

# Datación de documentos medievales castellanos según la «coocurrencia» de parámetros<sup>1)</sup>

Yoshifumi KAWASAKI

## 1. Introducción

El presente trabajo tiene como objetivo presentar un método de datación de documentos no fechados incluidos en el *Corpus de Documentos Españoles Anteriores a 1700 (CODEA)* del Grupo de Investigación de Textos para la Historia del Español (GITHE) de la Universidad de Alcalá, dirigido por Pedro Sánchez-Prieto Borja, y en la *Documentación de cancillería castellana del siglo XIII* del Grupo de Estudio de Documentos Históricos y Textos Antiguos de la Universidad de Salamanca (GEDHYTAS), dirigido por María Nieves Sánchez González de Herrero.<sup>2)</sup> Dichos corpus se componen, en conjunto, de más de dos mil documentos notariales en su versión paleográfica, datados entre los siglos XIII y XVII, la mayoría de los cuales van tanto con *data chronica* relativamente fiable, a diferencia de las fuentes literarias, como con *data topica*.

Existen dos estudios anteriores de Ueda y Kawasaki (2012a y 2012b) que, inspirados en el trabajo de Azofra (2009: 201-204), intentaron realizar la datación mediante el análisis multivariacional denominado «concentración». La presente contribución, por otra parte, pretende hacerlo mejorando el método propuesto en Kawasaki (en prensa), según la probabilidad de la «coocurrencia» de parámetros lingüísticos que presenta el texto en cuestión.

A fin de que los datos utilizados presenten mayor homogeneidad e intervengan lo menos posible rasgos dialectales, nos limitamos a tratar de documentos procedentes únicamente de las históricas Castilla la Vieja y Castilla la Nueva, reconociendo, claro está, la presencia de ciertas diferencias dentro de dichas regiones.

## 2. Parámetros

En primer lugar, hemos de explicar en qué consisten los parámetros lingüísticos aplicados en esta contribución. Son aquellos fenómenos gráfico-fonéticos y/o morfo-fonéticos abordados en la gramática histórica que, presentando una variación diacrónica más o menos nítida, servirán para distinguir una época de otra caracterizada por un haz de rasgos lingüísticos.

Dichos rasgos se establecen con base en las variantes cronológicas de una forma moderna. Por ejemplo, la grafía procedente de F- inicial del infinitivo latino FACERE, cuyo reflejo actual es *hacer*, se representaba bajo una de las cuatro formas, es decir, *f-*, *ff-*, *h-*, o *ø*, según la época de aparición. Como la aparición de cada forma se ve restringida en épocas determinadas, sin que faltase convivencia de algunas de las variantes (Figura 1), servirán de parámetros discriminadores de la cronología. Así, denominamos «parámetro» a cada una de las variantes (*fazer*, *ffazer*, *hazer* y *azer*) pertenecientes a una «agrupación», que en este caso es *hacer*.<sup>3)</sup> Hay que hacer notar que lo que nos concierne no es el valor fonético representado por *f-*, *ff-*, *h-*, o *ø*, sino la variación gráfica entre dichas formas.

A continuación mostramos el cuadro de frecuencia absoluta (Figura 1) y porcentual (Figura 2) de las cuatro variantes medievales del infinitivo *hacer* que hemos hallado en los corpus. La división cronológica, que abarca intervalos de veinte años (1221-1240, 1241-1260, ..., 1661-1680, 1681-1700), va indicada por el último año del periodo.

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700
<i>fazer</i>	11	127	43	103	15	0	0	15	28	81	12	28	25	15	27	4	1	0	0	0	0	0	0	0
<i>ffazer</i>	0	12	11	99	34	15	6	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>hazer</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	7	20	51	35	27	15	14	18	3	18	5
<i>azer</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1	5	1	0	4

Figura 1 Frecuencia absoluta de *fazer*, *ffazer*, *hazer* y *azer*

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700
<i>fazer</i>	100%	91%	80%	51%	31%	0%	0%	75%	97%	99%	100%	100%	100%	68%	57%	7%	3%	0%	0%	0%	0%	0%	0%	0%
<i>ffazer</i>	0%	9%	20%	49%	69%	100%	100%	25%	3%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
<i>hazer</i>	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	32%	43%	93%	95%	96%	88%	93%	78%	75%	100%	56%
<i>azer</i>	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	4%	12%	7%	22%	25%	0%	44%

Figura 2 Frecuencia relativa de *fazer*, *ffazer*, *hazer* y *azer*

Vemos claramente el paso gradual de *fazer* > *ffazer* > *fazer* > *hazer* con el transcurso del tiempo. Lo ideal es que registremos al menos una variante en cada periodo, tal como aparece arriba. Sin embargo, no siempre es posible, debido a la limitación material de la documentación, encontrar variantes en cada periodo, dando como resultado huecos en más de una celda de la figura correspondiente. Es el caso de la agrupación *tengáis*, donde nos vemos en la necesidad de introducir algunas modificaciones en la medida en que sea permisible (Figura 3 y Figura 4).

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700
<i>tengades</i>	0	0	1	1	5	0	0	0	0	3	1	1	0	1	0	0	0	0	0	0	0	0	0	0
<i>tengais</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	2	0	7	0

Figura 3 Frecuencia absoluta de *tengades* y *tengais*

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700
<i>tengades</i>	0%	0%	100%	100%	100%	0%	0%	0%	0%	100%	100%	100%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
<i>tengais</i>	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%	100%	0%	100%	0%

**Figura 4 Frecuencia relativa de *tengades* y *tengais***

Dada esta situación, intentamos, recurriendo a la gramática histórica, llenar los huecos producidos a raíz de la escasez de documentación. No será arriesgado suponer que hasta finales del siglo XV la variante predominante sería *tengades* en vez de *tengais*, de manera que suplimos los huecos correspondientes con el cien por cien. Del hecho de que el último registro de *tengades* date del periodo 1481-1500 y que la primera aparición de *tengais* sea del 1541-1560, juzgamos que el periodo 1501-1540 fue la época de convivencia entre ambas variantes. Por tanto, nos permitimos suplir provisionalmente el cincuenta por ciento en cada una de las cuatro celdas relevantes, suponiendo que la situación fue la de puro empate. Finalmente, llenamos con el cien por cien los huecos posteriores a 1581, época en la que la forma moderna ya había reemplazado, con toda probabilidad, a la medieval. Así, obtenemos la siguiente tabla modificada (Figura 5).

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700
<i>tengades</i>	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	50%	50%	0%	0%	0%	0%	0%	0%	0%	0%
<i>tengais</i>	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	50%	100%	100%	100%	100%	100%	100%	100%	100%

**Figura 5 Frecuencia relativa modificada de *tengades* y *tengais***

Ahora bien, una de las dificultades a la que se enfrentan los estudios históricos consiste en cómo interpretar un *hápax*, que, o bien puede ser forma corriente, o bien forma anormal. El fenómeno, por sí solo, no proporciona ninguna solución, pero sí lo hace el entorno o contexto en que se sitúa el único registro. He aquí las dos tablas correspondientes a la agrupación *juzgar* (< JUDICĀRE), bajo cuya etiqueta incluimos, aparte del infinitivo, sus formas conjugadas, así como el gerundio y el participio pasado (Figura 6 y Figura 7).

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700
<i>judgar</i>	0	41	13	49	1	0	0	1	5	1	0	3	1	1	7	0	3	0	0	0	0	0	0	0
<i>juzgar</i>	0	0	1	0	0	0	1	0	0	0	0	0	2	6	5	5	5	5	0	0	3	6	1	4

**Figura 6 Frecuencia absoluta de *judgar* y *juzgar***

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700	
<i>judgar</i>	0%	100%	93%	100%	100%	0%	0%	100%	100%	100%	0%	100%	33%	14%	58%	0%	38%	0%	0%	0%	0%	0%	0%	0%	0%
<i>juzgar</i>	0%	0%	7%	0%	0%	0%	100%	0%	0%	0%	0%	0%	67%	86%	42%	100%	63%	100%	0%	0%	100%	100%	100%	100%	100%

**Figura 7 Frecuencia relativa de *judgar* y *juzgar***

Si bien registramos un caso de *juzgar* en el periodo 1341-1360 en lugar de *judgar*, el conocimiento histórico de la lengua nos indica que fue una forma extraordinariamente innovadora para dicha época. Por consiguiente, anulamos la aparición de *juzgar* en el periodo en cuestión y, en cambio, suplimos el cien por cien en la celda relevante de *judgar*. Aplicando luego el procedimiento mencionado en el apartado anterior, obtenemos la siguiente tabla modificada (Figura 8).

	1240	1260	1280	1300	1320	1340	1360	1380	1400	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	1620	1640	1660	1680	1700	
<i>judgar</i>	100%	100%	93%	100%	100%	100%	100%	100%	100%	100%	100%	100%	33%	14%	58%	0%	38%	0%	0%	0%	0%	0%	0%	0%	0%
<i>juzgar</i>	0%	0%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	67%	86%	42%	100%	63%	100%	100%	100%	100%	100%	100%	100%	100%

**Figura 8 Frecuencia relativa modificada de *judgar* y *juzgar***

Hasta ahora hemos contado con más de cien parámetros, de los que, por razones de espacio, nos hemos limitado a tratar algunos en el presente trabajo. El porcentaje que presenta cada parámetro en una época determinada significa que este aparece en dicha época con la probabilidad indicada respecto a otras variantes de la misma agrupación dentro del universo limitado de los corpus de los que son extraídos los parámetros.<sup>4)</sup>

### 3. «Coocurrencia» de parámetros

En este apartado explicamos el procedimiento para realizar la datación. Nuestra suposición es que, cuando un documento sin fecha contiene tres parámetros  $\alpha$ ,  $\beta$  y  $\gamma$ , este procederá del periodo en que con mayor probabilidad aparecen los tres al mismo tiempo. Esta probabilidad, que denominamos «coocurrencia», se calcula multiplicando el valor porcentual de todos los parámetros que se hallan en el documento en cuestión, y se expresa en la siguiente fórmula:

$$\text{«Coocurrencia»} = x_1 * x_2 * \dots * x_n / 100^n$$

(donde  $n$  es el número de parámetros relevantes y  $x_n$  el porcentaje que presenta el parámetro $_n$ )

Nuestro método no se fija sino en los parámetros presentes en el documento, por lo que varía de un documento a otro el número de parámetros que intervienen. No es necesario que se encuentren en dicho

documento todos los parámetros establecidos de antemano.

A modo de ejemplo, vamos a demostrar de manera simplificada cómo efectuar la datación. He aquí una tabla (Figura 9), donde  $\alpha$ ,  $\beta$  y  $\gamma$  son los parámetros relevantes que se hallan en el documento en cuestión, y I-IV los periodos candidatos. Como se ve en la tabla, el valor de «coocurrencia» resulta más alto en el periodo II (0.168). De ahí que conjeturemos que este documento, desde el punto de vista lingüístico, procede probablemente de dicho periodo, sin poder descartar la posibilidad de que fuera escrito en los demás periodos. El periodo I, por su valor cercano al del periodo II, puede ser buen candidato, mientras que es poco probable que proviniera de los periodos III o IV por su valor inferior o hasta cero con respecto a I y II.

	$\alpha$ (%)	$\beta$ (%)	$\gamma$ (%)	Coocurrencia.	Ranking
I	50%	40%	80%	0.160	2
II	40%	60%	70%	0.168	1
III	10%	95%	50%	0.048	3
IV	0%	100%	25%	0.000	4

**Figura 9 Cálculo de «coocurrencia»**

De este modo podemos descartar periodos que presentan el valor cero de «coocurrencia» (es decir, casos en los que hay al menos un parámetro que presenta valor nulo en la época dada), porque significa que, desde el punto de vista probabilístico, es imposible que aparezcan al mismo tiempo todos los parámetros en cuestión. Por otra parte, este método nos permite jerarquizar los periodos que presentan un valor que no sea de cero según su mayor o menor valor de «coocurrencia», indicando el primer candidato, el segundo, el tercero, y así sucesivamente. La ventaja que presenta este modelo probabilístico es la de poder señalar, con un reducido número de parámetros, o incluso hasta con uno solo, periodos candidatos, lo que resultaba imposible con el análisis de concentración.

#### **4. Estudio de caso: documentos con fecha**

A fin de averiguar la aplicabilidad del método propuesto, ofrecemos aquí resultados de dos pruebas que hemos realizado con documentos fechados, tratándolos como si carecieran de *data chronica* y, por tanto, simular que la desconocemos, para comprobar si se obtienen predicciones satisfactorias. Son aquellos documentos que, antes de establecer parámetros, se han excluido de manera aleatoria para asegurar que estos no estén incluidos en la muestra y que el resultado sea independiente y persuasivo.

#### 4.1. CODEA 1424 (año 1521, Madrid)

Es un documento procedente de Madrid y escrito en 1521, en el cual encontramos los siguientes parámetros: *hasta* en vez de *fasta* o *ffasta*, *guadalajara* en vez de *guadalfajara* o *guadalhajara*, *auia* en vez de *auie*, *hazer* en vez de *fazer* o *ffazer*, *-adgo* en vez de *-azgo*, *segun* en vez de *segund* o *segunt*, *ciudad* con *-d* final en vez de *ciudat* con *-t*, *cibdad* con *-b-* interior en vez de *ciudad* con *-u-*. La aparición simultánea de estos parámetros predice como primer candidato el periodo 1521-1540, que coincide con la fecha verdadera (1521) de la composición del documento (Figura 10).

	hasta	guadalajara	auia	hazer	-adgo	segun	ciudad	cibdad	Coocurrencia	Ranking
1480	0%	54%	100%	0%	79%	6%	98%	98%	0.000	4
1500	9%	86%	100%	32%	90%	6%	99%	99%	0.001	3
1520	38%	100%	100%	43%	29%	19%	99%	88%	0.008	2
1540	74%	100%	100%	93%	50%	39%	100%	83%	0.109	1
1560	94%	100%	100%	95%	0%	53%	100%	33%	0.000	4

Figura 10 CODEA 1424 (año 1521)

#### 4.2. CODEA 147 (año 1613, Guadalajara)

Es un documento procedente de Guadalajara y compuesto en 1613, en el cual encontramos los siguientes parámetros: *conozco* en vez de *conosco*, *viere* con *v-* inicial en vez de *uiere* con *u-*, *juzgar* en vez de *judgar*, *ove* en vez de *uve*, *doy* en vez de *do*, *selo* en vez de *gelo*, *castilla* en vez de *castiella*, *para* en vez de *pora*, *ciudad* con *-d* final en vez de *ciudat* con *-t*, *cibdad* con *-b-* interior en vez de *ciudad* con *-u-*, *deuda* en vez de *debda*. La aparición simultánea de estos parámetros predice como primer candidato el periodo 1601-1620, que coincide con la fecha real (1613) de la composición del texto (Figura 11).

	conozco	viere	juzgar	ove	doy	selo	castilla	para	ciudad	cibdad	deuda	Coocurrencias	Ranking
1560	100%	100%	63%	83%	100%	83%	100%	100%	100%	67%	0%	0.000	6
1580	92%	100%	100%	57%	100%	100%	100%	99%	100%	87%	100%	0.447	2
1600	100%	100%	100%	31%	100%	100%	100%	98%	100%	100%	100%	0.306	3
1620	100%	100%	100%	75%	100%	100%	100%	100%	100%	100%	100%	0.750	1
1640	100%	100%	100%	7%	100%	100%	100%	100%	100%	98%	100%	0.070	5
1660	80%	100%	100%	33%	100%	100%	100%	100%	100%	100%	100%	0.267	4
1680	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	0.000	6

Figura 11 CODEA 147 (año 1613)

## 5. Estudio de caso: documentos sin fecha

A continuación, intentaremos datar un par de documentos no fechados. Hay algunos que, pese a la falta de *data chronica*, nos dan alguna pista para averiguar la época aproximada en que fueron escritos a través del hecho histórico mencionado en el documento, por ejemplo, sobre el rey, la reina o personajes destacados, etc., de los que se poseen conocimientos bastante seguros. Creemos que la coincidencia de la fecha indicada por la datación con el reinado al que se hace referencia en el documento asegurará todavía más la funcionalidad de nuestro método. Si bien, en principio, intentamos realizar la datación desde el punto de vista puramente lingüístico, estamos dispuestos, claro está, a consultar toda la información a nuestro alcance -paleografía, historia en general, archivología, entre otros-, porque el lenguaje empleado en el documento no es la única pista de la que disponemos.

### 5.1. CODEA 1439 (sin fecha ni lugar)

Se trata de un documento redactado durante el reinado de Enrique IV (1454-1474), en el cual hallamos los siguientes parámetros: *despues* en vez de *depues*, *regno* en vez de *reino*, *ove* en vez de *uve*, *fazer* en vez de *hazer* o *ffazer*, *fijo* en vez de *ffijo* o *hijo*, *mente* en vez de *miente* o *mientre*, *algund* en vez de *algunt* o *algun*, *segund* en vez de *segunt* o *segun*, *debda* en vez de *deuda*.

El resultado señala que dicho documento fue compuesto con mayor probabilidad entre 1461-1480, periodo este que coincide con el reinado de Enrique IV (1454-1474) (Figura 12).

	despues	regno	ove	fazer	fijo	mente	algund	segund	debda	Coocurrencia	Ranking
1380	100%	100%	100%	75%	50%	40%	33%	60%	100%	0.030	7
1400	78%	100%	100%	97%	82%	73%	20%	47%	100%	0.042	6
1420	100%	98%	100%	99%	97%	100%	50%	34%	100%	0.162	4
1440	100%	100%	100%	100%	100%	100%	50%	45%	100%	0.227	3
1460	100%	95%	97%	100%	100%	100%	63%	66%	67%	0.250	2
1480	100%	48%	100%	100%	100%	100%	100%	83%	100%	0.404	1
1500	100%	21%	100%	68%	75%	98%	100%	91%	100%	0.095	5
1520	100%	6%	100%	57%	38%	100%	73%	79%	0%	0.000	8
1540	100%	0%	97%	7%	5%	100%	63%	61%	50%	0.000	8

Figura 12 CODEA 1439 (sin fecha)

### 5.2. CODEA 373 (sin fecha ni lugar)

En este documento compuesto durante el reinado de Carlos I (1516-1556), hallamos los siguientes parámetros: *despues* en vez de *depues*, *viere* en vez de *uiere*, *reino* en vez de *regno*, *fasta* en vez de *fata*, *fasta* en vez de *ffasta* o *hasta*, *tuve* en vez de *tove*, *ove* en vez de *uve*, *auia* en vez de *auie*, *hazer* en vez de

fazer o ffazer, gelo en vez de selo, castilla en vez de castiella, mente en vez de miente o mientre, para en vez de pora, algund en vez de algunt o algun, segund en vez de segunt o segun, verdad en vez de verdat, salud en vez de salut, valladolid en vez de valladolit, cibdad en vez de ciudad, recabdar en vez de recaudar.

	después	viere	reino	faSta	fasta	tuve	ove	aula	hazer	gelo	castilla	mente	para	algund	segund	verdad	salud	valladolid	cibdad	recabdar	Cocurrencia	Ranking
1480	100%	100%	52%	100%	100%	0%	100%	100%	0%	90%	100%	100%	100%	100%	83%	100%	100%	100%	98%	100%	0.00000	5
1500	100%	100%	79%	100%	91%	4%	100%	100%	32%	90%	87%	98%	100%	100%	91%	89%	100%	100%	99%	100%	0.00561	2
1520	100%	100%	94%	100%	63%	15%	100%	100%	43%	42%	100%	100%	100%	73%	79%	100%	100%	100%	88%	57%	0.00466	3
1540	100%	100%	100%	100%	26%	28%	97%	100%	93%	40%	100%	100%	100%	63%	61%	100%	100%	100%	83%	100%	0.00824	1
1560	100%	100%	100%	100%	6%	37%	83%	100%	95%	17%	100%	100%	100%	27%	47%	100%	100%	100%	33%	50%	0.00006	4

Figura 13 CODEA 373 (sin fecha)

El análisis indica como primer candidato el periodo 1521-1540 que se incluye, de hecho, dentro del reinado de Carlos I (1516-1556), resultado que tenemos por bastante aceptable (Figura 13).

## 6. Consideraciones finales

A modo de conclusión trataremos de algunos problemas que suelen plantearse en una indagación de esta índole, y asimismo daremos noticia de nuestros futuros planes. Como señala Barra-Jovar (2001: 184-185), «una conclusión obtenida a partir de un conjunto de n textos no garantiza en ningún caso las predicciones que pueden hacerse sobre lo que ocurre en un n+1 texto». Pese a la gran cantidad de documentación que utilizamos, no deja de desaparecer esta aporía, de modo que nuestro método predice a veces una fecha desacertada, al hallarse en el documento rasgos lingüísticos demasiado conservadores o innovadores. Aspiramos a contrarrestar esta dificultad, aumentando tanto el número de documentos como el de parámetros.

Por otra parte, hemos de desarrollar una manera apropiada para suplir la carencia de datos, a la cual suelen enfrentarse los estudios históricos. Si bien es cierto que los investigadores diacrónicos debemos mostrar el mayor respeto posible a los datos existentes, creemos justificable, a menos que esta intervención perjudique gravemente la distribución original de los mismos, introducir ciertas modificaciones fundamentadas.

En nuestras futuras investigaciones, tenemos previsto incorporar todos los documentos incluidos en los corpus, hasta aquellos procedentes de las regiones dialectales asturleonese y navarro-aragonesa que han sido excluidos en el presente trabajo, para poder así tratar de la documentación de toda la Península. Al mismo tiempo, estamos barajando crear un modelo para localizar documentos, extrayendo del corpus

parámetros diatópicos, de igual manera que lo hemos hecho con los diacrónicos. Nuestra meta final es, tal como indican Sánchez-Prieto *et al* (2009: 36), establecer parámetros espacio-temporales y desarrollar con ellos una herramienta tanto para «datación» como para «localización» de documentos medievales españoles.

## Notas

<sup>1)</sup> Este artículo es una versión corregida y ampliada de la comunicación leída en el 50.º Congreso de la SOCIETAS JAPONICA STUDIORUM ROMANICORUM celebrado los días 19 y 20 de mayo de 2012 en la Universidad Sofía, de Tokio. Para la revisión del artículo en español se ha contado con la ayuda del profesor Carlos Rubio López de la Llave de la Universidad Complutense de Madrid y de la profesora Ana Isabel García de la Universidad de Tokio. Este trabajo ha sido realizado gracias a la financiación para las investigaciones de doctorado otorgada por la Universidad de Tokio.

<sup>2)</sup> Quiero expresar mi gratitud al profesor Pedro Sánchez-Prieto Borja de la Universidad de Alcalá y a la profesora Nieves Sánchez González de Herrero de la Universidad de Salamanca por haberme permitido acceder a la transcripción paleográfica de los corpus.

<sup>3)</sup> En esta agrupación no incluimos formas conjugadas (*hago, hacía, hice*, etc.), suponiendo que podrá existir cierta diferencia con respecto a la época en que se generalizó una grafía inicial en detrimento de otra, según la vocal que sigue. En caso afirmativo, obtendremos parámetros de otra agrupación que contribuirán a diferenciar un periodo de otro. Con el objetivo de aprovechar lo máximo posible la discrepancia cronológica entre las formas relacionadas, decidimos, en este caso, establecer parámetros a nivel de «tipos» en vez de a nivel de «lemas». Además, si resulta que no hay diferencia temporal, no nos encontramos con ningún problema.

<sup>4)</sup> Por tanto, la suma del valor porcentual en cada periodo de los parámetros de la misma agrupación es de cien.

## Referencias

- Azofra, María Elena (2009): *Morfosintaxis histórica del español: De la teoría a la práctica*, UNED, Madrid.
- Barra-Jovar, M. (2001): «Corpus diacrónico, constatación e inducción», En D. Jacob & J. Kabatek (eds.), *Lengua medieval y tradiciones discursivas en la Península Ibérica: Descripción gramatical -pragmática histórica- metodología*. Madrid/Frankfurt: Iberoamericana/Vervuert, págs. 177-197.
- Kawasaki, Y. (en prensa): «La determinación cronológica de cambios gráfico-fonéticos y la datación de

documentos no fechados en el CODEA», *Actas del II Congreso Internacional Tradición e innovación: Nuevas perspectivas para la edición y el estudio de documentos antiguos* celebrado en Neuchâtel, Suiza (septiembre de 2011).

Sánchez-Prieto Borja, P., Paredes García, F., Martínez Sánchez, R., Miguel Franco, R., Simón Parra, M. y Vicente Miguel, I. (2009): «El *Corpus de Documentos Españoles Anteriores a 1700 (CODEA)*». En A. Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas: Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt: Iberoamericana/Vervuert, págs. 25-38.

Ueda, H. & Kawasaki, Y. (2012a): «Datación estadística de los textos medievales españoles sin fecha», comunicación oral en el *Encuentro Científico Internacional, Periodización del CHARTA (Corpus Hispánico y Americano en la Red: Textos Antiguos) y Criterios Específicos de Edición*, Madrid, Consejo Superior de Investigaciones Científicas.

Ueda, H. & Kawasaki, Y. (2012b): «Datación de los documentos en el CODEA: Un estudio de filología de corpus», Comunicación leída en el *IX Congreso Internacional de Historia de la Lengua Española*, Cádiz, España (septiembre de 2012).

#### **Enlaces relacionados**

<http://demos.bitext.com/codea>, *Corpus de Documentos Españoles Anteriores a 1700 (CODEA)* del Grupo de Investigación de Textos para la Historia del Español (GITHE) de la Universidad de Alcalá.

<http://campus.usal.es/~gedhytas/>, *Documentación de cancillería castellana del siglo XIII* del Grupo de Estudio de Documentos Históricos y Textos Antiguos de la Universidad de Salamanca (GEDHYTAS).