

## La gramática interna de los diccionarios VOX

Manuel Alvar Ezquerra, M<sup>a</sup> José Blanco Rodríguez,

Juan Manuel López Guzmán, M. Fernando Pérez Lagos

Con la elaboración y la publicación, en septiembre de 1987, del *Diccionario General Ilustrado de la Lengua Española* VOX coronábamos un primer proyecto lexicográfico, que lograba un doble objetivo:

- Por una parte, desarrollar las directrices fundamentales de un diccionario general del español, en su momento trazadas por D. Samuel Gili Gaya, para transformarlo en la más extensa base de datos lexicográficos de nuestro idioma hasta entonces compilada.
- Por otra, a semejanza de otros proyectos automatizados de lexicografía que se acometían en múltiples países, aplicar esas directrices, por vez primera en español, bajo la férula material y lógica de un tratamiento computacional del texto del diccionario, cuyo desarrollo ya hemos explicado en otros trabajos.<sup>1</sup>

Como recordaba Manuel Alvar Ezquerra en el prólogo de la edición de 1987: «Para mantener la uniformidad del Diccionario fue preciso averiguar cuál era su estructura general, qué programa de informaciones se desarrolla en cada artículo: palabra de la entrada, etimología, categoría gramatical, nivel de uso en la lengua, lenguaje de especialidad en que aparece o del cual procede la acepción, definición, nombres científicos de animales y plantas, ejemplos, otras grafías posibles de la palabra en cuestión, remisiones a otros artículos [...]».<sup>2</sup>

Las limitaciones materiales que condicionaron todo el trabajo fueron considerables para nosotros entonces: el proyecto fue llevado a cabo en el Centro de Cálculo de la Universidad de Málaga, en el que a los precarios medios instrumentales existentes hubo que añadir el inconveniente de compartir el ordenador destinado a la investigación en la Universidad con otros muchos usuarios que realizaban diversos trabajos. A pesar de ello, la excelente acogida que tuvo el Diccionario en el ámbito lexicográfico,<sup>3</sup> así como las posibilidades que habían ido descubriéndose a medida que se avanzaba en la redacción, nos animaron a constituir una nueva base de datos lexicográfica del español más compleja, capaz tanto de satisfacer las modernas vías de investigación en el campo de la tecnología del lenguaje, como de colmar los intereses

1. «La rédaction automatisée du DGILE» y «Some programming aspects of the first computerized spanish dictionary», en *Proceedings of the Grosseto meeting held 23-25 November 1987*, actas en prensa; y Manuel Alvar Ezquerra, «El primer diccionario automatizado de español contemporáneo», en *Lingüística Española Actual*, IX/1, 1987, pp. 49-56.

2. Alvar Ezquerra, Manuel: «Prólogo de la nueva redacción», en *Diccionario General Ilustrado de la Lengua Española*, Barcelona, 1987, p. XXXVII.

3. Lázaro Carreter, Fernando: «La evolución de un gran diccionario», en *La Vanguardia*, 25-2-1988, p. 45; Marta C. Ayala Castro, reseña en *RFE*, LXVIII, 1988, pp. 169-172; Ken Benson, reseña en *Moderne Språk*, LXXXII, 2, 1988, pp. 175-177; David Mighetto, reseña en *Revista Española de Lingüística*, 19, 1, 1989, pp. 191-194; Ramón Cerdá, reseña en *Revista de la Asociación Europea de Profesores de Español*, XIX-XX, 36-37, 1989, pp. 209-214, etc.

editoriales de la empresa que sustenta el proyecto en la segunda etapa en que nos encontramos.<sup>4</sup>

Así pues, una vez iniciado el proyecto de la *Base de Datos* para la elaboración de los diccionarios VOX, el objetivo es claro y preciso: almacenar y organizar toda la información lexicográfica que existe en los diccionarios ya editados y la que se pueda acumular en el futuro para, de esta manera:

- a) Disponer de la información de un modo rápido y selectivo para la creación de otras obras, que podrán contener extracciones parciales del vocabulario total acumulado, y las informaciones lexicográficas que lo acompañan.
- b) Tener la posibilidad de corregirla siguiendo criterios de coherencia interna, por cuanto esas extracciones parciales, no sólo obedecerán a criterios comerciales, sino también a criterios léxicos, con lo que es fácil comprender las posibilidades de estudio, unificación y homologación que permite un sistema de trabajo basado en segmentos concretos de vocabulario relacionado entre sí por factores semánticos o morfosintácticos.
- c) Disponer de una estructura lógica que permita la introducción, enriquecimiento y reutilización de los datos, con independencia de su posible inclusión posterior en obras impresas.

La facultad de selección pasa necesariamente por la reunión de posibilidades, y así, para que todo lo anterior se tradujera en utilidad, la *Base de Datos* había de conjugar dos puntos fundamentales: su amplitud y su manejo. Si bien es indispensable que la *Base de Datos* sea amplia para poder almacenar y organizar el cúmulo de informaciones y de datos que supone cualquier diccionario general de la lengua, también es imprescindible que ofrezca comodidad en su manejo, rapidez en su consulta, múltiples facilidades para navegar por su interior y atender los aspectos más variados, requisitos que no siempre cumplía la primera que habíamos creado en el Centro de Cálculo de la Universidad de Málaga.

Como complemento ulterior de la *Base de Datos*, y como medios necesarios para conseguir el funcionamiento ideal en el proceso de elaboración de la diversa gama de diccionarios VOX, actuales y futuros, son imprescindibles algunos instrumentos, en los que ya estamos trabajando o están en proyecto:

- a) Un corpus de referencia de la lengua escrita tratado lexicográficamente, representativo de un estado de lengua actual, ya que nuestros diccionarios son sincrónicos y reflejan el último momento de la lengua. Ese corpus habrá de ir acompañado de un analizador morfológico válido para el español.
- b) Una base de datos lexicográfica monolingüe para cada una de las lenguas en las que se desee elaborar diccionarios monolingües. En nuestro caso, por ahora, sólo tratamos la lengua española. En un futuro cabe la posibilidad de poner en marcha una base de datos monolingüe para el catalán.
- c) Una base de datos bilingüe para cada par de lenguas en que se desee hacer

4. Véase M. Alvar Ezquerro, «Desarrollos actuales en lexicografía automatizada del español», en José Vidal Beneyto (coord.), *Las industrias de la lengua*, Madrid, 1991, pp. 219-240; y del mismo autor, «Una base de datos lexicográfica del español», en Antonio Zampolli (ed.), *Computational Lexicology and Lexicography / Lexicologie et Lexicographie assistées par ordinateur*, Pisa (en prensa).

dicionarios, fundamentalmente español-inglés, español-francés y español-catalán.

- d) Una base de datos terminológica plurilingüe con definiciones y contextos.

En la exposición que nos ocupa pretendemos exponer cada uno de los pasos que nos han llevado a la realización de la primera *Base de Datos Monolingüe de la Lengua Española*, desde la elección del material léxico de su contenido hasta el diseño de la gramática interna que le da forma, que, a su vez, configura al diccionario como producto de la lexicografía, esto es, como conjunto de informaciones léxicas archivadas magnéticamente, y es configurada por el diccionario como producto de la diccionárica.

Debido a los antecedentes editoriales, y al trabajo realizado, el punto de partida de nuestra base de datos era forzosamente un diccionario ya existente en forma impresa, lo que nos obligaba a asumir inicialmente todas las imperfecciones que pudiese tener, tanto las metodológicas como las inherentes a la propia descripción de una lengua. Como bien sabido es, en la elaboración manual de un diccionario, aun cuando las normas de redacción son severas, resultan inevitables las infracciones ocasionales del método de trabajo, provocadas tanto por la gran cantidad de información que debe ser manejada por el lexicógrafo, como por las excepciones a las normas que el redactor se ve obligado a introducir para resolver los problemas específicos de determinados artículos, imposibles de homogeneizar, pues no se puede conocer a priori, de manera omnisciente, la compleja estructura de lengua y las particularidades específicas de la forma, la función y la significación de algunas de sus piezas.

Para organizar informáticamente un diccionario consideramos que es preciso tomar lo más frecuente o probable como norma inicial a partir de la cual comenzar el trabajo. De esta manera lo que queda fuera de esa norma resulta marcado, al menos inicialmente, como error, inconsistencia, elemento marginal, o necesitado de un tratamiento especial, aunque no por ello asistemático. El redactor que trabaje sobre una versión informática de un diccionario deberá decidir si mantiene o no la excepción, que siempre será un caso aislado, no frecuente ni repetitivo, y siempre que no le sea posible sistematizarlo.

El diccionario que sirvió de punto de partida para el proyecto no podía ser otro que la mencionada nueva edición del *Diccionario General Ilustrado de la Lengua Española* (DGILE-VOX), en el que habíamos trabajado todos nosotros cuando se elaboraba en el Centro de Cálculo de la Universidad de Málaga, y en el que ya habíamos adquirido cierta experiencia al haberlo organizado en forma de base datos a partir de cintas de fotocomposición. Por circunstancias que ahora no vienen al caso, aquella estructura informática no valía para la nueva etapa que se comenzaba.

Del estudio de DGILE ya se había derivado una primera organización interna que respondía a los principios empleados por el estructuralismo en el análisis semiológico de textos. De esa organización surgiría el modelo de tratamiento informático que se iba a emplear.

El origen estaba, pues, en un diccionario que se puede caracterizar como conjunto abierto de artículos (por más que la aplicación práctica constituyese una cantidad cerrada), en el que cada uno de ellos puede tener anejo a las informaciones semiofuncionales un campo con la etimología y otro con las glosas. Además, cada entrada posee una o varias formas (como en los adjetivos de doble terminación, va-

riantes formales, etc.), y una o más acepciones; como en las entradas *centimano*, *centímáno*; *ensorino*, *na*, *ensorio*, *ria*; *cambur*, *cambure*, etc.

Las informaciones de la entrada son el lema y, en su caso, un número para diferenciar homógrafos, según la práctica habitual de la lexicografía; es lo que ocurre al separar los distintos artículos: *coca* (con los sentidos de 'arbusto', 'baya', 'fantasma' y 'embarcación'), *baca* (el accesorio del automóvil y la baya del laurel), o *coco* ('fruto', 'bacteria', 'larva', 'cuentecilla', 'fantasma').

La etimología, en caso de estar presente, contiene bien la indicación de la lengua de origen de la entrada, bien el modo de su composición, o bien un étimo. El étimo, a su vez, puede hallarse explícito en otro artículo, en cuyo caso aparece una remisión, o implícito en la misma morfología del lema, con lo que a veces sólo es necesario reseñar la lengua de origen o el modo de creación. La información etimológica debe recoger necesariamente, por la tradición en la lexicografía española, la forma etimológica original, de la que, además, pueden aportarse datos tales como su forma de creación, su lengua de origen, significado o informaciones complementarias de diverso tipo. Estas informaciones complementarias son las que pueden hallarse en *mecenas*, *voltío* o *zepelín*, donde se indican las fechas de nacimiento y muerte de los personajes directamente relacionados con la voz.

Los elementos principales de la acepción son la definición y las notas ligadas a ella: marca de localismo, tecnicismo, barbarismo y nivel de uso. También la categoría gramatical va ligada a la definición y no al lema, en contra de lo que podría pensarse al ser el diccionario una obra que refleja hechos de lengua, pero es que la forma de la entrada no es sino una manera de dar paso a lo que se dice sobre el signo, es únicamente una forma canónica, independiente de las significaciones y de las funciones. Por ello es la acepción la que lleva las marcas gramaticales. La acepción puede complementarse con la indicación de nombres científicos en las designaciones de animales y plantas, la presencia de ejemplos, y con la remisión a ilustraciones o cuadros donde pudiera aparecer la cosa nombrada, o el signo nombrante. Por último, pueden constar junto a la acepción las lexías complejas, y otras expresiones pluriverbales, como son realizaciones contextuales de sentidos especiales, que también quedan definidos.

La última parte de los artículos de nuestro diccionario está compuesta por una amplia serie de informaciones que agrupamos bajo la denominación de glosas. Su complejidad y diversidad hasta ahora sólo está esquematizada en la primera versión operativa de nuestra base de datos, a la espera de comprobar debidamente el funcionamiento inicial del resto del artículo ya organizado a partir de las reglas anteriores. Esa es la razón por la que esta parte de la base de datos está diseñada y existe, pero no se han cargado aún las informaciones, a pesar de disponer de su espacio.

Partiendo de la forma empleada en la fotocomposición de la obra, y teniendo en cuenta las variaciones de los tipos de letra, el método sistemático de puntuación y otros elementos formales externos, se desarrolló un sistema de división en campos de todos y cada uno de los artículos de la obra. Fueron traspasados a un sistema informático de tratamiento textual como banco de pruebas inicial de la organización mencionada, y como paso previo obligado para su adecuación y carga en el programa más complejo de la base de datos.

La entrada, la etimología, la definición, etc., aparecen en este sistema como campos independientes, aunque no están regidos por ningún tipo de relación jerárquica o restricción de contenido. Con este sistema informático han visto ya la luz tres diccionarios más de la editorial *Biblograf*,<sup>5</sup> y otro está en proyecto de redacción.

A partir de la organización inicial en forma de tratamiento de textos se creó la base de datos, cuyos esquemas de funcionamiento interno se cimentan sobre la gramática interna de los diccionarios VOX monolingües españoles. Las reglas de esta gramática son las que se exponen a continuación, y responden a un modelo arbóreo de registros ordenados por niveles: así un registro de nivel superior puede englobar varios registros relacionados por medio de enlaces (un artículo engloba varias acepciones o varias categorías gramaticales, etc.). Por otro lado, muchos lugares son recursivos para poder dar cuenta de todas las informaciones que deben acompañar a cada pieza léxica. Nos ahorramos más explicaciones que podrían llegar a ser prolijas, y que resultan innecesarias con la lectura atenta de nuestras reglas;<sup>6</sup> si bien más adelante nos detendremos en los aspectos más importantes.

### REGLAS DE LA GRAMÁTICA INTERNA

DICCIONARIOS	→	[diccionarios] diccionario
DICCIONARIO	→	[diccionario] artículos
ARTÍCULOS	→	{artículos} artículo
ARTÍCULO	→	[diccionario] [fuente] entradas [etimología] acepciones [glosas]
ENTRADAS	→	{entradas} entrada
ENTRADA	→	[número de homógrafo] lema [figura]
ETIMOLOGÍA	→	[diccionario] [fuente] [envío] { lengua de préstamo modo de creación étimos
ÉTIMOS	→	[étimos] étimo
ÉTIMO	→	[diccionario] [fuente] [información complementaria] [modo de creación] [lengua de origen] forma etimológica [significado] [nexo]
FORMA ETIMOLÓGICA	→	[prefijo/elemento prefijal] [término] [sufijo/elemento sufijal]
ACEPCIONES	→	[acepciones] acepción
ACEPCIÓN	→	[diccionario] [fuente] número notas contenidos [figuras] [glosas]
NOTAS	→	categoría gramatical [localismos] [tecnicismos]
CONTENIDO	→	[contenidos] contenido [barbarismos] [niveles] [figuras]
CONTENIDO	→	[diccionario] [fuente] [pluriverbal] definición [nombres científicos] [ejemplos]
DEFINICIÓN	→	[diccionario] [fuente] [limitación semántica] significado [figuras] [envío]

5. *Práctico-Vox. Diccionario Ilustrado de la Lengua Española*, Barcelona, 1990; *Diccionario Manual Ilustrado de la Lengua Española*, Barcelona, 1990 y *Diccionario Actual de la Lengua Española*, Barcelona, 1990.

6. Cfr. M. Antonia Martí e Irene Castellón, «Gramática para el análisis del diccionario Vox», en *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 9, 1991, pp. 123-143. En cualquier caso, el lector puede seguir el desarrollo práctico de estas reglas en cualquiera de las obras publicadas, en especial el *Diccionario Actual de la Lengua Española*.

N. CIENTÍFICOS	→	[nombres científicos] nombre científico
N CIENTÍFICO	→	género especie [autor de la nomenclatura]
EJEMPLOS	→	[ejemplos] ejemplo
EJEMPLO	→	[diccionario] [fuente] forma del ejemplo [figura] [notas] [significado] [n. científico] [figuras]
GLOSAS	→	[glosas] tipo de glosas
TIPO DE GLOSAS	→	[glosas A] [glosas B] [glosas C] [glosas D] [glosas E] [glosas F] [glosas G] [glosas H]
GLOSA A	→	modelo de conjug. { forma irregular conjugación
FORMA IRREGULAR	→	término [comentario] [ejemplo]
CONJUGACIÓN	→	[modos] [imperativo] [gerundio] [participios]
MODOS	→	[modos] tipo de modo
TIPO DE MODO	→	presente imperfecto [indefinido] futuro [condicional]
PARTICIOS	→	[participio regular] [comentario] [participio irregular] [comentario]
GLOSA B	→	[clases de B] clase de B
CLASE DE B	→	[palabras] palabra [comentario] [ejemplo]
GLOSA C	→	[palabras] palabra {[categoría gramatical] [envío]}
GLOSA D	→	[forma] comentario
GLOSA E	→	[clases de E] clase de E
CLASE DE E	→	[número] [remisión] [contenidos] contenido
CONTENIDO	→	{[envío] [comentario] términos [explicación] [etimología] [ejemplo]}
TÉRMINOS	→	[términos] término
TÉRMINO	→	palabra [notas]
GLOSA F	→	[clases de F] clase de F
CLASE DE F	→	[enlace] [palabras] [frases] [ejemplos] [enlace] [frases] [palabras] [ejemplos]
GLOSA G	→	[clases de G] clase de G
CLASE DE G	→	[número] [remisión] [contenidos] contenido
CONTENIDO	→	forma [notas] [significado] [ejemplo]
GLOSA H	→	[número] [remisión] [contenidos] contenido
CONTENIDO	→	explicación [ejemplo] [especificación]

Del mismo modo que un diccionario contiene diversos artículos, en la jerarquización de la exposición, una amplia base de datos debe permitir la agrupación de varios diccionarios, por lo que cada nivel podrá tener una marcación del diccionario específico al que pertenece, si no fuera general para todos. Es un aspecto que queda por desarrollar en nuestro sistema.

El nivel globalizador de DICCIONARIOS representa fundamentalmente la perpetuación de un único esquema gramatical para el desarrollo de cualquier proyecto con independencia del perfil lexicográfico que éste tenga. La falla de algún elemento estructural en este desarrollo supondría la inconsistencia del conjunto de la gramática. Sin embargo, para que este supuesto nuevo proyecto lexicográfico pudiera descabalar la gramática inicial, tendría que ofrecer a su vez las pautas de gramaticalidad necesarias para explicar y organizar todos los diccionarios comprendidos en ella.

El nivel inmediato de DICCIONARIO está constituido por el conjunto de los artículos y la marca opcional en éstos del diccionario al que se adscriben.

La estructura general del nivel ARTÍCULO consta esencialmente de unos apartados fijos (como ENTRADAS o ACEPCIONES), opcionales (como HOMÓGRAFOS o EJEMPLOS) y recursivos (como CATEGORÍAS GRAMATICALES o NIVELES DE USO), según las necesidades intrínsecas de sus contenidos y las exigencias generales del diccionario al que estén adscritos.

El referente operativo de toda entrada es, como vimos, el lema que la encabeza. Por ello, cuando el operador del sistema informático que contiene la base de datos necesite desplazarse por un diccionario determinado lo hará a través de sus lemas de dos formas: bien atendiendo al entorno de la nomenclatura que lo rodea, y que coincide con el conjunto de los denominados artículos, bien desde el interior de uno de ellos, mediante el acceso directo al contenido de otro cualquiera, de modo que se facilita la comprobación interna de los datos presentes, del encadenamiento de los hiperónimos, etc.

Otros campos anteriores se ven ahora descompuestos en nuevos formantes menores, como por ejemplo la ETIMOLOGÍA, a la que se dota de un nexa que permite multiplicarla tantas veces sea necesario sin alterar su estructura. Por otra parte, se permite la asociación del campo FIGURA con el de las NOTAS, para poder conjugar los cuadros que tengan una relación directa con marcaciones técnicas. De la misma manera, el NOMBRE CIENTÍFICO que debe acompañar a determinados nombres para la precisa identificación de lo designado por ellos, se estructura en otros campos independientes, tales como género, especie y autor de la nomenclatura técnica.

Como puede apreciarse, a medida que descendemos por la planta de cualquiera de los diccionarios, las notas características que rigen el comportamiento de todos los campos son la minimalización y la recursividad (un nivel o campo para cada unidad mínima de información y la posibilidad de que ésta se repita un número indefinido de veces), ya que un contingente tan vasto de información sólo puede mantenerse inscrito en una estructura rígida y por tanto sólida, desde el momento en que se establecen sus normas de relación y se aíslan sus componentes, pues no en vano toda relación presupone distancia.

Las marcas sintópicas, sincrónicas y sinestráticas, aun cuando tengan la posibilidad de aparecer pormenorizadamente en cualquiera de los campos con opción significativa (DEFINICIÓN, SIGNIFICADO DE LA CONSTRUCCIÓN PLURIVERBAL, SIGNIFICADO DE EJEMPLO, y otros), se organizan internamente como un bloque paralelo y ajeno al conjunto de informaciones, en listas cerradas, lo que permite en un primer momento un mejor control y seguimiento de la carga de información, evitando que surjan incoherencias y faltas de homogeneidad en el conjunto.

Los distintos campos que precisan el significado dentro de la acepción están relacionados horizontalmente, mientras que las informaciones adyacentes que conllevan se relacionan verticalmente ofreciendo la posibilidad de ser comunes a las de otros artículos.

Esas mismas directrices ven ampliado su campo de acción cuando nos detenemos a considerar las informaciones etimológicas o de las glosas, desde el momento en que pueden ser comunes en parte o en conjunto a uno o varios artículos diferentes. Para ello se habilita el campo original ENVÍO, que pasa ahora a poner de manifiesto la pluralidad de aplicaciones de la información, considerando verticalmente sus referentes en la nomenclatura, lo que facilita el control y el análisis tanto de los términos origen del envío como del conjunto de enviados de cada uno de ellos.

Por otra parte, ésta es una herramienta fundamental entre las varias empleadas para la determinación de la morfología del problemático conjunto de informaciones agrupadas bajo el epígrafe general de GLOSAS, tradicional cajón de sastre donde había ido a parar un vasto conjunto de elementos heteróclitos, imprescindibles para la comprensión global del artículo, pero testigos mudos de la imposibilidad de sistematizarlas y manifestación clara de la imposibilidad del lexicógrafo en su afán de encastrarlo todo, hasta la lengua.

En todos los diccionarios de lengua existentes en la actualidad encontramos, junto a la información tradicional y propia de su cometido (unidad de lengua + significado), una serie de informaciones complementarias, más o menos abundantes según la obra, que pretenden ofrecer la información necesaria para terminar de conocer todo el sentido de la lengua que les ocupa. Si consideramos que un *diccionario general de lengua* tiene la obligación de ofrecer al usuario todo el material léxico, y a veces también lingüístico en general, necesario para la correcta y completa utilización de una lengua, no puede quedarse en ofrecer una mera abstracción de sus unidades formales y acompañarlas de unos contenidos aislados, o lo que es lo mismo, no basta, no funciona, la tradicional fórmula «en qué se parece a» o «en qué se diferencia de» y «cuáles son sus rasgos propios». Siguiendo el conocido triángulo sémico planteado por Ogden y Richards, y su posterior perfeccionamiento en el conocido trapezoido de Klaus Heger, se hace evidente que para el conocimiento completo y perfecto de una lengua hay que tener en cuenta el mundo extralingüístico: la cosa designada es un componente del signo lingüístico y no podemos prescindir de ella sin más, por cómodo que ello pueda resultar, si es el conjunto de signos lingüísticos el que queremos dar a conocer, y ello entraña no pocas dificultades para el contenido del diccionario y para la clasificación de las obras lexicográficas.<sup>7</sup> No obstante, como reconoce Manuel Alvar Ezquerro, la tendencia de los diccionarios de lengua y de la propia tradición lexicográfica es la de querer abarcarlo todo. Para ello, en la obra se recurre a tantos tipos de informaciones como se consideren convenientes, sea cual sea el campo en el que tengan cabida, habilitando para ello cualquier tipo de estrategias, buena parte de ellas tipográficas, para así enmascarar la falta de sistematización en su uso. El cometido de nuestra gramática interna es el de estructurar el conjunto sin que se pierda la naturaleza específica de cada contenido. Así se explica que, sistematizadas y relacionadas, aparezcan informaciones de tipo enciclopédico, como fechas de la vida de personajes, nombres propios, nombres científicos, definiciones descriptivas, etc.; informaciones de uso normativo y gramatical, como conjugaciones y plurales irregulares, incorrecciones frecuentes, etc.; informaciones ortográficas o de pronunciación; informaciones de voces relacionadas, sinónimos, etc.; informaciones en imágenes o ilustraciones (terminan siendo agrupaciones temáticas o ideológicas, o por lo menos semasiológicas o sistemáticas); informaciones ya habituales sobre los niveles de uso, empleos técnicos, usos dialectales, barbarizantes, etc. Todo ello, junto a la presentación de la unidad de lengua en su forma y su significado, viene a ofrecer dicha unidad en la totalidad de su expresión y contenido; queda diseccionada en su integridad, dando cuenta de la forma, la función y la significación.

Bajo esta gramática las glosas se distribuyen según tres criterios: el tratamiento dado al contenido de sus campos, el grado de recursividad que tienen en el interior de

7. Véase Alvar Ezquerro, Manuel, «La forme des dictionnaires a la lumière du signe linguistique», en *Cahiers de Lexicologie*, 52, 1988-1, pp. 117-130.



la microestructura, y las relaciones que mantienen con la macroestructura. Tomando como ejemplos algunos tipos de glosas, según estos factores, las glosas de tipo normativo sobre la formación del plural no mantienen ninguna relación con otros artículos. Las glosas de tipo gramatical referidas a la conjugación, aun cuando sólo tienen operatividad en el interior del artículo en el que se hallan, pueden implementarse con un sistema de conjugación automática que permita su creación inmediata a partir de modelos insertos en otros artículos. Finalmente, las glosas adscritas a los sinónimos, contrarios y términos relacionados, a pesar de poseer el conjunto más rico de informaciones relacionadas con el artículo o acepción en el que aparecen, carecen de sentido si no se hallan vinculadas con todos aquellos artículos que se encuentran recogidos como sinónimos, contrarios o relacionados en su contenido.

Por todo lo anterior, se ha concedido similar importancia tanto al conjunto de informaciones que tradicionalmente han aparecido incluidas en el consabido artículo del diccionario impreso, como a todas aquellas informaciones complementarias que comúnmente han quedado a medio camino entre el diccionario y la enciclopedia, otorgándoles el mismo peso específico en la elaboración de la gramática que ahora nos ocupa, por cuanto ambas partes de la macroestructura debían estar regidas por idénticas reglas y sin que ello supusiera que una u otra perdiera su naturaleza específica. Logramos, de esta manera, que las diversas glosas funcionen como un conjunto unitario de información, relacionándose con los artículos sin perder una independencia capaz de convertirlas en un diccionario propio con unidad estructural y temática sistemáticamente organizada.

Cuando empezamos a diseñar la base de datos, el establecimiento de la gramática interna del diccionario (en general, no de un diccionario en particular) se mostraba como un paso necesario. Hoy que lo hemos dado, vemos que nos obliga a ir más adelante, pues si está bien concebida, como creemos, sus posibilidades operativas serán infinitas y nos permitirán integrar en la base de datos multitud de informaciones, lingüísticas, extralingüísticas, e incluso sobre el modo de trabajar, que enriquecerán nuestras informaciones sobre el léxico y nos permitirán un tratamiento avanzado de la lengua. El diccionario así concebido ya no es solamente la obra impresa, a la que no renunciamos, sino también el conjunto de informaciones archivadas electrónicamente. Hacemos, pues, lexicografía y diccionaria, según la concepción consagrada por Bernard Quemada.<sup>8</sup>

8. Cfr. Bernard Quemada, «Notes sur lexicographie et dictionnaire», en *Cahiers de lexicologie*, 51, 1987-2, pp. 229-242.