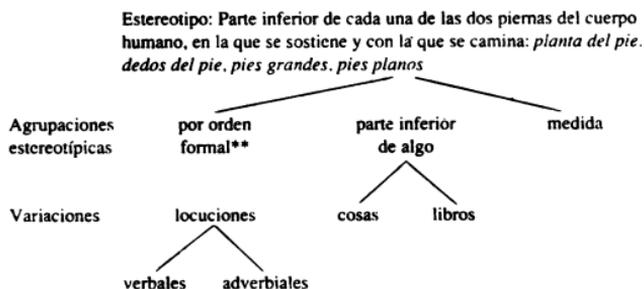


Notas: 1) Los resultados están condicionados por el hecho de que las acepciones de *cabeza*, tal como están escritas en el diccionario, no permiten otras formulaciones. La tarea de los informantes sólo consistió en ordenar esas acepciones *sin* conocer el orden original, sino a partir de una mezcla de acepciones. Las locuciones incluidas en el artículo de *cabeza* tuvieron el importante papel de dar complejidad al reordenamiento, además de haber dado lugar a las variaciones indicadas. 2) Los informantes fueron: 6 lexicógrafos, 6 estudiantes de lingüística y 3 matemáticos. 3) Las palabras entre comillas simples *no* son sememas, sino simples etiquetas para conjuntos de acepciones.

**pie** s m 1 Parte inferior de cada una de las dos piernas del cuerpo humano, en la que se sostiene y con la que se camina: *planta del pie, dedos del pie, pies grandes, pies planos* 2 Parte parecida en algunos animales: *los pies de un mono* 3 A *pie* Caminando, sin usar otro medio de transporte: *ir a pie, llegar a pie, andar a pie* 4 *De pie* En posición vertical, parado: *estar de pie, ponerse de pie* 5 Parte de un calcetín o de una bota que cubre esa parte del cuerpo 6 Parte inferior de algo, que le sirve de apoyo: *pie de una lámpara, pie de una copa* 7 *Al pie, a los pies de* En la parte inferior en donde comienza algo: *al pie de la montaña, a los pies de la sierra* 8 pl Parte de la cama opuesta a la cabecera 9 Parte inferior de un escrito o de una hoja de papel: *firmar al pie de la página, una nota de pie de página* 10 *Pie de imprenta* Indicación que se pone en un libro acerca de su editor, lugar y año de edición 11 Explicación o comentario que se pone en la parte inferior de un grabado

o de una fotografía 12 Medida de longitud usada principalmente en países de cultura inglesa, equivalente a 30.5 cm 13 Tallo o tronco de una planta, particularmente cuando se siembra o se usa para un injerto: *un pie de rosal* 14 Tratándose de versos, unidad rítmica de repetición formada por dos o más tiempos (débiles y fuertes o breves y largos) que no pueden dividirse en dos o más partes iguales 15 *Dar pie a algo* Dar ocasión o motivo para algo: "Su mala explicación *dio pie a* muchas confusiones" 16 *Creer algo a pie juntillas* Creerlo firmemente y por completo 17 *Buscarle tres o cinco pies al gato* Buscar dificultades o peligros por actuar con demasiada audacia o descuido: "Le estás *buscando tres pies al gato* si peleas con ese boxeador" 18 *Cojear alguien de un pie o del mismo pie que otro* Tener cierta debilidad, defecto o vicio, o los mismos que otro 19 *Irse o andarse con pies de plomo* Actuar con mucha prudencia y precaución 20 *Con (el) pie derecho* Con éxito, con buena suerte: *levantarse con el pie derecho* 21 *Con (el) pie izquierdo* Con mala suerte: *levantarse con el pie izquierdo* 22 *De pies a cabeza* Por completo: *mojado de pies a cabeza* 23 *Echar pie a tierra* Bajarse del caballo 24 *Estar uno al pie del cañón* Cuidar algo constantemente, vigilarlo, estar siempre dispuesto 25 *Estar uno con un o el pie en el estribo* Estar a punto de irse o de comenzar un viaje 26 *Estar algo en pie* Continuar algo siendo válido: "Mi ofrecimiento *está en pie*" 27 *Ir uno por su (propio) pie* Ir uno mismo a alguna parte sin ayuda o por su propia voluntad 28 *Írsele a uno los pies* Equivocarse, cometer una imprudencia 29 *No dar pie con bola* Equivocarse por completo, no entender nada 30 *Perder pie* Dejar de tocar el fondo de un río, del mar, etc 31 *Poner pies en polvorosa* Huir, escapar 32 *En pie de guerra* Preparado para la guerra o con aspecto amenazante.

Esquema estereotípico de *pie* según 15 informantes \*



Este arreglo se puede entender, de manera más abstracta, como un desplazamiento del estereotipo referido al sujeto hacia lo no referido al sujeto (de la acepción 1 a la 3), combinado con una oposición de actividad/no-actividad del sujeto (de la acepción 3 a la 5).

\* Mismas notas que para *cabeza*.

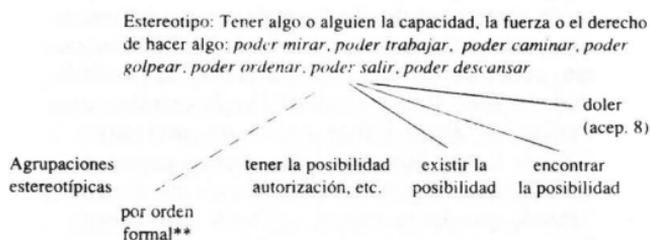
\*\* por ejemplo: 21: *con pie derecho*; 7: *al pie de*; 30: *perder pie*; 31: *poner pies en polvorosa*.

**poder<sup>1</sup>** v tr (Modelo de conjugación 11b; precede siempre a otro verbo en infinitivo) 1 Tener algo o alguien la capacidad, la fuerza o el derecho de hacer algo: *poder mirar, poder trabajar, poder caminar, poder golpear, poder ordenar, poder salir, poder descansar* 2 Tener algo o alguien la posibilidad, la autorización o el permiso de actuar de cierta manera: *poder entender, poder faltar a clases, poder resistir el cansancio, poder besar a la novia* 3 Haber la posibilidad de que algo suceda: "*Puede estallar una huelga*", "*Pudo haber caído un meteorito*", "*Puede llegar Susana el día menos esperado*" 4 *Puede que* Es posible que: "*Puede que llueva*", "*Puede que haga calor*", "*Puede que pierda*" 5 Encontrar o hallar la posibilidad, la oportunidad o el momento de hacer algo: "*Puede conocer a María Félix, pero llegué tarde al teatro*", "*¿Podrías traerme unas frutas del mercado?*", "*Si puedo, te llevaré el periódico*" 6 *Poder con alguien* o *con algo* Ser uno capaz de compren-

derlo, dominarlo o soportarlo: "Yo *puedo* con las matemáticas", "Veré si *puedo* con este niño tan travieso", "No *puedo* con el pesado de tu marido" 7 *No poder más* con Estar uno tan cansado o molesto con algo, que ya no es capaz de seguir con ello o continuarlo: "No *puedo más* con esta caminata", "No *puedo más* con los alumnos de la escuela" 8 *Poderle algo a uno* Serle doloroso moralmente: "Me *pudo* mucho que Antonio me abandonara" 9 *A más no poder* Con el máximo esfuerzo, en grado máximo: "Estudió *a más no poder*", "Es valiente *a más no poder*" 10 *No poder menos que* Ser incapaz de dejar de hacer algo o de actuar en cierta forma: "No *pude menos que* aceptar la invitación", "No *puedes menos que* agradecerse" 11 *¿Se puede?* Expresión con que se pide permiso para entrar en algún lugar.

*Diccionario Básico del Español de México*, El Colegio de México, México, 1986.

Esquema estereotípico del verbo *poder* según 15 informantes \*



Este arreglo se puede entender, de manera más abstracta, como un desplazamiento del estereotipo referido al sujeto hacia lo no referido al sujeto (de la acepción 1 a la 3) combinado con una oposición de actividad/no-actividad del sujeto (de la acepción 3 a la 5).

\* Mismas notas que para *cabeza*.

\*\* por ejemplo: 7: *no poder más con*; 9: *a más no poder*.

## DICCIONARIO DE LENGUA, ENCICLOPEDIA Y DICCIONARIO ENCICLOPÉDICO: EL SENTIDO DE SUS DISTINCIONES

### 1. *Definición preliminar*

La distinción entre diccionarios de lengua y enciclopedias se puede considerar como resultado de una antigua tradición lexicográfica iniciada en la Edad Media. Según Bernard Quemada (1968), en su origen está la diferencia entre los *glosarios* de voces latinas, que elaboraban los eruditos para comprender textos antiguos, y las *summas*, en las que coleccionaban los conocimientos a propósito de las cosas, que habían venido heredando del pasado (por ejemplo, las *Etimologías* de San Isidoro de Sevilla).

Quemada define el diccionario de lengua como aquel “dont les commentaires se limitent à assurer la compréhension des termes consignés et leur emploi dans le discours” (Quemada, 1968, 14); a la vez, define el diccionario enciclopédico —y con él la enciclopedia— como “répertoires de données extralinguistiques” (Quemada, 1968: 15, Zgusta, 1971: 198, Dubois, 1971: 7). Es decir, establece una diferencia clara entre el diccionario de lengua, orientado a las palabras, y el enciclopédico y la enciclopedia, orientados a los objetos. Sin embargo, ya desde el siglo XVIII la distinción presentaba la misma dificultad que presenta

hoy y que lleva a muchos a afirmar que no existe, que *los diccionarios son enciclopedias* (Haiman, 1980: 331). Pues la demarcación entre la *palabra* y el *objeto* al que refiere no es fácil de sostener: cada vez que interese hablar del objeto —y es lo que hace toda comunicación verbal— aparecerá la palabra, y cada vez que aparezca la palabra, referirá a un objeto (Wiegand, 1984). Pese a ello, varias tradiciones lexicográficas han conservado la distinción, que se ha visto justificada teóricamente desde que el estructuralismo vino a definir todo el horizonte de sentido de las disciplinas del lenguaje.

## *2. La defensa estructuralista de la distinción*

Si se considera que las características centrales de la lingüística estructural consisten en a) la idea de que toda lengua es un sistema de relaciones entre elementos, en el que el valor de cada elemento no está dado por otra cosa que por el conjunto de relaciones que establece con otros elementos de su misma clase, y b) la idea de que esas relaciones son fundamentalmente opositivas y binarias, se tendrá la base para comprender cómo el estructuralismo adoptó la vieja distinción entre el diccionario y la enciclopedia en unos términos que vinieron a permear la lexicografía de una manera importante. Pues lo que sostiene el estructuralismo es que, como consecuencia de su idea del sistema, el léxico no debe entenderse como una nomenclatura de referentes del mundo sensible, sino como un juego bien organizado de relaciones entre las palabras, que llega a definir por sí mismo sus significados, en independencia de sus referentes. El significado de las palabras no es una idea, una imagen mental o un concepto

del referente, sino la fórmula constituida por aquellos rasgos mínimos significativos de los signos, a los que suele llamarse *semas*. Para estos *semas* resulta indiferente tener o no tener alguna correspondencia en el referente mismo; aun en caso de haberla, será más por coincidencia (cf. A. J. Greimas, 1967: 13-14; Ullmann, 1962: 64-65; Coseriu, 1966). Según el estructuralismo, la organización interna del sistema lingüístico funda sus propios criterios de objetividad y de pertinencia (Coseriu, 1966, Baldinger, 1977); así por ejemplo, muchas lenguas distinguen entre *sol*, *luna* y *estrellas* sobre la base de características como el momento del día en que aparecen en el cielo, su brillo, sus movimientos, o su figura a la vista. De acuerdo con tales características, el *sol* y las *estrellas* son objetos que se excluyen mutuamente (cf. la definición de *estrella* del *DRAE* 70 en ilustración 1). Esas características son las pertinentes a la lengua y definen el carácter objetivo de las cosas a las que se refieren. En términos estructuralistas, son las que se descubren como *semas* en un análisis de campo semántico (Coseriu, 1966, Greimas, 1967) y, en consecuencia, las que componen el significado de esos vocablos.

Cabe preguntarse si verdaderamente las características del significado de un signo provienen del interior del sistema, con lo cual se daría el hecho milagroso de que las lenguas, a pesar de estar encerradas en ellas mismas (la noción del sistema autocontenido, que tanto cautivó al estructuralismo, cf. Lara, 1981), sirvan para hablar de las cosas y, por lo tanto, refieran a ellas. La mayor parte de los estructuralistas no se adhirieron a una concepción tan radical; más bien, explorando la forma en que la lengua define lo que le es pertinente de

**estrella.** (De lat. *stella*.) f. Cada uno de los innumerables cuerpos que brillan en la bóveda celeste, a excepción del Sol y la Luna. || **2.** Especie de lienzo.

**Ilustración 1a.** *Diccionario de la lengua española*, 19a. edición, Real Academia Española de la Lengua, Madrid, 1970.

**estrella.** (Del lat. *stella*.) f. En el uso común, cada uno de los innumerables cuerpos que brillan en la bóveda celeste, a excepción del Sol, la Luna, Aldebarán, Sirio, etc. || **2.** *Astron.* Cuerpo celeste que radia energía electromagnética (luminosa, calorífica, etc.), producida por las reacciones nucleares que ocurren en su seno. Una **estrella** típica es el Sol. || **3.** Especie de lienzo. || **4.** En el torno de la

**Ilustración 1b.** *Diccionario de la lengua española*, 20a. edición, Real Academia Española de la Lengua, Madrid, 1984.

**estrella** s f **1** Cuerpo celeste que emite energía, generalmente producida por reacciones nucleares en su interior, como el Sol o Sirio **2** Cualquiera de los cuerpos celestes que se ven brillar en la noche, excepto la Luna **3** Figura con la que se representa a ese cuerpo celeste,

**Ilustración 1c.** *Diccionario básico del español de México*, El Colegio de México, México, 1986, p. 221.

**ESTRELLA.** En esta dicción interpuso nuestra lengua castellana la R, como haze en otras muchas, con que se muestra áspera y fuerte; porque en latín se dize *stella, ab stando, quod stare stellae videantur, vel quia stant in caelo tanquam nodi in tabula inter se semper aequae distantes*. Estrellas fixas llamamos las del firmamento, y se mueven en él por su movimiento y guardan entre sí la distancia de lugares do están fixas; a diferencia de los planetas, que unas vezes están en conjunción, otras en oposición y en diferentes aspectos. La estrella, según los físicos y astrólogos, *est pars densior sui orbis*. Son las estrellas de diversas magnitudines: primera, segunda, tercera, etc. Magnitud, lo que influyen en los cuerpos inferiores y el dominio que tienen sobre ellos. Los astrólogos judiciarios lo dirán. Cielo estrellado, quando por la serenidad del aire se

**Ilustración 1d.** Sebastián de Cobarruvias, *Tesoro de la lengua castellana o española*, Ed. facsímil, Turner, Madrid, 1984.

un objeto o de una situación del mundo real, y comparando las clases de objetos que construye el sistema semántico con los objetos que define la ciencia, postularon que ambas, la lengua y la ciencia, toman por referente al mundo real pero lo hacen de diferente manera. La lengua, como receptáculo de las experiencias que tiene una comunidad lingüística a lo largo de la historia, privilegia y aun construye rasgos de los objetos de acuerdo con los intereses y las experiencias inmediatas de sus hablantes. En esa forma, crea una

**ESTRELLA.** f. f. Cuerpo celeste , esférico y denso ; que luce y resplandece con luz propia ò ajena. Es del Latino *Stella*. Lat. *Astrum*. CERV. Quix. tom.2. cap.18. Ha de ser Astrólogo, para conocer por las *estrellas* quantas horas son passadas de la noche. BURG. Son. 99.

*No digo yo que fuerzan las estrellas,  
Que inclinan digo ; pero tu no quieres  
Por tu eleccion, ni porque inclinan ellas.*

Ilustración 1e. Real Academia Española. Diccionario de autoridades, Ed. facsímil, Gredos, Madrid, 1969.

“visión del mundo” característica de cada comunidad lingüística que, a la vez que es resultado de su experiencia del mundo, es la que da forma a esa experiencia. Para Walther von Wartburg y Rudolf Hallig (1963) o Leo Weisgerber (1950), la lengua se convierte en un “mundo intermedio” entre el individuo y la realidad y es de ese mundo de lo que se trata cuando se estudia el significado de los signos. Siguiendo con el ejemplo del significado de *estrella*, resulta que la tradición lingüística española (pero como ella muchas otras tradiciones) concibe los cuerpos celestes sobre la base de claras distinciones entre el sol, la luna y las estrellas, distinciones que en un momento determinado de su historia fueron las únicas “objetivas”, pero que ahora, enfrentadas con el conocimiento que produce la astronomía, resultan científicamente falsas. De ahí que los estructuralistas hayan tenido que proponer que el significado de los signos corresponde a una etapa del desarrollo de la humanidad en la que el conocimiento es “precientífico” y “naif” (Wartburg/Hallig, 1963,

88). La ciencia, por el contrario, se considera como el conocimiento objetivo y verdadero por excelencia. Sólo que, como demostraron múltiples semantistas estructuralistas y como se puede ver hasta con el único ejemplo del vocablo *estrella*, ese conocimiento no corresponde al verdadero significado lingüístico, sino a la naturaleza de las cosas, que es lo que no toca la lengua. Nunca parece haberse ofrecido una explicación de cómo la experiencia de la realidad, así fuera precientífica y naif, penetra la estructura de la lengua como para ofrecerle rasgos significativos que, a pesar de todo, tienen alguna relación con esa realidad.

### 3. *La lexicografía del estructuralismo*

Como consecuencia del desarrollo de la semántica estructuralista, la distinción entre diccionario de lengua y enciclopedia se vio reforzada. Desde el punto de vista de la tipología de las obras lexicográficas hasta se vio radicalizada: aparecieron los “diccionarios lingüísticos”, los que Josette Rey-Debove define como “dictionnaires monolingues qui informent sur le signe à l'exclusion du contenu, ou ceux qui informent sur le contenu sans informer sur la chose (pas de définition). Ce sont en somme des ouvrages qui parlent du signe à l'état pur, sans que ce discours sur le signe révèle jamais rien de la chose” (1971: 29). Tales diccionarios son raros y difícilmente se encontrarían aunque fuera formando parte de una descripción lingüística de una lengua extraña para las culturas occidentales. El diccionario de lengua también se vio modificado teóricamente: dada la necesidad de distinguir los significados de las palabras de los casos en los cuales éstas son meros nombres de cosas (lo que se puede comprender

si uno piensa en los nombres de los elementos químicos), pues solamente las primeras forman parte de la lengua, ya que las segundas son puros soportes sonoros de terminologías, al diccionario de lengua le correspondería tratar únicamente ese léxico usual y tradicional que se organiza en el conocimiento precientífico de la comunidad lingüística, mientras que todas las terminologías le tocarían a la enciclopedia (Coseriu, 1966: 181). Términos como los del arado tradicional, los de la carpintería, los del caballo y su monta, etc. pasarían a formar parte de la enciclopedia.

La distinción entre conocimiento precientífico tradicional y científico entra rápidamente en contradicción, pues no se ve por qué esas terminologías no debieran incluirse en la lengua, dado su carácter histórico, anterior o ajeno a la ciencia y sus tecnologías y claramente interno a la cultura de la comunidad.

La definición lexicográfica debía sufrir modificaciones correspondientes al menos en uno de sus aspectos: en su formulación, debería poder transmitir la posición del vocablo en el interior de la estructura que le correspondiese; así por ejemplo, la definición de los nombres de los colores debería señalar solamente el lugar de cada uno de ellos en la estructura opositiva a la que pertenecieran; el *verde* sería el color que se encuentra entre el *amarillo* y el *azul*. Por ejemplo,

**verde.** (Del lat. «*víridis*»; v.: «berza, vergel; reverdecer; verdegay, verdemar, verdinegro».) ☉ Color simple que se encuentra en el espectro de luz blanca entre el amarillo y el azul. Es, por ejemplo, el color del follaje.

**Ilustración 2a.** María Moliner, *Diccionario de uso del español*, Gredos, Madrid, 1970.

**1. vert, e** | ver, vert | adj. 1° Se dit d'une couleur située entre le bleu et le jaune dans le spectre, et que l'on peut reproduire par la combinaison du bleu et du jaune : *Les feuilles vertes des arbres. Porter une robe verte. L'eau verte d'un étang. Avoir les yeux verts. Etre vert de peur (= avoir très peur).*

**Ilustración 2b.** *Dictionnaire du français contemporain Larousse*, Paris, 1971.

En el caso de palabras como los adverbios, los adjetivos o las preposiciones, la definición debería manifestarse más claramente como efecto de las estructuras opositivas a las que pertenecieran esas palabras (véase la ilustración 3).

Si, como se ve en los ejemplos anteriores, el estructuralismo dejó su huella en la lexicografía, ésta no obstante siguió conservando sus distinciones iniciales, aunque sin poder negar el aporte, al menos teórico, de la semántica estructural (cf. Zgusta, 1971: 103 ss.).

#### 4. *La distinción en la lexicografía contemporánea*

Las diferencias entre los diccionarios de lengua y las enciclopedias consisten en lo siguiente: a) La nomenclatura de los diccionarios de lengua se considera como un inventario del léxico *estructurado* de una lengua; en consecuencia, contiene todos los elementos de los llamados "paradigmas cerrados" (pronombres, preposiciones, conjunciones, etc.), elementos estructurales por excelencia, así como el mayor número posible de vocablos pertenecientes a "paradigmas abiertos" (Martinet, 1966: 117), como los verbos y los sustantivos. La nomenclatura de la enciclopedia, en cambio, no se interesa por los vocablos de "paradigmas cerrados", casi no incluye verbos, y de los sustantivos privilegia solamente los nombres propios (de persona-

après	avant
<p>1° Postériorité dans le temps.</p> <p><i>Un an après sa mort, on l'avait complètement oublié. Il est arrivé bien après moi. Ne vous déidez qu'après mûre réflexion. N'allez pas trop vite; il sera trop tard après pour regretter (syn. : ENSUITE). Après bien des difficultés, il réussit. Après le repas, nous trons au cinéma. Et que ferez-vous après? Vingt ans après, la guerre recommença.</i></p>	<p>1° Antériorité dans le temps.</p> <p><i>Il est arrivé avant moi. Il doutait de lui avant ce succès. Ne vous déidez pas tout de suite; réfléchissez avant (syn. : AUPARAVANT, PRÉALABLEMENT [langue soignée]). On a construit une nouvelle route; avant, il fallait faire un long détour (syn. : AUPARAVANT, AUTREFOIS, JADIS). N'attendez pas son retour, partez avant (partez devant signifierait « sur le chemin qu'il prendra »). Je l'ai vu avant dîner.</i></p>
<p>2° Postériorité de situation dans l'espace, dans le cours d'un mouvement réel ou figuré.</p> <p><i>La maison est juste après l'église (en partant d'ici; derrière l'église pourrait signifier « du côté opposé à l'entrée »). Allez jusqu'à l'angle de la rue, après vous verrez la rivière (syn. : ET PUIS, ENSUITE). Il traîne après lui deux enfants (= derrière lui). Voici la poste, l'épicerie est après.</i></p>	<p>2° Antériorité de situation dans l'espace, dans le cours d'un mouvement réel ou figuré.</p> <p><i>Le bureau de poste est juste avant le pont (en partant d'ici; devant le pont signifierait « en face du pont »). N'allez pas jusqu'à la place; arrêtez-vous avant.</i></p>
<p>3° Infériorité de rang.</p> <p><i>Dans la hiérarchie des grades, le lieutenant vient après (dernier) le capitaine. L'omusement passe après le travail. Qui mettez-vous après?</i></p>	<p>3° Priorité de rang.</p> <p><i>Il place son intérêt avant celui des autres. Il n'avait pas de rendez-vous; le fait-on cependant entrer avant?</i></p>
<p>4° L'hostilité ou l'attachement, le contact immédiat avec (comme prép. ou comme adv. avec un nombre limité de verbes) [généralement fam. en ces cas].</p> <p><i>Les chiens aboient après le facteur. Il crie après les enfants. Pourquoi toujours crier après? Elle est furieuse après son mari. Ils se mettent tous après lui. Il est toujours après son frère. La clef est après la porte. Il grimpe après l'arbre. Pourquoi es-tu monté après? Il est après son travail (il s'en occupe sans cesse). Il est toujours après. Il attend après lui (il ôtaire sa venue). Il court après elle. Il demande après lui.</i></p>	<p>4° L'éloignement du point de départ (employé seulement comme adv. et dans la langue soignée, avec les mots bien, plus, si, assez, fort, trop).</p> <p><i>Creusez plus avant. Je me suis engagé trop avant. Fort avant dans la nuit (la nuit étant fort avancée).</i></p>

**Ilustración 3. Dictionnaire du français contemporain Larousse, Paris, 1971.**

jes famosos, de lugares) o los comunes que refieren a conceptos importantes para el conocimiento del mundo. (Véase la ilustración 4, en la que se han hecho dos tablas con las primeras 25 entradas de dos diccionarios que se pueden reputar como de lengua —uno español y otro francés— y dos enciclopedias escritas en esas mismas lenguas). Esa misma concepción estructural del diccionario de lengua es la que explica la inclusión en esos diccionarios de palabras pertenecientes a un mismo paradigma de derivación (esp. *abacería*, *abacero*, *abacial*, fr., *abaissable*, *abaissant*, *abaisse-langue*, etc.). Los vocablos que forman parte de la nomenclatura de la enciclopedia, en cambio, no provienen de ninguna estructura, sino que constituyen una lista arbitraria desde el punto de vista lingüístico, pero adecuada a los intereses de conocimiento de las comunidades a las que se dirigen. b) Los diccionarios de

*DRAE 84*

a  
a-  
aarónico  
aaronita  
ab  
aba  
¡aba!  
abab  
ababa  
ababillarse  
ababol  
abacá  
abacal  
abacalero  
abacería  
abacero  
abacial  
ábaco  
abacorar  
abad  
abada  
abadejo  
abadengo  
abanderar  
abadesa

*Espasa*

A  
Á  
AA  
A.A  
A^A  
A.A.A.  
AAABAM  
A.A.A.F.F.  
A.AB  
AABAM  
Aabesch  
Aaboden  
Aabogen  
Aabora  
Aab's head  
Aabye  
Aacen  
Aaela  
Aacs  
Aaenabe  
Aacus  
Aaça  
Aach  
Aachen  
Aachenense

*PR*

a  
 a-  
 à  
 abaca  
 abaissable  
 abaissant  
 abaisse-langue  
 abaissement  
 abaisser  
 abaisseur  
 abajoue  
 abandon  
 abandonnataire  
 abandonné  
 abandonner  
 abaque  
 abasourdi  
 abasourdir  
 abasourdissant  
 abasourdissement  
 abat  
 abâtardir  
 abâtardissement  
 abat-jour  
 abat-son

*Universalis*

Aalto, A.  
 Aaron  
 'Abbadides  
 'Abbas ibn Al-Ahnaf  
 'Abbas 1er. Le grand  
 Abbasides  
 Abbaye  
 Abcès  
 'Abd Al-Hamid  
 Abd El-Kader  
 Abd El-Krim  
 Abdomen  
 Abulie  
 Abel, Niels H.  
 Abélard  
 Abencérage (Le)  
 Aberhart, William  
 Aberration astronomique  
 Aberration syst. optiques  
 Abhinavagupta  
 Abidjan  
 Abiétacées  
 Abiétique (acide)  
 Abkhazie  
 Ablation

lengua tratan *signos*, por lo que siempre los caracterizan lingüísticamente: con transcripciones de su pronunciación y su silabeo, con marcas que indican sus paradigmas flexionales o de conjugación, con breves o largas indicaciones etimológicas (cf. ilustración 1), con muestras de uso, reflexiones a propósito de su sintaxis, etc. Las enciclopedias suelen eliminar estas informaciones lingüísticas. c) El diccionario de lengua incluye observaciones o indicaciones acerca del uso de los signos en su nivel pragmático: para cada signo y cada acepción utiliza marcas de dialecto geográfico o social, de jerga, de estilo, etc. o llega a incluir pequeños textos reflexivos sobre esas características. La enciclopedia no ofrece nada al respecto. d) Los diccionarios de lengua suelen no incorporar dibujos, diagramas o fotografías como parte de sus artículos, precisamente porque domina en ellos la concepción de que el significado de un vocablo es de una naturaleza distinta a la de su referente. Así por ejemplo, el significado de *león* no tiene nada que ver, en principio, con la figura del animal. Una ilustración de ese tipo no refiere al signo, sino al objeto mismo. Las enciclopedias, en cambio, utilizan ilustraciones en profusión, precisamente porque su objetivo son las cosas. Siguiendo con el ejemplo, el diccionario de lengua dice *qué significa el signo león*, mientras que la enciclopedia dice y muestra *lo que es un león* (Rey- Debove, 1971: 35). e) Aunque no es una observación frecuente, también es un hecho que la *traducibilidad* de un artículo es un criterio más de distinción entre diccionario de lengua y enciclopedia, que resulta como corolario de su diferencia esencial de acercamiento a los signos o a las cosas. Los diccionarios de lengua no se pueden traducir. Un diccionario

bilingüe no es una doble lista de traducciones entre dos lenguas; la traducción se da en el plano de las proposiciones que transmite un discurso acerca de la realidad. La enciclopedia, por ser un discurso acerca del mundo, se puede traducir, como es el caso de las traducciones españolas de la *Encyclopaedia Britannica*, o la que se ha hecho de la enciclopedia soviética al inglés.

### 5. *El diccionario enciclopédico*

A diferencia del diccionario de lengua y la enciclopedia, la naturaleza del diccionario enciclopédico se puede concebir de dos maneras, según la tradición lexicográfica a la que pertenezca (la cual es, en última instancia, un producto de la concepción lingüística y social del lenguaje en su comunidad). Para las tradiciones que distinguen diccionarios de lengua y enciclopedia, los diccionarios enciclopédicos serán productos mixtos, ya que combinan el tratamiento característico de los signos en los diccionarios de lengua con informaciones acerca de las cosas designadas por esos signos. Hay diccionarios enciclopédicos que lo hacen claramente, separando, en el interior de cada artículo, la información lingüística de la información acerca de la cosa. En cambio, hay tradiciones lexicográficas que no tematizan el diccionario de lengua como tal, aunque distingan entre diccionarios a secas y enciclopedias. (Véase la ilustración 5, en que se listan las primeras 25 entradas del diccionario *Merriam-Webster 3rd.*, en comparación con las primeras 25 entradas de la *Encyclopaedia Britannica*).

La distinción entre éstos es la misma que se explicó antes (4), con las siguientes diferencias: a) en estos

*W3*

a  
 abelian group  
 abscisin  
 acceptable  
 acetylcolinesterase  
 acetylco A  
 actinospectacin  
 action painting  
 additive inverse  
 adenovirus  
 adrenocorticosteroid  
 aeronomy  
 aerospace  
 aflatoxin  
 aftertax  
 agora  
 agravic  
 aldosteronism  
 aleatoric  
 algol  
 algorithm  
 aliesterase  
 allosteric  
 alpha decay  
 alpha helix

*Britannica*

Aalto  
 Aaron  
 'Abbas I the Great  
 'Abd al Malik  
 'Abd al-Mu'mia  
 'Abd ar-Rahman  
 Abdelkader  
 'Abduh, Muhammad  
 Abel, Niels Henrik  
 Abelard, Peter  
 Abraham  
 Abrassises  
 Abu Hanifah  
 Acanthocephala  
 Acarina  
 Accelerators, Particle  
 Acclimatization  
 Accounting, Principles  
 Accra  
 Acid-base Reactions  
 Acoustical Engineering  
 Acre  
 Acting  
 Actinide Elements  
 Acton, Lord

diccionarios no hay una separación entre los vocablos propios de la lengua —definidos por la tradición y por la teoría estructuralista— y las terminologías. Vocablos usuales y tradicionales y términos científicos y técnicos aparecen unidos en la misma nomenclatura. b) La definición de estos diccionarios reúne sin distinción elementos de significado tradicionales de la lengua con elementos que provienen del conocimiento científico, como se ve en la ilustración 6. c) Es común que el diccionario enciclopédico apoye su definición con ilustraciones, tal y como lo hace la enciclopedia.

**'star** \ˈstär, ˈstá(r) n -s [ME *sterre*, fr. OE *steorra*; akin to OHG *sterro*, *sterno* star, ON *stjarna*, Goth *stairno*, L *stella*, Gk *astēr*, astron star, Skt *śrbhis* (instrumental pl.) by means of stars] **1 a** (1) : an object (as a comet, meteor, or planet) in the sky resembling a luminous point and usu. only bright enough to be seen at night; *specif* : FIXED STAR (2) : a heavenly body (as the sun or moon); *also* : EARTH (3) *obs* : POLESTAR (there's no more sailing by the ~ —Shak.) **b** (1) : a self-luminous gaseous celestial body of great mass whose own gravitation produces high internal pressure and temperature resulting in atomic and nuclear processes that cause the star to emit electromagnetic radiation and to be observable in the visible region of the spectrum if its surface temperature is about 2500° absolute or higher, whose shape is usu. spheroidal, whose size may be as small as the earth or larger than the earth's orbit, and that often is composed of two or more stars in close gravitational association — see BINARY STAR, MULTIPLE STAR, SUPERGIANT, WHITE DWARF (2) : one of the self-luminous bodies belonging to a star cluster, globular cluster, star cloud, or galaxy **2 a** (1) : a planet or a configuration of the planets influencing one's destiny or fortune — usu. used in pl. (2) : fortune or fame esp. with regard to its waxing and waning (during his lifetime he received no such acclaim . . . and . . . after his death his ~ had set apparently not to rise again —A.T. Davison) **b obs** : DESTINY, FATE (I was not born unto riches, neither is it I think my ~ to be wealthy —Sir Thomas Browne) **3 a** : a conventional figure with five or more points that represents a star (added another



stars 3a

**Ilustración 6.** *Webster's Third New International Dictionary of the English Language. Unabridged.* Merriam-Webster, Springfield, Mass., 1966.

## 6. *Sentido y conocimiento de las cosas*

Sin duda esta clase de diccionarios contradice al ideal lexicográfico del estructuralismo; sin embargo, resuelve la insatisfacción producida por el rechazo estructuralista a las terminologías y su distinción entre conocimiento científico y precientífico de las cosas. Pero, por otro lado, la existencia de una distinción entre el diccionario a secas y la enciclopedia en esta otra tradición lexicográfica vuelve a poner en el centro de la discusión la cuestión de sobre qué base se distinguen unos y otras. Ello se puede interpretar como la persistencia de un hecho característico de la lengua, que la tradición lexicográfica ha venido registrando desde sus inicios como disciplina del lenguaje: una diferencia real entre el significado de un vocablo y la naturaleza de la cosa significada. Pues si se regresa a la definición del inglés *star* (ilustración 6), se comprueba que también para un “diccionario enciclopédico” como el Merriam-Webster hay un significado (tradicional) de *star* y una naturaleza de ese cuerpo celeste definida por la ciencia. De ahí que la lexicografía moderna haya tratado de superar las distinciones estructuralistas sin negar del todo lo que esa corriente lingüística había destacado de las características del significado. Para la lexicografía contemporánea, el significado de un vocablo tiene un carácter histórico y antropológico cuyo principal objetivo es organizar el *sentido* del mundo real para la comunidad lingüística; es decir, el léxico *organiza*, sobre la base de la pertinencia que ha venido *determinando* la memoria de la experiencia histórica de la comunidad, una visión de la realidad que, *en efecto*, no es científica, pero que en cambio define tanto lo

inteligible del mundo en relación con la historia, como lo inteligible del propio conocimiento científico en relación con la sociedad (Rey 1965: 5-6, Rey-Debove 1971: 194). En consecuencia, el diccionario de lengua sí habla de las cosas, de la misma manera en que habla el diccionario enciclopédico; pero su definición no depende de la existencia de un reconocimiento científico de las cosas, sino que procede de la organización cultural del conocimiento de éstas (Dubois 1971: 36); dicho de otra manera, la definición del diccionario de lengua es una definición del *sentido* de las palabras en la cultura a la que pertenece, en tanto que la definición de la enciclopedia puede prescindir de la manifestación de ese sentido, a pesar de que en la inteligibilidad social es en donde tiene su propio fundamento, y orientarse exclusivamente a la definición de lo que, en su momento histórico, se considera como *naturaleza* de las cosas. Pero, en la medida en que al menos desde el siglo XVIII la ciencia ha venido constituyendo la manera privilegiada de conocer la realidad (Rey, 1965: 77ss.) y ha venido elaborando sus propios criterios de pertinencia y objetividad sobre la base de un extrañamiento de la cultura, su definición del conocimiento de las cosas como revelación de su naturaleza ha venido a imponerse como única forma posible de conocimiento. Así distinguido de la cultura, el conocimiento se define como universal, hace de las lenguas solamente vehículos de expresión, se abstrae de sus significados históricos y elabora conceptos como verdaderos discursos acerca del mundo; sobre esa base, naturalmente, es la enciclopedia la que tiene sentido y, como algunos piensan, el diccionario viene a ser una pequeña enciclopedia, para cuya definición el criterio metódico que

hay que establecer es el de dónde se detiene el discurso acerca de las cosas (Haiman, 1980:329). Los diccionarios de lengua, para los cuales el conocimiento científico de las cosas solamente se puede considerar después de que la cultura lo ha integrado como parte de su inteligibilidad, es decir, después de que el conocimiento se ha socializado, se ven por ello en la dificultad permanente de no encontrar, en términos metodológicos, la frontera entre el conocimiento científico de las cosas, socializado al punto en que comienza a operar como significado y el conocimiento científico en sí. En una u otra dirección, la distinción entre enciclopedia y diccionario continúa siendo conflictiva.



## LAGUNAS Y DEBILIDADES DE LA LEXICOGRAFÍA HISPÁNICA\*

La lexicografía española es una lexicografía imperialista. Se impone, bajo la unicidad del adjetivo, como actividad propia de lo que se hace en la Península Ibérica y aun define las variedades no peninsulares de la lengua española como desviaciones de un ideal cuyo núcleo está en Madrid. Para esa lexicografía, en parte simple tradición —pero no hay neutralidad en su simpleza— en parte simple hecho histórico de desarrollo, de imprenta, de educación, las otras variedades del español solamente pueden ser más o menos pintorescas, más o menos arcaicas, más o menos vulgares, pero siempre, constitutivamente, variedades coloniales: españoles de un *allá*, que sólo parece haberse asimilado a la comprensión que tiene la cultura española de sí misma en el caso de la literatura —desde Rubén Darío hasta García Márquez y Octavio Paz (pero... ¿Juan Rulfo?)—.

Correlativamente, la lexicografía de los hispanohablantes de allá, de América y, en menor medida —geográficamente hablando—, de África, se ha visto siempre como vocación por el desvío, incesante y repetido ciclo

\*Ponencia presentada como introducción a las discusiones sobre la lexicografía hispánica en la mesa redonda “Points noirs de la lexicographie romane”, durante el XVIII Congreso Internacional de Lingüística y de Filología Románicas, en Treveris, República Federal de Alemania, 1986, y publicada en las *Actes* correspondientes.

de documentación de las variedades coloniales de la lengua madre, seducido por la especificidad de sus aportes (sobre todo los provenientes de las lenguas amerindias) y por el pecado contra el casticismo que significa amar esos desvíos. En esa forma, la lexicografía española se escinde en dos concepciones, en dos tipos de diccionarios: la lexicografía de la metrópoli imperial y la lexicografía complementaria de la periferia colonial. Al lado de los diccionarios centrales, del español único, los diccionarios de americanismos y de cada una de las repúblicas que se desgajaron del imperio a principios del siglo XIX. Diccionarios de la lengua y diccionarios de *-ismos*: cubanismos, argentinismos, mexicanismos, etcétera.

El eje imperial no consiste solamente en la distinción entre una lengua española y sus *-ismos*; con él se aglutinan otros valores: en el orden temporal, el español peninsular como verdadera hora de la lengua, real contemporaneidad, frente a los españoles americanos detenidos en el pasado por sus arcaísmos (otros *-ismos*); en el orden social, el español de Madrid, ombli-go de la cultura, frente a los españoles de los antiguos virreinos y audiencias, en donde solamente pudieron anidar criollos desestimados como indianos, campesinos, artesanos y hampones.

La bibliografía de la lexicografía española es la clasificación ordenada por ese eje del imperio y sus concomitantes sociales. Peninsulares al igual que mexicanos o chilenos han contribuido a esa clasificación y extendido ese orden.

De ahí la debilidad más importante de nuestra lexicografía, a la que con tanta insistencia preferimos en América llamar hispánica y no española dejando este último adjetivo para calificar a España, una nación,

un Estado. Debilidad, digo, que consiste en su imposibilidad para considerar las variedades léxicas y la variación semántica de todos los españoles que se hablan en el mundo, pero no en cuanto desvíos, sino en cuanto lengua española; es decir, como existencia de una pluralidad lingüística que nunca niega su origen ni su historia peninsular, pero siempre reivindica su carácter de *lengua materna* y no prestada, ni robada. Esa debilidad se manifiesta precisamente en el carácter complementario, dependiente y siempre titubeante de los diccionarios de *-ismos*.

El agente del eje imperial de nuestra lexicografía es la Academia Española. Por un singular fenómeno social e ideológico acerca del cual los estudios de Fernando Lázaro Carreter (1949) e Hildegunde Henschel (1969) constituyen apenas un principio de comprensión, la Academia Española ha venido a suplantar el consenso que funda la lengua con su sanción normativa, siempre, más allá de la aparente federación de academias correspondientes, localizada y localizable en la historia real de sus miembros numerarios. En esa forma, el diccionario de la Academia no se nutre de la realidad histórica sino, fundamentalmente, de su propia refundición y de la limitación que le depara su carácter metropolitano, tanto en el orden geográfico, como en el más moderno y ahora preocupante del orden terminológico. Creo que ninguna de las lexicografías aquí representadas tiene un *point noir* tan concreto como el diccionario académico.

La Academia y su diccionario se nos presentan en todo su anacronismo: ni los académicos madrileños tienen alguna legitimidad que provenga de la realidad de nuestra lengua, ni su diccionario es capaz —ahora

menos que nunca— de registrar con la necesaria prolijidad y exhaustividad el vocabulario del español, extendido por más de 350 millones de hablantes en 23 países. De ahí dos debilidades más de la lexicografía hispánica: por un lado, la inexistencia de grupos y equipos de lingüistas que recorran todo el mundo hispánico y registren en él, con todo el rigor que la lingüística descriptiva ha alcanzado, el vocabulario del español; por el otro, el desinterés de nuestros lexicógrafos por propuestas lexicográficas alternativas que, ayudadas por el medio moderno de la computación, produzcan diccionarios regionales (y no de regionalismos, esos *-ismos* a los que me referí antes) que, ante todo, revelen una realidad léxica en relación con una identidad cultural o nacional.

Habría que preguntarse en qué momento, bajo qué condiciones históricas, a partir de qué clase de ideología España y con ella los países hispanoamericanos, casi en asunción del cliché europeo de lo ultrapirenaico, dejaron de reflexionar y de documentar su lengua y a la vez dejaron de atraer la atención de los estudiosos de otros países sobre su acervo léxico. Pues, en comparación con la lexicografía del francés, del italiano o del catalán, la lexicografía hispánica no cuenta con una obra etimológica de la importancia del *Französisches Etymologisches Wörterbuch* de Von Wartburg y Baldinger; no cuenta con un diccionario histórico que se acerque, siquiera, al *Gran Robert* o al *Litttré*, ya no al actual *Trésor*; no se ha interesado por la formación de nuestro vocabulario científico y técnico, desde el de las taxonomías botánicas y zoológicas hasta el de las disciplinas contemporáneas como la física, la ingeniería o la fisiología; no ha escrito su propia historia ni puede

evaluar sus diccionarios desde puntos de vista metódicos y rigurosos (pienso en las obras de Quemada 1968 y Rey-Debove 1971 para el francés). En cambio, a la bullente creatividad léxica de sus diversos pueblos, se ha concretado con oponer la sanción de la Academia madrileña que, como decía Antonio Machado, “desprecia cuanto ignora”. Así la lexicografía hispánica es, en su mayor parte, una lexicografía de *status quo*, abandonada por el intelecto a la oportunidad de las editoriales, desinformadora, rígida y solemne.



## BIBLIOGRAFÍA

- Autoridades* (1969): *Real Academia Española, Diccionario de Autoridades*, ed. facsímil, Gredos, Madrid, (1a. ed., 1726).
- Ávila, Raúl (1988): "Lengua hablada y estrato social: un acercamiento lexicoestadístico", *Nueva Revista de Filología Hispánica*, 36, pp. 131-148.
- Baldinger, Kurt (1977): *Teoría semántica. Hacia una semántica moderna*. Alcalá, Madrid.
- Barrera Vázquez, Alfredo *et al.* (1980): *Diccionario maya Cordemex*, Mérida.
- Bortolini, U., C. Tagliavini y A. Zampolli (1971): *Lessico di frequenza della lingua italiana contemporanea*, Garzanti, Milán.
- Buchanan, M.A (1927): *A graded Spanish Word Book*, Toronto.
- Britannica* (1974): *The new Encyclopaedia Britannica. Macropaedia*, Chicago.
- Cabrera, Luis (1974): *Diccionario de aztequismos*, rev. de J.I. Dávila Garibi, Oasis, México.
- Casares, Julio (1950): *Introducción a la lexicografía moderna*, C.S.I.C., Madrid (reimpr 1969).
- Castillo, R. del (Darío Rubio) (1919): *Estudios lexicográficos. Nahuatlismos y barbarismos*, prolog. J.J. Nuñez y Domínguez, México.

- Cobarruvias, S. de (1984): *Tesoro de la lengua castellana o española*, ed. facsímil, Turner, Madrid (1a. ed. 1611).
- Corominas, Joan (1954): *Diccionario crítico etimológico de la lengua castellana*, Gredos, Madrid.
- Coseriu, Eugenio (1955): *Sincronía, diacronía e historia. El problema del cambio lingüístico*. Gredos, Madrid.
- \_\_\_\_\_ (1966): "Structure lexical et enseignement du vocabulaire", en: *Actes du 1er. Colloque International de Linguistique Appliquée*, Nancy, pp. 175-217.
- Dávila Garibi, J.I. (1939): *Del náhuatl al español*, México.
- DBEM (1986): *Diccionario básico del español de México*, El Colegio de México, México.
- DFC (1971): *Dictionnaire du français contemporain*, Larousse, París.
- DFEM (1982): *Diccionario fundamental del español de México*, El Colegio de México/FCE, México.
- DRAE (1970): Real Academia Española, *Diccionario de la lengua española*, 19a. ed., Madrid.
- DRAE (1984): Real Academia Española, *Diccionario de la lengua española*, 20a. ed, Madrid.
- DUE (1970): María Moliner, *Diccionario de uso del español*, Madrid, Gredos.
- Dubois, J. y C. Dubois (1971): *Introduction à la lexicographie: le dictionnaire*, Larousse, París.
- Espasa (1908): *Enciclopedia universal ilustrada europeo-americana*, Espasa Calpe, Madrid.
- Fernández, Luz (1987): "Dialectalismos en el *Diccionario del español de México*; revisión documental", Simposio de lexicografía, Asociación Mexicana

- na de Lingüística Aplicada/UNAM, México (no publicado).
- Fradin, B. y J. M. Marandin (1979): "Autour de la définition: de la lexicographie à la sémantique", *Langue Française*, 43, pp. 60-83.
- García Hidalgo, I. (1981): "La formalización del analizador gramatical del DEM", en: Lara/García Hidalgo/Ham, pp. 87-155.
- \_\_\_\_\_ (1986): "El analizador gramatical del DEM", Simposio sobre presente y perspectivas de la lingüística computacional en México, AMLA, México (no publicado).
- García Hoz, Vicente (1953): *Vocabulario usual, vocabulario común y vocabulario fundamental*, Madrid.
- Geeraerts, D. (1985): "Les données stéréotypiques, prototypiques et encyclopédiques dans les dictionnaires", *Cahiers de Lexicologie*, 46, pp. 27-43.
- Gougenheim, G. et al. (1957): *L'élaboration du français fondamental*, Didier, París.
- Greimas, A.J. (1967): *Sémantique structurale, recherche de méthode*. Larousse, París.
- Gringoire, P. (A. Báez Camargo) (1982): *Repertorio de disparates*, Drago, México.
- Guiraud, Pierre (1960): *Problèmes et méthodes de la statistique linguistique*, París.
- Habermas, Jürgen (1962): *Strukturwandel der Öffentlichkeit. Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*, Frankfurt, Suhrkamp.
- Haensch, Günther (1989): "Der Wortschatz des amerikanischen Spanisch und seine Erfassung in lexiko-

- graphischen Inventaren", *Iberoromania* 30, pp. 1-25.
- Haensch, G. y R. Werner (1978): "Un nuevo diccionario de americanismos. Proyecto de la universidad de Augsburgo", *Thesaurus*, 33, pp. 1-40.
- Haensch, G. *et al.* (1983): *La lexicografía*, Gredos, Madrid.
- Haenschel, Helgunde (1969): *Zur Theorie und Praxis des "Diccionario de Autoridades" im Zusammenhang mit den Bemühungen um eine spanische Sprachreform*, Deutsche Akademie der Wissenschaften, Berlin, Dissertation, Institut für romanische Sprachen und Kulturen.
- Haiman, John (1980): "Dictionaries and encyclopedias", *Lingua* 50, pp. 329-357.
- Hallig, R. y W. v. Wartburg (1963): *Système raisonné des concepts pour servir de base à la lexicographie*. Berlín (1a. ed., 1952).
- Ham, Roberto (1981): "Del 1 al 100 en lexicografía", en: Lara/García Hidalgo/Ham, pp. 43-83.
- Herdan, Gustav (1966): *The advanced theory of language as choice and chance*, Berlín.
- Hurley Vda. de Delgaty, A. y A. Ruiz Sánchez (1978): *Diccionario tzotzil de San Andrés con variaciones dialectales*, ILV/SEP, México, Vocabularios Indígenas, núm. 22.
- Juilland, Alphonse (1961): *Outline of a general Theory of Structural Relations*, Mouton, La Haya.
- \_\_\_\_\_ y E. Chang Rodríguez (1964): *A frequency Dictionary of Spanish Words*, Mouton, La Haya.
- Keesing, Roger M. (1979): "Linguistic knowledge and cultural knowledge", *American Anthropologist*, 81, pp. 14-36.

- Kucera, Henry y Nelson Francis (1961): *A Computational Analysis of Present-day American English*, Providence, R.I..
- Lapesa, Rafael (1984): *Historia de la lengua española*, Gredos, Madrid.
- Lara, L.F. (1972): "Sobre la justificación de un diccionario de la lengua española hablada en México", *La Gaceta del Fondo de Cultura Económica*, 19, pp. 1-6.
- \_\_\_\_\_ y R. Ham (1974): "Base estadística del *Diccionario del español de México*", *Nueva Revista de Filología Hispánica* 23, pp. 245-267, editado en Lara/García Hidalgo/Ham, 1981, pp. 7-39.
- \_\_\_\_\_ (1976a): "Méthode en lexicographie: valeur et modalité du dictionnaire de machine", *Cahiers de Lexicologie*, 29, pp. 103-128.
- \_\_\_\_\_ (1976b): *El concepto de norma en lingüística*. El Colegio de México, México.
- \_\_\_\_\_ (1978): "Regional dictionaries: A lexicographical proposal for the Third World", *Actes du 5e Congrès de l'AILA*, Université Laval, Québec, 1981, pp. 313-321.
- \_\_\_\_\_ (1981), "Del análisis semántico en lexicografía", en L.F. Lara, I. García H. y R. Ham (1981), pp. 157-266
- \_\_\_\_\_, I. García Hidalgo y R. Ham (1981): *Investigaciones lingüísticas en lexicografía*, El Colegio de México, México.
- \_\_\_\_\_ y L. Verdugo (1982): *Vocabulario especializado de la publicidad en México*, SEP, México.
- \_\_\_\_\_ (1983): "Activité normative, anglicismes et mots indigènes dans le *Diccionario del español de México*", en J. Maurais y E. Bédard (ed.), *La*

- norme linguistique*, Conseil de la Langue Française, Québec, pp. 571-601.
- \_\_\_\_\_ (1984): "Características del *Corpus del español mexicano contemporáneo*", en: Actas del 1er. Congreso internacional sobre el español de América, San Juan, Puerto Rico (por publicarse).
- \_\_\_\_\_ (1985): Reseña a Haensch *et al* (1983), en *Anuario de letras*, 23, pp. 313-323.
- \_\_\_\_\_ (1986a): "Methodology in a non Spanish dictionary of the Spanish Language: the *Diccionario del español de México*" en: R. Ilson y M. Blanc (eds.), *A Spectrum of Lexicography. Papers from AILA Brussels 1984*, J. Benjamins, Amsterdam, 1987, pp. 11-28 y en *Polyglot*, 7, 1, B5 (microficha).
- \_\_\_\_\_ (1986b): "El léxico del español de México ¿le es privativo a México?", en: J.G. Moreno de Alba (ed.), *Actas del II congreso internacional sobre el español de América*, UNAM, México, pp. 566-570.
- \_\_\_\_\_ (1987a): "Une critique du concept de métalangage" *Folia Lingüística*, 24 (1990).
- \_\_\_\_\_ (1987b): "La Comisión para la defensa del idioma español du Mexique; chronique d'une politique linguistique avortée" in: J. Maurais (ed.), *Politique et aménagement linguistique*, Conseil de la Langue Française, Québec, pp. 317-358.
- \_\_\_\_\_ (1987c): "El objeto diccionario", *Simposio de lexicografía*, Asociación Mexicana de Lingüística Aplicada/UNAM, México.
- \_\_\_\_\_ (1987d): "La cuantificación en el *Diccionario del español de México*"
- \_\_\_\_\_ (1987e): "La normatividad en ortografía", en: *Escritura y alfabetización*, L.F. Lara y F. Garrido (eds.), El Ermitaño/AMLA, México, pp. 82-101.

- \_\_\_\_\_ (1988): "La question de la norme dans le *Diccionario del español de México*", *Revue québécoise de linguistique*, 17, pp. 61-93.
- Laughlin, Robert M. (1975): *The great Tzotzil dictionary of San Lorenzo, Zinacantán*, Smithsonian Institution, Washington.
- Lázaro Carreter, Fernando (1949): *Las ideas lingüísticas en España durante el siglo XVIII*, C.S.I.C., Madrid.
- León, Aurelio de (1937): *Barbarismos comunes en México*, Porrúa, México.
- Lope Blanch, J.M. (1968): "El proyecto de estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica", en: *El simposio de México*, UNAM, México, pp. 222-223.
- \_\_\_\_\_ (1969): *El léxico indígena en el español de México*, El Colegio de México, México.
- \_\_\_\_\_ (1971): "El léxico de la zona maya en el marco de la dialectología mexicana", *NRFH*, 20, pp. 1-59.
- \_\_\_\_\_ (1978): "Un falso nahuatlismo", *NRFH*, 27, pp. 296-298.
- \_\_\_\_\_ (1982): "Estado actual del español de México", en *El español hablado en México*, UNAM, México, pp. 11-34.
- \_\_\_\_\_ (1983): "El supuesto arcaísmo del español americano", en *Estudios sobre el español de México*, UNAM, México.
- \_\_\_\_\_, *Atlas lingüístico de México*, El Colegio de México, México (en preparación).
- Maldonado, Ricardo (1982): *Náhuatl: que suena bien, que es hermoso*, tesis, UNAM, México.

- Martinet, André (1966): *Eléments de linguistique générale*. A. Colin, París.
- Moreau, René (1962): "Au sujet de l'utilisation de la notion de fréquence en linguistique", *Cahiers de Lexicologie*, 3, pp. 140-159.
- Moreno de Alba, José G. (1988): *El español en América*, Fondo de Cultura Económica, México.
- Morínigo, M.A. (1963): "La penetración de los indigenismos en el español", en: *Presente y futuro de la lengua española*, Madrid, t. II
- Muller, Charles (1965): "Un dictionnaire de fréquence de l'espagnol moderne", *Zeitschrift für romanische Philologie*, 81, pp. 476-483.
- \_\_\_\_\_ (1968): *Initiation à la statistique linguistique*, Larousse, París (hay trad. española en Gredos, Madrid, 1973).
- Nebrija, E. Antonio de (1492): *Gramática de la lengua castellana*, ed. A. Quilis, Nacional, Madrid, 1980.
- Ontañón, Paciencia (1967): *La posible fragmentación del español en América. Historia de un problema*, tesis, UNAM, México.
- Patrón Peniche, Prudencio (1932): *Léxico yucateco, barbarismos, provincialismos y mayismos*, Tenoxtitlán, México.
- PR (1971): *Le Petit Robert. Dictionnaire alphabétique et analogique de la langue française*, París.
- Poirier, Claude (1986): "Les avenues de la lexicographie québécoise" en: L. Boivert et al., *La lexicographie québécoise. Bilan et perspectives*, Laval, Québec, pp. 269-285.
- Putnam, Hillary (1975): "The meaning of meaning" en: Putnam (1982), pp. 215-271.

- \_\_\_\_\_ (1970): "Is semantics possible?" en: Putnam (1982), pp. 139-152.
- \_\_\_\_\_ (1982): *Mind, language and reality*, Cambridge University Press, v. II, Cambridge.
- Quemada, Bernard (1962): "Actes du colloque international sur la mécanisation des recherches lexicologiques", en: *Cahiers de Lexicologie*, 3.
- Quemada, Bernard (1968): *Les dictionnaires du français moderne, 1539-1863*, Didier, Paris.
- Ramos i Duarte, Félix (1895): *Diccionario de mejicanismos, colección de locuciones i frases viciosas*, Imp. Dublán, Méjico.
- Revilla, Manuel G. (1910): "Provincialismos de expresión en Méjico; cuáles son los aceptables y cuáles los viciosos", *Memorias de la Academia Mexicana*, 6, pp. 353-367 [reimpresión facsimilar de 1975].
- Rey, Alain (1965): "À propos de la définition lexicographique", *Cahiers de lexicologie*, 6, pp. 67-80.
- \_\_\_\_\_ (1977): *Le lexique: images et modèles. Du dictionnaire à la lexicologie*, Paris.
- Rey Debove, Josette (1971): *Étude linguistique et sémiotique des dictionnaires français contemporains*. Mouton, La Haya.
- Robelo, Cecilio A. (1912): *Diccionario de aztequismos*. México.
- Rodríguez Bou, L. (1952): *Recuento de vocabulario español*, Puerto Rico.
- Romero, Mario G. (1972): *Epistolario de Rufino José Cuervo con los miembros de la Academia Colombiana*, Instituto Caro y Cuervo, Bogotá.
- Root Hulbert, James (1968): *Dictionaries; British and American*, A. Deutsch, Londres.







Este libro se terminó de imprimir  
en octubre de 1990  
en los talleres de Prisma Editorial,  
Patricio Sanz 1582, México, D.F.  
Se imprimieron 1 000 ejemplares  
más sobrantes para reposición.  
La edición estuvo a cargo del  
Departamento de Publicaciones de  
El Colegio de México.









EL COLEGIO DE MEXICO

308/188/DO 116/CE



\*3 905 0334499 -\*





**E**l trabajo del lexicógrafo oscila siempre entre la práctica del análisis, la redacción y la reflexión sobre su práctica. En el caso del presente libro, la reflexión sobre la lexicografía aborda cuestiones que no han sido tratadas anteriormente, sobre todo en relación con el español de México.

Los trabajos reunidos en este volumen tratan tres aspectos de la teoría y la práctica lexicográfica: el valor de los estudios estadísticos, la semántica del diccionario y las relaciones entre el diccionario, la tradición y la cultura.



Diccionario del español  
de México

**EL COLEGIO DE MÉXICO**