

Lematización y visualización cartográfica del corpus CODEA.

Formas de la conjunción ‘y’ en el norte de Castilla medieval

Hiroto Ueda / Pedro Sánchez-Prieto Borja / Antonio Moreno Sandoval

Universidad de Tokio / Universidad de Alcalá / Universidad Autónoma de Madrid

uedahiroto@jcom.home.ne.jp

pedro.sanchezp@uah.es

antonio.msandoval@uam.es

Resumen

A pesar de que en la lingüística de corpus existen controversias en cuanto a la manera de lematizar o agrupar las formas flexivas en una forma principal, hay consenso entre los investigadores sobre su utilidad en distintos niveles de estudio. Especialmente en el nivel gráfico, fonológico y morfológico, la lematización es fundamental sin ningún género de dudas. En la práctica, en el inventario de formas lematizadas, no se nos puede escapar ninguna variante rara, poco frecuente, en la observación total de la agrupación.

Como ejemplo de la importancia de la lematización de documentos antiguos españoles, hemos analizado distintas formas de la conjunción ‘y’, que es peculiar en castellano al lado de la forma general <e> en las lenguas históricamente relacionadas. Efectivamente, las variantes precursoras de la forma <y> se han encontrado con anterioridad al siglo XV, y nos sirven de importantes indicadores para aproximarnos a su aparición repentina en el mismo siglo, con la llegada de la nueva letra humanística. Anteriormente, la forma latente <y> estaba oculta bajo la letra gótica en forma del signo tironiano <&>. Veremos las razones del cambio histórico de <e> a <y> en el castellano medieval, que se caracteriza por la abundancia estadística del diptongo creciente, <ie>, idiosincrásico del castellano en contraste con las lenguas colindantes. Al finalizar el estudio, presentamos la historia de la conjunción en su dimensión geohistórica con visualizaciones cartográficas.

Palabras clave: lematización, conjunción ‘y’, letra gótica, letra humanística, diptongo creciente, signo tironiano, Burgos

Abstract

Although there are controversies in the corpus linguistics as to the way of lemmatizing or grouping the inflective forms in a main form, their usefulness at different levels of study

is admitted in the consensus of researchers. Especially at the graphic, phonological and morphological level, lemmatization is essential without any doubt. In practice, in the inventory of lemmatized forms, we do not miss a rare, infrequent variant, in the total observation of the grouping.

As an example of the utility of lemmatization of old Spanish documents, we have analyzed different forms of the conjunction <y>, which is peculiar in Spanish in comparison with the general form of <e> in historically related languages. Indeed, the precursor variants of the <y> form have been found before the fifteenth century, which serve as important indicators to approximate their sudden appearance in the same century, with the arrival of the new humanistic letter. Previously the latent form <y> was hidden under the Gothic letter in the form of the Tironian sign <&>. We look for the reasons of the historical change from <e> to <y> in medieval Castilian, which is characterized by the statistical abundance of the rising diphthong, <ie>, idiosyncratic of Castilian in contrast to the surrounding languages. At the end of the study, we present the history of the conjunction in its geohistorical dimension with cartographic visualizations.

Keywords: lemmatization, conjunction ‘y’, Gothic letter, Humanistic letter, rising diphthong, Tironian sign

1. Introducción*

La utilidad de un corpus lingüístico se puede medir por dos parámetros, su fiabilidad y las posibilidades de recuperación de la información que ofrezca. El primero resulta básico, pues un corpus será tanto más útil cuanto más fiables sean sus materiales textuales. En este sentido, la posibilidad de confrontar las transcripciones con el facsímil les otorga un respaldo que no podría alcanzarse por ninguna otra vía. Los materiales se han preparado expresamente para ese corpus, el CODEA, pues solo así es posible el control de todo el proceso de edición, empezando por la aplicación de unos criterios uniformes de transcripción. En el caso de los corpus que elaboramos en el Grupo de Investigación de Textos para la Historia del Español (GITHE) de la Universidad de Alcalá¹, y que son el «Corpus de Documentos Españoles a 1900 (CODEA+ 2020)»², el «Atlas Lingüístico Diacrónico e Interactivo de la Comunidad de Madrid»³ y el «Corpus Hispánico y Americano en la Red: Textos Antiguos (CHARTA)»⁴, los criterios aplicados son los de esta última red, y que han logrado una aceptación notable entre quienes publican fuentes

documentales. Estos criterios prevén una triple presentación de los textos editados: (1) transcripción paleográfica, (2) presentación crítica y (3) facsímil. La primera muestra los usos gráficos del manuscrito (*ssaber, conthenidos, ujsto*), mientras que la crítica unifica ciertas diferencias gráficas que no tienen valor fonético (*saber, contenido, visto*).

El segundo, para poder realizar el segundo parámetro de utilidad, las diferentes posibilidades de recuperación de la información, es necesario contar con un buscador que permita explotar todas las cuestiones previstas por los criterios de edición. En este sentido, las consultas al corpus pueden hacerse tanto dentro de la transcripción paleográfica como de la presentación crítica, aunque por defecto estas se realizan dentro de la última. En la paleográfica, pueden buscarse palabras gráficas y secuencias de caracteres en posición inicial, interior y final de “palabra”, lo que revelará el comportamiento gráfico de los textos en diferente época, lugar, escribano, tipo documental, etc. (p. ej., *ss-*, *-ns-*, *-rt*). La presentación crítica admitía hasta ahora búsquedas por formas, tanto palabras (*pascan*) como secuencias de caracteres dentro de una palabra (*-sci-*), ya coincidieran con elementos léxicos (*antig-*) o morfológico (*-ades*, *-áis*) o no (*-ebo*).

La aplicación no permitía la búsqueda por lemas. De esta manera, era relativamente fácil buscar el paradigma de un verbo regular como *aprovechar* (“*apro.ech-*”), pero no cuando la morfología era irregular, como la de *ir* (*voy, va, fue*). Por otra parte, si bien en un corpus de textos contemporáneos la variación formal dentro de un lema resulta previsible casi en su totalidad, no sucede de esta manera con los textos antiguos; así, quien rastree la historia de la palabra *constituir*, es bastante probable que no incluya, al menos en primera instancia, la variante *conestituir*, difundida en documentos de los ss. XV-XVI.

Resultaba, pues, del todo necesario para sacar el máximo partido al corpus CODEA llevar a cabo una lematización completa del mismo. Las ventajas son notorias, pues resulta así posible encontrar la variedad formal, no evidente, de verbos como *estar* (p. ej., *entudiésemos* ‘*estuviésemos*’). La ocasión de llevar a cabo este proceso de lematización, del que ahora se ofrecen los primeros resultados, surgió en el marco de una colaboración entre la Universidad de Tokio, la Universidad Autónoma de Madrid y la de Alcalá para el proyecto citado CODEA+ 2020. Hiroto Ueda y Antonio Moreno Sandoval ya habían desarrollado, bajo el nombre de LYNEAL, una aplicación para el tratamiento estadístico del corpus CODEA en su versión 2015⁵.

Por ahora, se ofrece el inventario léxico completo de los documentos de CODEA emitidos en las provincias de Castilla la Vieja: Ávila (AV), Palencia (P), Burgos (BU), Valladolid (VA), Segovia (SG) y Soria (SO), pero no las del antiguo reino de León. En fases sucesivas se incorporarán otras regiones hasta completar la geografía peninsular actualmente incluida en CODEA desde el s. XII al XIX, según lo previsto para la versión CODEA+ 2020.

El objetivo de este estudio es exponer la lista de los lemas más frecuentes (sección 2), junto con la aplicación posterior en la cartografía lingüística, tomando por ejemplos cambios cronológicos y variaciones geográficas de las formas de la conjunción castellana <y> en la región de Castilla la Vieja del siglo XIII al XVI (sección 3).

2. Lemas frecuentes

Uno de los provechos que podemos sacar de los datos lematizados es la capacidad de observación estadística de frecuencias del léxico. La estadística léxica siempre presenta una curva pronunciada de frecuencias acumuladas. Se trata de los lemas ordenados de manera descendente con porcentajes que acumulan las frecuencias precedentes de la lista. En nuestro caso, la preposición «de» presenta la mayor frecuencia, 24 760, que ocupa 8,6% de la totalidad. Seguidamente, con el artículo «el» (frec.: 23 819), el porcentaje acumulativo llega a 16,9%. Es curioso observar que, entre las tres palabras más frecuentes, «de» (prep.), «el» (art.) e «y» (conj.), se obtiene el 25,1%, es decir, una cuarta parte de la totalidad del léxico del corpus.

Para cubrir la mitad del corpus (50%), es suficiente observar los primeros 23 lemas, como se presenta la siguiente tabla y el gráfico:

Lema	Frecuencia absoluta	F. acumulada	% acumulado	Orden
<i>de</i> (prep.)	24 760	24 760	8.6	1
<i>el</i> (art.)	23 819	48 579	16.9	2
<i>y</i> (conj.)	23 695	72 274	25.1	3
<i>que</i> (comp.)	11 368	83 642	29.0	4
<i>en</i> (prep.)	6 921	90 563	31.4	5
<i>decir</i> (v.)	6 910	97 473	33.8	6
<i>a</i> (prep.)	5 886	103 359	35.9	7

<i>por</i> (prep.)	4 911	108 270	37.6	8
<i>ser</i> (v.)	4 282	112 552	39.1	9
<i>este</i> (dem.)	3 653	116 205	40.4	10
<i>hacer</i> (v.)	2 802	119 007	41.3	11
<i>lo</i> (clit.)	2 766	121 773	42.3	12
<i>su</i> (pos.)	2 492	124 265	43.2	13
<i>haber</i> (v.)	2 468	126 738	44.0	14
<i>él</i> (pro-s.)	2 467	129 205	44.9	15
<i>todo</i> (indef.)	2 339	131 544	45.7	16
<i>mi</i> (pos.)	2 131	133 675	46.4	17
<i>con</i> (prep.)	2 099	135 774	47.1	18
<i>o</i> (conj.)	1 831	137 605	47.8	19
<i>don</i> (n.)	1 776	139 381	48.4	20
<i>no</i> (av.)	1 763	141 144	49.0	21
<i>otro</i> (indef.)	1 750	142 894	49.6	22
<i>dar</i> (v.)	1 684	144 578	50.2	23

Tabla 1. Primeros lemas más frecuentes que ocupan la mitad del corpus

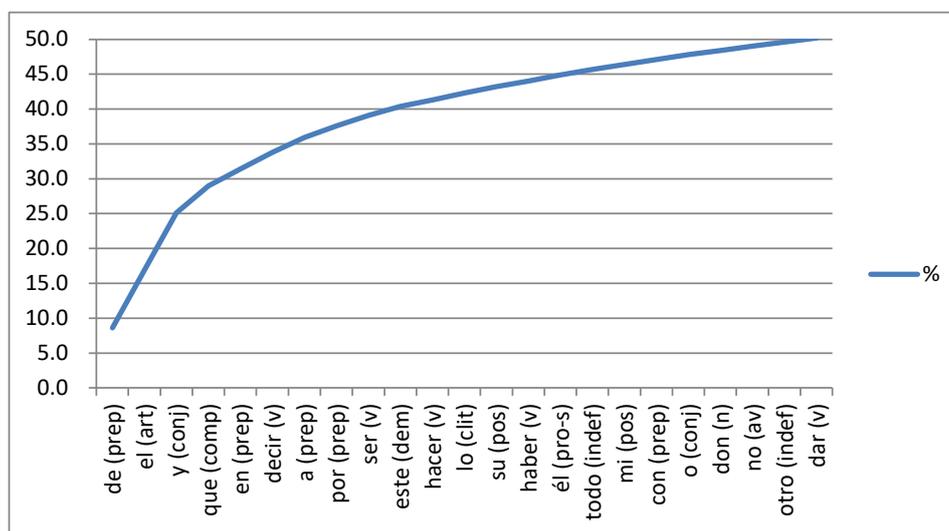


Fig. 1. Lemas más frecuentes que ocupan la mitad del corpus

Por la frecuencia destacada que poseen estos 23 lemas, creemos que merecen un tratamiento especial en los estudios estadísticos de las formas documentadas. Naturalmente, otras formas también son importantes para nuestro estudio histórico de la lengua. Por ejemplo, los lemas del nivel del 75% son:

(1) *carta* (n.) [frecuencia: 1676], *ni* (av.) [1636], *mandar* (v.) [1543], *para* (prep.) [1536], *nuestro* (pos.) [1472], *yo* (pro-s.) [1406], *se* (clit.) [1398], *le* (clit.) [1375], *un* (art.) [1372], *rey* (n.) [1218], *vosotros* (pro-suj.) [1168], *año* (n.) [1051], *tener* (v.) [1010], *señor* (n.) [1007], *como* (conj.) [977], *Juan* (npr.) [966], *parte* (n.) [931], *nosotros* (pro-s.) [914], *día* (n.) [900], *os* (clit.) [850], *santo* (a.) [818], *cualquiera* (a.) [801], *así* (av.) [766], *vuestro* (pos.) [763], *monasterio* (n.) [754], *maravedí* (n.) [749], *villa* (n.) [698], *cada* (indef.) [689], *cuanto* (rel.) [673], *ver* (v.) [655], *sobredicho* (a.) [654], *confirmar* (v.) [652], *mil* (nm.) [648], *poder* (v.) [636], *Pedro* (npr.) [630], *me* (clit.) [624], *más* (av.) [615], *estar* (v.) [603], *si* (conj.) [602], *alguno* (indef.) [598], *cosa* (n.) [591], *cual* (rel.) [591], *lugar* (n.) [583], *merced* (n.) [557], *saber* (v.) [552], *escribir* (v.) [548], *ciudad* (n.) [545], *bien* (n.) [534], *cumplir* (v.) [529], *dos* (nm.) [524], *escribano* (n.) [523], *hijo* (n.) [519], *casa* (n.) [508], *contra* (prep.) [489], *Dios* (npr.) [485], *ninguno* (indef.) [481], *Pérez* (npr.) [477], *testigo* (n.) [477], *pena* (n.) [474], *mí* (pro-p.) [472], *presente* (a.) [461], *otorgar* (v.) [453], *Fernando* (npr.) [443], *pagar* (v.) [442], *derecho* (n.) [435], *manera* (n.) [422], *convento* (n.) [417], *uno* (nm.) [416], *razón* (n.) [414], *aquí* (av.) [412], *Alfonso* (npr.) [406], *abad* (n.) [404], *cómo* (interrog.) [400], *nos* (clit.) [399], *Martín* (npr.) [395], *bien* (av.) [394], *ante* (prep.) [393], *vender* (v.) [389], *doña* (n.) [389], *tierra* (n.) [386], *venir* (v.) [386], *quien* (rel.) [385], *porque* (conj.) [381], *Domingo* (npr.) [381], *María* (npr.) [376], *público* (a.) [375], *pasar* (v.) [375], *ende* (av.) [375], *querer* (v.) [369], *real* (a.) [366], *reino* (n.) [366], *pedir* (v.) [365], *guardar* (v.) [364], *según* (prep.) [363], *sobre* (prep.) [357], *obispo* (n.) [356], *iglesia* (n.) [355], *hombre* (n.) [349], *persona* (n.) [346], *tiempo* (n.) [342], *ir* (v.) [330], *recibir* (v.) [328], *vecino* (n.) [327], *hasta* (prep.) [327], *moneda* (n.) [326], *gracia* (n.) [322], *adelante* (av.) [322], *sin* (prep.) [317], *poner* (v.) [310], *tres* (nm.) [307], *alcalde* (n.) [306], *valer* (v.) [305], *Castilla* (npr.) [305], *privilegio* (n.) [297], *bajo* (prep.) [296], *ahora* (av.) [296], *contener* (v.) [295], *bueno* (a.) [292], *concejo* (n.) [289], *mismo* (a.) [287], *García* (npr.) [285], *prior* (n.) [283] (Total: 216 061: 75,0%, rango: 145).

Para llegar al 80%, tenemos que contar con 213 lemas y para llegar a 90%, 577 lemas. El linde significativo de estadística, 95%, se consigue con 1 206 lemas.

Pensamos que los fenómenos que se presentan en estos lemas son cuantitativamente importantes para determinar la magnitud estadística de ellos, pues suelen ser influyentes en otras palabras que los siguen. Los cinco lemas más frecuentes, «de», «el», «y», «que» y «en», son tan ubicuos que no pueden faltar en casi ningún texto, de modo que sus variantes sirven como rasgos característicos fidedignos de cada documento, de cada época y de cada localidad. Nos interesa investigar la dinámica que se presenta en las relaciones entre las formas frecuentes y las poco frecuentes. También es interesante calcular la riqueza léxica de distintos documentos registrados. La relación que se supone existir entre la frecuencia y la estabilidad léxica es uno de los temas candentes en los estudios de variación lingüística. De esta manera, hay innumerables procesos históricos interesantes para estudios variacionistas, que ahora son factibles de abordar gracias al trabajo de lematización.

En lo que sigue, veremos como ejemplo la historia y la geografía de las formas de la conjunción «y», que es, en el plano cuantitativo, sumamente importante ya que figura dentro de los cinco lemas más frecuentes.

3. Conjunción ‘y’

La conjunción copulativa «y» es uno de los lemas más frecuentes que poseen formas variables a lo largo de la historia. Efectivamente, han existido distintas grafías, <et>, <e>, <y>⁶, al lado del signo tironiano que se transcribe con <&>. En ellas, la forma actual <y> es peculiar dentro de las lenguas románicas, puesto que fuera de la Península, todas mantienen la vocal [e], procedente del lat. ET⁷. Curiosamente, las tres lenguas ibéricas coinciden en el fonema /i/, incluido el portugués, no obstante su ortografía <e>.

Según nuestra investigación cuantitativa de los documentos castellanos (Ueda 2018), el aumento de frecuencia de la forma <y> no ha sido gradual, sino repentino y explosivo a mediados del siglo XVI. Lo repentino de los cambios lingüísticos no es raro: suelen observarse indicios precursores poco frecuentes pero significativos, seguidos de una adopción general masiva, lo que produce una curva acumulativa pronunciada de subida. Lo peculiar del cambio de <e> en <y> está en que la mayoría de veces esta conjunción no se representaba mediante la letra <e>, sino por el signo tironiano <&> desde el principio. En realidad, al lado de numerosas apariciones de <&>, encontramos casos sueltos de <e> e <y> en la historia, aunque con poca frecuencia. Esto indica que había

existido la variación entre las dos formas, la <e> tradicional y la <y>, latente bajo el signo de <&>. Cuando desaparece el signo <&> en el siglo XV, la nueva variante <y>, no nacida en el momento sino preparada con anterioridad, aparece con una frecuencia preponderante.

Es cierto que se ha evitado la secuencia en hiato, *e + e*, *e + a*, *e + o*, a nuestro modo de ver, no necesariamente por disimilación (Menéndez Pidal, 1976: 297; De Andrés Díaz 2013: 728-9), sino más bien por adaptación al modelo de diptongos, *ie*, *ia*, *io* (Alvar y Pottier 1983: 321-322). Hemos visto que la secuencia hiática, *ee*, *ea*, *eo*, es casi inexistente en español antiguo, mientras que los diptongos crecientes son sumamente numerosos (Ueda 2018: 19).

Somos conscientes de que los datos expuestos apoyarían tanto la teoría basada en la disimilación como la nuestra basada en la adaptación al modelo de diptongos. Sin embargo, existe una diferencia fundamental en las dos. La diferencia consiste en que en la disimilación se observa un intento de distinguir el sonido para que no se confunda con otro, mientras que la adaptación se trata de asimilar al modelo influyente. La dirección de las dos, disimilación y adaptación, es opuesta y, por tanto, completamente diferente. Otra diferencia está en que la disimilación se presenta en el plano sintagmático, /e/ + /a/ → /i/ + /a/, mientras que la adaptación, en el paradigmático, /ea/ vs. /ia/. Lo que proponemos es ofrecer otro punto de vista: adaptación a las formas de diptongos, sin negar completamente la validez de la teoría de disimilación.

Si el cambio de <e> en <y> se produce por la adaptación general al modelo de diptongos frecuentes y/o por la disimilación en la cadena de dos vocales, ¿por qué razón no aparece la nueva forma <y> en siglos medievales? Pensamos que los primeros indicios esporádicos de <y> son indicadores de la conjunción expresada con fonema de /i/. Durante siglos, la conjunción vacilaba entre la pronunciación antigua /e/ y la nueva /i/, bajo la apariencia del signo <&>. Estas dos variantes no funcionarían en oposición alguna, puesto que las secuencias /ea/ e /ia/ no se distinguirían fonológicamente por la escasez de la primera /e/. El rendimiento funcional de esta pareja era (y sigue siendo) mínimo. En el siglo XVI, con el cambio de estilo de letras, de la gótica a la humanística, desaparece el signo latino tradicional <&>, por lo que de repente se manifiesta la grafía <y> preparada desde hacía siglos⁸.

Tras repasar nuestro estudio anterior, en esta ocasión, vamos a observar más de cerca en Castilla la Vieja las frecuencias de <e> e <y>, cruciales para la historia de las formas de la misma conjunción. En primer lugar, hay que distinguir entre las dos tipologías documentales: los documentos cancillerescos y los no cancillerescos, por dos razones. Una por el carácter itinerante del escritorio canceleresco (es decir, no representa el dialecto del lugar de donde se emiten) y otra por su normatividad destacada, que lo hace apartarse de otros documentos, episcopales, jurídicos, municipales y privados. La Tabla 2 muestra la distribución de frecuencia absoluta de las cuatro formas en cuestión en los documentos cancelerescos⁹:

Cancilleresco	1250	1300	1350	1400	1450	1500
&	372	118	108	178	351	115
<i>et</i>	30	16	16	21		
<i>e</i>	9	4	21	19	76	51
<i>y</i>	3				25	93

Tabla 2. Frecuencia de las formas del lema «y». Documentos cancelerescos

Llamamos la atención sobre el hecho de que la forma <y> es casi inexistente antes de 1450. En cambio, en los documentos no cancelerescos, la situación es diferente. En la siguiente tabla se observan ciertas frecuencias de <y> durante los siglos medievales:

No canceleresco	1250	1300	1350	1400	1450	1500
&	326	402	171	230	305	440
<i>et</i>	34	65	7	16	3	1
<i>e</i>	13	16	12	46	73	96
<i>y</i>	14	4	1	5	42	132

Tabla 3. Frecuencia de las formas del lema «y». Documentos no cancelerescos

Creemos que la norma canceleresca no permitía la forma popular <y>, que es una nueva formación con respecto a <et> y <e> procedentes del latín ET. La diferencia entre los dos tipos de documentos parece apoyar nuestra hipótesis basada en la adopción más bien que en la disimilación. Efectivamente, la adopción de la forma general, es decir, la asimilación al diptongo creciente predominante de la época nace del pueblo, que sigue la tendencia general. Mientras que la disimilación es un producto del intento culto consciente de

distinción de las vocales continuas. Si la aparición de <y> es más temprana en los documentos no cancillerescos, es lógico pensar que el fenómeno consiste en la adopción propia del pueblo, más que la disimilación característica del escritorio cancellesco.

Ahora nos interesa observar la nueva forma <y> en su cronología y geografía en Castilla la Vieja. Veamos su progresión numérica en 19 casos de combinación cronogeográfica donde se registra la forma <y> en los documentos no cancellescos:

A50	Lugar	&	e, he, et	y, i, j	y/(y + e)
1250	P	96	1	0	0,000
1250	S	15	13	0	0,000
1250	BU	160	38	1	0,026
1250	AV	919	100	6	0,057
1250	LO	204	63	6	0,087
1250	VA	155	32	6	0,158
1300	AV	1198	189	0	0,000
1300	BU	125	32	0	0,000
1300	S	72	6	0	0,000
1300	LO	77	82	3	0,035
1300	VA	82	26	1	0,037
1350	AV	103	20	0	0,000
1350	LO	34	17	0	0,000
1350	VA	374	63	0	0,000
1350	S	295	48	1	0,020
1400	LO	500	132	0	0,000
1400	S	170	56	0	0,000
1400	SG	136	31	0	0,000
1400	SO	38	58	0	0,000
1400	VA	136	77	0	0,000
1400	P	114	13	11	0,458
1450	LO	20	24	0	0,000
1450	P	330	76	0	0,000
1450	S	322	34	7	0,171

1450	AV	367	114	29	0,203
1450	BU	308	57	101	0,639
1450	VA	168	17	32	0,653
1500	BU	47	145	44	0,233
1500	SG	84	124	44	0,262
1500	VA	554	170	122	0,418
1500	S	701	42	36	0,462
1500	P	257	18	46	0,719
1500	AV	421	10	37	0,787
1500	LO	25	18	122	0,871

Tabla 4. Relación de <y> con respecto a <e> en orden cronológico

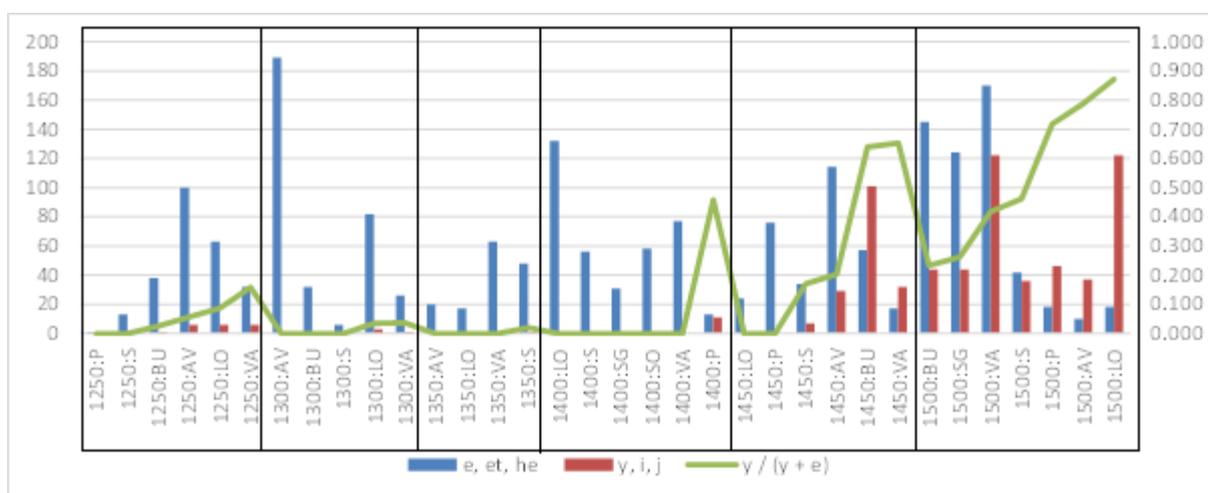


Fig. 2. Relación de <y> con respecto a <e> en orden cronológico

Por esta visualización gráfica se nota la tendencia cronogeográfica de la nueva forma <y>: en los años entre 1250 y 1350 (hasta 1400), la nueva forma <y> es minoritaria. A partir de 1400, la misma forma aparece primero en Palencia (P.1400) y posteriormente en Burgos (BU.1450) y Valladolid (VA.1450), que son espacios precursoros. En 1500, la nueva forma <y> es predominante en Palencia (P.1500), Ávila (AV.1500) y Logroño (LO.1500).

Vamos a repasar la historia en su dimensión geográfica. Empezamos en la franja cronológica de 1250 (Fig. 3), donde la forma <y> es todavía esporádica en Burgos (BU), Logroño (LO), Valladolid (VA) y Ávila (AV). En 1300 (Fig. 4) y 1350 (Fig. 5), la situación sigue prácticamente igual, no muy diferente de 1250. En 1400 (Fig. 6),

encontramos un indicio precursor de <y> en Palencia (P). En 1450 (Fig. 7), el aumento de <y> en Burgos (BU) es notable, seguida de Santander (S), Valladolid (VA) y Ávila (AV). En 1500 (Fig. 8), la nueva forma <y> se expande por todo el territorio de Castilla la Vieja. La visualización cronocartográfica nos facilita la observación espacial del fenómeno lingüístico en sucesivas etapas históricas, lo que es difícil de percibir en las simples tablas y gráficos de frecuencia. La clave del desarrollo de la conjunción <y> está en el Burgos (BU) de 1450, que es el centro de la geografía indicada, del que se produjo la expansión en 1500 en las comarcas colindantes, Santander (S), Logroño (LO), Palencia (P), Valladolid (VA), Segovia (SG) y Ávila (AV).

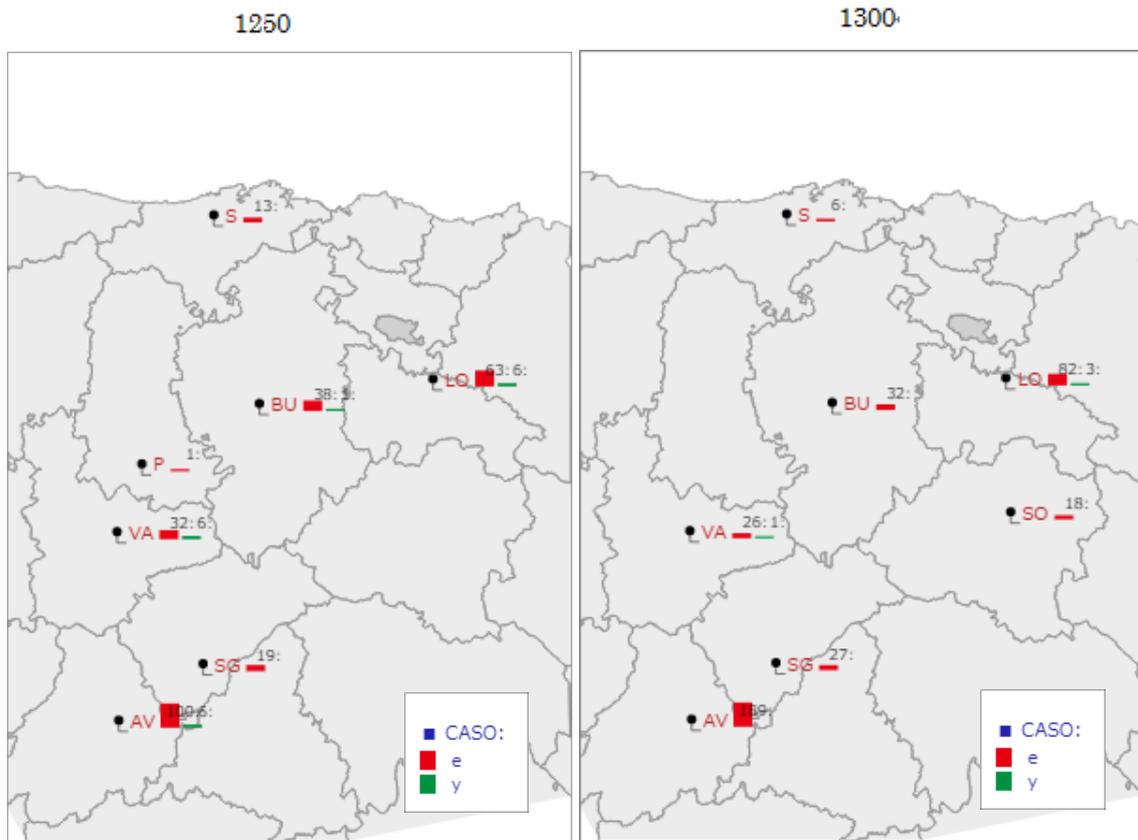


Fig. 3. <e> / <y> en 1250

Fig. 4. <e> / <y> en 1300

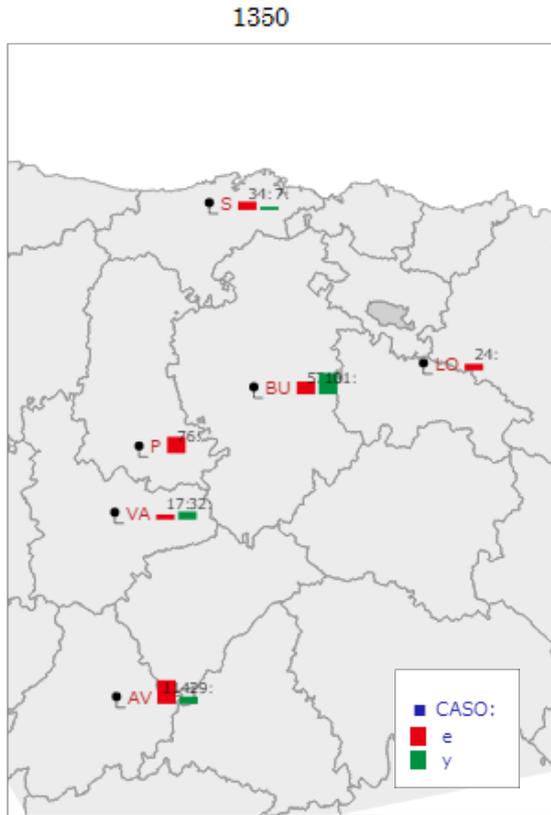


Fig. 5. <e> / <y> en 1350

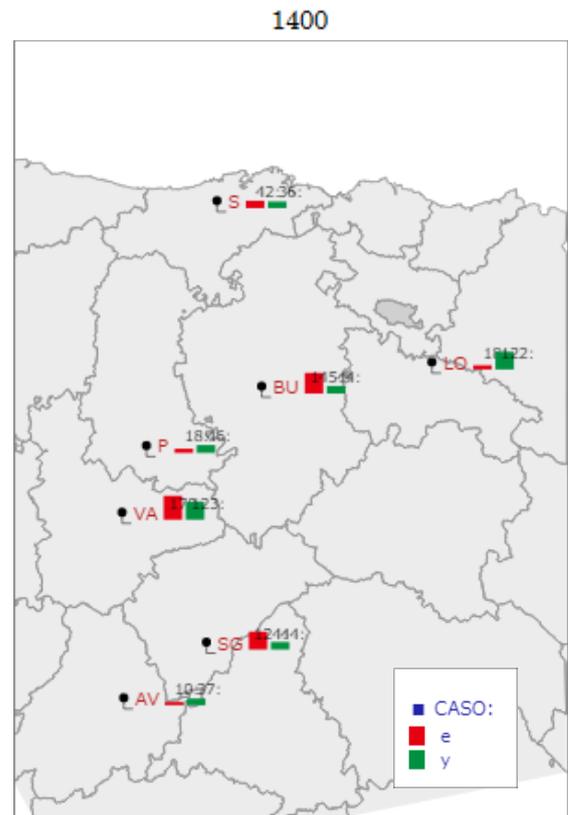


Fig. 6. <e> / <y> en 1400

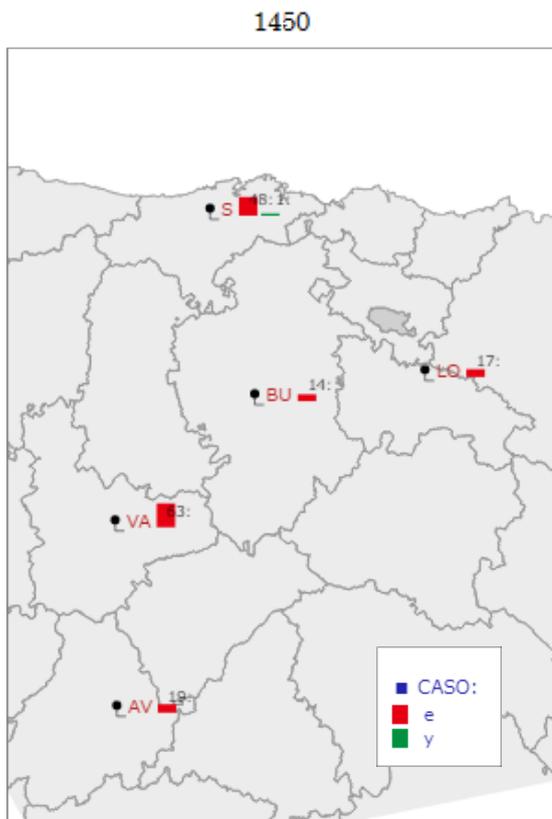


Fig. 7. <e> / <y> en 1450

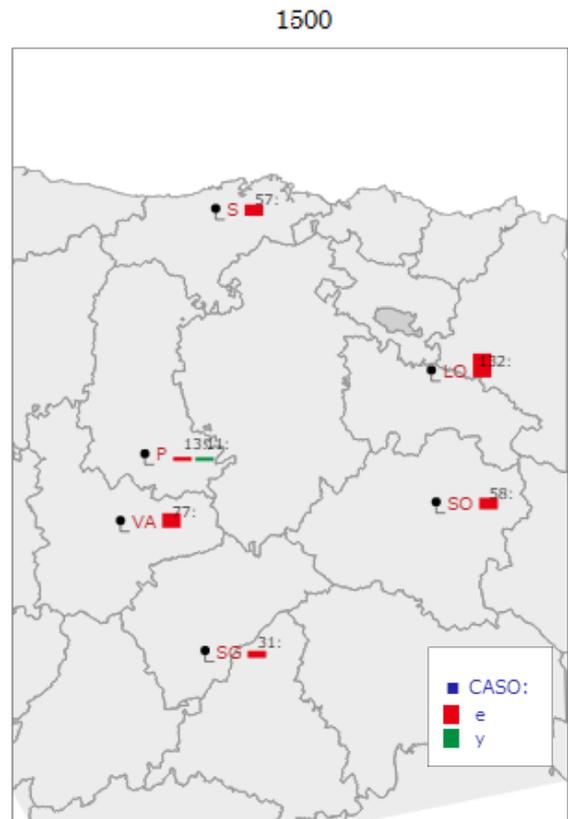


Fig. 8. <e> / <y> en 1500

Como resumen de estas observaciones, a continuación presentamos un mapa donde podemos observar la diacronía de cada sitio dentro de la geografía relevante:

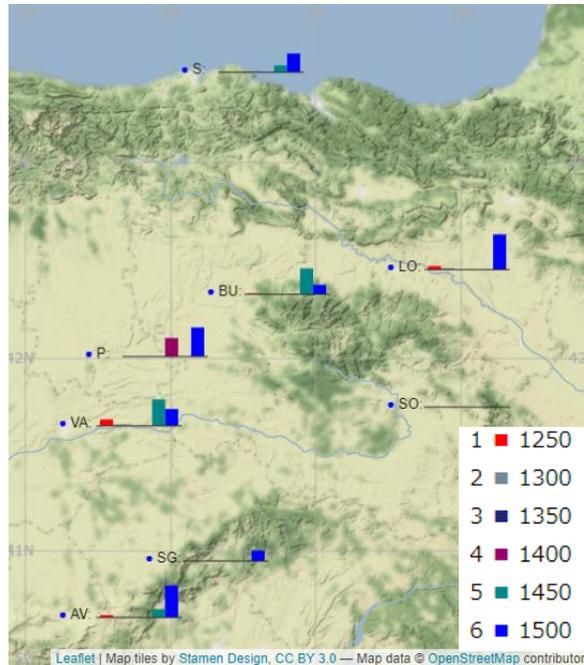


Fig. 9. Cambios diacrónicos y variaciones diatópicas de <y> en su ratio respecto a <e>¹⁰

De esta manera, podemos reconocer los cambios diacrónicos y variaciones diatópicas simultáneamente, ahora con los accidentes geográficos, además de las distancias físicas entre los lugares. El rápido crecimiento de la nueva forma <y> empieza en Palencia (P.1400) con su repercusión posterior en las dos ciudades vecinas, Valladolid (VA.1450) y Burgos (B.1450), a las que sigue el predominio en Logroño (LO.1500). Sandander (S.1500) se mantiene distante (Fig. 8, 9), probablemente debido a la Cordillera Cantábrica que constituye una barrera de comunicación, lo cual, sin embargo, de momento en esta etapa inicial de investigación, no podemos afirmar con certeza. Soria no presenta datos concernientes por falta de documentos no cancillerescos¹¹. Segovia (SG.1500) se muestra un tanto reacia, posiblemente por la distancia que hay con respecto a las localidades norteñas. Ávila (AV.1500), en cambio, adopta la nueva forma más fácilmente. La causa podemos buscarla, supuestamente, en su cercanía con Madrid. Estas interpretaciones geográficas son provisionales, que deben ser examinadas junto con otros factores lingüísticos y extralingüísticos. El más importante puede ser la diferencia del tipo de letras, góticas y humanísticas, que hemos mencionado anteriormente (Ueda 2018).

4. Conclusión

Desde el punto de vista lingüístico, todos los fenómenos de cambio histórico y variación geográfica son merecedores de ser observados. No se hace distinción sobre la importancia entre los rasgos frecuentes y los escasos. Sin embargo, la frecuencia debería ser tomada en consideración, puesto que es difícil o casi imposible sacar alguna conclusión con evidencias reducidas. Para comparar la magnitud cuantitativa en las distintas parcelas cronogeográficas, una de las condiciones de su viabilidad está en su elevada frecuencia, puesto que no se puede ni debe comparar los exiguos casos, ya que no son estadísticamente significativos.

Hemos visto las frecuencias sumamente altas de la conjunción copulativa ‘y’. Gracias al arduo trabajo de lematización (Sánchez-Prieto Borja / Ueda 2018), estamos en condición de llevar a cabo unas observaciones más precisas en la selección de casos y más fiables en el recuento de frecuencias de las formas en cuestión. Nuestro sistema de análisis lingüístico facilita la elaboración de tablas, gráficos y mapas, que son útiles para observar los fenómenos lingüísticos visualizados a lo largo del tiempo y a lo ancho del espacio (Ueda / Moreno Sandoval, 2017). Invitamos, por tanto, a los investigadores interesados a que consulten el corpus CODEA dentro de la herramienta LYNEAL¹². Nuestro trabajo futuro es ampliar la lematización a otras regiones de la Península para realizar estudios comparativos basados en evidencias documentadas, masivas y exhaustivas.

Referencias bibliográficas

- Alvar, Manuel; Pottier, Bernard. 1983. *Morfología histórica del español*. Madrid: Gredos.
- Corominas, Joan; Pascual, José Antonio. 1981. *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos
- De Andrés Díaz, Ramón. 2013. *Gramática comparada de las lenguas ibéricas*. Gijón: Ediciones Trea.
- Espinosa Elorza, Rosa María. 2008. Los conflictos de homónimos en el ámbito gramatical del español. Descripción de algunos casos y resoluciones en las épocas medieval y clásica. En C. Company y J. G. Moreno de Alba, eds. *Actas del VII Congreso Internacional de Historia de la Lengua Española*. Madrid: Fundación Banco de Santander, Arco/Libros y Asociación de Historia de la Lengua Española, pp. 1243-1261.

Espinosa Elorza, Rosa María. 2010. *Procesos de formación y cambio en las llamadas «palabras gramaticales»*. Fundación San Millán de Cogolla.

Menéndez Pidal, Ramón. 1976. *Cantar de Mio Cid. Texto, gramática y vocabulario. Primera parte. Crítica del texto - gramática*. 5a ed. Madrid: Espasa Calpe.

Meyer-Lübke, W. 1972. *Romanisches Etymologisches Wörterbuch*. Heidelberg: Carl Winter Universitätsverlag.

Sánchez-Prieto Borja, Pedro; Ueda, Hiroto. 2018. *Inventario léxico del corpus CODEA. I. Castilla la Vieja*. <http://shimoda.llf.uam.es/ueda/lyneal/ilc-cv.htm>

Ueda, Hiroto; Moreno Sandoval, Antonio. 2017. *LYNEAL: Letras y Números en Análisis Lingüísticos*. <http://shimoda.llf.uam.es/ueda/lyneal/>

Ueda, Hiroto. 2018. Formación histórica de la conjunción española ‘y’. Adaptación al modelo de los diptongos de /i/ + vocal y su latencia en siglos medievales. Comunicación oral presentada en el Círculo de Estudios Lingüísticos Hispánicos de Tokio, Universidad de Estudios Extranjeros de Tokio, 10 de noviembre de 2018. <https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/rekisi/e-y/e-y.pdf>

Notas

¹ <http://www.textoshispanicos.es>

² <http://www.corpuscodea.es>

³ <http://aldicam.blogspot.com/>

⁴ <http://www.corpuscharta.es> y <http://www.redcharta.es>

⁵ <http://shimoda.llf.uam.es/ueda/lyneal/codea.htm>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/codea.htm>.

⁶ Dentro de <y> incluimos las variantes minoritarias *i* y *j*, que son de muy reducida frecuencia.

⁷ Meyer-Lübke (1972, s. v. et) enumera: a.rum. *e*, it. *e(d)*, log., engad. *e*, frz. *et*, prov. *e(z)*, kat. *i*, sp. *y*, pg. *e*.

⁸ Sobre el tema histórico de ‘y’, remitimos la referencia general a nuestro trabajo anterior (Ueda 2018). Sin embargo, creemos necesario referirse a Espinosa Elorza (2008, 2010), que omitimos en la primera versión de este trabajo y generosamente nos ha indicado uno de los revisores anónimos. La autora (2010: 264-265) menciona precisamente el uso de <i> e <y> en obras literarias medievales, además de su relación con la forma de conjunción negativa ‘ni’ y el adverbio ‘y’ (‘allí’): “Aunque se representaba mayoritariamente como *e(t)*, también la encontramos *i*, *y*, con una distribución bastante regular. En la *Estoria de España*, de Alfonso X, se prefiere *y* cuando sigue palabra que comienza por *e*- (...) Un siglo antes, el *Auto de los Reyes Magos*, primer texto teatral, únicamente presenta esta variante (...)”. Su conclusión es que “la pronunciación de la conjunción copulativa medieval era [i]”, con la cual estamos de acuerdo. Su observación corrobora la nuestra, aunque se trata de otro tipo de evidencias de la misma Edad Media española.

⁹ El año representa el inicio de cada franja cronológica. Por ejemplo, 1250 comprende los años desde 1250 hasta 1299.

¹⁰ Véase la Tabla 4.

¹¹ Si en el futuro aparecen datos no cancllerescos sorianos y se siguiera manteniendo distante con respecto a las otras poblaciones, una de las posibles causas, de momento pura conjetura, podría ser la comunicación difícil debido a la altitud que hay en el noroeste

¹² Los datos de CODEA en textos y Castilla la Vieja en lemas el sistema de análisis están disponibles en los siguientes sitios:

<http://shimoda.llf.uam.es/ueda/lyneal/>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/index.html>