

# Un acercamiento cuantitativo a la periodización en la historia del español\*

ANTON GRANVIK

*Universidad de Gotemburgo*

CARLOS SÁNCHEZ LANCIS

*Universidad Autónoma de Barcelona*

*Resumen.* El objetivo de este trabajo es presentar una nueva solución a un viejo problema de la historia del español, a saber, su periodización. Con este fin, hemos aplicado un análisis cuantitativo denominado agrupamiento jerárquico (*cf.* Gries 2009) sobre una muestra extraída de un extenso corpus diacrónico del español, el *CORDE* de la RAE. La muestra incluye 27 fenómenos gramaticales (morfológicos y sintácticos), pues consideramos que son los datos gramaticales los que deben servir de base para la periodización de la lengua. El agrupamiento jerárquico es un cálculo matemático basado en la semejanza numérica entre diferentes categorías, agrupando juntas las categorías más semejantes entre sí y distanciando las categorías más diferentes. Por tanto los periodos de tiempo que presentan fenómenos lingüísticos semejantes se agruparán juntos, constituyendo una etapa diferenciada. Los resultados del análisis tanto corroboran observaciones anteriores sobre la periodización del español, como ofrecen nuevos detalles con respecto al español clásico y moderno.

*Palabras clave.* Periodización, agrupamiento jerárquico, historia de la lengua española, gramática histórica.

*Abstract.* The aim of this paper is to present a fresh solution to an old problem of the history of Spanish, namely its periodization. With this in mind we have applied a quantitative analysis called hierarchical clustering (*cf.* Gries 2009) on

---

\* La presente investigación ha sido parcialmente financiada con una ayuda del MICINN y FEDER (núm. de ref. FFI2014-56968-C4-2-P y FFI2017-87140-C4-1-P) y de la CIRIT del Comissionat per Universitats i Recerca de la Generalitat de Catalunya (núm. de ref. 2017 SGR 634).

an extensive sample of corpus data retrieved from the Spanish Royal Academy's *CORDE* corpus. The sample consists of 27 morphological and syntactic phenomena, since we believe that it is the linguistic data that must form the basis for establishing a periodization of the language. Hierarchical clustering is a technique based on the numerical similarity of different categories, grouping together the most similar categories and separating the most divergent ones. Time periods including similar grammatical phenomena are thus grouped together, forming a distinct period. The results of our analysis confirm many previous observations on the periodization of Spanish, but it also offers new detailed information on both classical and modern Spanish.

*Keywords.* Periodization, hierarchical cluster analysis, history of the Spanish language, diachronic grammar.

## 1. INTRODUCCIÓN

El objetivo de este estudio es presentar una nueva solución a un viejo problema de la historia del español, es decir, su periodización. Por periodización entendemos el establecimiento de lapsos de tiempo, de épocas, en los que se puede dividir el eje temporal de una lengua desde un punto de vista histórico. Esta labor que *a priori* parece no revestir una gran dificultad, ya que las diferentes historias de la lengua española han presentado desde siempre sin grandes problemas una serie de divisiones bastante consensuadas, ha sido muy cuestionada en las últimas décadas en el momento en que se ha visto la necesidad de incorporar criterios internos exclusivamente lingüísticos a los ya consabidos criterios externos de tipo histórico, literario, etc. (cf. Sánchez Lancis 2012). De esta manera, como se verá en el siguiente apartado, han ido apareciendo diferentes estudios con el objetivo de cuestionar, y a su vez replantear, los criterios que se habían ido utilizando para establecer fronteras lingüísticas temporales, al mismo tiempo que se buscaban nuevos elementos internos, más propios de la gramática histórica, que corroboraran los periodos existentes y permitieran, si fuera necesario, establecer otros nuevos.

Sin embargo, para poder avanzar en el tema de la periodización del español creemos que es necesario introducir nuevos elementos que nos permitan establecer y definir las etapas históricas de la forma más precisa posible. Somos conscientes de que todo periodo no deja de ser más que una aproximación hipotética a una realidad histórica de la que solo conservamos información a través de un material diverso recogido en forma escrita, con todo lo que esto último supone. Por todo ello, consideramos que la aplicación de un análisis cuantitativo denominado agrupamiento

jerárquico (*cf.* Gries 2009), explicado pormenorizadamente en el apartado 3, a un corpus diacrónico del español, como es el *Corpus diacrónico del español (CORDE)* de la Real Academia Española, nos puede permitir obtener una visión mucho más aproximada y clarificadora de los periodos más característicos de la historia de la lengua española (*cf.* Sánchez Lancis 2009). El espacio temporal estudiado, de una extensión suficientemente representativa, abarca desde el año 1000 hasta 1979, dividido en lapsos de décadas. Y este estudio se aplica en el presente trabajo a un total de 27 fenómenos gramaticales, tanto morfológicos como sintácticos, claros ejemplos del cambio gramatical experimentado por el español a lo largo de su historia. Se trata de un punto de partida para un estudio de una mayor envergadura que se pretende realizar en los próximos años.

La idea básica que así defendemos es que son los datos gramaticales los que deben servir de base para la periodización de la lengua y no fenómenos externos a ella, como puedan ser las corrientes intelectuales o los grandes acontecimientos históricos. La novedad de este estudio reside, por consiguiente, más que nada en la metodología, es decir, en el análisis cuantitativo de un gran número de casos, extraídos del mayor banco de datos de las distintas fases de la historia del español de que disponemos en la actualidad, el *CORDE*, analizados con una subdivisión muy detallada, la década.

## 2. LA PERIODIZACIÓN

### 2.1. *Estudios anteriores*

Como ya se ha señalado, son muchos los trabajos que han abordado en los últimos años el tema de la periodización. Sin ánimo de ser exhaustivos, vale la pena recoger aquí algunos de los más destacados. En primer lugar, merece especial atención el cuestionamiento que Menéndez Pidal (1942) realiza sobre las divisiones de la historia de la lengua centradas en periodos de cien años, los siglos, por lo que propugna estudios basados en lapsos de tiempo mucho más restringidos, idea que vamos a aplicar al presente estudio, como se verá más adelante. En segundo lugar se sitúa la división temporal que realiza Lapesa (1981), quien establece cinco etapas históricas: época de los orígenes (siglos IX-XI), castellano medieval (siglos XII-XIV), español preclásico (1474-1525), español clásico (siglos XVI-XVII) y español moderno (siglo XVIII-español actual), subrayando sobre todo la existencia de una etapa de transición entre el castellano medieval y el español clásico. En tercer lugar destacan princi-

palmente los artículos de Eberenz (1991 y 2009), y de manera especial el primero de ellos por ser el precursor de una serie de trabajos posteriores de distintos autores en los que no solo se ha replanteado el concepto de periodización, sino que además se han ido buscando elementos exclusivamente lingüísticos (*parámetros esenciales*), que permitan establecer los diferentes periodos del español: *fase antigua de estabilidad* (1200-1450), *etapa media de transformación* (1450-1650) y *fase moderna de estabilidad* (1650-actualidad). Como complemento de los estudios anteriores y nacido uno de ellos casi de forma simultánea al primero, tenemos los trabajos de Marcos Marín (1992 y 1995), en los que se critica la utilización de criterios externos para la delimitación de etapas históricas y se propugna el empleo básicamente de criterios internos.

Posteriormente, otros estudios, como los de Cano Aguilar (1992), inciden de nuevo en la existencia de ese español preclásico, denominado en este caso *época del Descubrimiento*, en una línea similar a la propugnada también por Ridruejo Alonso (1993). Además, el establecimiento de las distintas etapas históricas basadas principalmente en criterios internos continúa en los trabajos de Martínez Alcalde/Quilis Merín (1996), Sánchez Lancis (1997-1998, 1999, 2001), Buenafuentes (2002a y 2002b), entre otros; así como también la mayor precisión de etapas ya existentes, como es el caso del español clásico por parte de Girón Alconchel (2005), y del español moderno por parte de Melis/Flores/Bogard (2003), Octavio de Toledo y Huerta (2007), Girón Alconchel (2008), Sánchez Lancis (2012) y Flores/Melis (2015).

Todas las reflexiones aportadas por los autores anteriores han sido asumidas en el presente estudio, cuya finalidad principal es cuestionar y establecer los distintos periodos de la lengua española a partir única y exclusivamente de criterios internos, es decir, de datos gramaticales, como se explicará a continuación.

## 2.2. *Los fenómenos incluidos*

Para poder comprobar la idoneidad de la metodología utilizada, basada en el agrupamiento jerárquico aplicado a la periodización, se ha seleccionado un total de 27 fenómenos gramaticales, tanto morfológicos como sintácticos, divididos en cuatro grupos, claramente característicos o definitorios de las diferentes etapas históricas del español. A continuación se recogen todos estos elementos tal y como han sido buscados en el corpus, con todas sus posibles variantes:

- 1) Pronombres y determinantes (8):
  - a) *nosotros* o *nos otros*;

- b) *os*;
  - c) *este* o (*aqueste*) o *est* o (*aquest*);
  - d) *las* o *elas*;
  - e) *su*;
  - f) *qui* o *qi* o *ki*;
  - g) *quienes* o *qienes* o *kienes* o *qujenes*;
  - h) *unas*.
- 2) Adverbios y preposiciones (8):
- a) *ende* o *end* o *ent*;
  - b) *allende* o *allend* o *alliende* o *alliend* o *alende* o *alend* o *aliende* o *aliend* o *allent* o *alient* o *alent* o *hallende*;
  - c) *suso* o *susso* o *desuso* o *desusso* o *dessusso* o *dessusso* o *asuso* o *asusso* o *assuso* o *assusso*;
  - d) *delant* o *delante* o *denant* o *denante*;
  - e) *ahora*;
  - f) *cras*;
  - g) *fuerte mente* o *fuerte ment* o *fuerte miente* o *fuerte mientre*;
  - h) *ante*.
- 3) Verbos (6):
- a) *pienso*;
  - b) *soy* o *soi*;
  - c) *sodes*;
  - d) *serian* o *serien* o *serían* o *seríen* o *seryan*;
  - e) *fuessedes* o *fossedes* o *fuesedes* o *fosedes* o *fuéssedes* o *fóssedes* o *fuésedes* o *fósedes*;
  - f) *fuere* o *fore* o *fuére* o *fóre*.
- 4) Construcciones sintácticas (5):
- a) *la su*;
  - b) *'lo no'* o *'lo non'*;
  - c) *\*r lo he* / *\*r lo hé* / *\*r lo e* / *\*r lo é*;
  - d) (*dixolo* o *díxolo* o *díjolo* o *dijolo*) + *lo dixo* o *lo dijo*;
  - e) *lo faria* o *lo faría* o *lo haría* o *lo haria*.

Por lo que respecta a los pronombres y los determinantes, se ha buscado la aparición de la forma sintética del pronombre personal de primera persona del plural *nosotros*, del clítico de segunda persona del plural *os*, el contraste entre las dos variantes del demostrativo de primer grado *este* / *aqueste*, la difusión del determinante *las*, la forma del posesivo de tercera persona *su*, el relativo de persona *qui*, la presencia del plural en el relativo *quienes* y la difusión del determinante / indefinido *unas*.

En relación a los adverbios y las preposiciones, como se puede comprobar, los elementos adverbiales son mayoritarios (7 casos) frente a las preposiciones (1 caso), debido a que los primeros presentan muchos menos casos de ambigüedad. Así, se ha analizado la presencia en el corpus de los adverbios espaciales siguientes: el pronominal locativo *ende*; el adverbio *allende*, a caballo entre los pronominales y los preposicionales; el direccional / prepositivo (*a*)*suso*; y el prepositivo *delante*. También se han estudiado los adverbios temporales *ahora* y *cras*; el adverbio analítico de modo *fuertemente*; y la preposición *ante*.

Respecto a los verbos, se ha estudiado la primera persona del singular del presente de indicativo del verbo *pensar* (*pienso*), por ser de uso muy frecuente a partir de un determinado momento en un tipo particular de construcciones, así como una serie de diferentes formas del verbo *ser*, por tratarse de un verbo de empleo general en el corpus: la primera persona del singular del presente de indicativo *soy*; la segunda persona del plural del presente de indicativo *sodes*; la forma sintética del condicional *serían*; la segunda persona del plural del imperfecto de subjuntivo *fuésedes*; y la forma singular del futuro imperfecto de subjuntivo *fuere*.

Finalmente, en relación a las construcciones sintácticas analizadas, se ha seleccionado la estructura [artículo + posesivo + sustantivo] representada por la secuencia *la su*; la interpolación de la negación entre el clítico y el verbo a partir de la secuencia *lo no*; la inserción del clítico en las formas analíticas de futuro en *\*r lo he*; la anteposición del clítico al verbo en *lo dijo* y a la forma condicional en *lo haría*.

Consideramos que estos 27 fenómenos, claramente característicos de diferentes etapas del español, pueden ser suficientes, tanto por su variedad como por su representatividad, para poder establecer un primer esbozo de periodización en función del corpus estudiado, como se podrá constatar en el apartado 4.

### 3. MÉTODO: EL AGRUPAMIENTO JERÁRQUICO APLICADO A LA PERIODIZACIÓN

El agrupamiento jerárquico es un método de análisis descriptivo, cuyo fin es identificar semejanzas y diferencias entre grupos. Mediante el cálculo matemático se establece un modelo descriptivo basado en la semejanza numérica entre diferentes categorías, agrupando las categorías más semejantes entre sí y distanciando las más diferentes. Un grupo está constituido por un conjunto de datos que son más similares entre sí que los datos que componen otros grupos. Así, el objetivo es encontrar aquellas similitudes que, a partir de las variables, permiten agrupar los

datos dentro de grupos que internamente sean lo más homogéneos posible (cf. Cardillo 2015).

En lingüística, este procedimiento ha sido utilizado tradicionalmente en el establecimiento de familias lingüísticas dentro de la lingüística histórico-comparada (glotocronología), donde la existencia de cognados en dos o más lenguas se interpreta como indicio de un parentesco entre ellas. Sin embargo, ya que el agrupamiento jerárquico agrupa categorías que comparten rasgos o propiedades, y la periodización lingüística impone que se respete la cronología de los datos, su aplicación a la periodización no es directamente posible. Es decir, para que la agrupación tenga sentido desde la perspectiva de la historia de la lengua, solo pueden agruparse categorías que sean cronológicamente sucesivas.

Solo recientemente el agrupamiento jerárquico ha sido aplicado a la periodización lingüística en el sentido que aquí nos interesa. Los lingüistas alemanes Stefan Th. Gries y Martin Hilpert (cf. Gries/Hilpert 2008, 2012 y Hilpert 2012) han desarrollado una variante del agrupamiento jerárquico aplicable a la periodización, llamado *Variability-based Neighbor Clustering* (VNC), cuya traducción aproximada al español sería *agrupamiento vecinal basado en la variabilidad*. La idea básica del VNC es justamente que aquellos periodos de tiempo sucesivos que presenten fenómenos lingüísticos semejantes se agruparán juntos, constituyendo así un periodo diferenciado. Así, en el análisis que presentamos en este trabajo hemos recurrido al algoritmo de VNC (elaborado por Gries y Hilpert para usarse en el entorno de análisis estadístico R), aplicándolo sobre los datos gramaticales de la historia del español extraídos del *CORDE* (vid. el apartado 2, arriba).

En la práctica, dadas las características del corpus base de este trabajo, cada uno de los 27 fenómenos gramaticales está representado por 98 puntos de información numérica: una cifra por década. Invertiendo el esquema, puede decirse que cada década, que es la unidad mínima de periodización empleada, se caracteriza por 27 cifras. La Tabla 1 incluye los datos de las diez décadas del siglo XIII a modo de ejemplo. Mediante el algoritmo VNC, las 98 décadas se comparan entre sí según la semejanza / diferencia relativa de sus cifras, agrupándose juntas las décadas más semejantes, y respetando siempre la cronología de los datos. Cabe hacer hincapié en que las cifras empleadas corresponden a la frecuencia relativa de cada fenómeno gramatical estudiado por millón de palabras. Usando frecuencias relativas se asegura que las cifras de las 98 décadas sean comparables entre sí, al mismo tiempo que se conserva el peso cuantitativo de los fenómenos más

frecuentes. En todo caso, hay que tener muy presente que dado que se trata de un análisis cuantitativo, los fenómenos más frecuentes tienen mayor peso a la hora de realizarse el agrupamiento.

	casos/millón	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290
pronombres	<i>su</i>	3313	3735	1126	2613	4489	6632	4008	6210	6729	6812
	<i>qui</i>	1295	1926	1267	2477	2795	1673	495	568	741	295
	<i>quienes</i>	0	0	0	0	0	0	0	0	2	1
	<i>las</i>	4409	3046	1838	3501	5416	9837	7347	9193	9008	6428
	<i>unas</i>	95	28	47	91	119	106	33	171	103	43
	<i>este</i>	477	230	1038	394	494	1198	948	2174	2282	1066
	<i>nosotros</i>	0	0	10	0	2	1	0	1	0	0
	<i>os</i>	310	153	202	152	136	86	35	86	56	201
	adverbios y preposición	<i>ante</i>	294	1008	202	671	950	1188	1171	1014	1147
<i>allende</i>		8	49	21	30	19	14	23	69	74	71
<i>delante</i>		1954	118	67	224	322	132	153	143	171	235
<i>ahora</i>		0	0	0	4	2	0	0	0	0	1
<i>cras</i>		64	0	0	19	23	12	6	16	21	3
<i>suso</i>		310	403	363	182	362	333	211	286	214	167
<i>ende</i>		246	417	301	463	447	825	1003	1143	1043	1135
<i>fuerte mente</i>		0	0	0	0	0	1	1	1	4	6
<i>pienso</i>		0	0	0	0	0	2	0	0	0	1
verbos	<i>soy</i>	0	14	0	4	11	11	10	19	10	68
	<i>serian</i>	16	7	5	4	6	17	18	23	41	41
	<i>sodes</i>	32	0	21	49	70	34	125	72	67	38
	<i>fuere</i>	143	2796	78	523	1093	3548	3224	506	859	1497
	<i>fuessedes</i>	0	7	0	8	0	0	5	7	4	4
	<i>lo non</i>	0	56	5	80	53	118	132	176	198	146
sintaxis	<i>lo haría</i>	8	0	0	0	6	2	0	1	2	5
	<i>lo dixo</i>	16	0	0	0	4	9	7	17	12	21
	<i>rlo he</i>	8	0	5	0	6	7	1	6	4	4
	<i>la su</i>	56	139	202	258	437	337	141	506	718	477

TABLA 1. Extracto de la muestra de datos cuantitativos correspondiente al siglo XIII. Las cifras indican frecuencias relativas de cada forma en el corpus.

#### 4. ANÁLISIS Y RESULTADOS

El resultado del análisis de VNC es un árbol, también llamado *dendrograma*, en el que cada ramificación indica una relativa divergencia cuantitativa entre décadas contiguas. Así, las agrupaciones permiten interpretarse como espacios de tiempo en los que los fenómenos investigados demuestran frecuencias de uso lo suficientemente semejantes entre sí como para constituir un periodo propio. En este apartado comentaremos los detalles más importantes de cinco árboles jerárquicos, empezando por el árbol general que incluye todos los fenómenos analizados. A continuación presentaremos cuatro árboles correspondientes a cada grupo de fenómenos gramaticales incluidos, a saber, los pronombres y los determinantes, los adverbios y las preposiciones, los verbos y las construcciones sintácticas.



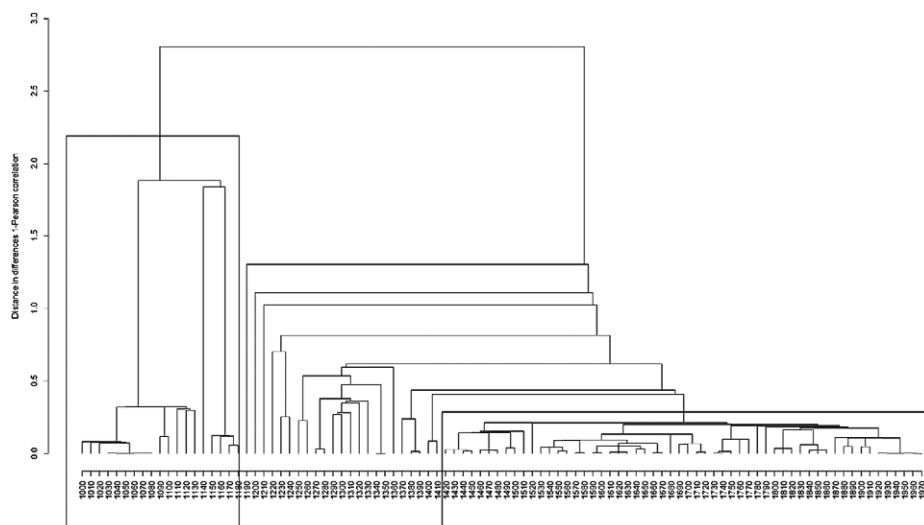


GRÁFICO 1. Árbol general basado en las frecuencias relativas de los 27 fenómenos gramaticales.

Como ilustra el Gráfico 1, el árbol global de nuestro análisis revela una clara bipartición de los datos. Así, destaca un primer periodo de orígenes, que va desde el año 1000 hasta 1180, y un segundo periodo, mucho más amplio, que va desde 1190 hasta 1970. Sin embargo, adentrándonos un poco más en las ramificaciones, parece claro que pueden hacerse algunas subdivisiones más. El primer periodo puede dividirse en dos (o incluso tres) sub-periodos: 1000-1130, seguido por 1140 (fecha del *Poema de Mio Cid*, que formaría un grupo aparte) y 1150-1180. El segundo periodo, por su parte, puede subdividirse en una primera época de unos 200 años (1190-1410), caracterizada por una gran variación. A esta le sigue una segunda época relativamente homogénea en comparación con todo el tiempo anterior, que empieza en 1420 y llega hasta finales del siglo XX. Lo que se desprende de estos datos es que, si bien parecen existir más subdivisiones dentro de este último periodo de 450 años, que incluye tanto el llamado español clásico como el moderno (*cf.* Eberenz 1991), la diferencia entre la década de 1450 y la de 1950, solo por poner un ejemplo, es considerablemente menor que la existente entre 1410 y 1450.

Para precisar la importancia de los cuatro grupos de fenómenos incluidos, se verá a continuación qué tipo de árboles se establecen si en el análisis jerárquico solo se incluyen fenómenos pertenecientes al mismo tipo gramatical. Si se empieza por los pronombres, se observa que el árbol resultante (Gráfico 2) es muy semejante al primero (Gráfico 1), estableciendo límites

entre periodos en casi los mismos puntos que aquel. Aparte de sugerir que esa primera periodización es adecuada, lo que indica la gran semejanza entre ambos árboles es que los pronombres parecen jugar un papel importante en el agrupamiento general. Es decir, como veremos a continuación, parece que son los datos de los pronombres los que mayor peso tienen sobre el agrupamiento general.

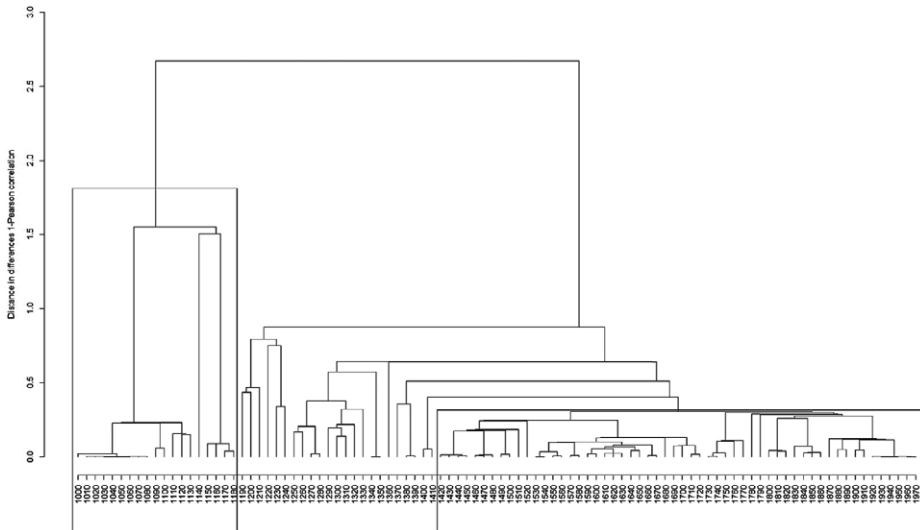


GRÁFICO 2. Periodización basada en las frecuencias relativas de los ocho pronombres.

Con respecto al primer periodo del Gráfico 2, de 1000 a 1130, lo que lo distingue de las décadas posteriores es, por un lado, la baja frecuencia de uso de los pronombres *las* y *su* así como la ausencia de *quienes* y *nosotros*, y el empleo poco más que esporádico de las formas *este* y *os*. A partir de 1420, sin embargo, los cambios más destacables que se observan son, por un lado, la aparición de *quienes* así como la reducción de la frecuencia de *qui*. Por otro lado, desde 1470 el uso de *os* se multiplica, y a partir de 1550 el de *nosotros*.

Sin embargo, si miramos el grupo de los adverbios y las preposiciones, la imagen es completamente diferente. Así, en el árbol jerárquico del Gráfico 3 (abajo) destacan cinco periodos distintos. Tenemos, pues, un primer periodo que va desde 1030 hasta 1110, y un segundo periodo, que va de 1230 a 1500. Sin embargo, el que parece más relevante comentar es el tercero, de 1560 a 1670, con respecto al cual se observa que los adverbios *cras* y *fuerte mente* se encuentran prácticamente desaparecidos ya, mientras que las frecuencias de *suso* y *ende* disminuyen claramente.

También la frecuencia de *ante* se reduce, mientras que este periodo ve el establecimiento de *delante* y *ahora*. En el cuarto periodo, que va de 1720 a 1890, *cras*, *suso*, *ende* y *fuerte mente* son ya inexistentes, y se consuma el ocaso de *allende*. Finalmente, en las décadas del siglo XX, 1910-1970, encontramos ya el sistema actual, pero lo que distingue este último periodo del anterior es, quizá sorprendentemente, la renovada frecuencia de *ende*..., y una menor frecuencia de *delante*.

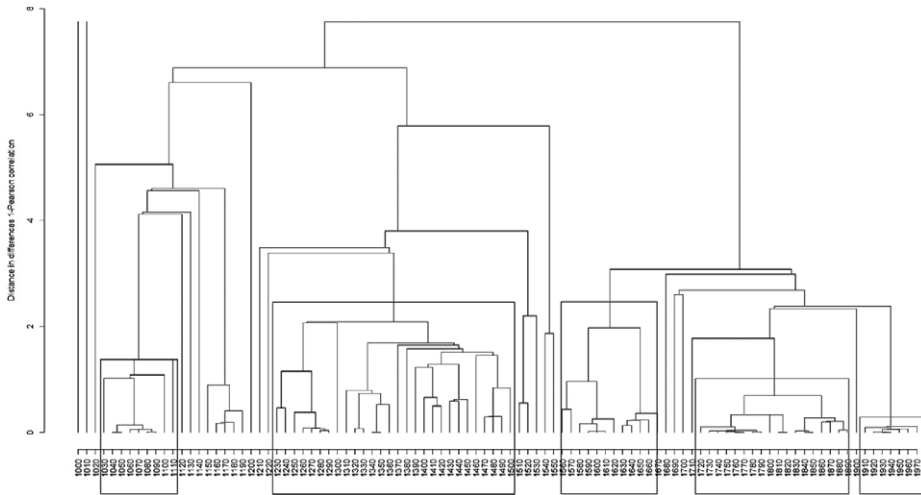


GRÁFICO 3. Periodización basada en las frecuencias relativas de los ocho adverbios y preposiciones.

También las seis formas verbales muestran una situación muy particular. Como revela el Gráfico 4 (abajo), se diferencian cuatro periodos principales, cuyos márgenes se sitúan en puntos completamente distintos comparado con los árboles anteriores (Gráficos 1, 2 y 3). La excepción es la primerísima época que también en el caso de los verbos termina con el *Poema de Mio Cid*, en 1140. A partir de ahí tenemos cuatro periodos más: el que va de 1150 a 1580 y se caracteriza por mucha variación; el situado entre 1590 y 1670, en el que aumenta claramente la frecuencia de la forma *soy*, mientras que experimentamos el ocaso de *fuessedes*, y la forma *sodes* acaba siendo prácticamente inexistente; el periodo comprendido entre 1690 y 1750, en el que *fuésedes* desaparece definitivamente; y un último periodo, que va de 1760 a 1970, en el que encontramos ya establecido el sistema actual, con *soy* y *pienso* deviniendo más frecuentes, mientras que nuestros datos revelan cómo la forma *fuere* pervive hasta bien entrado el siglo XX.

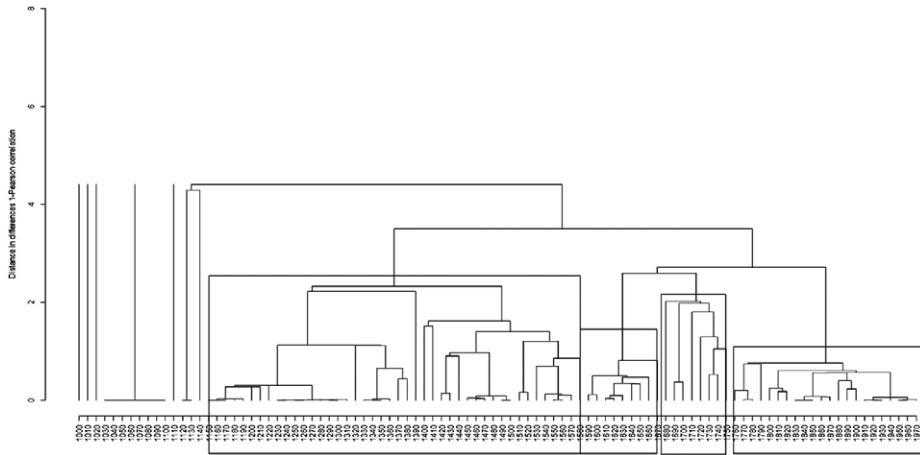


GRÁFICO 4. Periodización basada en las frecuencias relativas de las seis formas verbales.

Finalmente, asimismo los fenómenos sintácticos trazan un árbol muy particular (Gráfico 5), que permite establecer dos periodos principales, pero que incluyen bastante variación. En el primer periodo, que va de 1170 a 1520, prevalece *lo non*, es decir, los casos de interpolación de la negación entre el clítico personal y el verbo, además de la posposición del pronombre de complemento directo e indirecto al verbo, y el uso del artículo ante posesivo, *la su*. En la segunda época, de 1530 a 1970, encontramos que hasta 1680 se mantiene el futuro analítico, con construcciones del tipo *hacer lo he*, mientras que a partir de entonces tenemos solo la anteposición del pronombre con la forma condicional, *lo haría*.

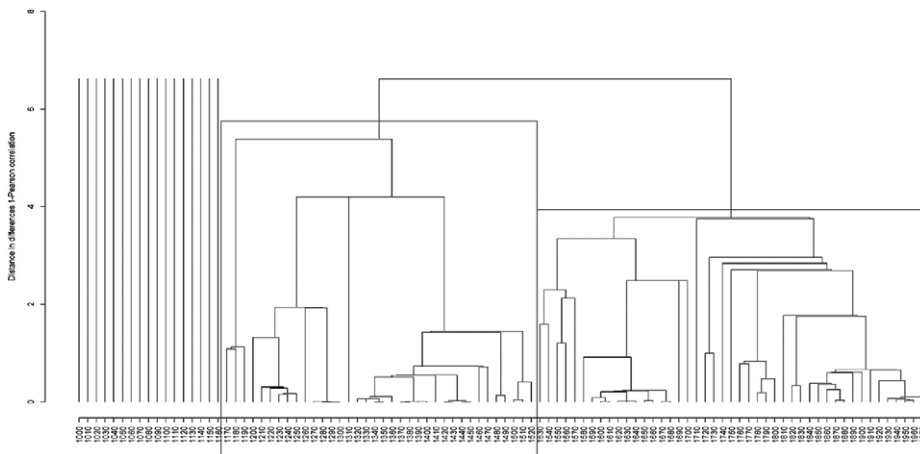


GRÁFICO 5. Periodización basada en las frecuencias relativas de las cinco construcciones sintácticas.

Después de analizar cinco árboles jerárquicos, contruidos a partir de diferentes datos, podemos acabar este apartado constatando, en primer lugar, que la periodización general parece obedecer más que nada a los datos de los pronombres, lo cual se deduce de la comparación del árbol general con los de los fenómenos particulares. Cabe notar que la correspondencia casi completa entre los pronombres y la totalidad de los datos se debe a que este sea el grupo con el mayor número de casos. De hecho, los pronombres representan el 89 por ciento de todas las frecuencias, motivo por el cual el árbol general no puede considerarse realmente mucho más que una versión algo decorada del de los pronombres. A consecuencia de esto, el análisis de las diferentes configuraciones de los árboles presentados en los Gráficos 2 a 5 nos hace plantearnos el verdadero significado del concepto de periodización. Al hablar de la periodización en el sentido cuantitativo que adoptamos en este estudio es esencial ser conscientes de que los fenómenos observados condicionan los resultados, aspecto que no siempre se ha tenido en cuenta a la hora de establecer las distintas etapas históricas de la lengua. Es decir, la periodización es diferente dependiendo de los fenómenos que se analicen, lo cual resulta totalmente evidente a partir de los datos estudiados. A la inversa, y ya no tan obvio, esto puede llegar a significar que la periodización en términos globales quizá sea poco más que un ideal.

## 5. CONCLUSIONES

Como se ha podido comprobar, los resultados presentados se basan únicamente en los datos de los fenómenos, es decir, en criterios lingüísticos. Ello nos permite operar con cifras globales, con frecuencias totales, con frecuencias relativas, etc., siempre a partir del análisis del corpus. Por todo ello, debemos reconocer también que los resultados obtenidos en el presente trabajo dependen completamente de los fenómenos seleccionados para su investigación. Ciertamente, hemos escogido fenómenos «clásicos» de los estudios sobre periodización, fenómenos que sabemos que cambian durante la historia del español, por lo que los resultados obtenidos están anclados en la historia de los estudios sobre periodización. Sin embargo, por contra, gracias a la aplicación metodológica del agrupamiento jerárquico, ofrecen un panorama, quizá, más exacto y más claramente basado en los datos que los análisis anteriores que aparecen en las historias de la lengua y en las gramáticas históricas, ya que describen detalladamente la «historia conocida».

Pero no son ni mucho menos la última palabra. Ya que los diferentes fenómenos ofrecen resultados (= periodizaciones) diferentes entre sí,

destacan la necesidad de hacer análisis más precisos y más específicos. Una posibilidad sería «normalizar» los datos con el fin de evitar que los pronombres predominen tan claramente como lo hacen en el árbol general. Además, continúan existiendo factores que en un futuro deberían poder ser incorporados al análisis. Y aparte de los elementos puramente gramaticales, se debería tener también en cuenta la existencia de los géneros discursivos, las áreas dialectales, incluso los idiolectos, entre otros, si bien en este último caso el hecho de operar en franjas de diez años nos permite recoger de forma aproximada esta información. Pero todo esto es trabajo para futuras investigaciones.

#### BIBLIOGRAFÍA

- Buenafuentes, Cristina (2002a): «Cambio gramatical y periodización en español», *Res Diachronicae*, 1, 96-105.
- (2002b): «Sintaxis y periodización en español», *Res Diachronicae*, 1, 393-398.
- Cano Aguilar, Rafael (1992): «La sintaxis española en la época del Descubrimiento», en José Antonio Bartol Hernández, Juan Felipe García Santos y Javier de Santiago Guervós, eds., *Estudios filológicos en homenaje a Eugenio de Bustos Tovar*, Salamanca, Universidad de Salamanca, 183-197.
- Cardillo, Marcelo (2015): *Análisis de agrupamiento jerárquico*. Material de enseñanza publicado en Academia.edu, <<https://conicet.academia.edu/MarceloCardillo/Teaching-Documents>> [julio de 2015].
- CORDE: Real Academia Española, Banco de datos. *Corpus diacrónico del español*, <<http://www.rae.es>> [junio-julio de 2015].
- Eberenz, Rolf (1991): «Castellano antiguo y español moderno: reflexiones sobre la periodización en la historia de la lengua», *Revista de Filología Española*, 71, 79-106.
- (2009): «La periodización de la historia morfosintáctica del español: propuestas y aportaciones recientes», *Cahiers d'Études Hispaniques Médiévales*, 32, 181-201.
- Flores, Marcela y Chantal Melis (2015): «Periodización en español. Evidencia para una tercera etapa evolutiva», *Études Romanes de Brno*, 36/2, 11-28.
- Girón Alconchel, José Luis (2005): «Cambios gramaticales en los Siglos de Oro», en Rafael Cano Aguilar, coord., *Historia de la lengua española*, Barcelona, Ariel, 2.<sup>a</sup> ed. actualizada, 859-893.
- (2008): «La lengua de un embajador y un marino del siglo XVIII: ¿español moderno ya, o todavía clásico?», en Concepción Company Company y José G. Moreno de Alba, eds., *Actas del VII Congreso Internacional de Historia de la Lengua Española*, Madrid, Arco Libros, vol. II, 2243-2253.
- Gries, Stefan Thomas (2009): *Statistics for Linguistics with R*, Berlín-Boston, De Gruyter Mouton.

- y Martin Hilpert (2008): «The identification of stages in diachronic data: variability-based neighbor clustering», *Corpora*, 3, 59-81.
  - (2012): «Variability-based neighbor clustering: A bottom-up approach to periodization in historical linguistics», en Terttu Nevalainen y Elizabeth Closs Traugott, eds., *The Oxford Handbook of the History of English*, Oxford, Oxford University Press, 134-144.
- Hilpert, Martin (2012): «Diachronic collocation analysis: how to use it and how to deal with confounding factors», en Kathryn Allan y Justyna A. Robinson, eds., *Topics in English Linguistics: Current Methods in Historical Semantics*, Berlín, De Gruyter Mouton, 133-160.
- Lapesa, Rafael (1981): *Historia de la lengua española*, Madrid, Gredos, 9.<sup>a</sup> ed. corregida y aumentada.
- Marcos Marín, Francisco (1992): «Spanisch: Periodisierung», en Günther Holtus, Michel Metzeltin y Christian Schmitt, eds., *Lexikon der Romanistischen Linguistik*, Tübinga, Max Niemeyer, vol. VI/1, 602-607.
- (1995): «La periodización», en M.<sup>a</sup> Antonia Martín Zorraquino y Túa Blesa, coords., *Homenaje a Félix Monge. Estudios de lingüística hispánica*, Madrid, Gredos, 325-333.
- Martínez Alcalde, M.<sup>a</sup> José y Mercedes Quilis Merín (1996): «Nuevas observaciones sobre la periodización en la historia de la lengua española», en Alegría Alonso González, Ladislao Castro Ramos, Bertha Gutiérrez Rodilla y José Antonio Pascual Rodríguez, eds., *Actas del III Congreso Internacional de Historia de la Lengua Española*, Madrid, Arco Libros, tomo I, 873-886.
- Melis, Chantal, Marcela Flores y Sergio Bogard (2003): «La historia del español. Propuesta de un tercer periodo evolutivo», *Nueva Revista de Filología Hispánica*, 51, 1-56.
- Menéndez Pidal, Ramón (1942): «El lenguaje del siglo XVI», en Ramón Menéndez Pidal, *La lengua de Cristóbal Colón, el estilo de Santa Teresa y otros estudios sobre el siglo XVI*, Madrid, Espasa-Calpe, 47-84.
- Octavio de Toledo y Huerta, Álvaro S. (2007): «Un rasgo sintáctico del primer español moderno (ca. 1675-1825): las relaciones interoracionales con *ínterin* (que)», en Marta Fernández Alcaide y Araceli López Serena, coords., *Cuatrocientos años de la lengua del Quijote: estudios de historiografía e historia de la lengua española. Actas del V Congreso Nacional de la Asociación de Jóvenes Investigadores de Historiografía e Historia de la Lengua Española*, Sevilla, Universidad de Sevilla, 421-442.
- Riduejo Alonso, Emilio (1993): «¿Un reajuste sintáctico en el español de los siglos XV y XVI?», en Ralph Penny, ed., *Actas del Primer Congreso Anglo-Hispano*, Madrid, Castalia, t. I, 49-60.
- Sánchez Lancis, Carlos (1997-1998): «Cambio morfológico y periodización en español preclásico», *Estudi General*, 17-18, 173-199.

- (1999): «Sintaxis histórica, informática y periodización del español», en José Manuel Blecua, Gloria Clavería, Carlos Sánchez y Joan Torruella, eds., *Filología e Informática. Nuevas tecnologías en los estudios filológicos*, Lleida-Bellaterra, Milenio-Seminario de Filología e Informática (Universidad Autónoma de Barcelona), 415-446.
- (2001): «Historia de la lengua, gramática histórica y periodización en español», *Estudi General*, 21, 395-412.
- (2009): «Corpus diacrónicos y periodización del español», *Cahiers d'Études Hispaniques Médiévales*, 32, 159-180.
- (2012): «Periodización y cambio gramatical: el siglo XVIII, ¿frontera temporal del español?», en M.<sup>a</sup> Teresa García-Godoy, ed., *El español del siglo XVIII. Cambios diacrónicos en el primer español moderno*, Berna-Berlín-Bruselas-Fráncfort am Main-Nueva York-Oxford-Viena, Peter Lang, 21-51.