

Inventario léxico del corpus CODEA

I. Castilla la Vieja

Pedro Sánchez-Prieto Borja / Hiroto Ueda
2018

EXPLICACIÓN

Hemos elaborado un inventario de todas las formas léxicas que aparecen en los documentos provenientes de la región de Castilla la Vieja, recogidos en el corpus CODEA («*Corpus de documentos Españoles Anteriores a 1800*»)¹. Al abrir la página, nos encontramos con una hoja de Google Sheet, que ofrece funciones parecidas a las de Microsoft Excel²:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	I	Lema	C.G.	F.moderna	F.critica	F.paleográfic	Contexto mínimo	Año	A.50	Pr.	Id.	T	H	L
2	A	a	prep.	a	a	a	& {a}	1295	1250	VA	496	C	1r	15
3	A	a	prep.	a	a	a	& {a}	1298	1250	VA	353	C	1r	17
4	A	a	prep.	a	a	a	& {a}	1301	1300	AV	107	P	1r	24

En la fila aparecen los títulos siguientes:

I:	Letra inicial de Lema
Lema:	Véase 1. y 2.
C.G.:	Categoría gramatical
F. moderna	Forma moderna del español actual
F. crítica	Forma crítica en la Presentación Crítica de CODEA
F. paleográfica	Forma paleográfica en la Transcripción Paleográfica de CODEA
Contexto mínimo	Un elemento anterior y posterior de la forma léxica
Año:	Año de publicación del documento
A.50:	Franja cronológica con intervalo de 50 años
Pr.:	Provincia
Id.:	Identificador de CODEA

¹ <http://corpuscodea.es/> [24 de junio, 2018]

² <http://shimoda.lllf.uam.es/ueda/lyneal/ilc-cv.htm> [20 de junio de 2019]

T:	Tipo de documento
H:	Hoja
L:	Línea

Para llevar a cabo la agrupación de las formas léxicas por Lema, que corresponde aproximadamente a la entrada del *Diccionario de la lengua española* de la Real Academia Española (DRAE)³, hemos utilizado los dos textos paralelos del Corpus: la Transcripción Paleográfica (TP) y la Presentación Crítica (PC) de CODEA. Para lematizar las formas paleográficas hemos preparado una Edición Moderna (EM) a partir de la Presentación Crítica (PC) siguiendo la ortografía académica actual. La siguiente tabla muestra las cuatro formas con un ejemplo del lema *abad*:

Forma de transcripción:	Ejemplo
1. Transcripción paleográfica (TP)	<i>abbad<e>s</i>
2. Presentación crítica (PC)	<i>abades</i>
3. Edición moderna (EM)	<i>abades</i>
4. Lema (L)	<i>abad</i>

1. Unión y separación de la transcripción paleográfica

Para hacer corresponder una por una las cuatro formas (TP, PC, EM, L) arriba mencionadas, hemos necesitado llevar a cabo la operación manual semiautomática de la unión y separación de textos de la Transcripción Paleográfica. Hemos utilizado tres signos: el signo más (+) para unir las formas separadas, el signo igual (=) para unir formas divididas en las dos líneas, de arriba y de abajo, el signo menos, guión, (-), para separar las formas unidas. Veamos los ejemplos:

(1) Correspondencia completa entre TP y PC:

TP: 1277 BU CV C Pri. Cas A 1r 1 *Sepan q<ua>ntos este P<ri>uilegio uieren & oyeren. Como nos don ALFONSO por la gr<aci>a de dios Rey de Castiella de Toledo de Leon de Gallizia de Seuilla de Cordoua de Murçia de Jah<e>n & del Algarue. En uno con la Reyna*

PC: 1277 BU CV C Pri. Cas B 1r 1 *Sepan cuantos este privilegio vieren e oyeren*

³ <http://dle.rae.es/> [24 de junio, 2018]

cómo nós don Alfonso, por la gracia de Dios rey de Castiella, de Toledo, de León, de Gallizia, de Sevilla, de Córdoba, de Murcia, de Jaén e del Algarve, en uno con la reína

(2) Unión por el signo de plus (+)

TP: 20 1450 P CV P Cor. Cas A 1r 5 *ferra<n>do ortega me dixo q<ue> desde la feria de **me+djna** escriujera al sen<n>or marq<ue>s*

PC: 20 1450 P CV P Cor. Cas B 1r 5 *Ferrando Ortega me dixo que desde la feria de **Medina** escriviera al señor marqués,*

(3) Cambio de línea por el signo de igualdad (=)

TP: 20 1450 P CV P Cor. Cas A 1r 22 *q<ue> enbiase ma<n>dar A j<oa>n ma<n>rriq<ue> q<ue> pre<n>diese A ferrand<o> malo plazeme d<e>- -lo **ma<n>=dar***

PC: 20 1450 P CV P Cor. Cas B 1r 22 *que embiase mandar a Joán Manrique que prendiese a Ferrando Malo, plázeme de lo **mandar**;*

(4) Separación por guion (-)

TP: 1277 BU CV C Pri. Cas A 1r 2 *don<n>a YOLANT mi mugier & con n<uest>ros fijos el Jnffante don Sancho fijo mayor & con don Pedro & don Joh<a>n & don Jaymes. Por grand sabor q<ue> auemos de fazer bien & merçed al Conçeio de Guadalffaiara tan bien **a-los de-la villa***

PC: 1277 BU CV C Pri. Cas B 1r 2 *doña Yolant mi mugier; e con nuestros fijos el infante don Sancho, fijo mayor; e con don Pedro, e don Joán e don Jaimes, por grand sabor que avemos de fazer bien e merced al concejo de Guadalfajara, tan bien **a los de la villa***

De esta manera, la TP **me+djna**, por ejemplo, representa la forma unida de la TP original separada: **me djna**. La TP unida por =, **ma<n>=dar**, representa **ma<n>** al final de la línea y **dar** al principio de la línea siguiente. La TP **a-los de-la villa** representa las

TP originales unidas: *alos dela villa*.

2. Lematización

Hemos hecho corresponder las formas críticas a las formas modernas cuando las dos coinciden tanto en su raíz como prefijos, sufijos y flexiones. Por ejemplo, la forma verbal *toviéredes* corresponde a la forma moderna *tuviereis*, que figura en el lema *tener*.

Cuando discrepa uno de los elementos morfológicos, independizamos las formas modernas, aunque no se encuentren en el español actual, por ejemplo la forma crítica *atal* 'tal' no la incluimos en el lema *tal*, sino en el lema *atal* 'tal'.

El criterio adoptado para el método de lematización, es decir, agrupación de las formas modernas por lema, obedece a la clasificación de la Categoría Gramatical (CG) de cada lema.

Abreviatura	Categoría Gramatical	Ejemplo de <i>lemas</i> y <i>formas</i>
adj.	adjetivo	<i>blanco, blancos, blanca, blancas</i>
adv.	adverbio	<i>abajo, adelante, antiguamente</i>
art.	artículo	<i>el, los, la, las, lo, l; un, unos, una, unas</i>
conj.	conjunción	<i>y; o; maguer; pero</i>
conj. relat.	conjunción / relativo	<i>que</i>
demos.	demonstrativo	<i>este, estos, esta, estas, esto</i>
indef.	indefinido	<i>otro, otros, otra, otras</i>
interj.	interjección	<i>amen</i>
interrog.	interrogativo	<i>cómo; dónde; qué; quién, quiénes</i>
n. prop.	nombre propio	<i>Fernando, López</i>
num.	numeral	<i>uno, dos, tres, cuatro</i>
poses.	posesivo	<i>mi, mis, mío, míos,</i>
prep.	preposición	<i>a, con, de, en, por</i>
pron. clit.	pronombre clítico	<i>me; nos; os; lo, los, la, las, le, se [dat.], se [ref.]</i>
pron. suj.	pronombre de sujeto	<i>yo; nosotros, nosotras, nos; vosotros, vos; él, ellos, ella, ellas, ello</i>
pron. prep.	pronombre de	<i>mí, -migo; sí, -sigo; nos, -nosco; vos, -vusco</i>

	preposición	
relat.	relativo	adonde; cual, cuales; cuanto, cuantos, cuanta, cuantas, cuan; cuyo, cuyos, cuya, cuyas; donde, quien, quienes.
sus.	sustantivo	abad, abades, abadesa, abadesas; abadía; abuelo, abuelos, abuela, abuelas; rey, reina.
vb.	verbo	confirmar, confirmo, confirma

- Hemos dividido las formas contractas de *al* y *del* en *a-l* y *d-el*.
- Hemos agregado el significado entre comillas en los lemas desusados en el español actual, por ejemplo, **ál** ‘otro’, **aqueste** ‘este’, etc.
- No hemos incluido las formas latinas, por ej. *dei gratia rex*.
- No hemos incluido los numerales romanos, *i, ii, iii, v, x, d, c, m*.
- En cuanto al pronombre clítico, hemos independizado **me, nos, os** en tres lemas diferentes, mientras que agrupamos bajo el lema **lo** todos los pronombres de tercera persona, singular y plural, incluyendo *se* dativo, *se* [dat.] y *se* reflexivo, *se* [ref.]. Hemos hecho la distinción entre las formas del artículo definido **el**, concretamente, *lo, la, las* y los pronombre clíticos homófonos.
- Los pronombre de sujeto, **yo, nosotros, vosotros**, se independizan en los lemas diferentes, mientras que los pronombres de sujeto de tercera persona, **él, ellos, ella, ellas, ello**, se agrupan en el mismo lema **él**.
- Los pronombres especiales que aparecen solo en contracción con la preposición *con*, en forma de *-migo, -sigo, -nosco, -vusco* son agrupados bajo los lemas, **mí, sí, nos, vos**, respectivamente.
- Las formas plurales del sustantivo están bajo el mismo lema del singular. Las formas femeninas de persona figuran en el mismo lema de la forma masculina, por ejemplo, **abuelo, abuelos, abuela, abuelas; abad, agadesa; príncipe, princesa; rey, reina**. Las formas femeninas de la no persona pertenecen a los lemas diferentes: **huerto, huertos; huerta, huertas**.
- El lema de un verbo en forma de infinitivo agrupa todas las formas conjugadas, incluyendo los gerundios y los participios pasados, masculino y femenino, singulares y plurales: **volver, volvería, volvían, volviendo, volviere, volviesen, volvió, vuelta, vuelto, vueltos, vuelva, vuelvan, vuelven**.

3. Frecuencia

A continuación ofrecemos las frecuencias de formas correspondientes a cada grupo.

(1) Grupo de letra inicial:

Letra inicial	Frec. forma
A	16 817
B	2 683
C	21 556
D	42 993
E	39 605
F	3 362
G	3 171
H	8 183
I	1 660
J	2 950
L	8 868
M	16 125
N	8 483
O	7 375
P	20 015
Q	12 886
R	5 987
S	18 116
T	8 647
U	2 101
V	7 678
Y	25 635
Z	88
Total	284 984

(2) Grupo de categoría gramatical:

Abreviatura	Categoría Gramatical	Frec. forma
adj.	adjetivo	10 703
adv.	adverbio	10 021
art.	artículo	22 240

conj.	conjunción	27 942
conj. relat.	conjunción. relativo	11 368
demos.	demostrativo	4 053
indef.	indefinido	6 178
interj.	interjección	4
interrog.	interrogativo	558
n. prop.	nombre propio	20 289
num.	numeral	4 678
poses.	posesivo	6 860
prep.	preposición	48 951
pron. clit.	pronombre clítico	7 402
pron. prep.	pronombre de preposición	575
pron. suj.	pronombre de sujeto	5 965
relat.	relativo	1 935
sus.	sustantivo	51 749
vb.	verbo	43 513
-	Total	284 984

(3) Grupo de franja cronológica con intervalo de 50 años:

A.50	Frec. forma
1200	5 003
1250	52 536
1300	34 869
1350	21 443
1400	32 925
1450	46 789
1500	48 223
1550	16 684
1600	8 968
1650	4 298
1700	6 962
1750	6 284
Total	284 984

(4) Grupo de provincia:

Para abreviar el nombre de la Provincia, hemos utilizado el Código del Ministerio del Interior:

Código del Ministerio del Interior	Provincia	Frec. forma
AV	Ávila	44 497
BU	Burgos	44 133
LO	La Rioja (Logroño)	20 535
P	Palencia	13 577
S	Cantabria (Santander)	21 793
SG	Segovia	29 095
SO	Soria	9 274
VA	Valladolid	102 080
-	Total	284 984

(5) Grupo de tipo del documento:

Código	Ámbito del documento	Frec. forma
C	Cancilleresco	119 094
E	Eclesiástico	67 482
J	Judicial	19 406
M	Municipal	10 108
P	Particular	68 894
-	Total	284 984

FIN