

# HEIDELBERGER SPÄTLESE

AUSGEWÄHLTE TROPFEN AUS VERSCHIEDENEN  
LAGEN DER SPANISCHEN SPRACH-  
UND ÜBERSETZUNGSWISSENSCHAFT

Festschrift anlässlich des  
70. Geburtstages von  
Prof. Dr. Nelson Cartagena

herausgegeben von

Jörn Albrecht

und Frank Harslem

Sonderdruck

Romanistischer Verlag  
Bonn 2008



MINISTERIO  
DE CULTURA

EMBAJADA DE ESPAÑA  
BOTSCHAFT VON SPANIEN

**ProSpanien.de**



DEUTSCHER HISPANISTENVERBAND e. V.  
ASOCIACION ALEMANA DE HISPANISTAS



Romanistischer Verlag  
Hochkreuzallee 46, 53175 Bonn

Copyright by Jörn Albrecht, Frank Harslem

Alle Rechte vorbehalten  
ISBN 978-3-86143-183-1

## De nuevo sobre la frecuencia de las formas *llegara* y *llegase*

Guillermo Rojo (Santiago de Compostela)

1. El paso de la forma heredera del pluscuamperfecto de indicativo del verbo latino al subjuntivo, "posiblemente el aspecto más llamativo de la diacronía del verbo español en comparación con otros sistemas románicos" en palabras de A. Veiga (1996: 7), es un fenómeno relativamente bien conocido, aunque todavía quedan –al menos, esa es mi impresión– algunos factores sobre los que hay que seguir trabajando. La migración de *llegara* a valores subjuntivos sin abandonar del todo los correspondientes al indicativo, la consiguiente confluencia con *llegase*, la aparición de la compuesta *hubiera llegado* y su esperable coincidencia con *hubiese llegado* produce, en efecto, una situación peculiar, que, al lado de las de corte más teórico, presenta cuestiones de gran interés relacionadas con el origen del cambio, su ritmo de generalización en los últimos siglos y también la distribución que muestra actualmente en las diversas variedades del español. Dado que se trata de uno de los muchos temas sobre los que todos nos hemos beneficiado del trabajo realizado por el Prof. Cartagena (cf., entre otros, Cartagena 1978 y 1999), me propongo en estas páginas contribuir a su mejor conocimiento enfocándolo ahora desde una perspectiva muy general, una perspectiva casi exclusivamente cuantitativa procedente de la consideración de los datos que podemos obtener del análisis de los grandes corpus textuales de los que disfrutamos en la actualidad. El enfoque que voy a adoptar pretende compensar la falta de atención a las cuestiones gramaticales implicadas con la fiabilidad derivada de la posibilidad de tener en cuenta un gran número de casos distribuidos en un período de tiempo suficientemente amplio.<sup>1</sup>

2. Reducido al mínimo, y con una cierta superficialidad, el proceso del que nos vamos a ocupar aquí consiste en que se parte de una situación en la que dos formas lingüísticas, A y B, cada una de ellas con un conjunto propio de valores, sufren un proceso

<sup>1</sup> Sería injusto dejar únicamente en el ámbito más íntimo el reconocimiento y la gratitud por la inestimable colaboración de María José Ceinos, que se ha ocupado, con paciencia infinita, de hacer cientos de consultas al CREA y el CORDE.

mediante el cual la forma A va adquiriendo la posibilidad de expresar, además de los valores que le corresponden inicialmente, alguno de los que en la situación de partida corresponden a la forma B, o incluso todos ellos. Si la forma B no adquiere nuevos valores, su uso se irá reduciendo a medida que el empleo de A en las nuevas funciones se generalice, con lo que B podría llegar finalmente a desaparecer o conservarse únicamente como, por ejemplo, un rasgo de estilo. Dado lo anterior, puesto que el conjunto total de valores de las dos formas se mantiene a lo largo del período, la frecuencia relativa de A y B en diferentes cortes temporales nos dará, sin necesidad de análisis más detallados, una visión válida de las grandes líneas del proceso.

Por supuesto, eso no significa que no haya que proceder a una investigación mucho más detallada y detenida del modo en que esa forma, *llegara* en nuestro caso, va adquiriendo las funciones de *llegase* en diferentes entornos estructurales, como se ha hecho ya en varias ocasiones. Sin embargo, esos trabajos tienen, por necesidad, que restringirse al estudio de lo que sucede en unos pocos textos, seleccionados con diferentes criterios, con lo que el panorama obtenido resulta, cuando menos, incierto. Eso es lo que sucede, por ejemplo, con el trabajo en el que hace ya algunos años analicé la distribución de las formas del tipo *llegara* y *llegase* en español actual utilizando para ello la totalidad de los datos contenidos en el corpus ARTHUS, que es el que reunimos en la Universidad de Santiago de Compostela para construir la *Base de datos sintácticos del español actual* (BDS).<sup>2</sup> Ese corpus, formado por 34 textos, con un total de algo menos de millón y medio de formas y un conjunto de casi 160 000 cláusulas, resulta adecuado para la finalidad con la que fue construido, la de analizar los fenómenos relacionados con el régimen verbal y la estructura de la cláusula en español contemporáneo, pero tiene un tamaño excesivamente reducido para poder proyectar los datos obtenidos en su análisis a la caracterización general de las diferencias en el uso de estas dos formas en español actual. La distribución que presentan los 3948 casos de las cuatro formas implicadas (dos simples y dos compuestas) parecen confirmar la opinión general según la cual el proceso de sustitución de *llegase* y *hubiese llegado* por *llegara* y *hubiera llegado* está más avanzado, en general, en América que en España, pero hay fenómenos específicos, que surgen en el análisis de lo que sucede en algunos textos concretos, que oscurecen el panorama general y dan un

<sup>2</sup> Disponible en <http://www.bds.usc.es>.

cierto aire de inseguridad a la hora de hacer formulaciones generales.<sup>3</sup>

Me propongo aquí complementar el trabajo anterior enfocando la cuestión en un aspecto distinto y con una metodología bastante diferente. Trataré de identificar las grandes líneas de cambio en la relación entre estas dos formas desde 1700 hasta la actualidad y proporcionar una idea aproximada de su distribución actual a partir de los datos contenidos en el *Corpus del español*, construido por Mark Davies, de una parte (<http://www.corpusdel-espanol.org>) y, de otra, en el *Corpus diacrónico del español* (CORDE) y el *Corpus de referencia del español actual* (CREA) construidos por la Real Academia Española (<http://www.rae.es>). El primero de ellos contiene alrededor de cien millones de formas de todas las épocas del español, está parcialmente lematizado y permite la recuperación de información cuantitativa con diferenciación de épocas. Los corpus académicos son, como es bien sabido, dos grandes conjuntos textuales que contienen en total más de cuatrocientos millones de formas de todas las épocas del español.<sup>4</sup> En su versión pública actual, CREA y CORDE no están lematizados, de modo que las búsquedas han de ser exclusivamente textuales, pero es posible, en cambio, hacer recuperación selectiva en función de la época, el país, el tipo de texto, el autor o la obra. Ello permite, si el procedimiento diseñado para la recuperación textual resulta fiable, obtener cifras correspondientes a datos de diferentes países (España y México, por ejemplo) a lo largo de un período amplio, con diferenciación de tramos temporales suficientemente finos, según mostraré a continuación.

3. Como he indicado ya, el *Corpus del español* está lematizado, al menos parcialmente<sup>5</sup>, de modo que, puesto que esa lematización incluye también la indicación de categoría y subcategoría gramatical, es posible pedir la totalidad de los casos de la forma conocida habitualmente como pretérito (o pretérito imperfecto) de subjuntivo, afortunadamente diferenciando entre la forma en *-ra* y la forma en *-se*. La opción de agrupar por lemas pro-

<sup>3</sup> Cf. infra, § 8 para los resultados generales obtenidos. Examinando los datos con más detalle se tropieza, por ejemplo, con que la idea de que las formas en *-se* son muy escasas en México encaja perfectamente con el 100 % que suponen las formas en *-ra* en *Querido Diego, te abraza Quiela y otros cuentos*, de Elena Poniatowska, pero se compadece mal con el 92,31 % que las formas en *-se* alcanzan en *Tiempo nublado*, de Octavio Paz, el porcentaje más alto de todos los textos incluidos en ARTHUS. Para más detalles, vid. Rojo 1996: 685.

<sup>4</sup> Para todos los detalles de construcción, textos incluidos, sistema de codificación, etc., vid. <http://www.rae.es>.

<sup>5</sup> Cf. Davies 2005.

porciona datos cuantitativos muy fáciles de manejar y, además, proporciona totales generales distribuidos por siglos. Las formas compuestas, en cambio, no han sido caracterizadas como tales, por lo que ha sido necesario recurrir a la recuperación de datos utilizando la fórmula correspondiente a la forma auxiliar utilizada (*hubiera* o *hubiese*) seguida de un participio cualquiera. También aquí los datos aparecen agrupados según el verbo de la forma auxiliada y, por supuesto, con los totales correspondientes y diferenciados por siglos. Para mantener el paralelismo con la técnica utilizada en CORDE y CREA, hice las consultas únicamente con las formas *hubiera* y *hubieran* de una parte y *hubiese* y *hubiesen* de otra, esto es, las de primera y tercera de singular y tercera de plural.<sup>6</sup>

En el caso de CREA y CORDE no es posible recurrir al lema ni a la subcategoría gramatical, de modo que fue necesario diseñar otro procedimiento, puesto que no parecía razonable pensar en obtener por separado los casos correspondientes a todas y cada una de las personas de todos y cada uno de los verbos documentados en los corpus, con sus posibles variantes morfológicas y ortográficas. Decidí, en consecuencia, optar por una aproximación no excesivamente costosa en la recogida de los materiales y que, no obstante, produce un número de casos que, a mi modo de ver, resulta realmente significativo y traza un panorama suficientemente claro y congruente de lo que ha sido la evolución de estas formas desde 1700 hasta la actualidad. El procedimiento adoptado consiste en solicitar el número de apariciones en diferentes tramos temporales, como se verá, de los casos de primera y tercera de singular y tercera de plural de las formas del tipo *llegara*, *llegase*, *hubiera llegado* y *hubiese llegado* de once de los verbos más frecuentes en español actual en uso independiente, según los datos de la BDS.<sup>7</sup> Dado que estos dos corpus permiten una selección temporal muy fina, decidí, tras las pruebas iniciales, trabajar en bloques de veinticinco años desde 1700 hasta 1999. Ejemplificando, para mayor claridad, con el verbo *tener*, he realizado todas las consultas correspondientes a las formas

<i>tuviera</i>	<i>tuvieran</i>
<i>tuviese / tuviese</i>	<i>tuviesen / tuviesen</i>
<i>hubiera tenido</i>	<i>hubieran tenido</i>
<i>hubiese / hubiese tenido</i>	<i>hubiesen / hubiesen tenido</i>

<sup>6</sup> En el caso de *hubiese*, incluí también las grafías con *-ss-*, para más seguridad en las primeras épocas estudiadas en este trabajo. No hice, en cambio, búsquedas con otras variantes gráficas.

<sup>7</sup> Son once por la confluencia de algunas de las formas correspondientes a *ser* e *ir*. No son exactamente los más frecuentes para evitar las complicaciones morfológicas y ortográficas de, por ejemplo, *decir*, o *hacer*.

en periodos de veinticinco años (1700-1724, 1725-1749, ... , 1950-1974 y 1975-1999) para textos de España por un lado y de México por otro. El resultado arroja un total de 141 637 casos de estas formas en textos de España y de 12 454 en textos mexicanos. A pesar de la fortísima reducción a unos pocos verbos, las cifras parecen suficientes para proporcionar una perspectiva general de lo ocurrido con estas formas en los últimos trescientos años.

4. La línea general de la evolución de la frecuencia relativa de uso de las dos formas aparece, con bastante claridad, en la tabla número 1, en la que resumo los datos obtenidos del análisis del *Corpus del español* mediante una búsqueda referida a la subcategoría (con las expresiones de búsqueda \*.v\_subj\_ra y \*.v\_subj\_se) y con los resultados agrupados por lemas. Los totales correspondientes a 600 verbos –supongo que son los de mayor frecuencia general–, que parece ser lo máximo que proporciona el sistema, son los que figuran a continuación.

	-ra	-se	Total	% -ra	% -se
<b>1700-1799</b>	14 274	25 730	40 004	35,68	64,32
<b>1800-1899</b>	45 319	33 961	79 280	57,16	42,84
<b>1900-1999</b>	42 550	6 363	48 913	86,99	13,01
<b>Totales</b>	102 143	66 054	168 197		

Tabla 1. Frecuencias absolutas y relativas de las formas del tipo *llegara* y *llegase* reconocidas como tales en el *Corpus del español*. Elaboración propia a partir de los datos obtenidos en el corpus ([www.corpusdelespanol.org](http://www.corpusdelespanol.org)). [Comprobado en agosto de 2007].

Dado que los datos vienen agrupados por siglos, el panorama resultante tiene una apariencia bastante tosca, pero la distribución de esas casi 170 000 formas proporciona una impresión realmente clara de la evolución experimentada por estas dos formas en los últimos trescientos años de historia de la lengua: la forma en -se, dominante en el siglo XVIII, se ve reducida a un 13 % en el conjunto del siglo XX. La malla de la red establecida para la obtención de los datos es, sin duda, excesivamente ancha, pero el peso de las frecuencias absolutas y la claridad de los porcentajes hacen que los datos sean realmente significativos.

En la misma línea general, pero con ciertas diferencias de interés con respecto a los anteriores, se mueven los datos correspondientes a las formas compuestas,<sup>8</sup> que incluyo en la tabla 2.

	<i>hubiera</i> <i>-do</i>	<i>hubiese</i> <i>-do</i>	Total	% <i>hubiera</i> <i>-do</i>	% <i>hubiese</i> <i>-do</i>
<b>1700-1799</b>	1166	1045	2211	52,74	47,26
<b>1800-1899</b>	6018	2491	8509	70,73	29,27
<b>1900-1899</b>	5089	1108	6197	82,12	17,88
<b>Totales</b>	12 273	4644	16 917		

Tabla 2. Frecuencias absolutas y relativas de las formas del tipo *hubiera llegado* y *hubiese llegado* en el *Corpus del español*. Elaboración propia a partir de los datos obtenidos en el corpus ([www.corpusdelespanol.org](http://www.corpusdelespanol.org)). [Comprobado en agosto de 2007].

También aquí, como era de esperar, las formas en *-se* experimentan un notable descenso a lo largo del período estudiado, pero los extremos de partida y llegada están mucho más próximos: las formas en *-ra* son predominantes ya en el xviii y, de otro lado, su porcentaje general no es tan alto como el de las simples en el conjunto del xx. Los dos aspectos quedan visibles en el gráfico 1.

<sup>8</sup> Como ya he indicado, puesto que el sistema de anotación utilizado por Davies no identifica como tales las formas compuestas, las expresiones de búsqueda empleadas utilizan las formas de primera y tercera del singular y la tercera del plural de *haber* seguidas de un participio. También aquí el sistema devuelve los datos correspondientes a los 600 verbos más frecuentes. Lo anterior implica que las frecuencias de *hubiera*, *hubieran*, *hubiese* y *hubiesen* están en los dos bloques. No es difícil retirarlos del bloque de las simples, pero eso significaría perder los casos de uso independiente del verbo y, además, el peso sobre los totales no es relevante.

Porcentaje de uso de las formas simples y compuestas en el Corpus del español

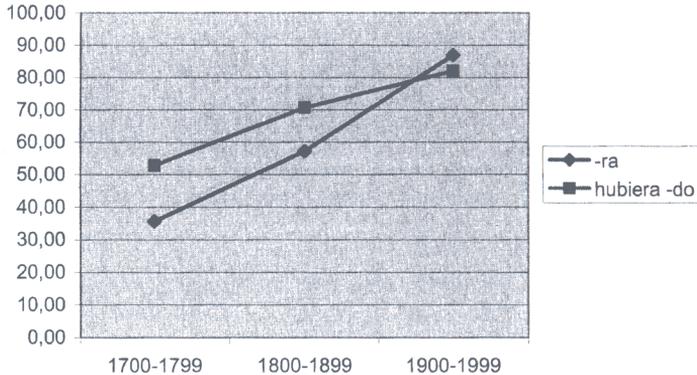


Gráfico 1. Evolución de los porcentajes de uso de las formas simples y compuestas según los datos del *Corpus del español*. Elaboración propia a partir de los datos obtenidos en el corpus ([www.corpusdelespanol.org](http://www.corpusdelespanol.org)).

5. Las versiones públicas de CORDE y CREA no admiten todavía la búsqueda por lemas ni por categorías o subcategorías gramaticales, por lo que he tenido que recurrir, como he indicado ya en el § 3, a la extracción de datos de ciertas personas (primera y tercera de singular, tercera de plural) de algunos de los verbos más frecuentes según los datos obtenidos en el análisis de la BDS (*tener, dar, ser/ir, ver, estar, hablar, quedar, llevar, mirar y pensar*).<sup>9</sup> Con el fin de asegurar que los datos resultantes de estos dos procedimientos distintos resultan comparables, he hecho la búsqueda paralela en el *Corpus del español*. Los resultados obtenidos para las formas del tipo *llegara*, que son los que figuran en la tabla 3, muestran con toda claridad que las diferencias son muy reducidas, de modo que es posible utilizar el segundo procedimiento en la recogida de datos de los corpus académicos.

<sup>9</sup> Como es de suponer, los recuentos incluyen también todas las apariciones del adverbio *fuera*. Las calas que he hecho parecen indicar que ese hecho no distorsiona los datos correspondientes a los verbos *ser* e *ir* y mucho menos los generales. De todas formas, es un factor que habrá que tener en cuenta en análisis más refinados.

	600 verbos	11 verbos (= CORDE y CREA)
1700-1799	35,68	38,18
1800-1899	57,16	56,20
1900-1999	86,99	87,69

Tabla 3. Comparación de los porcentajes correspondientes a las formas de tipo *llegara* en el *Corpus del español* en las consultas correspondientes a los 600 verbos más frecuentes y los once verbos utilizados en la recogida de datos de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en el corpus ([www.corpusdelespanol.org](http://www.corpusdelespanol.org)). [Comprobado en agosto de 2007].

La indudable desventaja de no poder formular las consultas a la totalidad de las formas de un cierto tipo se ve compensada por la organización de CORDE y CREA en otros aspectos, gracias a lo cual, en términos generales, es posible lograr un refinamiento considerablemente más alto en la obtención de los resultados. En concreto, para lo que aquí nos ocupa, he podido hacer recogida selectiva de datos en dos aspectos distintos. Por una parte, para tratar de obtener una aproximación cuantitativa a las diferencias de uso entre España y América, he extraído por separado los datos correspondientes a España y México. Por otra, para poder calibrar mejor los posibles altibajos de la distribución de frecuencias, he segmentado las búsquedas en períodos de veinticinco años entre 1700 y 1999. La conjunción de ambos filtros consigue un panorama que combina la claridad de la evolución general con el detalle de lo que sucede en los diferentes períodos considerados.<sup>10</sup>

<sup>10</sup> No me ha parecido pertinente para esta visión panorámica de la evolución de las formas, pero, como es bien sabido, la organización de CORDE y CREA permite la obtención de datos diferenciados también según otros parámetros, como el tipo de texto, género literario, autor, etc. En el § 6 utilizo la posibilidad de diferenciar la obtención de resultados por tipos de textos.

	-ra	-se	Total	% -ra	% -se
1700-1724	823	1879	2702	30,46	69,54
1725-1749	1546	2144	3690	41,90	58,10
1750-1774	1314	1490	2804	46,86	53,14
1775-1799	1191	1547	2738	43,50	56,50
1800-1824	1271	1805	3076	41,32	58,68
1825-1849	2975	3484	6459	46,06	53,94
1850-1874	3658	2694	6352	57,59	42,41
1875-1899	10 258	7394	17 652	58,11	41,89
1900-1924	5158	4952	10 110	51,02	48,98
1925-1949	7305	3441	10 746	67,98	32,02
1950-1974	8798	3681	12 479	70,50	29,50
1975-1999	42 058	9039	51 097	82,31	17,69
<b>Totales</b>	<b>86 355</b>	<b>43 550</b>	<b>129 905</b>		

Tabla 4. Frecuencias absolutas y relativas de las formas del tipo *llegara* y *llegase* de once verbos muy frecuentes en los textos españoles de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)). [Comprobado en marzo de 2006].

La distribución de estos casi 130 000 casos contenidos en CORDE y CREA es, me parece, bastante clara. Las líneas generales muestran, como era de esperar después de lo obtenido en el *Corpus del español*, un marcadísimo descenso en la utilización de la forma *llegase* desde los primeros años del XVIII hasta los últimos del XX. Según la tabla 4, para el conjunto de los verbos considerados se pasa de casi un 70 % a algo menos del 18 %. Ahora bien, el ascenso de *llegara* y el descenso de *llegase* no presentan, como también era de esperar, líneas que muestren una tendencia estable. Es fácil apreciar que *llegara* va aumentando desde 1700 hasta 1774, pero desciende en los dos períodos siguientes, de forma que entre 1825 y 1849 presenta unos porcentajes de utilización semejantes a los que tenía ya entre 1750 y 1774. La línea ascendente se quiebra otra vez en el tramo 1900-1924, pero desde ese momento sube sin parar y, además, se observa un fortísimo aumento entre el último tramo correspondiente al CORDE y los veinticinco años considerados del CREA.

De nuevo, la línea seguida por las formas compuestas resulta diferente a la que hemos podido observar en las simples. Es evidente la diferencia en las frecuencias generales, puesto que pasamos de unos 130 000 casos en las simples a menos de 12 000 en las compuestas. Además, según muestra la tabla 5, el punto de partida de *hubiera llegado* es excepcionalmente alto y el de *llegada* queda ligeramente por debajo del correspondiente a la forma simple, con lo que la horquilla resulta bastante más reducida que en el caso anterior, lo mismo que hemos podido observar

con los datos procedentes del *Corpus del español*. Aun sin tener en cuenta el efecto producido por el inesperado porcentaje correspondiente al primer tramo (que habrá que atribuir de momento a lo escaso de los ejemplos procedentes de ese período), la línea de las formas compuestas presenta ligeras diferencias en los altibajos que hemos observado en las simples: no se da exactamente del mismo modo en el final del XVIII y comienzos del XIX (lo cual resulta perfectamente comprensible si tenemos en cuenta los porcentajes en que está situada desde el principio), pero presenta también la curiosa caída observada en el primer cuarto del siglo XX.

	<i>hubiera</i> -do	<i>hubiese</i> -do	Total	% <i>hubiera</i> -do	% <i>hubiese</i> -do
<b>1700-1724</b>	34	16	50	68,00	32,00
<b>1725-1749</b>	23	18	41	56,10	43,90
<b>1750-1774</b>	82	47	129	63,57	36,43
<b>1775-1799</b>	170	73	243	69,96	30,04
<b>1800-1824</b>	218	84	302	72,19	27,81
<b>1825-1849</b>	497	199	696	71,41	28,59
<b>1850-1874</b>	655	196	851	76,97	23,03
<b>1875-1899</b>	1218	500	1718	70,90	29,10
<b>1900-1924</b>	427	244	671	63,64	36,36
<b>1925-1949</b>	786	264	1050	74,86	25,14
<b>1950-1974</b>	893	231	1124	79,45	20,55
<b>1975-1999</b>	3902	955	4857	80,34	19,66
<b>Totales</b>	8905	2827	11 732		

Tabla 5. Frecuencias absolutas y relativas de las formas del tipo *hubiera llegado* y *hubiese llegado* de once verbos muy frecuentes en los textos españoles de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)). [Comprobado en marzo de 2006].

Semejanzas y diferencias en el comportamiento de simples y compuestas quedan de manifiesto en el gráfico 2, que muestra de forma muy nítida lo ocurrido con cada tipo de formas.

**Evolución de las formas en -ra en España**

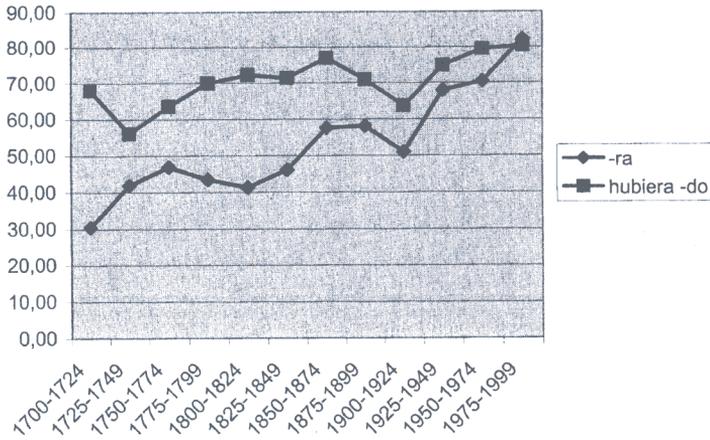


Gráfico 2. Evolución de los porcentajes de uso de las formas simples y compuestas según el CORDE y el CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)).

La igualación aparente que produce la utilización de los valores porcentuales no puede hacer olvidar que las formas compuestas tienen en todos los períodos considerados una frecuencia general muy inferior a la que presentan las formas simples, como muestran las tablas anteriores. Además de ello, conviene tener en cuenta que la relación entre las formas simple y compuesta de la variante *-ra* ha experimentado un ligero cambio en el período considerado, consistente en que la simple ha cedido algo de terreno a la compuesta, que es lo esperable por lo que sabemos de la evolución del sistema verbal. Los datos están en la tabla 6.

	<i>-ra</i>	<i>hubiera -do</i>	Total	% <i>-ra</i>
<b>1700-1724</b>	823	34	857	96,03
<b>1725-1749</b>	1546	23	1569	98,53
<b>1750-1774</b>	1314	82	1396	94,13
<b>1775-1799</b>	1191	170	1361	87,51
<b>1800-1824</b>	1271	218	1489	85,36
<b>1825-1849</b>	2975	497	3472	85,69
<b>1850-1874</b>	3658	655	4313	84,81
<b>1875-1899</b>	10 258	1218	11 476	89,39
<b>1900-1924</b>	5158	427	5585	92,35
<b>1925-1949</b>	7305	786	8091	90,29
<b>1950-1974</b>	8798	893	9691	90,79
<b>1975-1999</b>	42 058	3902	45 960	91,51
<b>Totales</b>	86 355	8905	95 260	

Tabla 6. Frecuencias absolutas y relativas de las formas del tipo *llegara* y *hubiera llegado* de once verbos muy frecuentes en los textos españoles de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)). [Comprobado en marzo de 2006].

6. Como es bien sabido, la estadística de las categorías y subcategorías gramaticales está siempre determinada –en un grado que a veces puede confundir– por lo que sucede con los lemas que tienen mayor frecuencia en todas sus formas. Los datos manejados hasta el momento son el resultado de sumar para cada período la frecuencia que alcanzan los once verbos considerados. Esa agrupación podría, sin duda, ocultar fuertes diferencias en el comportamiento individual de los verbos estudiados, de modo que, para llevar la posibilidad al extremo, el ascenso global fuera el resultado del aumento de las formas en *-ra* en los tres o cuatro verbos más frecuentes combinado con el mantenimiento o incluso el descenso en todos los demás, de menor frecuencia y, por tanto, de menor incidencia en la visión de conjunto. La tabla 7, que aparece a continuación, desglosa los datos correspondientes a cada uno de estos verbos en la forma del tipo *llegara* y añade la media de los porcentajes individuales, de forma que los diferentes pesos de cada uno de estos verbos quedan igualados y todos ellos tienen el mismo efecto sobre la media. Como era de esperar, el porcentaje de utilización de la forma *llegara* resulta diferente según los verbos, pero la media de los porcentajes muestra una línea que sigue la misma tendencia, con los mismos altibajos que hemos observado con los datos agrupados, de modo que lo lógico, al menos mientras no dispongamos de datos más detalla-

dos, es atribuir las diferencias entre los distintos verbos a factores no directamente vinculados a los que afectan a la redistribución de funciones con las formas del tipo *llegase*. El gráfico 3 produce, de forma más intuitiva, esta misma impresión de evolución paralela en líneas generales.

	<i>tener</i>	<i>dar</i>	<i>ser/ir</i>	<i>ver</i>	<i>hablar</i>	<i>estar</i>	<i>quedar</i>	<i>llevar</i>	<i>mirar</i>	<i>pensar</i>	Media de %
1700-1724	21,91	13,20	41,94	24,79	11,11	21,59	13,79	14,06	0,00	0,00	16,24
1725-1749	24,60	13,25	51,89	24,04	8,51	18,48	9,68	8,57	10,00	33,33	20,24
1750-1774	31,69	21,94	55,44	40,00	28,26	32,18	35,85	40,91	66,67	65,22	41,82
1775-1799	40,47	34,50	49,92	35,34	21,05	33,08	14,61	14,12	17,65	43,75	30,45
1800-1824	28,13	32,22	49,67	49,43	18,92	28,51	9,52	11,43	12,50	18,52	25,89
1825-1849	37,23	40,00	52,59	44,21	22,22	26,86	22,93	15,63	31,03	34,69	32,74
1850-1874	56,47	54,66	60,99	64,19	31,33	49,16	32,61	39,62	37,21	63,64	48,99
1875-1899	54,09	46,66	63,16	62,16	40,45	46,27	43,85	37,94	29,35	55,07	47,90
1900-1924	40,31	37,64	57,97	44,06	27,62	34,96	31,31	33,18	31,03	50,82	38,89
1925-1949	60,78	62,79	73,24	60,82	52,20	56,85	36,07	41,67	36,36	67,65	54,84
1950-1974	66,15	65,06	74,13	67,80	50,83	60,64	58,75	54,87	36,84	70,97	60,60
1975-1999	81,19	86,07	83,92	80,82	71,25	76,63	69,55	71,02	54,31	78,51	75,33

Tabla 7. Frecuencias relativas de las formas del tipo *llegara* de los once verbos muy frecuentes en textos españoles de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)). [Comprobado en marzo de 2006].

## RA España

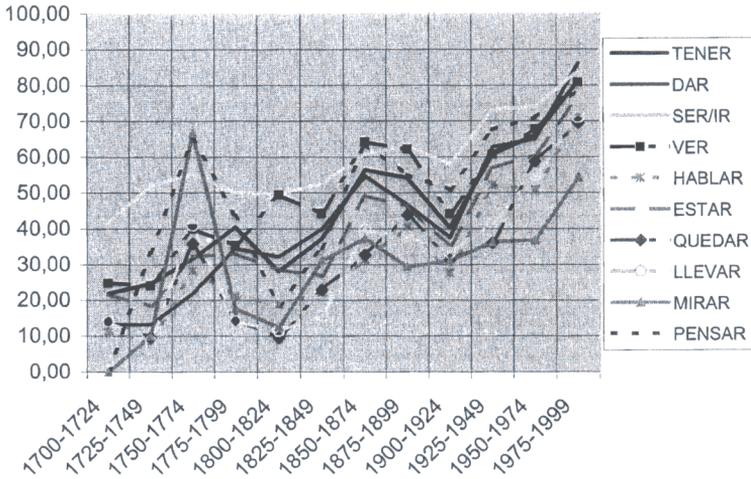


Gráfico 3. Frecuencias relativas de las formas del tipo *llegara* de los once verbos muy frecuentes en textos españoles de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)).

7. En un trabajo anterior, al que ya he hecho alusión (Rojo 1996), afirmaba que los datos obtenidos en el análisis de las frecuencias de estas cuatro formas en la BDS no permitían confirmar la validez de las diferentes hipótesis parciales manejadas para tratar de explicar las diferencias en su uso y que los factores individuales –cuyo efecto era reforzado por lo relativamente escaso de la muestra estudiada– parecían tener un peso superior al deseado. En cambio, sí daban pie a seguir manteniendo que el proceso de sustitución de las formas en *-ra* por las formas en *-se*, muy visible en cualquier enfoque general, estaba más avanzado en América que en España, como puede observarse en la tabla 8.

	Simple	Compuesta
<b>España</b>	79,78	83,06
<b>América</b>	92,39	96,39

Tabla 8. Porcentajes correspondientes a las formas del tipo *llegara* y *hubiera llegado* en España y América según los datos de la BDS. Fuente: BDS (www.bds.usc.es). Elaboración propia (cf. Rojo 1996: 683).

Con la intención de profundizar en el estudio de esas diferencias, he aplicado a los textos mexicanos contenidos en CORDE y CREA el mismo procedimiento de obtención de datos utilizado para los españoles. Los correspondientes a las formas simples aparecen en la tabla 9.

	-ra	-se	total	% -ra	% -se
<b>1700-1724</b>	27	2	29	93,10	6,90
<b>1725-1749</b>	74	30	104	71,15	28,85
<b>1750-1774</b>	3	6	9	33,33	66,67
<b>1775-1799</b>	128	322	450	28,44	71,56
<b>1800-1824</b>	183	72	255	71,76	28,24
<b>1825-1849</b>	79	17	96	82,29	17,71
<b>1850-1874</b>	105	77	182	57,69	42,31
<b>1875-1899</b>	245	54	299	81,94	18,06
<b>1900-1924</b>	126	49	175	72,00	28,00
<b>1925-1949</b>	272	81	353	77,05	22,95
<b>1950-1974</b>	747	107	854	87,47	12,53
<b>1975-1999</b>	7459	1222	8681	85,92	14,08
<b>Totales</b>	9448	2039	11 487		

Tabla 9. Frecuencias absolutas y relativas de las formas del tipo *llegara* y *llegase* de once verbos muy frecuentes en textos mexicanos de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus (www.rae.es). [Comprobado en marzo de 2006].

Salta a la vista que la evolución de las frecuencias relativas de estas dos formas resulta bastante más irregular que la que hemos podido observar en las españolas. Sin duda, el escaso número de textos existentes en algunos de los tramos establecidos distorsiona una tendencia que no debería seguir un camino muy alejado del que hemos podido observar en España. En el período correspondiente al CREA, donde el número de casos es suficientemente amplio, la distribución que encontramos no resulta demasiado diferente de la que se ve en España: el porcentaje de *llegara* en los textos mexicanos es solo 3,6 puntos porcentuales superior al

que presenta en los textos españoles y se queda casi 6,5 puntos porcentuales por debajo del que produce el conjunto de los textos americanos –incluidos los mexicanos– que figuran en la BDS. Están más lejos todavía de los obtenidos por Moreno de Alba en su trabajo sobre las formas verbales en las encuestas realizadas para el estudio de la norma culta mexicana. Sobre un total de 210 casos, “solo siete (3,3 %) fueron pretéritos en *-se*, de los cuales, cuatro corresponden a un mismo informante, cuya voz fue grabada durante una conferencia, estilo evidentemente más formal [...]. En el corpus adicional, de 63 apariciones de pretérito de subjuntivo, cuatro fueron formas en *-se* (6,3 %), todas ellas pertenecientes a un informante que se caracteriza por un estilo de habla poco natural”<sup>11</sup>.

La diferencia entre ‘lengua escrita’ y ‘lengua oral’ oculta un conjunto complejo de elementos diversos, pero, tomada en general, puede ser, sin duda, uno de los factores que intervengan en la distribución de frecuencias relativas de estas dos formas, puesto que, como se señala en la cita de Moreno de Alba, las formas en *-se* suelen ser consideradas más elegantes, más cultas. Aprovechando la posibilidad que brinda el CREA de hacer búsquedas selectivas por tipos de textos, en la tabla 10 incluyo los datos correspondientes a las frecuencias relativas de las formas *tuviera(n)* y *tuviese(n)* en los textos mexicanos del CREA. Es evidente que no se les puede conceder demasiada fiabilidad, puesto que se trata únicamente de algunas de las formas de un solo verbo, pero no parece casual la marcada gradación que se observa.

	<i>tuviera(n)</i>	<i>tuviese(n)</i>	total	% <i>-ra</i>	% <i>-se</i>
<b>Textos orales</b>	59	4	63	93,65	6,35
<b>Periódicos</b>	66	5	71	92,96	7,04
<b>Revistas</b>	78	10	88	88,64	11,36
<b>Libros</b>	472	152	624	75,64	24,36
<b>Totales</b>	675	171	846		

Tabla 10. Frecuencias absolutas y relativas de las formas *tuviera(n)* y *tuviese(n)* en los diferentes tipos de textos mexicanos del CREA. Elaboración propia a partir de los datos obtenidos en el corpus ([www.rae.es](http://www.rae.es)). [Comprobado en agosto de 2007].

<sup>11</sup> Moreno de Alba 1978: 148-149; cf. también Moreno de Alba 2001: 130.

Con las mismas precauciones señaladas, y sólo para que sea posible adelantar una comparación de datos, en la tabla 11 figuran los correspondientes a textos españoles. Las diferencias entre los distintos tipos son bastante menores y se produce el curioso y un tanto inesperado hecho de que la forma en *-se* es mucho más frecuente en los textos orales, periódicos y revistas en España, mientras que el porcentaje de *-se* en libros resulta más alto en libros mexicanos que en libros españoles.

	<i>tuviera(n)</i>	<i>tuviese(n)</i>	total	% <i>-ra</i>	% <i>-se</i>
<b>Textos orales</b>	187	30	217	86,18	13,82
<b>Periódicos</b>	1142	228	1370	83,36	16,64
<b>Revistas</b>	174	33	207	84,06	15,94
<b>Libros</b>	2580	631	3211	80,35	19,65
<b>Totales</b>	4083	922	5005		

Tabla 11. Frecuencias absolutas y relativas de las formas *tuviera(n)* y *tuviese(n)* en los textos españoles del CREA. Elaboración propia a partir de los datos obtenidos en el corpus ([www.rae.es](http://www.rae.es)). [Comprobado en agosto de 2007].

Puede pensarse, pues, que las frecuencias generales de las formas en *-ra* y en *-se* en los textos mexicanos actuales (y no sólo actuales) pueden estar perturbadas por la incidencia que en el tipo de texto tiene en la elección de la variante. Como es fácil suponer, los datos correspondientes a las formas compuestas muestran una oscilación todavía mayor. Los datos del CREA, que son los que tienen un mayor número de casos, son muy próximos a los que hemos visto en las formas simples, pero muestran un porcentaje inferior para *hubiera llegado*, en contra de lo que se da en el conjunto de la BDS.

	<i>hubiera</i> -do	<i>hubiese</i> -do	total	% <i>hubiera</i> -do	% <i>hubiese</i> -do
1700-1724	0	0	0	--	--
1725-1749	0	0	0	--	--
1750-1774	0	0	0	--	--
1775-1799	24	10	34	70,59	29,41
1800-1824	5	2	7	71,43	28,57
1825-1849	8	0	8	100,00	0,00
1850-1874	9	7	16	56,25	43,75
1875-1899	11	5	16	68,75	31,25
1900-1924	5	6	11	45,45	54,55
1925-1949	20	7	27	74,07	25,93
1950-1974	35	4	39	89,74	10,26
1975-1999	673	136	809	83,19	16,81
<b>Totales</b>	<b>790</b>	<b>177</b>	<b>967</b>		

Tabla 12. Frecuencias absolutas y relativas de las formas del tipo *hubiera llegado* y *hubiese llegado* de once verbos muy frecuentes en textos mexicanos de CORDE y CREA. Elaboración propia a partir de los datos obtenidos en los corpus ([www.rae.es](http://www.rae.es)). [Comprobado en marzo de 2006].

8. No parece, pues, que las diferencias entre los porcentajes de España y América, a los que tantas veces se ha aludido, estén suficientemente claras en lo que se refiere al contraste de los datos españoles con los mexicanos. Con intención de dar las líneas generales de la situación actual, he obtenido también las frecuencias de las formas simples y compuestas correspondientes al verbo *tener* en los textos del CREA en los diferentes países hispánicos. Dado que se trata únicamente de un verbo, aunque muy frecuente, las cifras resultantes y los porcentajes de uso derivados no son más que una aproximación, todavía muy superficial, a lo que un estudio más profundo deberá poner de manifiesto. Teniendo eso en cuenta, creo que los datos reflejados en la tabla 13, que se mueven entre el 80 % de Argentina y el 93,48 % de Estados Unidos, muestran, de nuevo, que la situación es muy cambiante. Cuatro países (Colombia, Estados Unidos, Nicaragua y Panamá) presentan porcentajes superiores al 90 % para *tu-viera*, diez se sitúan entre el 85 % y el 89,9 % y siete lo hacen entre el 80 % y el 84,62 %. Nótese, además, que Argentina, México y El Salvador presentan porcentajes inferiores a los que se encuentran en España. Ya en pruebas este trabajo, veo que Pitloun (2006) obtiene, para el habla culta de San José de Costa Rica, unos porcentajes de uso que, como era de esperar, están mucho

más próximos a los paralelos de México mencionados antes que a los que se obtienen del CREA. Las formas en *-ra* alcanzan en estos materiales orales el 98,7 % de los 232 casos registrados para ambas variantes.<sup>12</sup> En opinión de este autor, el uso de la forma en *-se* “implica, por lo menos en la variación idiolectal, menos compromiso por parte del hablante y/o mayor grado de duda o de hipótesis”.<sup>13</sup>

	<i>tuviera</i> (n)	<i>tuviese</i> (n)	total	% <i>-ra</i>	% <i>-se</i>
<b>Argentina</b>	500	125	625	80,00	20,00
<b>Bolivia</b>	45	7	52	86,54	13,46
<b>Chile</b>	352	60	412	85,44	14,56
<b>Colombia</b>	231	23	254	90,94	9,06
<b>Costa Rica</b>	59	8	67	88,06	11,94
<b>Cuba</b>	211	47	258	81,78	18,22
<b>Ecuador</b>	33	6	39	84,62	15,38
<b>El Salvador</b>	13	3	16	81,25	18,75
<b>España</b>	4165	945	5110	81,51	18,49
<b>Estados Unidos</b>	86	6	92	93,48	6,52
<b>Guatemala</b>	52	7	59	88,14	11,86
<b>Honduras</b>	24	3	27	88,89	11,11
<b>México</b>	685	171	856	80,02	19,98
<b>Nicaragua</b>	46	5	51	90,20	9,80
<b>Panamá</b>	24	2	26	92,31	7,69
<b>Paraguay</b>	19	4	23	82,61	17,39
<b>Perú</b>	183	21	204	89,71	10,29
<b>Puerto Rico</b>	111	19	130	85,38	14,62
<b>Rep. Dominicana</b>	50	7	57	87,72	12,28
<b>Uruguay</b>	100	17	117	85,47	14,53
<b>Venezuela</b>	282	43	325	86,77	13,23
<b>Totales</b>	7271	1529	8800		

Tabla 13. Frecuencias absolutas y relativas correspondientes al pretérito de subjuntivo de *tener* en textos del CREA pertenecientes a diferentes países. Elaboración propia a partir de los datos obtenidos en el corpus ([www.rae.es](http://www.rae.es)). [Comprobado en agosto de 2007].

<sup>12</sup> Cf. Pitloun 2006: 345.

<sup>13</sup> Pitloun 2006: 370. Para más datos sobre la distribución de estas dos formas en el habla culta de diversas ciudades hispanas, cf. DeMello 1993.

Aunque su fiabilidad es todavía mucho menor por lo escaso de muchas de las frecuencias halladas, incluyo también los datos correspondientes a las formas compuestas de estos verbos en la tabla 14.

	<i>hubiera(n)</i> tenido	<i>hubiese(n)</i> tenido	Total	% -ra	% -se
<b>Argentina</b>	58	22	80	72,50	27,50
<b>Bolivia</b>	6	2	8	75,00	25,00
<b>Chile</b>	39	12	51	76,47	23,53
<b>Colombia</b>	34	5	39	87,18	12,82
<b>Costa Rica</b>	13	2	15	86,67	13,33
<b>Cuba</b>	24	2	26	92,31	7,69
<b>Ecuador</b>	5	1	6	83,33	16,67
<b>El Salvador</b>	0	3	3	0,00	100,00
<b>España</b>	517	121	638	81,03	18,97
<b>Estados Unidos</b>	12	2	14	85,71	14,29
<b>Guatemala</b>	6	3	9	66,67	33,33
<b>México</b>	1	19	20	5,00	95,00
<b>Honduras</b>	91	4	95	95,79	4,21
<b>Nicaragua</b>	5	2	7	71,43	28,57
<b>Panamá</b>	2	0	2	100,00	0,00
<b>Paraguay</b>	3	0	3	100,00	0,00
<b>Perú</b>	24	5	29	82,76	17,24
<b>Puerto Rico</b>	6	4	10	60,00	40,00
<b>Rep. Dominicana</b>	4	0	4	100,00	0,00
<b>Uruguay</b>	8	2	10	80,00	20,00
<b>Venezuela</b>	36	10	46	78,26	21,74
<b>Totales</b>	894	221	1115		

Tabla 14. Frecuencias absolutas y relativas correspondientes al pretérito pluscuamperfecto de subjuntivo de *tener* en textos del CREA pertenecientes a diferentes países. Elaboración propia a partir de los datos obtenidos en el corpus ([www.rae.es](http://www.rae.es)). [Comprobado en agosto de 2007].

9. A pesar de las evidentes distorsiones creadas por los problemas de la lematización, el efecto de las homografías, la amplitud de la organización temporal en la devolución de datos, la limitación a unos pocos verbos o la escasez de textos en algunos

períodos y países, creo que los resultados obtenidos tienen validez general y nos proporcionan un panorama relativamente claro de lo que ha sucedido y sucede actualmente con la frecuencia de uso de estas dos formas. El aumento de las formas en *-ra* es muy nítido desde comienzos del siglo XVIII, aunque hay algún quiebro en la línea ascendente que, dado que se produce en casi todos los verbos considerados, no parece ser el simple efecto de un sesgo en los materiales del CORDE correspondientes a ese período.

En la actualidad, el predominio de las formas en *-ra* es muy marcado en todos los países, pero no se aprecia una diferencia fuerte entre España y América. Es muy probable, en cambio, que la consideración de 'forma culta' para la variante *-se*, reflejada en el tipo de texto, tenga una incidencia realmente importante.

### Referencias bibliográficas

- CARTAGENA, NELSON (1978): "Acerca de las categorías tiempo y aspecto en el sistema verbal del español", *RSEL*, 8/2, 373-408.
- CARTAGENA, NELSON (1999): "Los tiempos compuestos", cap. 45 de Bosque, Ignacio y Violeta Demonte (eds.) *Gramática descriptiva de la lengua española*, Madrid: Espasa-Calpe, 2935-2975.
- DAVIES, MARK (2005): "The advantage of using relational databases for large corpora: Speed, advanced queries, and unlimited annotation", *IJCL*, 10/3, 307-334.
- DEMELLO, GEORGE (1993): "*-ra* vs. *-se* Subjunctive: A New Look at an Old Topic", *Hispania*, 76, 235-244.
- MORENO DE ALBA, JOSÉ G. (1978): *Valores de las formas verbales en el español de México*, México, D.F.: UNAM, 1985<sup>2</sup>.
- MORENO DE ALBA, JOSÉ G. (2001): "Valores de algunas formas verbales", en Moreno de Alba, J. *El español en América*, México, D.F.: FCE, 2001. Cito por su reed. en Moreno de Alba, J.: *Estudios sobre los tiempos verbales*, México, D.F.: UNAM, 2003, 121-132.
- PITLOUN, PETR (2006): "El uso del imperfecto de subjuntivo en el habla culta de Costa Rica", *RFE*, 86/2, 339-371.

- ROJO, GUILLERMO (1996): "Sobre la distribución de las formas *llegara* y *llegase* en español actual", en Casado Velarde, Manuel et alii (eds.): *Scripta Philologica in memoriam Manuel Taboada Cid*, Universidade da Coruña, II, 677-691.
- ROJO, GUILLERMO Y ALEXANDRE VEIGA (1999): "El tiempo verbal. Las formas simples", cap. 44 de Bosque, Ignacio y Violeta Demonte (eds.) *Gramática descriptiva de la lengua española*, Madrid: Espasa-Calpe, 2867-2934.
- VEIGA, ALEXANDRE (1996): *La forma verbal española cantara en su diacronía*, Univ. de Santiago de Compostela.
- VEIGA, ALEXANDRE (2006): "Las formas verbales subjuntivas. Su reorganización modo-temporal", en Company Company, Concepción (dir.): *Sintaxis histórica de la lengua española*, México, D.F.: FCE, I, 93-240.

#### **Direcciones electrónicas**

- BDS: [www.bds.usc.es](http://www.bds.usc.es)  
CORDE y CREA: [www.rae.es](http://www.rae.es)  
Corpus del español: [www.corpusdelespanol.org](http://www.corpusdelespanol.org)