

## Tema 2. La lexicografía retroactiva frente a la lexicografía proactiva. Corpus y bases de datos

Tipos de proyectos lexicográficos:

- autor único;
- autor anónimo;
- autor plural o colectivo (con autores explicitados o anonimizados) [p. ej. DLE]

Teoría funcional de la lexicografía – un diccionario como un elemento para satisfacer necesidades puntuales de información en diversas situaciones sociales.

- Situaciones comunicativas donde puede presentarse la necesidad de resolver un problema de comunicación;
- Situaciones cognitivas donde puede presentarse la necesidad de obtener conocimientos sobre algún tema o disciplina;
- Situaciones operativas donde puede presentarse la necesidad de realizar una acción física, cultural o mental;
- Situaciones interpretativas donde puede presentarse la necesidad de interpretar y comprender un signo, señal, símbolo, etc., para determinar si algo es importante.

Principales modelos lexicográficos:

- Retroactivos: caracterizados por el continuismo con respecto a las obras lexicográficas precedentes. Se retroalimentan de trabajos anteriores;
- Proactivos: determinados por la capacidad de ofrecer una usabilidad mejorada, cierta neutralidad ideológica, introducción de un componente intercultural plural en los nuevos productos lexicográficos con respecto a los productos convencionales.

Conflictos entre los planes editoriales – necesidades de usuarios – sostenibilidad de las editoriales en la era de diccionarios gratuitos en línea:

- Patrocinio de una empresa privada; RAE (DLE-2014);
- Acceso abierto a algunos repertorios, pero restringido a otros de la misma editorial; (SM – Clave);
- Los diccionarios de instituciones públicas suelen ser de libre acceso (DEM);

## Lexicografía retroactiva: modelos continuistas

### La macroestructura

Fuentes primarias: orales y escritas, i.e. documentos textuales recogidos de primera mano;

Fuentes secundarias: metalingüísticas, i.e. diccionarios anteriores;

G. Haensch: “los verdaderos progresos de la lexicografía se deben al aprovechamiento de las fuentes primarias, es decir, de textos en sentido más amplio, donde las unidades léxica aparece, en general, en un contexto.”

La planta del diccionario:

El proyecto al que ha de ajustarse la realización de un diccionario o de una obra enciclopédica, e implica desde la lematización hasta el uso o no de los bancos de datos, y la propia distribución de trabajo en la realización.

Ejemplo de un diccionario retroactivo: DRAE (DLE-2014)

- Opciones limitadas de búsqueda en comparación con DRAE-2001 en CD-ROM;
- Dispone de un enorme fichero de 10 millones de papeletas lexicográficas (Fichero General) – para investigaciones diacrónicas;
- Corpus de referencia: CREA, CORDE, CDH, CORPES XXI;
- Tiene una “Nueva planta” que sigue sin ser aplicada del todo: se revisa, pero la revisión ni es completa ni es sistemática, la respuesta a las necesidades de los usuarios llega de forma muy tardía en comparación con otros repertorios;

- La macroestructura da cuenta de la retroalimentación endogámica y no llega a sacar el partido de los datos de sus propios corpus;

Ejemplos concretos:

*almóndiga, matrimonio, calagozo, dolorioso, empezamiento, murciégalo*, etc. => al recogerse en un diccionario usual adquieren legitimación normativa;

Desactualización debida a la exclusión de los neologismos y préstamos ampliamente presentes en el uso.:

Lema	DLE-2014	DEA	Clave
background	no	sí	sí
backup	no	sí	sí
body milk	no	sí	sí
boom	sí	sí	sí
cool	no	sí	sí
Leggin(s)	no	sí	sí
spam	sí	no	sí

Resumen del carácter retroactivo en la macroestructura:

- La ejecución inicial del proyecto sin partir de nueva planta;
- La ausencia de guías de uso y de materiales didácticos complementarios;
- Las limitaciones de búsqueda (en las obras digitales);
- El mantenimiento de las voces arcaicas y desusadas propias más bien de diccionarios históricos;
- La ausencia de neologismos y préstamos presentes en el uso común (diccionarios generales) o en el especializado (diccionarios con fines profesionales)

### La microestructura

Requisitos que deben cumplir las definiciones (frecuentemente ausentes o fallidos en los repertorios retroactivos):

- La unidad definida no debe aparecer en la definición;
- La definición no debe traslucir ninguna ideología;
- La definición debe participar de las características del lenguaje de la época y debe ser formulada con palabras sencillas, claras y precisas.

Otras características:

- En el primer enunciado deben incluirse informaciones sobre la pronunciación, etimología, gramaticales, y pragmáticas, junto a explicaciones recogidas en los prolegómenos del diccionario (p.ej. Salamanca);
- En el segundo enunciado: se ha defendido la inclusión de elementos culturales o interculturales como un “rasgo definidor de los diccionarios del siglo XXI”, pero no debe convertírselos en diccionarios enciclopédicos; la relación de conocimientos los conocimientos culturales con la lengua deberían ceñirse a etapas de aprendizaje → diccionarios escolares y de E/LE.

Ejemplos de retroalimentación (DLE-2014):

Lema	Año de edición	Definición
Airoso (2)	1884 2014	Garboso ó gallardo Garboso o gallardo
Amargo (4)	1884 2014	Áspero y de genio desabrido Áspero y de genio desabrido
Amistad (2)	1780 2014	Amancebamiento Amancebamiento
Amistad (3)	1780 2014	Merced, favor Merced, favor

Bragueta (1)	1780 2014	La abertura que tienen los calzones por delante La abertura que tienen los calzones o pantalones por delante
Matahúmos	1884 2014	Despabiladeras Despabiladeras
Piropo (3)	1853 2014	Lisonjas, requiebros Lisonja, requiebro

Orden de acepciones: corresponde a modelos retroactivos si sigue la frecuencia de uso

Lema	DRAE-2014 en papel	DRAE-2014 en línea	DEA
Fe	1. En la religión católica, primera de las tres virtudes teologales, asentimiento a la revelación de Dios, propuesta por la Iglesia.    2. Conjunto de creencias de una religión. [...]	1. Conjunto de creencias de una religión. [...] 9. En el cristianismo, virtud teologal que consiste en el asentimiento a la revelación de Dios, propuesta por la Iglesia	1. Creencia [en algo de lo que no se tienen pruebas o evidencia] <i>Frec. sin compl. b) (Rel. catól.)</i> Virtud teologal que consiste en creer todo lo que la Iglesia enseña. <i>c) Esp:</i> creencia en que Dios existe

Los modelos retroactivos actuales no se rigen por el uso de corpus de referencia, sino por el uso de fuentes secundarias o de ficheros léxicos => fuentes primarias pero desactualizadas.

Resumen de tendencias de la lexicografía retroactiva:

- Utilizan fuentes secundarias y competencia individual;
- Realizan revisiones parciales poco sistematizadas;
- Van muy por detrás del cambio social;
- Son reflejos de la ideología dominante;
- No ofrecen una respuesta óptima a las necesidades de usabilidad y navegabilidad.

### Lexicografía proactiva: hacia la funcionalidad lexicográfica

Sven Tarp (2013, 2014): todo diccionario que se realice en la actualidad debe responder a las 4 categorías fundamentales de funciones:

- Comunicativas;
- Cognitivas;
- Operativas;
- Interpretativas.

#### *Un diccionario actualizado*

Deben partir de un corpus:

- Las contextualizaciones proporcionadas en el diccionario son reales;
- La ampliación y actualización de los corpus se reflejará en los diccionarios;
- Reflejará los cambios del léxico, las acepciones o sus frecuencias, etc., que se plasmarán en la edición nueva del diccionario concreto o se pueden incorporar inmediatamente en su versión en línea.

#### *Diccionario del español de México (DEM):*

- Basado en el *Corpus del español mexicano contemporáneo* (1921-1974), y un conjunto de datos posteriores que abarcaron hasta poco antes de la publicación;
- El registro es independiente del nivel empleado;
- Un baremo estadístico para incluir una determinada voz o acepción en el diccionario con 3 ocurrencias y más ocurrencias en el corpus;

#### *Diccionario del español actual (DEA):*

- Basado en un conjunto equilibrado de muestras textuales del español europeo entre 1955-1993;
- Independiente del nivel de lengua o registro;

#### *Clave (ediciones SM):*

- Basado en el español documentado en los medios de comunicación;
- En continua renovación desde 1997

#### *REDES:*

- Se sustenta sobre un corpus extraído de prensa de España y América;
- Se fundamenta en las relaciones semánticas que se establecen entre las palabras.

#### *Diccionario general de la lengua española (Larousse-Vox):*

- La última edición en papel da acceso a la edición en línea;
- Basado sobre un corpus propio de la editorial de más de 10 millones de palabras;
- Preinstalado en los Mac;

- Disponible en línea junto con otros productos de la editorial, pero con acceso restringido

#### Un diccionario accesible (y asequible)

- precio asequible;
- definiciones formuladas con un lenguaje o estilo fácilmente comprensible (accesible);
- solución a la opacidad de las abreviaturas, etc. (accesible);

Principales diccionarios de español que permiten acceso en línea, gratuito y sin suscripción previa (selección subjetiva):

- DLE: <https://dle.rae.es/?id=DglqVCc>
- Clave: <http://clave.smdiccionarios.com/app.php>
- Diccionarios Larousse-Vox: <https://www.diccionarios.com>
- DEM: <http://dem.colmex.mx/Default.aspx>

#### Un diccionario que no toma partido

Problemas pragmáticos:

- Tratamiento del contexto cultural (monolingües);
- Tratamiento de equivalencias en los bilingües;
- Redacción de las definiciones en los diccionarios monolingües.

Causados por las relaciones entre lengua(s) – cultura(s) e ideología(s).

#### Proactividad y presencia de las ideologías

Lema	DLE-2014 en papel	DLE-2014 en línea	Clave
Castidad	2 Virtud de quien se abstiene de	2 Virtud de quien se abstiene de	2 Renuncia a todo placer sexual.

	todo goce carnal.	todo goce carnal.	
Frágil	3. Dicho de una persona: Que cae fácilmente en algún pecado, especialmente contra la castidad	3. Dicho de una persona: Que cae fácilmente en algún pecado, especialmente contra la castidad.	1. Que se quiebra o rompe con facilidad    2. Débil, poco resistente o que se estropea con facilidad
Gitano	5. <b>trapacero</b> U.t.c.s.	5. <b>trapacero</b> U. como ofensivo o discriminatorio U.t.c.s.	4 <i>col desp.</i> Que estafa o que actúa con engaño
Hechicería	1. Arte supersticioso de hechizar	1. Arte supersticioso de hechizar	Conjunto de conocimientos y poderes sobrenaturales con los que se pretende dominar los acontecimientos y las voluntades
Masoquismo	1. Perversión sexual de quien goza con verse humillado o maltratado por otra persona.	1. Perversión sexual de quien goza con verse humillado o maltratado por otra persona.	Tendencia sexual que consiste en obtener disfrute erótico al someterse alguien a los malos tratos o a las humillaciones de otra persona.

- La presencia de cultura e ideología es constante.
- Dubois (1971): discurso lexicográfico como un discurso sobre la lengua y sobre la cultura;

- Forgas (2001): diccionario presenta una visión tamizada y normalmente orientada por las élites ilustradas y la parte de la sociedad que detenta el poder.

### *Un diccionario más didáctico*

- Accesibilidad;
- Claridad;
- Usabilidad;

Concepción Maldonado (2012):

- “un buen diccionario es siempre un buen instrumento educativo”;
- Debe haber “una relación de compromiso” entre los diccionarios y la educación.

Orientación didáctica:

- desterrar clichés
- ofrecer guías de ejercicios
- adaptarse a la edad y el nivel de los aprendices
- permitir la aplicación en el aula
- adaptar los repertorios a las nuevas tecnologías.

### *Corpus y bases de datos*

Corpus lingüísticos: son una colección de textos orales y escritos reunidos con la finalidad de ser una muestra de uso de una o más lenguas durante una temporada dada, y en general anotados (aunque no necesariamente).

Criterios para la clasificación de los corpus:

- Lenguas que son objeto de recopilación (monolingües – multilingües);
- Modalidades discursivas que se recogen (orales – escritos);
- Perspectiva temporal (sincrónicos – diacrónicos)

### *Corpus generales (o de referencia)*

- Una muestra representativa del uso de una lengua en un periodo determinado;
- Representativo = suelen ser amplios y recogen textos orales y escritos procedentes de distintos ámbitos.

Ejemplos:

- CEMC – 66,80% lengua culta, 11,70% lengua inculta, 21,50% lengua no estándar (Lara 1974);
- CORPES XXI – 90% proviene de la lengua escrita (Rojo 2013); <https://www.rae.es/recursos/banco-de-datos/corpes-xxi>
- CREA: <https://www.rae.es/recursos/banco-de-datos/crea>
- Corpus del Español (CdE) – Mark Davies <https://www.corpusdelespanol.org>

### *Corpus específicos*

No siempre están dispuestos a la consulta libre del público.

- COSER: <http://www.corpusrural.es>
- CEMC: <http://cemc.colmex.mx>
- Cordiam: <http://www.academia.org.mx/Cordiam>
- Corpus CHARTA: <http://www.corpuscharta.es>
- Corpus Cumbre: base para los diccionarios de SGEL