

CARMEN QUIJADA VAN DEN BERGHE
BORJA ALONSO PASCUA
FRANCISCO ESCUDERO PANIAGUA
CAROLINA MARTÍN GALLEGO
GEMA BELÉN GARRIDO VÍLCHEZ (Eds.)

DE ESTEPA A SALAMANCA MIRADAS EN TORNO A LA LENGUA

Admirable. Rigor. Exhaustive

Prof. J. Gón. Azuero

Gracias por los regalos



DE ESTEPA A SALAMANCA
MIRADAS EN TORNO A LA LENGUA



José J. Gómez Asencio (Estepa, 1953 - Salamanca, 2022)

CARMEN QUIJADA VAN DEN BERGHE
BORJA ALONSO PASCUA
FRANCISCO ESCUDERO PANIAGUA
CAROLINA MARTÍN GALLEGO
GEMA BELÉN GARRIDO VÍLCHEZ (Eds.)

DE ESTEPA A SALAMANCA
MIRADAS EN TORNO A LA LENGUA



Ediciones Universidad
Salamanca

AQUILAFUENTE, 351

© Ediciones Universidad de Salamanca
y los autores

Esta obra ha sido posible gracias al GIR GramHisGram; al Centro Internacional del Español, al Departamento de Lengua Española, a la Facultad de Filología y a la Agencia de Investigación de la Universidad de Salamanca, al Centro de Estudios Salmantinos y a Cursos Internacionales de la Universidad de Salamanca.

1ª edición: septiembre, 2023
ISBN: 978-84-1311-830-7 (PDF)
ISBN: 978-84-1311-831-4 (POD)
DOI: <https://doi.org/10.14201/0AQ0351>

Ediciones Universidad de Salamanca
Plaza San Benito s/n
E-37002 Salamanca (España)
<http://www.eusal.es>
eusal@usal.es

Hecho en UE-Made in EU

Maquetación y realización:
Cícero, S.L.U.
Tel.: +34 923 12 32 26
37007 Salamanca (España)

Impresión y encuadernación:
Nueva Graficesa S.L.
Teléfono: 923 26 01 11
Salamanca (España)



Usted es libre de: Compartir – copiar y redistribuir el material en cualquier medio o formato
Ediciones Universidad de Salamanca no revocará mientras cumpla con los términos:

i Reconocimiento – Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.

€ NoComercial – No puede utilizar el material para una finalidad comercial.

= SinObraDerivada – Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

«Obra sometida a proceso de evaluación mediante sistema de doble ciego»

Ediciones Universidad de Salamanca es miembro de la UNE
Unión de Editoriales Universitarias Españolas www.une.es



Catalogación de editor en ONIX accesible en <https://www.dilve.es/>

Índice

PRÓLOGO.....	13
PUBLICACIONES DE JOSÉ J. GÓMEZ ASENCIO	17
Historiografía de las ideas gramaticales. Estudios longitudinales.....	17
Epistemología y metodología de la Historiografía Lingüística	19
Gramática e Historiografía Lingüística de los Siglos de Oro	20
Ideas gramaticales y lingüísticas en España a partir del siglo XVIII.....	21
Ideas gramaticales y lingüísticas en Hispanoamérica a partir del siglo XIX.....	25
El español fuera de la hispanofonía y otras tradiciones lingüísticas	26
Historia de la lengua y sociedad.....	28
Gramática descriptiva y dialectología	28
Enseñanza del español como lengua extranjera.....	29
Fonética y análisis del discurso	31

GRAMÁTICA E HISTORIOGRAFÍA LINGÜÍSTICA DE LOS SIGLOS DE ORO

HUGO E. LOMBARDINI (<i>Università di Bologna</i>)	
«J. J. Gómez Asencio por J. J. Gómez Asencio: un historiógrafo de la gramaticografía»	35
VICENTE BÉCARES BOTAS (<i>Universidad de Salamanca</i>)	
«Gramáticas y lenguas en la Salamanca áurea».....	53
FRANCISCO ESCUDERO PANIAGUA (<i>Universidad de Salamanca</i>)	
«La ejemplificación de Nebrija (vive)»	69
MIGUEL ÁNGEL ESPARZA TORRES (<i>Universidad Rey Juan Carlos</i>)	
«De Nebrija y la gramatización de las lenguas filipinas: ¿viajó Nebrija en el Galeón de Manila?»	85
ALEJANDRO DÍAZ VILLALBA (<i>Université Sorbonne Nouvelle / Université Paris Cité, CNRS</i>)	
«El tratamiento de las formas compuestas del verbo en <i>Leys d'amors</i> (s. XIV), Nebrija (1492) y Ramus (1572)»	101
MARÍA DOLORES MARTÍNEZ GAVILÁN (<i>Universidad de León</i>)	
«A vueltas con la originalidad de Correas: los conceptos de <i>norma</i> y de <i>propiedad</i> en el <i>Arte Grande</i> »	119

IDEAS GRAMATICALES Y LINGÜÍSTICAS EN ESPAÑA
A PARTIR DEL SIGLO XVIII

MARÍA JOSÉ MARTÍNEZ ALCALDE (<i>Universitat de València</i>) «Diversidad lingüística y selección léxica en el <i>Diccionario de autoridades</i> »	141
MARISA MONTERO CUIEL (<i>Universidad de Extremadura</i>) «Las tres “Voces nuevas” del <i>Diccionario de autoridades</i> »	155
MERCEDES QUILIS MERÍN (<i>Universitat de València</i>) «El <i>Diccionario Geográfico Universal</i> de Antonio de Capmany (1783) en la lexicografía de especialidad del siglo XVIII»	171
GERDA HAßLER (<i>Universität Potsdam</i>) «Discusiones metalingüísticas en los Espectadores en la España del siglo XVIII»....	187
CARLOS VILLANUEVA GARCÍA (<i>Universidad de Salamanca</i>) «El desplazamiento de la propedéutica orientada al latín de las gramáticas del español durante el siglo XVIII»	203
GEMA BELÉN GARRIDO VÍLCHEZ (<i>Universidad de Salamanca</i>) «De la norma (y su ausencia) en la <i>GRAE</i> decimonónica»	217
MARÍA JOSÉ GARCÍA FOLGADO (<i>Universitat de València</i>) «La terminología gramatical en la enseñanza: repertorios decimonónicos».....	237
CAROLINA MARTÍN GALLEGO (<i>Universidad de Salamanca</i>) «“Unas gramáticas pueden ayudar a bien comprender otras”: sobre las líneas de influencias (de nuevo) en el tratamiento de la <i>conjunción</i> »	253
MARÍA TERESA ENCINAS MANTEROLA (<i>IES José María Pereda de Santander</i>) «Aproximación a los <i>Elementos de gramática castellana</i> de Yllas y Figuerola».....	275
VICTORIANO GAVIÑO RODRÍGUEZ (<i>Universidad de Cádiz</i>) «Fernando Gómez de Salazar y sus obras filológicas en la segunda mitad del siglo XIX».....	291
MARÍA LUISA CALERO VAQUERA (<i>Universidad de Córdoba</i>) «El léxico de los prejuicios lingüísticos en las diatribas sobre el euskera»	307

IDEAS GRAMATICALES Y LINGÜÍSTICAS
EN HISPANOAMÉRICA A PARTIR DEL SIGLO XIX

MANUEL RIVAS ZANCARRÓN (<i>Universidad de Cádiz</i>) «Los espacios de opinión pública como fuente para el estudio de la recepción gramatical en la Nueva España y en los primeros años del México independiente»	321
ALFONSO ZAMORANO AGUILAR (<i>Universidad de Córdoba</i>) «La difusión de la teoría gramatical de Luis de Mata i Araujo en América Latina. A propósito de las ediciones publicadas en Colombia y Cuba»	339

JUAN PEDRO SÁNCHEZ MÉNDEZ (<i>Université de Neuchâtel</i>)	
«Andrés Bello y la aportación ilustrada hispanoamericana a la historiografía lingüística española»	355
DARÍO ROJAS (<i>Universidad de Chile</i>)	
«Los bordes de la lengua española en el <i>Diccionario de chilenismos</i> (1875) de Zorobabel Rodríguez»	369
ESTEBAN T. MONTORO DEL ARCO (<i>Universidad de Granada</i>)	
«Ideología y ejemplificación en la <i>Gramatiquilla infantil</i> (1885) del ecuatoriano Francisco Febres Cordero, el “hermano Miguel”»	387
JUAN ANTONIO ENNIS (<i>Universidad Nacional de La Plata / CONICET</i>)	
«El lingüista alemán y la tradición: algunas observaciones sobre <i>La oración y sus partes</i> , y otros trabajos de Rodolfo Lenz»	405
MARÍA MARTÍNEZ-ATIENZA DE DIOS (<i>Universidad de Córdoba</i>)	
«Canon literario en la gramática colombiana del siglo xx: el caso de Martín Restrepo Mejía (1928)»	419
ELVIRA NARVAJA DE ARNOUX (<i>Universidad de Buenos Aires</i>)	
«Análisis de una serie glotopolítica: documentos y propuestas didácticas en el proceso de reforma de los programas sobre la enseñanza de la lengua en la escuela secundaria (Argentina, 1933-1939)»	435

EL ESPAÑOL FUERA DE LA HISPANOFONÍA Y OTRAS TRADICIONES LINGÜÍSTICAS

PIERRE SWIGGERS y MARÍA CRUZ CASÁÑEZ (<i>Katholieke Universiteit Leuven / Université de Liège & CLT Leuven</i>)	
«La gramática castellana del licenciado Villalón (1558)»	455
CARMEN CASTILLO PEÑA (<i>Università degli Studi di Padova</i>)	
«Una gramática minorada: el <i>Compendio di Massimo Troiano tratto dalle Osservazioni di M. Giovanni Miranda</i> »	469
DIANA ESTEBA RAMOS (<i>Universidad de Málaga</i>)	
«Las glosas en la tradición de los diálogos para aprender español como lengua extranjera: aproximación a la serie de J. Minsheu (1599)»	483
EMILIO RIDRUEJO (<i>Universidad de Valladolid</i>)	
«El tratamiento del verbo en las <i>Institutiones in linguam hispanicam</i> de Doergangk» ..	503
ROGELIO PONCE DE LEÓN ROMEO (<i>Universidade do Porto</i>)	
«La terminología sintáctica en la gramaticografía de la lengua portuguesa del siglo xviii»	519
M. ^a ÁNGELES GARCÍA ARANDA (<i>Universidad Complutense de Madrid</i>)	
«L'orthographe espagnole a été long-temps incertaine; mais depuis l'établissement de l'Académie espagnole, cette société savante s'est constamment occu-	

pée avec le plus grand succès de la fixation et de la simplification de l'orthographe": la Academia en los manuales de enseñanza de español».....	537
DANIEL M. SÁEZ RIVERA (<i>Universidad de Granada</i>)	
«Un libro de lecturas para la enseñanza del español en Alemania en el siglo XIX: <i>El nuevo lector español: Spanisches Lesebuch</i> de Mariano Velázquez de la Cadena adaptado por Friedrich Funck».....	555
JUAN CARLOS BARBERO BERNAL (<i>Università degli Studi dell'Aquila</i>)	
«La gramatización de las formas <i>canté</i> y <i>he cantado</i> en la tradición gramatical italiana dirigida a hispanohablantes (1870-1900)».....	571
CARMEN GALÁN RODRÍGUEZ (<i>Universidad de Extremadura</i>)	
«Academias y lenguas artificiales: el latino sine flexione»	589
LOLA PONS RODRÍGUEZ (<i>Universidad de Sevilla</i>)	
«Una carta de Navarro Tomás antes del bombardeo de Barcelona. El legado hemerográfico y epistolar de Hayward Keniston».....	603

HISTORIA DE LA LENGUA Y SOCIEDAD

JOSÉ A. PASCUAL RODRÍGUEZ (<i>Real Academia Española</i>)	
«Una nota sobre la voz <i>pomelado</i> de la <i>Comedieta de Ponça</i> del marqués de Santillana»	619
MARÍA DE LAS NIEVES SÁNCHEZ GONZÁLEZ DE HERRERO (<i>Universidad de Salamanca</i>)	
«El pronombre personal sujeto de 1ª persona en la documentación medieval de El Bierzo».....	625
ÁLVARO S. OCTAVIO DE TOLEDO Y HUERTA (<i>Consejo Superior de Investigaciones Científicas</i>)	
«Gramatización y dialectología histórica: a propósito de dos relacionantes locativos codificados por Nebrija».....	639
JOSÉ ANTONIO BARTOL HERNÁNDEZ (<i>Universidad de Salamanca</i>)	
«Gramaticalización y uso de <i>en tal caso</i> como conector conclusivo hipotético» ...	653
MARÍA TERESA ECHENIQUE ELIZONDO (<i>Universitat de València</i>)	
«La fraseología histórica como marco filológico para el estudio de locuciones: nuevos apuntes sobre <i>redor</i> y los cognados de su familia fraseológica».....	669
MARÍA JESÚS MANCHO DUQUE (<i>Universidad de Salamanca</i>)	
«Testimonios cultos en el ámbito de la Filosofía Natural y de la Física a partir del <i>Diccionario de la Ciencia y de la Técnica de Renacimiento</i> »	687
RAFAEL CANO AGUILAR (<i>Universidad de Sevilla</i>)	
«Los nombres de la lengua en el primer español moderno»	703
EVA BRAVO-GARCÍA (<i>Universidad de Sevilla</i>)	
«Las cartas de Isabel y Eulalia Figueredo. Escritura cotidiana femenina en Cuba (siglo XIX)»	719

MARTA TORRES MARTÍNEZ (<i>Universidad de Jaén</i>)	
«Léxico culinario en prensa histórica: el caso de los <i>mantecados</i> y <i>polvorones</i> de Estepa».....	735
EFREM YILDIZ (<i>Universidad de Salamanca</i>)	
«Las fuentes documentales antiguas e imperiales de lengua aramea»	751
RACHID EL HOUR (<i>Universidad de Salamanca</i>)	
«La cofradía de los <i>māğariyyūn</i> de Safí y la creación de caravana de los peregrinos (<i>Rakb al-Huğğāğ</i>) en Marruecos».....	771

GRAMÁTICA DESCRIPTIVA

JULIO BORREGO NIETO (<i>Universidad de Salamanca</i>)	
«El concepto de “réplica de función” o “función-r” y su utilidad en sintaxis»	787
IGNACIO BOSQUE (<i>Universidad Complutense de Madrid</i>)	
«Verbos que expresan acciones intermedias o incompletas»	801
FELISA BERMEJO CALLEJA (<i>Università degli Studi di Torino</i>)	
«Verbos subordinantes de voluntad o de influencia: gramatización y valoración»...	817
SALVADOR GUTIÉRREZ ORDÓÑEZ (<i>Real Academia Española</i>)	
«Sobre los relativos complejos».....	835
MIGUEL ÁNGEL AIJÓN OLIVA (<i>Universidad de Salamanca</i>)	
«La variabilidad entre la construcción pasiva y la impersonal de <i>se</i> . Isomorfismo e interpretación contextual»	857
GUILLERMO ROJO (<i>Universidade de Santiago de Compostela</i>)	
«Distribución de las clases de palabras en los corpus textuales».....	873
TERESA AMORES SIERRA (<i>Universidad de Salamanca</i>)	
«Sobre <i>adelante</i> ».....	883
SUSANA AZPIAZU TORRES (<i>Universidad de Salamanca</i>)	
«Observaciones en torno al perfecto de indeterminación temporal»	901
MARÍA DEL ROSARIO LLORENTE PINTO (<i>Universidad de Salamanca</i>)	
«Porcentajes de uso de los pasados perfectos en España y América y su combinación con <i>hoy, ya, nunca</i> y <i>siempre</i> »	917
FÉLIX SAN VICENTE (<i>Università di Bologna</i>)	
«PPS y PPC en gramáticas universitarias de español para itálofonos (1997-2014)»...	935
BORJA ALONSO PASCUA (<i>Universidad de Salamanca</i>)	
«Sobre el aspecto en el sistema verbal del español, otra vez».....	953
ELENA BAJO PÉREZ (<i>Universidad de Salamanca</i>)	
«El imperativo y la posición de los pronombres personales»	969

GRAMÁTICA PARA LA ENSEÑANZA DEL ESPAÑOL
COMO LENGUA EXTRANJERA

CARLOS CABRERA MORALES (<i>Universidad de Salamanca</i>) «La enseñanza del subjuntivo en ELE. Una propuesta multicausal».....	987
MANUEL IGLESIAS BANCO (<i>Universidad de León</i>) y MARIO DE LA FUENTE GARCÍA (<i>Fundación Sierra Pambley</i>) «La estructura argumental y su aplicación a la enseñanza de ELE: el caso de los pronombres átonos de complemento»	1007
MARTA SESEÑA GÓMEZ (<i>Universidad de Salamanca</i>) «La herencia gramatical en la didáctica de ELE: el caso de las oraciones condicio- nales con <i>si</i> »	1021

FONÉTICA Y ANÁLISIS DEL DISCURSO

CARMEN QUIJADA VAN DEN BERGHE (<i>Universidad de Salamanca</i>) «Vitalidad de la fricativización de las consonantes implosivas en Salamanca».....	1041
JOSÉ MANUEL BUSTOS GISBERT (<i>Universidad de Salamanca</i>) «Las construcciones temporales y su rendimiento en el texto escrito»	1069
JAVIER DE SANTIAGO GUERVÓS (<i>Universidad de Salamanca</i>) «Análisis de un discurso político: lengua y cognición»	1089
<i>TABULA GRATULATORIA</i>	1107

DISTRIBUCIÓN DE LAS CLASES DE PALABRAS EN LOS CORPUS TEXTUALES¹

GUILLERMO ROJO

Universidad de Santiago de Compostela

RESUMEN

El presente trabajo se propone examinar los aspectos generales de la distribución de las clases de palabras léxicas tanto en lo que se refiere al inventario de unidades como en lo relativo a su uso. En este último aspecto, se analiza también si existen diferencias apreciables según diferentes parámetros relacionados con los tramos de frecuencias, los tipos de texto o el área temática a la que pertenecen.

ABSTRACT

The aim of this paper is to examine the general aspects of the distribution of lexical word classes both in terms of the inventory of units and in terms of their use. In the last aspect, it is also analyzed if there are appreciable differences according to different parameters related to the frequency ranges, the types of text or the thematic area to which they belong.

1. INTRODUCCIÓN

ENTRE OTROS MUCHOS EFECTOS sobre las diversas disciplinas y subdisciplinas existentes en nuestro campo de trabajo, la lingüística de corpus ha producido un más que notable aumento de los estudios sobre la configuración estadística de los textos. La posibilidad de utilizar computadoras, tanto para el almacenamiento de cantidades ingentes de textos como para la realización de cálculos,

¹ El presente trabajo ha sido realizado en el marco del proyecto de investigación «El corpus ESLORA de español oral: enriquecimiento, análisis lingüístico y aplicaciones» (ref. PID2020-118133GB-I00), financiado por el Ministerio de Ciencia e Innovación (2021-2024).

nos ha permitido profundizar en aspectos en los que hace unos pocos años era muy laborioso o incluso imposible trabajar. En el caso de los elementos léxicos hay que resaltar la importancia de los diccionarios de frecuencias, pero también es preciso reconocer sus limitaciones. Como he tratado de mostrar recientemente (*cf.* Rojo en prensa b), están basados habitualmente en corpus muy reducidos e incluso en los casos en los que el corpus es mucho mayor, como sucede en Davies y Davies (2018), el diccionario, pensado para su publicación en forma impresa, se limita a los 5000 lemas con índices de uso más elevados.

Como es de esperar, esas limitaciones se proyectan también sobre caracterizaciones estadísticas más abstractas, como es la referida a las clases de palabras. En este trabajo, dedicado a la memoria de José Gómez Asencio, me he propuesto estudiar los aspectos más generales de esta cuestión, utilizando para ello los datos procedentes de la última versión del CORPES XXI (la 0.99, publicada en enero de 2023). Es esta una tarea realizable con cualquier corpus que haya sido lematizado y anotado morfosintácticamente. Llevarla a cabo con un corpus de referencia bien codificado nos permite tratar de detectar diferencias entre los distintos tipos de texto que lo componen, con lo que visión obtenida resulta mucho más rica.

2. FRECUENCIA DE USO Y FRECUENCIA DE INVENTARIO

En el análisis de las frecuencias léxicas es tradicional trabajar con la diferencia entre los que la lingüística inglesa denomina *types* y *tokens*. No es fácil dar con términos equivalentes en español, pero la diferencia es clara. Un *token* es, por ejemplo, cada una de las formas ortográficas que aparecen en un texto. El número de *tokens* es lo que se mide cuando se solicita un ensayo que tenga no más de 5000 palabras. Un *type* es, para seguir con el ejemplo anterior, cada una de las palabras ortográficas distintas que forman un texto. La relación entre estas dos cifras (número de *types* dividido entre el número de *tokens*) es lo que se conoce como *type token ratio* (TTR) y fue utilizado como un índice de la «riqueza léxica» de un texto ya desde los primeros tiempos de empleo de computadoras en análisis lingüísticos y literarios.

Trabajar con las formas ortográficas es, por supuesto, un nivel muy bajo de análisis lingüístico, pero la distinción puede ser proyectada sobre elementos de carácter más abstracto, como, por ejemplo, lemas, sufijos o esquemas sintácticos. Es fundamental en este aspecto la propuesta realizada por Bybee (2007), retocada y generalizada en la distinción entre inventario y uso establecida en Rojo (2011). La frecuencia de inventario es, por ejemplo, el número de lemas verbales distintos que contienen un diccionario o un texto o bien el número de esquemas triactanciales diferentes que se documentan en un texto o un corpus. La frecuencia de uso es la cantidad de verbos que figuran en un texto o en un corpus, la cantidad de secuen-

cias que presentan un determinado esquema sintáctico triactancial. La importancia de mantener diferenciados ambos conceptos y la utilidad de su diferenciación se hace evidente cuando, por ejemplo, se contrasta lo escaso del tamaño del inventario de artículos o preposiciones del español con la gran amplitud de su uso en los textos. En una variación de esta perspectiva, es evidente también que con los 10 lemas más frecuentes de un corpus (que son una pequeñísima parte del inventario léxico) se cubre aproximadamente un tercio de lo que podemos encontrar en los textos (*cf.* Rojo 2022, tabla 3).

3. CLASES DE PALABRAS EN CORPUS Y DICCIONARIOS

Como es evidente, la utilización de los diccionarios para conocer la distribución de las clases de palabras se limita a lo relacionado con el inventario (podría usarse el diccionario como un texto, pero en esa opción no aporta nada importante). Obtener el inventario de elementos pertenecientes a una determinada clase de palabras requiere, por supuesto, la posibilidad de trabajar con un diccionario en formato electrónico del que se pueda recuperar esa información. Hay que tener en cuenta que esta caracterización categorial figura en cada una de las acepciones. En consecuencia, no es válido trabajar directamente con las entradas, que suelen ser clasificadas por la clase de su primera acepción (elegida de acuerdo con ciertos criterios convencionales de organización de la microestructura). En el caso del *DLE*, el recurso Enclave RAE proporciona los datos correspondientes al número de acepciones que llevan indicación de pertenencia a una determinada clase. Está claro que estos datos tampoco sirven directamente, puesto que el recuento requerido para el inventario de lemas (con indicación de la clase) no puede trabajar con todas las acepciones del mismo lema. Es necesario, por tanto, partir de los datos facilitados por Enclave RAE y filtrarlos de forma automática para obtener el recuento de lemas diferentes adscritos a una determinada clase de palabras².

Dado que el objetivo de esta extracción es obtener los datos proporcionados por el diccionario para compararlos con los que podamos extraer de un corpus, hay que considerar que pueden aparecer diferencias entre la caracterización categorial que figura en un determinado diccionario y lo que resulta de la anotación automática añadida en el tratamiento de un corpus concreto. Para tratar de reducir esos inconvenientes, vamos a restringir la comparación a las clases de palabras «llenas»: adjetivos, adverbios, sustantivos (comunes) y verbos. Las frecuencias de inventario

² Téngase en cuenta que la recuperación de elementos con Enclave RAE devuelve también los que llevan indicaciones del tipo «usado también como sustantivo» y similares. Por tanto, la primera acepción del lema *abasi* es recuperada al pedir las adjetivas y también las sustantivas.

y los porcentajes correspondientes a la última edición del *DLE* (la 23.5, publicada en diciembre de 2022) y la versión 0.99 del CORPES (publicada en enero de 2023) figuran en la tabla 1³.

TABLA 1. *Frecuencias y porcentajes de lemas de algunas clases de palabras en el DLE y el CORPES*

	DLE 23.5		CORPES 0.99	
	Frecuencia	%	Frecuencia	%
Adjetivos	26 593	22,21	57 878	22,82
Adverbios	5479	4,58	5940	2,34
Sust. com.	75 315	62,90	171 864	67,77
Verbos	12 351	10,32	17 934	7,07
Total	119 738	100,00	253 616	100,00

Fuente: DLE (con Enclave RAE) y CORPES. Elaboración propia.

Sin duda, quienes no estén familiarizados con estadísticas de este tipo se sorprenderán de, por lo menos, dos aspectos distintos. En primer lugar, el número de lemas adscritos a alguna de estas cuatro clases de palabras supera las cifras manejadas habitualmente sobre el tamaño del lecionario del *DLE*. La vigésimotercera edición, publicada en formato impreso en 2014, contiene 93 111 entradas (*cf.* DLE23, xi) y ha experimentado aumentos anuales en cada una de las cinco actualizaciones digitales posteriores. La razón ha sido ya mencionada: lo que figura en la tabla 1 no son entradas, sino acepciones adscritas a alguna de esas clases de palabras. Por tanto, *regular* cuenta cuatro veces, una por cada clase de palabras a las que son adscritas las diferentes acepciones distribuidas en dos entradas distintas. El segundo aspecto consiste en que, a pesar de lo anterior, el volumen de lemas con clase correspondiente al *DLE* es menos de la mitad del que se puede documentar en el CORPES. También en este caso existen razones que justifiquen la diferencia. El lecionario de un diccionario implica siempre un filtro conectado con diferentes factores. La frecuencia y dispersión del uso es el más importante de ellos. Los lemas con frecuencia igual a 1 vinculados a una de estas cuatro clases son 103 972 en esta versión del CORPES y es comprensible que estos elementos no figuren en el diccionario. Por otro lado, en las zonas en las que lo léxico se combina con lo gramatical, los diccionarios pueden evitar, por ejemplo, la inclusión de elementos

³ Los datos correspondientes al CORPES han sido obtenidos de las listas de frecuencia de lemas (<https://apps2.rae.es/CORPES/org/publico/pages/estad/estad.view>).

cuyo significado es deducible del que tienen la base y el elemento derivativo. En la anotación de un corpus, en cambio, es necesario atribuir una etiqueta a cada elemento segmentado. Con un ejemplo claro, en el DLE figuran 1789 adverbios terminados en *mente*, mientras que en el CORPES hay 5188 (1552 de ellos con frecuencia igual a 1). Algo parecido sucede con muchos derivados.

Teniendo en cuenta estas peculiaridades, la tabla 1 muestra unos porcentajes bastante próximos, como era de esperar, pero hay también algunas diferencias de interés, que se pueden resumir en la indicación de que el DLE parece dar menos peso a los sustantivos (casi 5 puntos porcentuales por debajo) y que esa diferencia se compensa en el mayor peso de adverbios y verbos.

Dado que, como se muestra en Rojo (2006 y 2017), hay aspectos generales del léxico que presentan características diferentes según el rango de frecuencias tomado en consideración, puede ser de interés estudiar la distribución de estas cuatro clases de palabras en distintos cortes. En la tabla 2 figuran los porcentajes correspondientes a estas cuatro clases de palabras entre los 5000, 10 000, 25 000 y 50 000 lemas más frecuentes del CORPES. Incluye también, para facilitar la comparación, los datos de la totalidad del CORPES y el DLE.

TABLA 2. *Porcentajes de las clases de palabras en distintos tramos de frecuencias del lemario del CORPES*

	CORPES					DLE
	5000	10 000	25 000	50 000	Total CORPES	
Adjetivos	17,62	19,81	23,12	24,40	22,82	22,21
Adverbios	4,04	3,64	3,34	3,38	2,34	4,58
Sust. com.	56,67	57,22	57,71	59,41	67,77	62,90
Verbos	21,68	19,33	15,82	12,82	7,07	10,32
Total	100,00	100,00	100,00	100,00	100,00	100,00

Fuente: DLE (con Enclave RAE) y CORPES. Elaboración propia.

Los datos de la tabla 2 muestran que la distribución relativa de las clases de palabras cambia según los tramos de frecuencia tomados en consideración y también las clases en las que hay mayores diferencias. En términos generales, adjetivos y sustantivos aumentan sus porcentajes a medida que los tramos considerados se hacen más amplios, mientras que los adverbios y los verbos siguen la línea contraria. Resulta muy llamativo el caso de los verbos, cuyo porcentaje en la totalidad del corpus es la tercera parte del que muestran entre los 5000 lemas más frecuentes. Hay que resaltar también el caso de los adjetivos, que crecen considerablemente en

los tramos de frecuencias más bajas. Estas diferencias cuestionan la conveniencia de generalizar los datos incluidos en los diccionarios de frecuencias léxicas (*cf.* Rojo, en prensa b), que se quedan con los aproximadamente 5000 lemas de mayor frecuencia o mayor índice de uso. La tabla muestra con claridad que la configuración que sale de este tramo, importante, pero reducido, del léxico es bastante diferente del que se obtiene con la consideración de todos los lemas documentados.

4. FRECUENCIA DE USO EN CORPUS

El segundo aspecto que vamos a analizar es el referido a la frecuencia de uso de las diferentes clases de palabras en los textos. Los datos generales son los que muestra la tabla 3.

TABLA 3. *Frecuencia general y normalizada del uso de las clases de palabras en el CORPES*

	F	FN	% s/FN
Adjetivos	25 145 740	65 780,44	13,55
Adverbios	18 299.099	47 869,84	9,86
Sust. com.	85 569.202	223 846,23	46,11
Verbos	56 555.533	147 947,42	30,48
Totales	185 569.574	485 443,93	100,00

Fuente: CORPES. Elaboración propia

Antes de entrar en el análisis de lo que sucede con cada una de estas cuatro clases de palabras es interesante observar que la suma de frecuencias de todas ellas en un corpus general, representativo del español actual y constituido por unos 380 millones de formas, se queda ligeramente por debajo del 50 % del total del corpus, puesto que la frecuencia acumulada es de algo menos de 500 000 casos por millón de elementos⁴. No es un dato inesperado por lo que sabemos ya acerca de la frecuencia general de elementos como los artículos, las preposiciones, las conjunciones o los determinantes en general. Añádase a lo anterior que en el recuento que reproduce la tabla 3 tampoco entran pronombres personales ni sustantivos propios⁵.

⁴ Los porcentajes que figuran en la tabla se refieren al total de las cuatro clases, no al total del corpus.

⁵ Los nombres propios son, en esta versión del CORPES, 1 274 064 (frecuencia de inventario) y tienen una frecuencia normalizada de 33 255 casos por millón de elementos. Un nuevo caso de discrepancia fuerte entre lo que supone la frecuencia de inventario y la frecuencia de uso.

La cuestión siguiente consiste en analizar la distribución de estas cuatro clases de palabras en algunos de los diferentes tipos de texto incluidos en el CORPES: textos de ficción, textos de no-ficción publicados en forma de libro (ensayos, divulgación científica, etc.), prensa y textos orales (que suponen todavía un porcentaje muy reducido del CORPES). Los datos aparecen en la tabla 4.

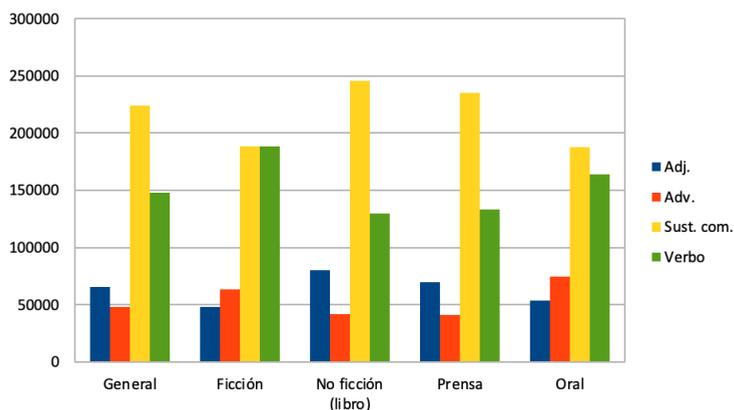
TABLA 4. *Frecuencias normalizadas de uso de las clases de palabras en distintos tipos de texto*

FN	General	Ficción	No ficción (libro)	Prensa	Oral
Adj.	65 780	48 294	80 204	69 371	53 821
Adv.	47 870	63 670	41 979	40 944	74 803
Sust. com.	223 846	188 102	245 309	235 105	187 441
Verbo	147 947	188 588	129 388	133 240	163 826
Totales	485 443	488 653	496 879	478 662	479 891

Fuente: CORPES. Elaboración propia

Tampoco aquí hay diferencias de gran entidad, diferencias que pudieran permitirnos identificar automáticamente, por ejemplo, textos de ficción, textos ensayísticos o periodísticos. Es una tipología muy general, que contiene muchos subtipos en su interior, con lo que es comprensible que se reduzcan las diferencias detectables en la comparación de textos concretos. No obstante, el gráfico 1, construido con los datos de la tabla 4, muestra algunas tendencias que convendría analizar con más detención.

GRÁFICO 1. *Frecuencia normalizada del uso de palabras en diferentes tipos de texto*



Fuente: CORPES. Elaboración propia

Se observa, por ejemplo, que los sustantivos comunes muestran una frecuencia normalizada más alta o incluso mucho más alta en todos los tipos de texto salvo en los correspondientes a ficción, en los que están al nivel de los verbos.⁶ Por otra parte, los adjetivos pesan más que los adverbios en textos de no-ficción (libros) y prensa, mientras que en ficción y orales se da la situación contraria.

Veamos, para terminar, si hay diferencias importantes en la distribución de las clases de palabras según las distintas áreas temáticas en el mismo tipo de texto. Con la estructuración del CORPES, el subcorpus más adecuado para un análisis de este tipo es, sin duda, el constituido por los textos de prensa, que se distingue con claridad de otros tipos y, al tiempo, contiene en su interior una gran variedad que el CORPES estructura en seis áreas temáticas distintas, como muestra la tabla 5.

TABLA 5. *Frecuencias normalizadas de las clases de palabras en las distintas áreas temáticas de los textos de prensa*

FN	Actualidad	Artes	Ciencia y tecn.	Cienc. Soc.	Política	Salud
Adjetivo	58 496	67 367	75 011	73 506	70 150	78 441
Adverbios	42 310	41 729	38 300	42 651	38 640	40 474
Sust. com.	223 221	226 393	248 425	232 374	237 273	252 887
Verbos	136 765	131 945	126 388	134 949	130 990	131 652
Totales	460 792	467 433	488 124	483 480	477 054	503 454

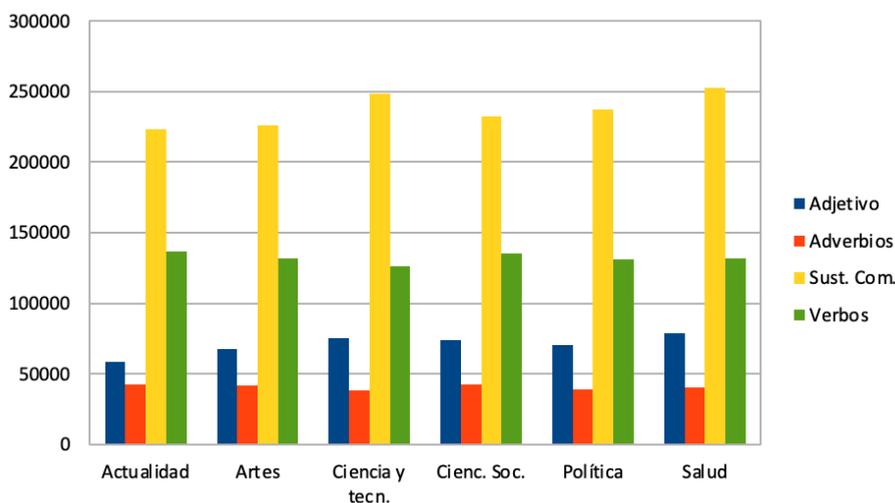
Fuente: CORPES. Elaboración propia

Los datos de la tabla 5 indican que las frecuencias normalizadas correspondientes a estas seis agrupaciones temáticas son bastante similares. Lo son en lo que suponen conjuntamente con relación al total de cada subcorpus y también en la distribución interna de las clases de palabras. Por supuesto, la diferencia en este punto entre los textos de actualidad (460 000 cpm) y los de salud (503 000) tiene importancia y convendría profundizar en este punto, pero no parece excesiva. La distancia entre estas dos áreas temáticas se ve también, en la distribución interna,

⁶ Es importante tener en cuenta que, en esta versión del CORPES, las formas compuestas de los verbos reciben un análisis unitario. *Había ofrecido*, por ejemplo, aparece directamente como pluscuamperfecto de indicativo del verbo ofrecer y cuenta como un elemento único. No se aplica este tratamiento, en cambio, a construcciones perifrásticas del tipo *ir a* + infinitivo, *estar* + gerundio, etc., que cuentan como dos apariciones de verbos.

en la existente en el peso de los adjetivos. Las demás clases resultan bastante próximas y la configuración general es la misma, como muestra con claridad el gráfico 2.

GRÁFICO 2. *Frecuencias normalizadas de las clases de palabras en las distintas áreas temáticas de los textos de prensa*



Fuente: CORPES. Elaboración propia

5. CONCLUSIÓN

Las secciones anteriores, que son un estudio muy preliminar de la estadística de las clases de palabras léxicas, muestran que su distribución en el uso no presenta grandes diferencias en los aspectos que hemos tomado en consideración. No obstante, es posible detectar algunas tendencias que merecen un análisis más detallado. Entre ellas, el descenso del porcentaje que corresponde al inventario de verbos según los diferentes tramos de frecuencia del CORPES y su relación con el que presenta el *DLE*. En cuanto a las frecuencias de uso, es destacable que los sustantivos comunes se acercan al 50 % del total del corpus analizado, que ese porcentaje desciende en los textos de ficción y aumenta en los libros de ensayo y divulgación; aumenta también en los textos que el CORPES clasifica como pertenecientes a las áreas temáticas ciencia y tecnología y salud. En términos muy generales, los adjetivos muestran un comportamiento similar, mientras que verbos y adverbios parecen moverse en dirección contraria.

REFERENCIAS BIBLIOGRÁFICAS

- BYBEE, Joan (2007): *Frequency of Use and the Organization of Language*, Oxford: Oxford University Press.
- DAVIES, Mark y Kathy H. DAVIES (2018): *A Frequency Dictionary of Spanish. Core Vocabulary for Learners*. Nueva York / Londres: Routledge.
- ROJO, Guillermo (2006): «Sobre las frecuencias verbales en español». *Haciendo Lingüística. Homenaje a Paola Bentivoglio*. Comp. por Mercedes Sedano, Adriana Bolívar y Martha Shiro. Caracas: Universidad Central de Venezuela, 309-324.
- ROJO, Guillermo (2017): «Sobre la configuración estadística de los corpus textuales». *Lingüística*, 33/1, 121-134.
- ROJO, Guillermo (2022): «Sobre algunos rasgos estadísticos del léxico de la lengua oral». *Studia philologica in honorem José Antonio Samper*. Madrid: Academia Canaria de la Lengua / Arco/Libros, 877-894.
- ROJO, Guillermo (en prensa b): «Hacia un nuevo concepto de diccionario de frecuencias». *Perspectives de recherche en linguistique et philologie romanes (I et II)*. Ed. por Dolores Corbella, Josefa Dorta y Rafael Padrón. París: Éditions de Linguistique et de Philologie, Bibliothèque de Linguistique Romane (BiLiRo), n.º 18.1-18.2, [2023], 45-63.

CORPUS Y OTROS RECURSOS ELECTRÓNICOS MENCIONADOS

- CORPES XXI: *Corpus del Español del Siglo XXI*. Versión 0.99. <<https://www.rae.es/banco-de-datos/corpes-xxi>>.
- DLE23 = Real Academia Española y Asociación de Academias de la Lengua Española: *Diccionario de la lengua española*. Madrid: Espasa-Calpe (2014). <<https://dle.rae.es/>>.
- Enclave RAE: Real Academia Española. *Enclave RAE*. <<https://enclave.rae.es/>>.