

ARDIDES: Aplicación de Redacción de un Diccionario Diacrónico del Español*

PILAR SALAS QUESADA

Fundación Instituto de Investigación Rafael Lapesa

ABELARDO TORRES MORCILLO

Fundación Instituto de Investigación Rafael Lapesa

1. INTRODUCCIÓN

Tal y como se ha señalado en diversos trabajos (Pascual y García Pérez 2007, 2008), el *Nuevo diccionario histórico de la lengua española* (en adelante, *NDHE*) se ha concebido desde sus inicios como una obra relacional. Esta era una aventura que hubiera resultado imposible sin contar, no ya con un proyecto asentado en los pilares de la lingüística, sino, sobre todo, sin disponer de la tecnología informática.¹ En efecto, esa condición relacional de esta obra ha sido posible al haberse ideado como un repertorio electrónico, sin las constricciones que impone el formato tradicional de libro;² sin embargo, esta ventaja (que suponía no tener que adaptar a las posibilidades de la informática un diccionario planificado para su redacción y consulta en papel), al concebirlo de nueva planta y destinado a su publicación electrónica ha supuesto (y continúa suponiendo) un reto estimulante —y a veces peligroso—, que ha obligado a diseñar (o

* Todo el esfuerzo y la labor que relatamos a continuación no habría sido posible sin la colaboración entre los filólogos e informáticos que trabajan y han trabajado con nosotros desde el comienzo de este proyecto y a los que dedicamos este artículo. Del mismo modo, queremos agradecer sinceramente a Mar Campos, José Antonio Pascual y Octavio Pinillos su confianza en nosotros para escribir estas notas y también su inestimable ayuda con la lectura, corrección y sugerencias en este artículo.

¹ La concepción del diccionario como obra relacional, donde los valores se conectan con sus sinónimos, antónimos, hiperónimos, etc., e incluso con otros miembros de su misma familia, conecta este proyecto con el *Diccionario histórico de la lengua española* que la Real Academia Española publicó entre los años 1960 y 1996, puesto que ya R. Lapesa afirmaba en su Prólogo que «el conocimiento exacto del vocabulario solo puede lograrse teniendo en cuenta la situación de cada palabra dentro de su respectivo campo semántico y estudiando a la vez los valores de sus sinónimos, satélites y antónimos» (*DHLE*, 1972: VII-VIII).

² No se trata de contraponer, como adversarios irreconciliables, al diccionario-libro frente al diccionario electrónico, puesto que, por ejemplo, una versión abreviada de los artículos redactados para el *NDHE* podría convertirse en el futuro en una publicación en papel, aunque con una reducción notable de los datos (y, particularmente, de los ejemplos) contenidos en ellos. Acerca de los diccionarios digitales y analógicos puede verse, entre otros Gelpí (2003); para un primer acercamiento a la relación entre la informática y la lexicografía práctica puede consultarse Atkins y Rundell (2008: esp. 106-117).

adaptar) aplicaciones informáticas específicas para la consulta de los materiales y para las tareas de redacción, factibles solo si se lograba establecer una colaboración estrecha entre lexicógrafos e informáticos.³

No es necesario demorarse en señalar las considerables ventajas que las versiones digitales (o digitalizadas) de algunos diccionarios históricos (como la del *Oxford English Dictionary* o la del *Trésor de la langue française*) ofrecen a los usuarios, pues su mayor flexibilidad en la consulta permite obtener un amplio abanico de informaciones sin obligar a que el consultante deba adquirir, por otra parte, unas destrezas diferentes de las precisas para desplazarse por internet.⁴ Entre las características de un diccionario electrónico ha de estar el aumento de posibilidades de acceso al inventario, así como dar cuenta de la rica red de relaciones que se teje entre las palabras de una lengua, sin que los usuarios sientan el vértigo de quedar atrapados en un enmarañado laberinto. Somos conscientes de que probablemente en este momento solo nos estamos asomando a un pequeño porcentaje de la riqueza y variedad de consultas que se podrá realizar a una obra como el *NDHE*; sin embargo, en la confección de las aplicaciones se ha procurado tener presentes, en la medida de lo posible, las necesidades de esos usuarios potenciales, de acuerdo con las facilidades que nos brinda la técnica informática. Se han puesto al servicio del quehacer lexicográfico los medios que suministra la informática para intentar potenciar la calidad de la obra (o, si se prefiere, reducir el margen de error), facilitar las tareas de redacción y, finalmente, dotar a la consulta de una gran versatilidad.

Si perseguimos la satisfacción de cualquier tipo de usuario (y no solo del filólogo que conoce a la perfección las características de este género lexicográfico), la variedad

³ En cualquier caso, consideramos que conviene mantenerse alejados de triunfalismos (y, por tanto, mantener una actitud prudente) ante las innegables aportaciones de la informática y de la lingüística de corpus, pues, como advierte Kirkness, aunque es obvio que «the future of lexicography is undoubtedly electronic» y que uno de los factores más importantes en la evolución de la lexicografía en los últimos años se ha debido al «dramatic impact of the computer», también ha de asumirse que «however much the computer can aid lexicographers as dictionary writers, it will not replace them» (2006: 57). Por consiguiente, el lexicógrafo continúa siendo la pieza esencial del proceso: «The advent of electronic corpora and media can make lexicographer's work better, but not necessarily easier. Computers can store and process quantities of textual data quite unmanageable by humans. Where several million manually and painstakingly excerpted citation slips were once considered a sufficient basis for a multi-volume scholarly dictionary, now even one-volume trade dictionaries rest on hundreds of millions of rapidly and automatically entered running words. The differences are not only in quantity, but more importantly in quality. Lexicographers now have at their disposal vastly superior language data. [...] Lexicographers have been at the Forefront in utilizing language corpora as applying the findings of corpus linguistics [...] to good effect in their analysis and description of lexis and hence to the benefit of their users. The corpus revolution is very real; computerphoria would be misplaced, however. There may be huge savings in storage space and processing time, but it is humans who continue to choose the texts and analyze the vastly increased data, which can now in fact require more time, experience, and skill to process than before» (2006: 56).

⁴ La versión electrónica del *Diccionario histórico de la lengua española* (1960-1996) que se ha realizado en la Fundación Rafael Lapesa no solo facilita el acceso a una obra cuyos fascículos solo se pueden consultar en algunas bibliotecas (o adquirir hoy en librerías de viejo), sino que, además, permiten efectuar un amplio abanico de búsquedas en el cuerpo del repertorio.

de modos de consulta, así como la capacidad de albergar multitud de datos y mostrarlos de forma escalonada, estableciendo relaciones entre las palabras o sus acepciones, resulta imprescindible afrontar el proyecto contando, como hemos dicho ya, con el trabajo común de informáticos y lexicógrafos, para diseñar una herramienta que soporte toda esta información y gestión de datos, garantizando la comodidad, la rapidez, la fiabilidad y la seguridad del producto.

2. LA CONEXIÓN ENTRE LEXICOGRAFÍA E INFORMÁTICA

Por ello hemos abierto y cerrado el apartado anterior con esta lograda colaboración entre lexicógrafos e informáticos, pues una obra de estas características no puede desarrollarse sin el trabajo conjunto y coordinado de ambos. De este esfuerzo común procede la herramienta de redacción que se presenta aquí, *ARDIDEs* que, partiendo de la planta del diccionario, ha ido tomando forma a medida que los lexicógrafos exponíamos nuestras necesidades y los informáticos iban respondiéndonos con propuestas y sugerencias, hasta que con los esfuerzos de unos y otros se ha conseguido un artificio sin el que no sería posible un diccionario electrónico de este tenor. Veamos brevemente cómo se ha logrado.

Mientras se iban preparando los materiales que habrían de servir de base para la elaboración del *NDHE* (véase, a este respecto, el apartado 3) y tras llevar a cabo un examen exhaustivo de algunos diccionarios históricos, se estableció el esquema básico del artículo lexicográfico, procurando que se adaptase a la variedad de entradas del diccionario, a la vez que intentando entrelazar algunos de los hilos de la trama que habría de relacionar las palabras y acepciones incorporadas a la obra. En definitiva, se redactó una primera versión de la planta, en la que se describían las características esenciales de la macroestructura, de la microestructura e incluso de la hiperestructura del *NDHE*.

Con el fin de probar lo expuesto en la planta, comenzamos experimentando una aplicación de redacción comercial y personalizada para la elaboración del *NDHE*; evidentemente, aunque tuviese una vocación provisional (casi cabría decir efímera), la elaboración de esta herramienta consumió notables recursos, pues si bien se partía de un editor XML (*InfoPath* de Microsoft Office), hubo que adaptarlo a las exigencias derivadas de la estructura del artículo. De este modo, desde finales del año 2009 se pudieron iniciar las tareas de prerredacción del diccionario; hablamos deliberadamente de prerredacción, puesto que la herramienta no permitía, en ese estado, aspirar a confeccionar versiones completas de los artículos (pues entre otras cosas, no soportaba un número elevado de ejemplos ni resultaba factible establecer algunos tipos de relaciones entre las palabras, como las genéticas), ni mucho menos garantizaba la integridad referencial de los materiales (asunto al que volveremos más adelante). Pese a las rigideces de esa aplicación, que enseguida nos hizo ser conscientes de la necesidad de sustituirla por otro artefacto mucho más potente y más acorde con la planta del diccionario, consideramos que ha sido un peldaño sin el cual no habríamos podido llegar hasta *ARDI-*

DEs. Comprobadas, pues, la falta de flexibilidad y las deficiencias de una herramienta que, por otra parte, había nacido como un recurso temporal de apoyo a la labor lexicográfica, se empezó a desarrollar el embrión de la actual herramienta de redacción, que se ha convertido en la verdadero trasunto de la planta del diccionario.

A nadie se le oculta que no podría existir un diccionario histórico sin unas amplias bases documentales; por ello, antes de internarnos en las características de la herramienta de redacción haremos una breve incursión en las aplicaciones que se han desarrollado para la consulta y gestión del *Corpus del Nuevo diccionario histórico* (en adelante, *CDH*). No es nuestro objetivo detallar el funcionamiento de todas y cada una de las aplicaciones y utilidades informáticas construidas para la elaboración y consulta del diccionario, cuya explicación requeriría mucho más espacio (como se puede apreciar, por ejemplo, en los manuales de usuario realizados para la interfaz del corpus), pretendemos de una manera sencilla explicar el propósito, el alcance, las pretensiones y las características principales de nuestro trabajo. Veamos, pues, los rasgos esenciales de la interfaz de consulta del *CDH*.

3. LA ELABORACIÓN DE LOS MATERIALES: EL *CDH*

Aunque, como ya se ha apuntado en alguna ocasión, el *NDHE* no es un diccionario de corpus (entendiendo por tal aquel que se basa de modo exclusivo en los datos suministrados por un corpus concreto), una de las primeras tareas que se planteó en el marco del proyecto del diccionario fue la necesidad de confeccionar un corpus restringido, de textos controlados, que albergara, en una primera fase, «un repertorio de textos representativos para el léxico general de todas las épocas, géneros y temas» (Pascual y Domínguez 2009: 79); en resumen, un corpus nuclear de la lengua española —idea, por otra parte, nada revolucionaria, pues respondía al canon pidaliano—, con aproximadamente cincuenta millones de palabras. La tarea, no obstante, fue costosa, pues la nómina del corpus se construyó tomando en consideración la representatividad de las obras, según su importancia, difusión, influjo en otras obras, la variedad o especificidad de su léxico e, incluso, la calidad de las ediciones disponibles. El número total de ocurrencias es de más de 53 millones, de los cuales 32 pertenecen a textos españoles y más de 20 millones a obras americanas; la distribución temporal también es significativa, con una ponderación en los períodos medieval, Siglo de Oro y moderno, en la proporción siguiente: en la Edad Media 12 millones, en el Siglo de Oro 10 y en los siglos XVIII y XIX, 11 millones. En el siglo XX, en este corpus orientativo, se ha decidido forzar la proporción para dar entrada a un número adecuado de datos americanos: 19 millones, de los que 15 son americanos. En los siguientes gráficos se muestran estos porcentajes:

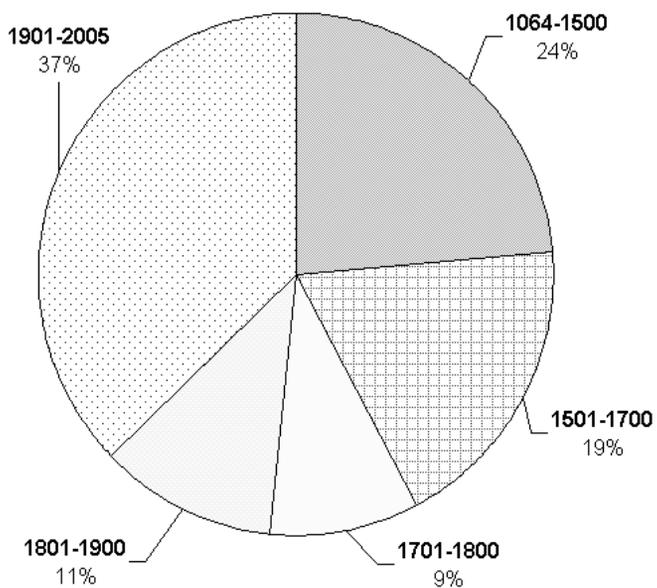


Gráfico 1. *Distribución por períodos temporales*

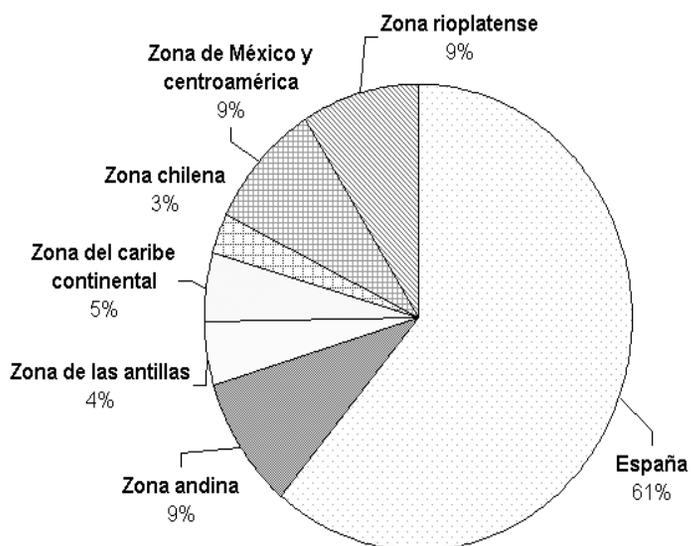


Gráfico 2. *Distribución por áreas lingüísticas*

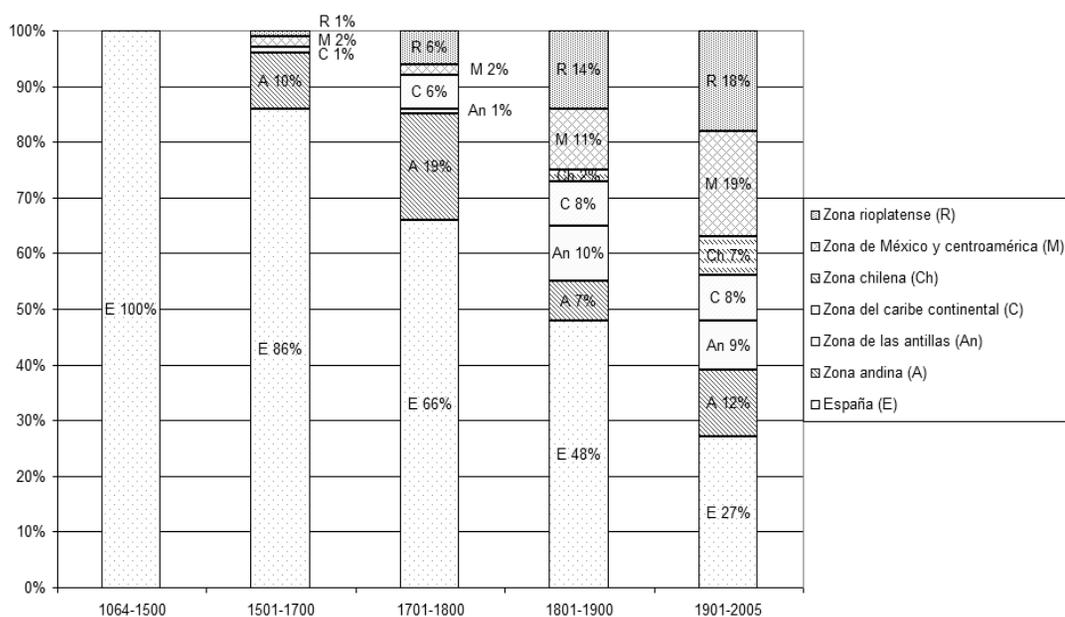


Gráfico 3. Distribución por áreas lingüísticas y periodos temporales

En buena medida, el *CDH* es un subconjunto del *CORDE* y el *CREA* (si bien se incluyen algunas obras que estos corpus no habían incorporado), pero lo novedoso no es este conjunto provisional de obras —que en el futuro requerirá algunos cambios a los que no podemos referirnos aquí—, sino el tratamiento que han recibido los textos para poderlos integrar posteriormente en la herramienta de redacción, ya que se han sometido a un proceso semiautomático de anotación lingüística; esta lematización, efectuada por el Departamento de Tecnología de la Real Academia Española, requiere aún de un refinamiento que solo podrá lograrse con las actualizaciones que se van derivando del uso del corpus en las tareas de redacción. En una fase posterior, se afrontó el diseño de una aplicación de consulta del *CDH*, que permitiese buscar y gestionar la valiosa información textual y lingüística que contiene. Esta interfaz, que se ha ido desarrollando a través de diferentes prototipos, se presenta al usuario de un modo gradual, pues parte de una vista básica que permite realizar consultas sencillas, para ir desplegando nuevos menús, aumentando escalonadamente los criterios de búsqueda (y, por consiguiente, incrementando el nivel de complejidad de la consulta) en función de los intereses del consultante.

La aplicación de consulta del *CDH* ofrece, por otra parte, la posibilidad de ordenar las concordancias obtenidas mediante diversos criterios de clasificación. Esto resulta de una indudable utilidad para los redactores del diccionario, pues permite incorporar criterios cronológicos, diatópicos, temáticos, gramaticales o de combinatoria léxica. Además de estas búsquedas (o de las que restringen la consulta por autor u obra), la interfaz permite descartar (o seleccionar) los testimonios del vocablo contenidos en

ciertas zonas del texto (previamente marcadas) que muestran unas características especiales; así, por ejemplo, se pueden discriminar las documentaciones incluidas en una glosa, una cita o en un cambio de mano (que aparecen destacadas en otro color).

Concordancias		Coapariciones		Estadísticas		Configuración		Ayuda		Sugerencias	
Concordancia simple		Concordancia ampliada									
Lema	adarga	Forma	Clase de palabra (Todos)			<input type="checkbox"/> Grafía original	+ Proximidad	Buscar	Nueva consult		
Subcorpus		Título	Autor	Fecha de creación							
<input type="button" value="Ocultar"/> <input type="button" value="Eliminar"/>		Origen	(Todos) América España								
		Tema	(Todos) Actualidad Aficiones								
		Codificación del corpus	<input checked="" type="radio"/> Buscar solo en <input type="checkbox"/> Todos <input type="radio"/> No buscar en <input checked="" type="checkbox"/> cambio de mano <input type="checkbox"/> cita <input type="checkbox"/> corrección <input type="checkbox"/> desarrollo <input type="checkbox"/> glosa <input type="checkbox"/> idioma extranjero <input type="checkbox"/> leyenda <input type="checkbox"/> nota								
REFERENCIA		CONCORDANCIA						Ordenar por: Año ascendente			
1	1435 Es	cavallo, desarmado, como oyó la voz, vino a mi en su cauallo, e solo en vna adarga ; e estovo en todo el fecho en el mayor peligro fasta el v									
2	1435 Es	fecho en el mayor peligro fasta el vencimiento, e fué ferido de vn viratón en el adarga que le pasó a los pechos, el qual estubo con asaz trau									
1 - 2 <input type="button" value="Imprimir"/>								página: 1 Ir a página: <input type="text"/> <input type="button" value="Ir"/>			
1435 ANÓNIMO, fragmento (a1454 Carrillo de Huate, Pedro, <i>Crónica del halconero de Juan II</i>) [España] [Juan de Mata Carriazo, Madrid, Espasa Calpe, 1946] Historia											
<small>e Gonçalo Fernandes, fijo del alcaide de los donzeles, e Alfonso de Valençuela, e Juan de Daça, e Fernando de Cárdenes, alcaide de Agular, que fué ferido de vna saeta por la pierna, e Pero Nuñes de Canbrana fué ferido; e asy a este como a los otros mataron e ferieron cauillos. E faziendo por vuestro seruicio, señor, el vuestro seruidor Alfonso Gonçales de León, andando faziendo screuir las fazas de las tierras, en vn cavallo, desarmado, como oyó la voz, vino a mi en su cauallo, e solo en vna adarga; e estovo en todo el fecho en el mayor peligro fasta el vencimiento, e fué ferido de vn viratón en el adarga que le pasó a los pechos, el qual estubo con asaz trauajo, con mucho deseo de servir a vuestra merced.</small>											

Figura 4. Consulta de adarga en una cita

Del mismo modo, podemos efectuar consultas teniendo en cuenta el contexto (lemas, formas o categorías de los términos vecinos al buscado) en el que se inserta el término. Y esta búsqueda se puede combinar, a su vez, por medio de operadores lógicos *booleanos* (Y, O y NO) con otras consultas complejas sobre el mismo término o términos distintos, ampliando las posibilidades que acabamos de describir para las consultas simples.

El avance más significativo, frente a otros corpus, radica en la lematización, pues podemos buscar una palabra acudiendo a su forma canónica (el lema), con el correspondiente ahorro de comodines o, incluso, de imaginación, que en ocasiones debía expresirse al máximo para deducir todas las variantes gráficas posibles de un vocablo, con sus respectivas formas flexivas e incluso con sus derivados apreciativos. Así, con *arcabuz*, cuando buscamos ejemplos a través de la casilla de Lema, la herramienta nos muestra casos de *alcabuz*, *arcabuces*, *arcabuz*, *arcabúz*, *arcabuzes* y *arcabuzicos*; del mismo modo que de *traer*, obtenemos todos los ejemplos que responden a las 923 formas registradas del verbo: *traher*, *traigáis*, *truxo*, *traxesse*, *troxieron trayo*, *traya*, etc.

CONCORDANCIA		Ordenar por:	Año ascendente
De omne qui abuerit suspecta de	omicidio	.	
de uicino pectet C morabetinos in auro, & pectet el	homizilio	, et diuidant per tres partes istos C morabet	
s fiadores paguet; altero uernes paguet al azor & el	homizilio	; & si non inuenerint C morabetinos, illum q	
ien suas casas in terra, & exeat inimico, & pectet el	omizidio	, & si fuerit ei probatum de dia cum	
si non, sua iura; et qui lo matare, suo senior coiata el	homizidio	.	
el, plena de iudicios. Justicia manie en ella, e agora	homicidio	. To argent es tornado escoria, to vino es bu	
racion non vos ayré. Vuestras manos son plenas de	homycidios	". *	
e con él conteçel grant error. Esta faz a los omnes	omeçidios	obrar, fázeles a las madres los sus hijos me	
or: «¡Vía, dixo, daquende, falso & traydor!, as fecho	omeçidio	& muy gran trahición, non te prendré	
	Omizidio	demandar. .xxxvj.	

Figura 5. *Concordancias del sustantivo homicidio*

Las ventajas —no haría falta decirlo— son mayores aún para los verbos, especialmente para los irregulares, ya que evita la búsqueda de cada grafía y de los comodines, y simplemente buscando el infinitivo nos conduce a todas las formas conjugadas con testimonios en el *CDH*, tal y como se aprecia en la figura 6.

CONCORDANCIA		Ordenar por:	Año ascendente
, tanto avién el dolor, de las sus bocas todos	dizían	una razón: — ¡Dios, qué buen vassallo, si c	
ianas, ascóndense de mio Cid, ca no l'osan	dezir	nada. El Campeador adeliñó a su posada, i	
on todas sus virtudes santas.— Esto la niña	dixo	e tornós" pora su casa. Ya lo vee el Cid, qu	
io. Fabló Martín Antolínez, odredes lo que á	dicho	: — ¡Ya Canpeador, en buen ora fuestes nac	
ngar las metamos que non sea ventado. Mas	dezidnos	del Cid, ¿de qué será pagado o qué ganaru	
nguados, á menester seiscientos marcos.—	Dixo	Rachel e Vidas: — Dárgelos hemos de gra	
rebos avemos que nos dedes los marcos.—	Dixo	Rachel e Vidas: — Non se faze assí el mer	

Figura 6. *Concordancias de un verbo*

Conviene advertir, sin embargo, que ha de considerarse la lematización del corpus como una operación progresiva, paralela al avance de redacción de los artículos lexicográficos. No vamos a ocultar que la anotación lingüística del *CDH* dista mucho de la perfección, como, por otra parte, era esperable, pues el corpus no deja de ser un recurso para realizar distintos productos o investigaciones sobre la lengua, que había que lograr de una manera semiautomática, con el menor esfuerzo y mayor rapidez posibles. De ahí que un pequeño porcentaje de las formas del *CDH* carece de lema (si bien están localizadas mediante una etiqueta concreta, que facilita su ulterior estudio), en tanto que una parte considerable de las ocurrencias recibe más de un análisis (es decir, su anotación es ambigua); este hecho se explica porque parece preferible brindar a los

usuarios (y, en concreto, a los redactores del *NDHE*) las opciones de análisis que sugiere una forma concreta a optar en exclusiva por una de ellas.⁵ La corrección se va haciendo de un modo gradual a partir de dos niveles, en virtud de las tareas de redacción del diccionario: en el primer caso, se trata de un proceso que conduce a quitar ruido al corpus, dado que con frecuencia los nombres propios, por ejemplo, se asignan a lemas erróneos (cfr. *tremecer=tremecen* tercera persona en plural del verbo y *Tremecén*, topónimo; o el nombre *Calila* como resultado de la búsqueda del verbo *caler*). En el segundo caso, se persigue que la integración de los materiales en la herramienta de redacción redunde en beneficio de la anotación lingüística del corpus; de este modo, cuando un redactor se enfrenta al artículo del sustantivo *lanza* y asigna los ejemplos del corpus a la acepción correspondiente del sustantivo está, al tiempo, refinando la lematización, pues resuelve la ambigüedad de *lanzas* y *lanza* (y, por tanto, solo quedarán libres los ejemplos de estas formas como verbo, que se vincularán a *lanzar*).⁶

Una de las utilidades más valiosas para los redactores del diccionario consiste en la consulta de las coapariciones, que agrupa los resultados en bloques atendiendo a los vocablos con que la palabra buscada figura en los textos con una frecuencia mayor de la que sería explicable por el azar. Se trata de una herramienta singular, porque hace una búsqueda de introspección automática sobre el propio corpus, es decir, que, mientras las otras búsquedas son guiadas por el usuario, esta consulta la dirige la propia aplicación. Este sistema de búsqueda proporciona un primer índice de la concordancia, esto es, da cuenta de las particularidades del comportamiento combinatorio de una voz en este corpus. En el estudio de las palabras de una frecuencia elevada, las coapariciones son fundamentales como punto de partida, ya que pueden orientar al redactor y darle pistas sobre cómo organizar las acepciones o incluso indicarle el camino hacia otras acepciones que hayan podido pasar desapercibidas; del mismo modo que, las combinaciones anómalas pueden ser la pista de un error o de una ruptura «literaria» de lo esperable.⁷ En la figura 7 tenemos un ejemplo de consulta sobre la utilidad de las coapariciones.

⁵ Dejando a un lado casos complejos, como el de los participios (cuyo carácter híbrido aconseja una anotación ambigua), probablemente podría sugerirse un análisis preferido (entre los posibles), de acuerdo con medidas estadísticas.

⁶ Dentro de este segundo caso, habría que mencionar, en un apartado especial, otras atribuciones erróneas, perfectamente esperables, a nuestro entender, en un corpus diacrónico que se anota mediante un procedimiento semiautomático: no es sorprendente que *yogo* no se localice bajo *yacer* y, en cambio, aparezca bajo un lema *yogo* (sustantivo) o *yogar* (verbo).

⁷ Las coapariciones se pueden ordenar por frecuencia o de acuerdo con tres medidas estadísticas de asociación: *MI* (*Mutual Information*), *T-score* o *LL* simple (*Log likelihood*) (Curran 2004, Evert 2008, Hoffmann 2008). No se nos escapa el hecho de que un corpus pequeño como el *CDH* no permite extraer conclusiones de gran alcance sobre las llamadas, a veces, colocaciones; simplemente se presentan aquí como una ayuda más al trabajo lexicográfico.

Coaparición básica				
<input checked="" type="radio"/> Lema	<input type="radio"/> Forma	<input type="text" value="romper"/>	<input type="checkbox"/> Combinaciones de clase léxica	Clase de palabra
				(Todos) adjetivo adverbio
Número de resultados	<input type="text" value="40"/>	Ordenar por:	<input type="text" value="LL Simple"/>	Tolerancia:
				<input type="text" value="5%"/>
<input type="button" value="Coapariciones"/>		<input type="button" value="Nueva consulta"/>		
	f	MI	LL SIMPLE	T-SCORE
silencio	226	3.82	777.02	52.51
vidrio	112	4.89	543.81	55.82
cadena	116	4.47	498.8	48.59
cristal	117	4.33	480.35	46.14
romper	140	2.85	313.56	27.48
lazo	77	4.29	312.12	36.87
pañó	89	3.69	291.77	31.34
pedazo	85	3.71	280.51	30.84
dique	35	6.75	258.24	60.84
cabeza	200	1.95	245.28	20.67
hilo	74	3.57	231.01	27.18
lanza	82	3.28	226.55	25.4
saco	58	4.04	216.13	29.04
velo	53	4.09	201.04	28.32
llorar	124	2.28	196.26	19.58
ola	57	3.82	196.08	26.38
el	11269	0.19	193.42	14.96
vínculo	38	4.96	188.02	33.35

Figura 7. Consulta de las coapariciones del lema romper

La herramienta permite que el usuario pueda acceder a las concordancias concretas de cada coaparición, como se puede apreciar en la figura 8:

Tüencio ovieron con sue cuita a	romper	el <i>silencio</i> ; dissieron: «Compañeros,
ado de tancto rey e de mi señor,	rompí	el <i>silencio</i> , poniéndome por señal pa
n su plazentería abra su boca e	rompa	el <i>silencio</i> fasta esa hora guardado, e
s palabras, aquel Patróculo non	ronpe	el <i>silencio</i> , mas calla y menospreçia l
ándola por entrañal sentimiento,	ronper	el <i>silencio</i> a la péñola, sea verdat mu
por ser al suyo conformado, que	rompió	el <i>silencio</i> que tenla de su voto, y má
i tienen la lengua fria "col parlar	roto	e subito <i>silencio</i> ". * Y así por ventura
mi vida y costumbres; entonces	romperé	yo mi <i>silencio</i> , quando sepa dezir cos
a paciencia del amor libidinoso,	rompió	el <i>silencio</i> de lo que callaba mucho ti
divia calla y teme el punto, pero	rompió	el <i>silencio</i> y pena junto diciendo: «¡C
las vencedoras armas presto y	romped	el <i>silencio</i> ya escusado saliendo a to
¡ Fadrique, y, rato después, assi	rompió	el <i>silencio</i> :

Figura 8. Acceso a las concordancias del lema romper a partir de las coapariciones con silencio

Es imponderable el valor que tiene para el redactor el módulo de estadísticas que posee la herramienta de consulta del Corpus, al ofrecer la dispersión diacrónica y diatópica de una palabra de forma gráfica. Esto proporciona una visión general de cómo se distribuye dicha palabra a lo largo de los siglos o, ya en el período moderno, en los distintos países de habla hispana. Por otro lado, para facilitar la parte interpretativa de nuestro trabajo, las estadísticas proporcionan todas las formas gráficas de un lema o los lemas bajo los que puede aparecer una determinada forma; desde la estadística, además, es posible acceder a las concordancias del vocablo. En la figura 9 tenemos un ejemplo de consulta de las estadísticas.

«flechar» como «lema»

Frecuencia absoluta: 451

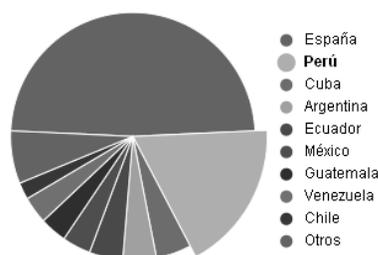
Frecuencia normalizada: 6.8

Dispersión diatópica: 81.081 %

País	f	f Norm
España	219	5.36
Perú	81	20.39
Cuba	21	10.79
Argentina	20	4.47
Ecuador	20	14.43
México	17	5.31
Guatemala	16	26.72
Venezuela	16	13.32
Chile	10	5.31
Nicaragua	10	24.34

1 - 10 de 14 página: 1 2

Dispersión diatópica

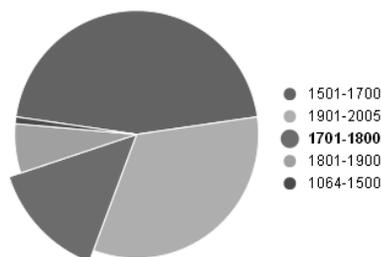


Dispersión diacrónica: 79.279 %

Periodo	f	f Norm
1501-1700	205	16.83
1901-2005	149	6.04
1701-1800	64	11.16
1801-1900	29	3.81
1064-1500	4	0.24

1 - 5 de 5

Dispersión diacrónica



Formas de este lema			
Forma	Categoría	f	f Rel
flecha	verbo	189	41.9
flechas	verbo	103	22.83
flechar	verbo	23	5.09
flechado	verbo	14	3.1
flechando	verbo	12	2.66
flechados	verbo	10	2.21
flecharon	verbo	10	2.21
flechada	verbo	9	1.99
flechassen	verbo	9	1.99
Flecha	verbo	8	1.77

1 - 10 de 43 página: 1 2 3 4 5



Figura 9. Módulo de estadísticas del lema flechar

El *CDH* no constituye, como se ha apuntado antes, la única fuente del diccionario; consiguientemente, no solo se ha integrado este corpus en la herramienta de redacción, sino que se está trabajando para incluir igualmente otras fuentes documentales. Así, en el primer trimestre del año 2012 se prevé que se hayan incorporado a *ARDIDES* el *CREA* (en una versión lematizada facilitada por el Departamento de Tecnología de la Real Academia Española) y una considerable proporción de las obras del *CORDE* (excluidos los textos latino-romances y aljamiados, que exigen un tratamiento particular). Para ello, se ha efectuado una anotación semiautomática de los 199 064 288 de registros contenidos en 3 767 obras (correspondientes a 27 860 documentos), mediante herramientas libres y recurriendo a los lexicones derivados de la anotación de otros corpus o bases de datos léxicas. En un horizonte algo más lejano se aspira a incorporar a ese motor de búsqueda otros recursos empleados de forma sistemática en la redacción del *NDHE*, como el *Nuevo tesoro de la lengua española* de la Real Academia Española, el *Fichero general* de la Real Academia Española u otros corpus específicos (como el *Corpus Léxico de Inventarios* dirigido por J. R. Morala) que están en proceso de conformación. Para ello es imprescindible que pasen el mismo proceso de control y anotación lingüística para garantizar siempre la fiabilidad y calidad de la información.⁸ Nos hemos referido más arriba a la imperfección de partida con que se han construido estos corpus; era una imperfección que estimábamos iría adelgazando progresivamente a lo largo de la redacción del *NDHE*. Añadamos que unos corpus concebidos para un trabajo que da lugar a obras lexicográficas consultables *on line*, no solo han de mejorar por el trabajo de los lexicógrafos, sino también por la participación de los consultantes.

⁸ El uso y la integración de una cantidad tan enorme de materiales documentales así como su procedencia tan diversa y su naturaleza tan diferente tiene un problema adicional que consiste en la posibilidad de que una misma obra esté incluida en varios recursos documentales, incluso que cada recurso documental tenga una información bibliográfica diferente. Para solucionarlo la informática nos proporciona una ayuda impagable, ya que se puede generar una nómina de todas las obras de todos los recursos documentales y cotejarlos de forma automática usando algoritmos de semejanza gráfica para detectar posibles obras repetidas o cuya información bibliográfica no sea ni siquiera idéntica. De esta forma evitamos el uso de obras repetidas, mejorando así la calidad del propio diccionario.

4. HERRAMIENTA DE REDACCIÓN

Hemos indicado antes, aunque de un modo tangencial, que la tecnología usada para almacenar la información en *ARDIDEs* (incluso de la herramienta provisional, realizada con *InfoPath*), así como la aplicación de visualización del diccionario, está basada en el metalenguaje XML, que usamos para almacenar y recuperar tanto la información manejada para la redacción del diccionario, como para albergar los recursos documentales asociados a él; para ello, necesitábamos un tipo específico de base de datos XML que pudiera asegurar el rendimiento de unas aplicaciones que operarían no solo con un gran volumen de información, sino también con datos de una notable complejidad.

Las clásicas bases de datos relacionales tienen el inconveniente de la rigidez de su estructura basada en tablas, frente a la flexibilidad del XML. La herramienta elegida como gestor de base de datos XML nativa (*Tamino*, de Software AG) ha permitido obtener una aplicación perfectamente estructurada, con la que se puede trabajar con seguridad y comodidad, pero que, al mismo tiempo, ofrece la posibilidad de adaptarse sin grandes esfuerzos a las nuevas necesidades de los lexicógrafos, por lo que es posible realizar modificaciones y extensiones de la planta de forma casi dinámica. Este carácter flexible ha sido especialmente útil en los inicios del diccionario, ya que, en esta fase, ha estado sujeto a ajustes y ampliaciones de la planta; además, esa peculiaridad nos ha brindado la posibilidad de disponer en un plazo razonable de tiempo de una herramienta básica que, como explicaremos, contenía los aspectos más importantes del artículo del diccionario (prescindiendo de muchos otros), pero que, de este modo (y con esa versión de mínimos), permitió reemprender los trabajos lexicográficos antes de que la dotáramos de toda una serie de ventajas, que hemos ido incorporando de forma gradual.

La tecnología XML nos permite definir un *esquema XML* para describir la estructura y las restricciones de los contenidos de la planta del diccionario de una forma precisa, de modo que, de esta manera tan cómoda, tenemos la seguridad de que no habrá información en la base de datos que viole estas restricciones, al tiempo que nos supone un ahorro de tiempo de programación y de posibles errores de validación. Así tenemos una equivalencia absoluta entre la planta del diccionario y el *esquema XML*, lo que también ayuda a la comunicación entre informáticos y lexicógrafos, ya que el traspaso de la información es más sencillo al plasmarlo a través del *esquema XML*.

Dado que el *NDHE* es un diccionario electrónico (que se publicará, por consiguiente, en red, acompañado de buena parte de los materiales usados para su elaboración), todo el software creado para el *NDHE* se basa en las últimas tecnologías para la generación de aplicaciones web (aplicaciones que los usuarios utilizan accediendo a un servidor mediante un navegador web); esto es, todas nuestras aplicaciones y recursos informáticos son susceptibles de aparecer en internet. De un modo particular, el uso de tecnología AJAX (acrónimo de *Asynchronous JavaScript And XML*) para el desarrollo de las interfaces web de nuestras aplicaciones permite a los lexicógrafos realizar cam-

bios sobre las páginas sin necesidad de recargarlas y, por lo tanto, aumenta la interactividad, velocidad y usabilidad —llamémoslo así— de nuestras herramientas.

En esta dirección, conviene advertir que en el diseño de *ARDIDES* se ha tenido muy en cuenta que la redacción del diccionario no tiene por qué estar centralizada en un determinado lugar, sino que pueden existir grupos de trabajo diferentes con redactores dispersos geográficamente, circunstancia que condiciona la manera de configurar este recurso. Esos grupos deberían poder acceder a las fuentes documentales del diccionario del mismo modo que accedemos quienes trabajamos en la Fundación Lapesa. Precisamente, el hecho de que, como ya dijimos anteriormente, todo el *software* generado para el *NDHE* esté basado en las tecnologías actuales para la generación de aplicaciones web, tiene como feliz consecuencia que los redactores solo necesiten, para efectuar su labor, una conexión a internet y un dispositivo con un navegador web (ordenador personal, ordenador portátil, teléfono móvil, tableta, etc.), lo que les permite, de hecho, redactar artículos del diccionario desde cualquier lugar del mundo (en la oficina, en casa, durante un algún viaje, etc.), con la ventaja adicional de poder usar todas las aplicaciones desarrolladas para el diccionario sin requerir la instalación adicional de ningún *software* en sus dispositivos.

En definitiva, para el diseño y elaboración de los diferentes programas informáticos que ayudan en la redacción del *NDHE* hemos aplicado los principios, métodos y técnicas más actuales en ingeniería del software, adaptados a las necesidades particulares de este proyecto. No nos importa insistir —es la tercera vez que lo hacemos—, en que la estructuración, planificación y control del proceso de desarrollo de software de todas las aplicaciones del *NDHE* ha supuesto una continua comunicación entre informáticos y lexicógrafos, en todas y cada una de las fases del desarrollo (análisis de requisitos, diseño, codificación, pruebas, implantación, mantenimiento, documentación, etc.) de cada aplicación creada en el marco de este proyecto. Hemos de reconocer que esta fue una condición impuesta explícitamente por los directores de la obra con respecto a nuestro trabajo.

La sección de la planta lexicográfica correspondiente a la microestructura y a la hiperestructura se ha trasladado —no sin problemas de todo tipo, que no es el momento de exponer—, a nuestra herramienta de redacción. Para diseñar la interfaz de usuario de *ARDIDES* de la forma más sencilla y amigable —según hoy se dice— para los redactores, se optó por construir una interfaz WYSIWYG (acrónimo del inglés *What You See Is What You Get*, que significa ‘lo que ves es lo que obtienes’), para que el lexicógrafo acceda fácilmente a la estructura del artículo, evitando de este modo las farragosas ventanas para redactar y visualizar las distintas partes de un artículo lexicográfico. De esta manera, cada redactor tiene a la vista en cada momento el esquema estructural del artículo y, al mismo tiempo, puede desplegar una sección concreta de la monografía y trabajar sobre ella, sin que su labor se vea entorpecida por la presencia simultánea y permanente, en la pantalla, de todos los apartados y subapartados de la ficha. Por otra parte, el lexicógrafo dispone de opciones de visualización básica del artículo, lo que le permite revisar con mayor fiabilidad y rapidez aquello que está redactando.

Antes de desgranar algunas de las características de *ARDIDEs*, desearíamos dejar constancia de que una aplicación como esta obliga a los redactores a ceñirse por completo a lo establecido en la planta. Esto, por un lado, sistematiza el trabajo y, por otro, entraña la obligación de tomar decisiones explícitas (teniendo que hacer patente todo lo que es dudoso, problemático o contradictorio), a la vez que, como veremos, ayuda a la automatización de algunos procedimientos en la redacción de los artículos, lo que supone un ahorro de tiempo en el trabajo de redacción.

Mientras los lexicógrafos empleaban la aplicación provisional de redacción (o mejor, como hemos dicho antes, de prerredacción) elaborada con un editor XML, se comenzó, pues, a esbozar la herramienta definitiva, para lo que se partió de un módulo *software* nuclear con los aspectos más importantes de la estructura del artículo, dado que la prioridad radicaba en brindar a los lexicógrafos, cuanto antes, un artefacto no solo más flexible y potente, sino también estable. Así, se construyó en un breve período de tiempo un prototipo funcional e incremental muy básico (y con algunas incomodidades en su manejo), pero que indudablemente estaba llamado a ser el germen de *ARDIDEs*: una herramienta de redacción que fuera evolucionando paulatinamente para ir incluyendo todos los aspectos funcionales imprescindibles para la elaboración del diccionario. Planteada de este modo, la propia herramienta de redacción constituyó un utilísimo instrumento (y una piedra de toque) para que los redactores fueran probando, refinando y mejorando la planta y, además, para que esa mejora sirviese simultáneamente para retroalimentar la evolución y perfeccionamiento de la aplicación de redacción.

En los próximos párrafos expondremos someramente algunas de las características esenciales de *ARDIDEs*. Al igual que la aplicación de consulta del *CDH*, la herramienta de redacción ofrece una primera pantalla de trabajo muy sencilla que permite ir seleccionando (y, en consecuencia, desplegando) las distintas tareas que se pueden realizar en el *NDHE*, desde la mera consulta de un artículo ya redactado a una vista para proceder a su corrección, pasando, por ejemplo, por la selección de las autoridades para cada una de las acepciones de un vocablo. En esta ocasión, nos detendremos en primer lugar en la pestaña de *Redacción del diccionario*, pues constituye la puerta de acceso a la estructura del artículo lexicográfico, dividido, a su vez, en distintas secciones (detailladas en una barra en la herramienta), como la de lema, donde se fichan las variantes gráfico-fonéticas y flexivas de una voz, así como sus derivados apreciativos, o la de repertorios lexicográficos, en la que se cargan automáticamente las referencias de los diccionarios más significativos en la historia del lexicografía del español que consignan el vocablo estudiado en la monografía.

Lema | [Etimología](#) | [Familia de la palabra](#) | [Acepciones](#) | [Formas complejas](#) | [Repertorios lexicográficos](#)

Resumen | Ocultar / Mostrar todo | Crear nuevo | **Guardar** | Lema en edición: arcabuz

↓ Artículo guardado en la colección definitivos

Alerta

LEMA Ordenar formas flexivas

Lema*	Encabezamiento Original	Fuente*	Orden*	Relev.*	Nombre*	Formas flexivas
arcabuz	<input type="checkbox"/> arcabuz	DRAE2001	<input type="checkbox"/>	significativo	alcabuz	Categoría* sustantivo Género masculino Número singular
	<input type="checkbox"/> arcabuz	NDHE	1			<input type="checkbox"/> Enviar a ... Cambiar a "Sin documentación"
	<input type="button" value="Ocultar enc. original"/>					
			2			Categoría* sustantivo Género masculino Número singular
						<input type="checkbox"/> arcabuces Categoría* sustantivo Género masculino Número plural
						<input type="checkbox"/> arcabuzes Categoría* sustantivo Género masculino Número plural

Figura 10. Herramienta de redacción: lemas y sublemas de arcabuz

Sin duda, uno de los apartados más complejos de la herramienta es el que se dedica a las acepciones; si bien no podemos desmenuzar la nutrida lista de datos que se asocian a cada una de ellas, nos gustaría destacar algunos de los aspectos que juzgamos de mayor interés. Si el *NDHE* ha de prestarse a múltiples consultas, en función de los diversos intereses de los usuarios, necesita incorporar una variada gama de informaciones que sustenten esas consultas; por citar un caso concreto, es obvio que si se va a ofrecer, en la consulta, la posibilidad de agrupar las acepciones del artículo de acuerdo con un criterio cronológico o en función de otro lógico-lingüístico, por ejemplo, habrán de consignarse esos datos en el artículo. Por defecto, las acepciones se ordenan cronológicamente en cada monografía, pero se pueden disponer atendiendo a otros criterios:

— Por frecuencia, partiendo de la acepción con mayor número de ejemplos; el empleo de esta pauta obliga a asignar todas las autoridades que se manejan en la redacción a una acepción o subacepción concreta del vocablo redactado.

— Por la relación genética que se puede establecer entre las acepciones del vocablo y las de su palabra base de derivación o, si nos hallamos ante una palabra de raigambre latina o ante un préstamo, por la vinculación etimológica entre esas acepciones y las que posee la voz en la lengua que la ha suministrado al español. Por ejemplo: *pavés*¹ como ‘escudo’, prestada del italiano *pavese* hacia mediados del siglo XIV, y *pavés*², tomado del francés *pavé*, ‘pavimento’, de finales del XX.

— Por un criterio lógico-lingüístico, que atiende a los mecanismos de cambio lingüístico (y, particularmente, de cambio semántico) que explican la creación de nuevos significados (como la metonimia, la metáfora o la analogía). Este es el caso de los

nombres de tipos de escudos, como *adarga*, *broquel*, *pavés* o *rodela*, pues todos ellos dan lugar a una acepción abstracta de ‘defensa o protección’. En la figura 11 se muestra una captura de pantalla del apartado de codificación de las relaciones semánticas entre acepciones.

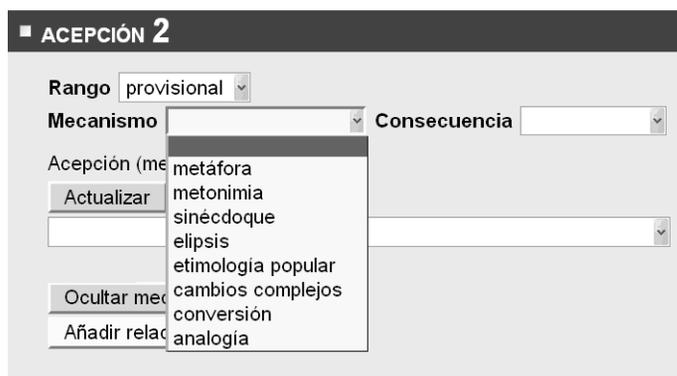


Figura 11 *Conexión entre acepciones: mecanismos de cambio*

En último lugar, dentro del artículo, figuran destacadas (con letra menor) las acepciones que solo arrojan testimonios lexicográficos, pues, si bien no permiten establecer este rico sistema de relaciones, constituyen una valiosa fuente de información.

Una de las tareas más costosas para los redactores del diccionario (por la considerable inversión de tiempo que supone) consiste en la clasificación de los ejemplos; debemos recordar que, en una primera etapa, el *NDHE* aspira a redactar las voces documentadas en el *CDH* (y, por tanto, a asignar cada una de sus ocurrencias a un vocablo y acepción determinados).⁹ Una de las novedades de *ARDIDEs*, que está facilitando enormemente la redacción de las fichas, consiste en la integración de los materiales en la herramienta de redacción. En la práctica, esto implica que, desde la interfaz (y a través de la pestaña de *Selección de ejemplos del CDH*), se efectúa una consulta al corpus, y se pueden ir seleccionando, marcándolos con el ratón, todos los ejemplos obtenidos para distribuirlos en las distintas acepciones de la palabra que se está redactando (y que aparecen alineadas en una columna en el margen derecho de la aplicación). De este modo, al volver a la página de redacción, los ejemplos se despliegan bajo cada acepción, ilustrándola, evitando así las farragosas operaciones de copiar y pegar (a las que hubimos de dedicarnos —con no poco esfuerzo— en la herramienta provisional de *InfoPath*), que no solo ralentizan considerablemente el trabajo y lo

⁹ Resulta obvio que no sería ni factible ni razonable pretender alcanzar esa meta en palabras como la preposición *de*, que muestra 3 181 806 de ejemplos en el *CDH*.

hacen tedioso, sino que además hacen planear siempre sobre el ánimo de los lexicógrafos el temor a incurrir en errores de teclado. Por otra parte, lo que obtenemos no es una copia limitada del texto original, sino una referencia al propio texto original, por lo que en cualquier momento el redactor podrá ampliar el contexto sin necesidad de tener que volver a buscarlo en la obra a la que pertenece la autoridad.¹⁰ Con este procedimiento el redactor también extrae de forma automática y transparente toda la información bibliográfica del ejemplo que está asignando a una acepción (autor, obra, año de creación, editorial, etc.), ya que no solo se establece un enlace con el texto original del *CDH*, sino con toda la información bibliográfica que contiene. En la figura 12 se pueden ver dos ejemplos de distintas fuentes en una misma acepción.

The screenshot shows the ARDIDES redaction tool interface. It displays two examples of authority data for the word 'mosquetazo'. Each example includes a text snippet, a 'Foco' button, and a metadata form.

Example 1 (ID: 28102):

- Rango:** publicableB
- Fuente:** CNDHE
- Carácter:** lexica
- Uso figurado:**
- Foco:** [Botón]
- verificado:**
- modo edición:**
- Text snippet:** Porque, llegando una tarde al campo de la verdad en busca del enemigo, le hallamos atrincherado entre unos castillos donde el triste Rostubaldo, que era un capitán valentísimo, con cincuenta arcabuceros llegó a reconocer el puesto y a ganalle dos trincheras al contrario, y al retirarse le dieron un *mosquetazo*.
- Metadata form:**
 - Autor:** Rojas Villadrando, Agustín de
 - Título:** El viaje entretenido
 - Localización:** [Campo]
 - País:** [Campo]
 - M.ini:** [Campo]
 - A.ini:** 1603
 - M.fin:** [Campo]
 - A.fin:** [Campo]
 - Orden:** 0

Example 2 (ID: 28131):

- Rango:** publicableB
- Fuente:** CORDE
- Carácter:** lexica
- Uso figurado:**
- Foco:** [Botón]
- verificado:**
- Text snippet:** ¡Que habiendo sido un furioso rayo de colera, impacientísimo en los tiernos años de su mocedad, sujetase con grande paciencia su robusta condición a servir en Flandes con tantas ventajas que templase la furia de los amotinados y pusiese su valeroso pecho a recibir los *mosquetazos* con que querían escalar y saquear su casa!
- Metadata form:**
 - Autor:** Espinel, Vicente
 - Título:** Vida del escudero Marcos de Obregón
 - Localización:** [Campo]
 - País:** [Campo]
 - M.ini:** [Campo]
 - A.ini:** 1618
 - M.fin:** [Campo]
 - A.fin:** [Campo]
 - Orden:** 0

Figura 12. Herramienta de redacción: autoridades del artículo mosquetazo

Tanto los ejemplos que proceden del *CDH* como los tomados de otras fuentes, se caracterizan en función de diversos parámetros: además de la fuente (*CORDE*, *CREA*, *Hemeroteca digital*, etc.) se determina su rango (es decir, si juzgamos que debe ofrecerse al consultante en primera instancia, en una consulta ampliada o si ha de someter-

¹⁰ La necesidad de preservar la integridad referencial obliga, no obstante, a someter las autoridades a un proceso de edición para que se ajusten a una vista convencional en el artículo lexicográfico, dado que la asignación que se efectúa desde el corpus se basa en las *sentences* identificadas en la anotación textual. La imposibilidad de cortar y pegar, que muestra su cara positiva en todo lo referido previamente, ofrece el lado menos amable (y más costoso, en lo que respecta al tiempo empleado) en el hecho de que en la herramienta de redacción se ha de seleccionar el fragmento del texto que se desea publicar y, si se considera necesario incluir marcas de supresión o comentarios editoriales, han de incorporarse mediante un sistema de marcas flotantes. Otra consecuencia poco grata de esta integración es que, si bien los artículos prerredactados con *InfoPath* (la herramienta comercial personalizada de los inicios del proyecto de la que ya hemos hablado) empleaban también el metalenguaje XML, deben someterse a un proceso de validación para poder ser integrados en *ARDIDES* y aprovechar el trabajo ya realizado en su momento, puesto que, por poner solo un caso concreto, las autoridades se cortaban y pegaban en cada acepción a mano (es decir, no se establecía la asociación biunívoca entre corpus y diccionario). En cualquier caso, la utilización de algoritmos de semejanza gráfica desarrollados para el *Mapa de diccionarios* facilita estas operaciones de validación y verificación para poder integrar de forma automática dichas autoridades en *ARDIDES*.

se a un examen más detallado), se indica si nos hallamos ante un testimonio léxico o lexicográfico (información que está predeterminada en función de la fuente, pues es evidente que el *CORDE* suministrará preferentemente testimonios léxicos), así como los datos bibliográficos completos que, como se ha apuntado en el párrafo anterior, se extraen directamente en el caso del *CDH* (y, por consiguiente, se arrastrarán del mismo modo cuando los otros materiales se hayan integrado en *ARDIDEs*). El lexicógrafo dispone, además, de la posibilidad de incluir comentarios sobre el ejemplo (anotaciones que, por otro lado, están sometidas a una notable formalización).

La integridad referencial asegura, por otra parte, que cualquier futura modificación de los textos originales (corrección de alguna errata, ampliación de información bibliográfica, modificación de la caracterización, etc.) revertirá automáticamente en todos los ejemplos de ese texto insertados en nuestro diccionario, sin intervención del redactor (aunque no sin su conocimiento). Además, este procedimiento de selección de los ejemplos garantiza que esa palabra pivote no vuelva a asignarse a otro lema (o al mismo en otra acepción o subacepción), de tal manera que los ejemplos quedan bloqueados y se evita el riesgo de repetir ejemplos.¹¹ Por fin, este mecanismo facilita también la elaboración de los artículos de voces de alta frecuencia en el corpus, puesto que se puede abordar también su clasificación mediante los subcorpus (buscando por épocas temporales, por zonas geográficas, etc. es decir, con el criterio que mejor nos convenga) o con el recurso a las coapariciones.

The screenshot shows a software interface with a menu bar at the top: 'Consulta', 'Redacción del diccionario', 'Selección de ejemplos del CNDHE', 'Vista NDHE', 'Vista revisión', and 'Configuración'. Below the menu bar is a sub-menu: 'Concordancias', 'Coapariciones', 'Estadísticas', 'Configuración', 'Ayuda', and 'Sugerencias'. The main window is titled 'Concordancia simple' and 'Concordancia ampliada'. It features a search bar with 'Lema clarinete', a 'Forma' field, and a 'Clase de palabra' dropdown set to '(Todos)'. There are also buttons for 'Grafía original', '+ Subcorpus', and '+ Proximidad'. The main area contains a table with columns for 'REFERENCIA', 'CONCORDANCIA', and 'Ordenar por: Año ascendente'. The table lists 15 entries with checkboxes, each showing a reference number, year, country, and a snippet of text with the word 'clarinete' highlighted. To the right of the table is a panel for 'Aceptaciones' (Senses) with a list: 'Instrumento musical de viento madera', 'Persona que toca el clarinete.', 'Esp. Pe. Coloq. Expresa afirmación o...', and 'Ho. «Casualidad, combinación de circ...'. Below this list are buttons for 'Forma Compleja', 'Aceptaciones', 'clarinete requinto', 'Clarinete pequeño de tono agudo.', and 'Aceptar'.

REFERENCIA	CONCORDANCIA	Ordenar por: Año ascendente
1 1791 Es	capilla de los niños dotinios, que tiene un yerno que toca la chirimía como un clarinete	
2 1817 Me	cuidado de proporcionar a los naturales, quien los enseñe al manejo de violines, bajo clarinetes, flautas y trompas, para mayor solemnidad	
3 1836 Es	se muda de casa se aparecen delante del balcón o en la escalera con el bombo y el clarinete	a dar y tomar en la canción de Carnicer. L
4 1839 Cu	mulatos músicos, tres violines, un contrabajo, un flautín, un par de timbales y un clarinete	El último de los instrumentos aquí mencio
5 1839 Cu	de su cuidado, o de su cariño, que como jóvenes quedaban en la sala. Pimienta, el clarinete	se mantenía en pie a la cabeza de la orqu
6 1839 Cu	hacia un arco con su cuerpo y sacaba los bajos más profundos, imaginables, y el clarinete	ejecutaba las más difíciles y melódicas vi
7 1839 Cu	animación y al movimiento difundidos por todas partes. Al pasar ella por junto al clarinete	Pimienta, lo tocó con el ahorcico en el braz
8 1839 Cu	grupo de jóvenes blancos, de seguidas pasó a la sala donde, después de amar el clarinete	.tocó algunos registros a fin de que entenc
9 1839 Cu	rumor en toda apariencia duplicaba el número de la orquesta. Bien claro decía el clarinete	en sus argentinas notas: caramelo vend,
10 1839 Cu	inquietud y su pasión no parece sino que encontraron desahogo por las llaves del clarinete	: se exhalaban, por decirlo así, según lo per
11 1839 Cu	verdica historia papel menos importante. Nos referimos ahora al célebre tocador de clarinete	. José Dolores Pimienta.
12 1839 Cu	que hacerse gran violencia, cosa que no podía echar a puerta agena, para trocar el clarinete	. su instrumento favorito, por el dedal o la a
13 1839 Cu	puerta de su casa, a tiempo que salía de ella su querido hermano Jose Dolores con el clarinete	en la funda debajo del brazo y un rollo de p
14 1839 Cu	algunos literatos habaneros; a Jose Dolores Pimienta, sastre y diestro tocador de clarinete	. tan agraciado de rostro como modesto y s
15 1858 Co	- Mi palabra. Manuela. ¿Oye usted la tambora y las trompas, y los clarinetes	y los flautines, y los cornabacetes?

Figura 13. Selección de ejemplos y asignación a acepciones del artículo clarinete

¹¹ Si en la revisión del artículo se descubre que se ha producido una asignación errónea, la herramienta permite seleccionar todas las autoridades que se deseen cambiar de ubicación y enviarlas a una acepción o subacepción diferente.

En la imagen superior se puede ver la ventana de integración, en la que, tras la búsqueda de un lema se van seleccionando los ejemplos, marcándolos en la casilla de la izquierda, y después asignándolos a las acepciones de la palabra que se muestran en la columna de la derecha y que, previamente se han establecido. Tras esta operación los ejemplos se visualizan del modo que se muestra en la figura 14:

Figura 14. Acceso a algunos ejemplos del artículo clarinete tras su asignación

El corpus sale asimismo beneficiado de ese proceso de integración: ahora se entenderá lo que decíamos antes respecto a la corrección de los corpus a partir del proceso de redacción. Empezando por su anotación lingüística, que se va aquilatando a la par que progresan los trabajos lexicográficos; y continuando por la corrección de erratas que perciba el redactor.¹² Incluso, en el futuro, está previsto trasladar al *CDH* la anotación semántica, de tal modo que pudiesen efectuarse consultas en el corpus no solo de un vocablo con un análisis gramatical determinado, sino incluso con una acepción concreta.

En el apartado de la acepción se han dispuesto otra serie de desplegables que permiten tejer la red de relaciones entre los significados de las palabras (hiperonimia, meronimia, sinonimia, antonimia...) e incluso dibujar las jerarquías taxonómicas. Estos

¹² Véase, a este respecto, Campos Souto y Pascual (en prensa).

vínculos se establecen también de forma controlada, pues se han de proponer entre vocablos y acepciones consignados en el *NDHE*, para evitar las pistas perdidas.

The screenshot shows a software interface titled "RELACION SEMÁNTICA". It features a table with columns: "Tipo*", "Nivel", "Clase", "Relacionado*", and "Acepción". The "Tipo*" column has a dropdown menu open, listing various semantic relations: hiperonimia, sinonimia, hiponimia, cohiponimia, meronimia, holonimia, antonimia, oposicion, complementariedad, and TAXONOMIA. There are also checkboxes and a "+" sign next to some items. To the right of the table, there are buttons labeled "Consulta" and "OK".

Figura 15. Relaciones entre los significados

La automatización de algunos de los procedimientos (como la asignación automatizada de los ejemplos del corpus, las remisiones cruzadas o la red de sinonimia de las voces estudiadas) se ha convertido en una ayuda inestimable para los lexicógrafos, que padecemos hasta hace poco tener que desarrollar a mano estas tareas, con un enorme consumo de tiempo y con el desánimo, a que nos hemos referido ya, que provocaba la constatación de un incómodo porcentaje de errores. Ahora la herramienta propaga automáticamente y de forma transparente esta información por el diccionario, evitando así estas enojosas tareas consistentes en repetir un proceso a lo largo de los diferentes artículos implicados, como ocurre cuando se asigna en una palabra una relación de sinonimia en una acepción «A» contra otra acepción «B» que ya tenía varios sinónimos «C» y «D», pues entonces la herramienta establece automáticamente que la acepción «A» tiene ahora a «C» y «D» como sinónimos (además de «B»), que es lo que había indicado el redactor), pero señala, además, en sus respectivas fichas, que tanto «B» como «C» y «D» tienen como nuevo sinónimo a la acepción «A». Es decir, que si *panderetista* ya estaba marcado como sinónimo de *panderetólogo* y *panderetero*, al hacer la ficha de *pandereta* encontramos ejemplos metonímicos y, por tanto, se indica que es sinónimo de *panderetista* y entonces automáticamente la herramienta establece el vínculo también con *panderetólogo* y *panderetero*, y en las cuatro fichas se establece la relación de sinonimia, como se observa en la figura 16.

Lema en edición: **panderetero**

RELACIÓN SEMÁNTICA

Tipo* Nivel Clase Relacionado* Aceptión

+ 

Refrescar sinonimias

Sinónimos

- panderetista: [] Persona que toca la pandereta.
- pandereta: [] Persona que toca la pandereta.
- panderetólogo: [] Persona que toca la pandereta.

Figura 16. Automatización de las relaciones establecidas entre las voces: sinonimia del artículo panderetero

Dejemos de lado algún aspecto que nos distanciaría excesivamente del hilo que estamos siguiendo, como la etimología o la relación genética entre las palabras, que están pendientes de desarrollo, para pasar a algunos elementos que, a nuestro juicio, enriquecen notablemente las posibilidades de acceso al diccionario; nos referimos a la presentación o resumen de la historia de la palabra. Estamos persuadidos de que un diccionario histórico debería ser un recurso de fácil acceso para cualquier persona con la curiosidad suficiente para preguntarse, en cualquier momento, por las peripecias biográficas de una palabra y las de cuantas la acompañan en su deambular por la historia; este convencimiento no alienta en exclusiva nuestra preocupación por los modos de acceso al inventario, sino que se traslada a la concepción misma de la obra. Por ello, el resumen actúa a modo de diccionario histórico abreviado, pues condensa los principales episodios en la vida de las voces. Supone, no obstante, un considerable riesgo — buscado—, pues implica una interpretación particular de los datos e incluso la manifestación explícita de los problemas que pueden quedar pendientes de solución; solo siendo explícitos y mostrando el estado actual de nuestros conocimientos o los problemas que se nos plantean podrá cualquier otra persona proponer una mejora. En este momento estamos trabajando en la elaboración de un programa que permita enviar al resumen, desde la ficha de cada palabra redactada, los datos básicos para su ulterior elaboración.

BROQUEL

Voz tomada del francés antiguo *bocler* (francés actual *bouclier*) 'escudo', derivada a su vez de *bocle* 'guarnición de metal que el escudo llevaba en su centro', vocablo que remonta al latín *būcula*, derivado apreciativo de *bucca*, 'mejilla' (DECH, s. v. *broquel* y DHLF, s. v. *bouclier*). Según el DECH, se documenta por primera vez en la *Gran Conquista de Ultramar*, pero el término no se encuentra en la sección conservada por el ms. BNE 1187, copiado a finales del XIII o comienzos del XIV, cuya transcripción recoge el CORDE, sino en un fragmento conservado por el impreso salmantina de Hans Giesser de 1503 (véase R. Ramos 2002).

Las variantes *bloquer*, *broquell*, *broquen*, *bronquel*, *broquer* se registran en diversos textos aragoneses, como diferentes inventarios (el más antiguo data de 1368) de los siglos XIV y XV, o las *Ordinaciones y paramientos de la ciudad de Barbastro* (1396). No obstante, el máximo esplendor de esta voz se vivió en los siglos XVI y XVII.

En el siglo XVI experimenta la actuación de un proceso metafórico y adquiere el significado de 'defensa o protección', acepción atestiguada por primera vez en la *Historia general y natural de las Indias* (1535-1557) de Fernández de Oviedo, y que se mantiene hasta nuestros días. Covarrubias, en su *Tesoro* (1611: s. v.) advertía que

la expresión «Todo es dar en los broqueles» se emplea «quando en las disputas y altercaciones se entretienen en lo que no importa, sin tocar la dificultad ni el punto sustancial del negocio».

Con el nombre de *broquel* se designa, en la *Ciencia popular* de Echegaray (1870-1905), una máquina tuneladora; esta nueva acepción llega a nuestra lengua a través del francés, si bien el origen último se halla en el sistema y la máquina que el ingeniero Mark I. Brunel emplea construir un túnel bajo el Támesis, a la que se denomina *shield*, vocablo que en francés se traduce como *bouclier* (como prueba, por ejemplo, el artículo «Tunnel sous la Tamise», publicado en *Le Magasin pittoresque*, 1835, n.º 3, p. 38).

Con la acepción de ‘muro que rodea la boca de un pozo’ se registra por vez primera en 1899, en *Prometeo & Cía* de E. Wilde; previamente, Rosal, en su *Origen y etimología de todos los vocablos originales de la Lengua Castellana* (1601) había relacionado ambos vocablos («Broquel, como Brocal, por ser instrumento acomodado para cubrir boca de vaso. Y de aquí Broche al botón, como diminutivo de broquel, porque tiene hechura de broquelete, o escudillo») y, aunque en algunos diccionarios de americanismos se caracteriza como «barbarismo por brocal», los textos muestran que se ha consolidado en el uso.

Marcado siempre como vocablo náutico o marino, *broquel* se documenta con la acepción de ‘posición en que quedan las vergas cuando se maniobra con ellas para que reciban el viento por la cara de proa’ desde 1766 [1806] (en el *Tratado instructivo, y práctico de maniobras navales* de Zuloaga). Por último, algunos diccionarios recogen la acepción, propia de México, de ‘pendiente para oreja que a modo de pequeño perno atraviesa el lóbulo y se asegura por el lado interior del mismo con una tuerquecilla’, como muestra el *Diccionario general de americanismos* de Santamaría (1942).

Gráfico 16. Resumen de la historia de la palabra

Sin ser los seres más satisfechos del mundo, nos movemos en un equilibrio entre el encorsetamiento que impone la informática y la libertad que permite la edición de textos. Ciertamente, a ello ha contribuido la posibilidad que ofrece la herramienta de usar «marcas flotantes» en algunos campos de texto libre (en el resumen, en las definiciones), estructurando de una manera muy flexible la información. Por otra parte, en el Manual de redacción se establecen las normas concretas que rigen la elaboración de cada sección del artículo, para intentar garantizar la coherencia.

Pese a la complejidad que entraña esta ficha de trabajo, al final la tarea del lexicógrafo se ha visto beneficiada y facilitada por tener un molde razonable al que ajustarse, garantizando la grabación de toda la información volcada y las relaciones apuntadas entre los lemas. Sirve así nuestra herramienta para comparar y corregir mientras se va elaborando el trabajo y, en consecuencia, se van enmendando los errores. Esta comodidad se complementa con la seguridad que proporciona la progresiva forma de control sobre la redacción y la integridad de los resultados obtenidos, tanto para los que consultan como para los que redactan. Como muestra, seleccionamos uno de los mecanismos automáticos de control que proporciona esta herramienta: el de la integridad referencial de las relaciones entre distintas partes del *NDHE*, ya que se impide que, por ejemplo, se elimine una acepción de un artículo si tiene alguna relación con algún otro elemento de la obra, del mismo modo que es imposible crear relaciones con partes del diccionario que han sido eliminadas, etc.

Para ir finalizando este apartado, explicaremos también que, para el seguimiento del proyecto, *ARDIDEs* sirve de gran ayuda en el control y gestión del diccionario, ya que, con su módulo sobre información administrativa, los responsables podrán dirigirlo de

forma más eficiente. Este módulo permite conocer el ritmo de redacción del diccionario para poder detectar problemas y, consiguientemente, habilitar soluciones para ellos, evitando desviaciones en tiempo y en la calidad del trabajo realizado. Proporciona, a la vez, a los responsables del proyecto, una asistencia para distribuir el trabajo de una forma razonable entre los distintos grupos de trabajo y entre los redactores de dichos grupos. En definitiva, ayuda a mejorar la cantidad y la calidad del trabajo realizado.

Durante los últimos años se viene haciendo especial hincapié en la seguridad de las aplicaciones informáticas, y este factor se ha tenido muy en cuenta especialmente en el diseño de la herramienta de redacción (aunque también en el resto de aplicaciones desarrolladas para la elaboración del *NDHE*). Los módulos y estándares de seguridad de la tecnología usada en el desarrollo de la herramienta de redacción (*j2ee*, *acegi*, *ldap*, etc.) permiten hacer fácil lo que se expresa con la voz «autenticación», así como otras tareas más llevaderas, como es el control de acceso para los redactores; facilita y simplifica mucho, al tiempo, el trabajo de los desarrolladores de las aplicaciones, pero, sobre todo, evita la aparición de agujeros de seguridad incontrolados y cumple con los estándares y con las leyes que afectan a la seguridad de los datos personales, como la *LORTAD* (*Ley Orgánica de Regulación del Tratamiento Automatizado de los Datos de Carácter Personal*).

Para adaptarnos a esta ley, se crearon unos perfiles de usuarios que facilitan el control y la organización del trabajo (colaborador, coordinador, redactor, asesor y administrador), basándonos en los módulos de seguridad de la tecnología usada. Cada perfil cuenta con unos permisos concretos para acceder a las zonas de la aplicación donde tiene competencia; así, por ejemplo, un administrador puede asignar la redacción de una familia de palabras a un equipo de trabajo formado por un coordinador del equipo y un grupo de colaboradores y, posteriormente, disponer que la revise un asesor que solo tenga permiso de lectura sobre el grupo de palabras redactado. De esta forma tan sencilla la propia aplicación facilita la tarea de organización, coordinación, revisión y aprobación del trabajo realizado en la redacción del diccionario.

5. HACIA EL FUTURO INMEDIATO: LA HERRAMIENTA DE VISUALIZACIÓN

Uno de los fines perseguidos al presentar este diccionario en red consiste en la posibilidad de ir ofreciendo al público, desde la web, los materiales listos para su publicación, sin esperar a la redacción final de toda la obra; porque el criterio de la perfección no se contradice con lograr gradualmente nuestros objetivos, con la enmienda y mejora progresiva de los artículos, en lo que parece razonable, además, contar con las observaciones de los usuarios.

Precisamente, el objetivo que nos planteamos para los primeros meses del año 2012 consiste en diseñar una herramienta de visualización capaz de satisfacer las diferentes necesidades de los usuarios del *NDHE*. La finalidad perseguida radica en ofrecer una obra de consulta sencilla y rápida, accesible no solo al especialista, sino a cualquier persona interesada en la historia del léxico español; al usuario le corresponde, por lo tanto, decidir el grado de complejidad de sus búsquedas o la cantidad de datos que desee manejar. Como hemos indicado anteriormente, el resumen de la historia de la

palabra sirve fundamentalmente a este fin, pues, a nuestro juicio, es un elemento básico para acercar el diccionario a los usuarios no especializados. De este modo, un consultante podrá acceder al *NDHE* por distintos caminos, unos más sencillos (como la visualización solo del resumen), otros más completos y complejos (como la visualización de todo el artículo, con todos los materiales recopilados y tratados en él), pasando por otras perspectivas intermedias (como la visualización del artículo sin autoridades, la consulta con el esquema del artículo y los testimonios más significativos, etc.). Por otra parte, los usuarios podrán determinar, en su consulta, el orden en que quieren acceder a las acepciones del artículo (cronológico, de frecuencia, lógico-lingüístico, geográfico...), desplegar la familia de palabras, las jerarquías taxonómicas o los árboles de sinónimos de cada acepción.

Evidentemente, cuando disponemos de una gran cantidad de información, lo fundamental es su correcta ordenación y disposición, y, por supuesto, que la cantidad no oculte la calidad de los datos; por consiguiente, la aplicación de visualización permitirá explotar la potencia relacional del diccionario, ofreciendo la posibilidad de navegar a través de las relaciones existentes entre artículos, acepciones (de un mismo artículo o de distintos artículos del diccionario), etc. En definitiva, el usuario que consulte el *NDHE* podrá bucear en él desde lo general hasta los detalles más pequeños, de una forma sencilla, guiado por la propia herramienta de visualización.

La herramienta de visualización del *NDHE* será también una aplicación web, de la misma manera que el resto del *software* generado para el diccionario. Precisamente la informática y los últimos avances en ingeniería del *software* y en controles gráficos de *interfaces* de usuario para aplicaciones *web* posibilitan una consulta muy potente y alejada de la clásica búsqueda en diccionarios en papel, o en diccionarios electrónicos con una perspectiva básicamente plana, es decir, diccionarios electrónicos con una lectura en una sola dimensión, muy similar a la consulta de un diccionario en papel. De esta manera se podrá ver y consultar el diccionario no solo como una serie artículos independientes, sino como un todo constituido por artículos interrelacionados, lo que lo convierte en un auténtico diccionario multidimensional.

6. CONSIDERACIONES FINALES

El resultado de esta labor de conversión de la planta del *NDHE* en una herramienta de redacción y consulta, hecha —como hemos repetido hasta la saciedad— en colaboración entre lexicógrafos e informáticos, permitirá —nos lo está permitiendo ya a quienes trabajamos en él— disponer un diccionario explícito, que admitirá una variada gama de enfoques, así como el acercamiento de cualquier usuario, con sus particulares intereses o necesidades, que podrá moverse por él atendiendo a cualquiera de ellos. Por otra parte, no debemos olvidar —algo que, lamentablemente, se olvida demasiadas veces— que un diccionario histórico no mira únicamente hacia el pasado del idioma, sino que es una pieza fundamental para alcanzar una mejor comprensión de los usos actuales de la lengua.

En suma, *ARDIDES* intenta integrar toda la información del *CDH* y de otras bases documentales en la herramienta de redacción, de tal modo que los materiales están analizados y controlados (pues garantiza contenerlos todos), y permite, al tiempo, que se vayan añadiendo nuevas fuentes documentales, dentro de un proceso que debe estar continuo progreso. Nos parece reseñable, además, la automatización de procesos, pues evita a los redactores mucho trabajo rutinario, proporcionando a su vez un control de la integridad de la información, hecho que, en definitiva, evita errores. En este sentido, la herramienta de redacción ayuda a la dirección del proyecto, a la organización y distribución de las tareas, a la mejora de la cantidad y la calidad del trabajo realizado y a la detección de posibles problemas durante la redacción, para así ponerles solución lo antes posible.

Como se ha expuesto con antelación, la redacción de los artículos irá mejorando a la vez los materiales, en un proceso de ida y vuelta, de retroalimentación. Los materiales, que no se conciben como un ente estático, se van calibrando a medida que se elabora el diccionario y se refinan no solo mediante la corrección de erratas, sino también a través de la disminución del grado de ambigüedad de la lematización y de la incorporación de nuevas funcionalidades (como, por ejemplo, de un repositorio bibliográfico al servicio de la caracterización filológica de los textos).

En fin, los artículos redactados no se contemplan como un trabajo culminado y cerrado, sin posibilidad de modificación, sino como una hipótesis de partida, una columna vertebral a partir de la cual se abre un marco de discusión. Porque no debemos perder de vista que este diccionario no es una mera recopilación de una palabra tras otra, sino que su principal objetivo consiste en su capacidad de plantear y, en su caso, resolver problemas semánticos, fonéticos, morfológicos, etc. Algunas cuestiones se dilucidarán en el seno del *NDHE*, pero otras hallarán su solución en las propuestas de los estudiosos que se acerquen al diccionario. Somos plenamente conscientes de que ningún repertorio podría aspirar a dar respuesta a todas las preguntas que se pueden formular en la historia del léxico, pero sí debería brindar, en nuestra opinión, una hipótesis de partida, así como dar acceso a todos los materiales y funcionalidades de que dispone para que otros expertos consigan recorrer, con sus nuevos datos, una etapa más de esta senda del conocimiento. Igual que cualquier investigación en otro campo de la ciencia, el diccionario está, pues, abierto a la colaboración en su mejora y progreso. Esta anhelada colaboración ha sido el motivo de la creación de una aplicación de redacción como *ARDIDES*.

REFERENCIAS BIBLIOGRÁFICAS

ATKINS, Sue y Michael RUNDELL (2008): *The Oxford Guide to Practical Lexicography*, Oxford, Oxford University Press.

- CAMPOS SOUTO, Mar y José Antonio PASCUAL (en prensa): «*Dalle que dalle*: la filología como intermediaria en el salto de la cantidad a la calidad».
- CURRAN, James R. (2004): *From Distributional to Semantic Similarity*, Tesis doctoral, University of Edinburgh.
- GELPÍ ARROYO, Cristina (2003): «El estado actual de la lexicografía: los nuevos diccionarios», en A. M. Medina Guerra, coord., *Lexicografía española*, Barcelona, Ariel, pp. 307-309.
- HOFFMANN, S., S. EVERT, N. SMITH, D. LEE y Y. BERGLUND PRYTZ (2008): *Corpus Linguistics with BNCweb: A Practical Guide*, Frankfurt-Main, Peter Lang.
- KIRKNESS, Alan (2006): «Lexicography», en A. Davies y C. Elder, eds., *The Handbook of Applied Linguistics*, Malden, MA, Blackwell, pp. 55-81.
- STEFAN, Evert (2008): «Corpora and collocations», en A. Lüdeling y M. Kytö, eds., *Corpus Linguistics: An International Handbook*, Berlin-New York, Walter de Gruyter, pp. 1212-1248.
- PASCUAL, José Antonio y Carlos DOMÍNGUEZ CINTAS (2009): «Un corpus para un *Nuevo Diccionario Histórico del Español*», en A. Enrique-Arias ed., *Diacronía de las lenguas iberorrománicas: Nuevas aportaciones desde la lingüística de corpus*. Madrid-Frankfurt, Vervuert-Iberoamericana, pp. 79-93.
- PASCUAL RODRÍGUEZ, José Antonio y Rafael PÉREZ GARCÍA (2007): *Límites y horizontes en un diccionario histórico*, Salamanca, Diputación de Salamanca.
- (2008): «Estado del nuevo diccionario histórico de la Real Academia Española», en P. Garcés García, coord., *Diccionario histórico: Nuevas perspectivas lingüísticas*, Madrid-Frankfurt, Vervuert-Iberoamericana, pp. 11-15.
- REAL ACADEMIA ESPAÑOLA (1960-1996): *Diccionario histórico de la lengua española*, Madrid, Imprenta Aguirre.