

Análisis cuantitativo del uso real de los verbos pronominales estrictos del castellano utilizando un corpus diacrónico (Google Books)¹

IRENE RENAU, ROGELIO NAZAR

Institut Universitari de Lingüística Aplicada

Universitat Pompeu Fabra, Barcelona

Correspondencia: Irene Renau

Infolex - IULA (C/Roc Boronat, 138, 08018 Barcelona)

irene.renau@upf.edu

Resumen

Este trabajo presenta un análisis cuantitativo del uso de los verbos pronominales estrictos del castellano empleando un corpus diacrónico de grandes dimensiones (una parte de Google Books). Los verbos pronominales estrictos, comunes en todas las lenguas románicas, son aquellos que no pueden prescindir del pronombre que los acompaña, como en «Se ha fugado de nuevo». Tras recoger los lemas verbales pronominales de tres diccionarios generales de lengua española, se buscan en el corpus mencionado. Los resultados apuntan a que muchos verbos marcados en las fuentes lexicográficas como pronominales inherentes tienen variedad de estructuras transitivas e intransitivas en el uso. También se observa que muchas de las formas no se encuentran, lo que revela una distancia de la representación de estos verbos en los diccionarios, con respecto a los datos empíricos.

Palabras clave: corpus diacrónico, Google Books, metalexigrafía, verbos pronominales estrictos

Quantitative Analysis of the Real Use of Spanish Strict Pronominal Verbs

Using a Diachronic Corpus (Google Books)

Summary

This paper presents a quantitative analysis of the use of Spanish strict pronominal verbs using a very large diachronic corpus (a part of Google Books). Strict pronominal verbs, common in all Romance languages, are those that cannot be used without the pronoun, as

¹ Este trabajo ha recibido una subvención de los siguientes proyectos del MEC: «Agrupación semántica y relaciones lexicológicas en el diccionario», dir.ª: J. DeCesaris (HUM2009-07588/FILO); APLE: «Procesos de actualización del léxico del español a partir de la prensa», período 2010-2012, dir.ª: M. T. Cabré (FFI2009-12188-C05-01/FILO). También ha recibido subvención de la Fundación Comillas en relación con el proyecto de *Diccionario de aprendizaje del español como lengua extranjera*.

in “Se ha fugado de nuevo” (= ‘S/He has escaped again’). After collecting the verbal pronominal lemmas of three Spanish general dictionaries, we looked for them in the mentioned corpus. The results show that a lot of verbs which appear in the lexicographical sources as strict pronominals have variety of transitive and intransitive structures in the real use. Also, it can be observed that many of the forms are not found, which reveals a divergence in the representation of these verbs in dictionaries, regarding the empirical data.

Key words: diachronic corpus, Google Books, strict pronominal verbs, theoretical lexicography

1. INTRODUCCIÓN Y OBJETIVOS

Este trabajo presenta un análisis cuantitativo del uso de los verbos pronominales estrictos del castellano empleando un corpus diacrónico de grandes dimensiones (una parte de Google Books). Los verbos pronominales estrictos (también llamados *puros*, *inherentes*, *intrínsecos* o de otros modos), comunes en todas las lenguas románicas, son aquellos que no pueden prescindir del pronombre que los acompaña:

Si de algo puedo *jactarme* es de haber trabajado con intensidad.
Se ha vuelto a escapar. *Se ha fugado* de nuevo.
Se personó en casa de los Hernández.

Los verbos *jactarse*, *fugarse* o *personarse* de estos ejemplos no entran en el español actual en construcciones sin clítico, como **Él jacta de haber trabajado con intensidad*, **Ha fugado de nuevo* o **El problema personó al hombre en casa de los Hernández*.

Se parte de la evidencia de que, en varios diccionarios actuales, los verbos pronominales estrictos han sido lematizados en algunos casos en su forma pronominal (*jactarse*) y en otros en su forma no pronominal (*jactar*). La finalidad del trabajo es estudiar estos verbos en un corpus diacrónico voluminoso que permita averiguar a qué se deben dichas divergencias, y que permita contrastar la información de las fuentes lexicográficas con datos reales de uso. Así pues, el objetivo que nos proponemos es a la vez lexicológico y metalexigráfico.

El artículo tendrá la siguiente estructura: en primer lugar, se expondrá la noción de verbo pronominal estricto tal como se ha descrito en las gramáticas, y también su

representación en tres diccionarios generales de lengua española; a continuación, se establecerán las hipótesis de las que partirá el análisis de corpus; se describirá seguidamente este análisis, con sus resultados; el trabajo se cerrará con las conclusiones y trabajo futuro.

2. LOS VERBOS PRONOMINALES ESTRICTOS EN LOS ESTUDIOS GRAMATICALES

Los verbos pronominales —es decir, aquellos que van acompañados de pronombre clítico—² existen desde los inicios de las lenguas románicas como una de las consecuencias del paso del sistema de casos latino al nuevo sistema flexivo, y han sido estudiados ya en el comienzo de la tradición gramatical española (Nebrija menciona ya en su *Gramática castellana*, de 1492, el uso del pronombre reflexivo *se* como recurso del castellano para suplir la voz media o impersonal). En concreto, aparecen ya lemas solo pronominales en el *Diccionario de Autoridades* (1726-1739), por ejemplo, *abstenerse*, del latín *abstinere*, o *dignarse*, de *dignari*.

Pese a su documentación temprana, los verbos que no pueden prescindir del pronombre no han sido específicamente muy estudiados, posiblemente por considerarse que son el final de un proceso de gramaticalización que prácticamente ha lexicalizado al pronombre, convirtiéndolo en un rasgo morfológico del verbo y, por tanto, sin características sintácticas o semánticas propias. Un estudio más detallado de estas unidades léxicas muestra sin embargo su diversidad y especificidad³. De todos modos, se presta atención a los verbos pronominales estrictos en Sánchez López (2002: 96-102), Otero (1999: 1465-1472) y la *Nueva gramática de la lengua española*, RAE (2010: 3102-3103). En estos estudios, se explica, entre otras pocas cuestiones, que acostumbran a ser inacusativos y que suelen llevar complemento de régimen (*jactarse de*, *atreverse a*, *obstinarse en...*).

La variación en estos verbos «invariables» puede ser diacrónica (*arrogar*, *atrever*, *desvivir* y otros son formas antiguas del castellano, mientras que la lengua actuar solo admite *arrogarse*, *atreverse*, *desvivirse*, etc.; véase la figura) y diatópica (como en este ejemplo de Cuba: «Al presentarse los refuerzos por él solicitados, *adentró* al ejército

² Un estudio más amplio de este tipo de verbos pronominales y de otros se encuentra en la tesis doctoral de Renau (en preparación), de la que parte el presente estudio.

³ Queda fuera del alcance del presente estudio profundizar en esta cuestión, que ha sido abordada para los verbos pronominales estrictos del catalán por Lorente (2010 y en prensa).

más allá de lo deseado»⁴; en España y otros países se emplea solo *adentrarse*). También puede haber explotaciones expresivas⁵ que conviertan un verbo empleado solo con pronombre en uno con estructura no pronominal: en un caso como «No creo que se suicidase. [...] *Le suicidaron*»⁶, el verbo es explotado sintácticamente para sostener un cambio semántico (‘matar a alguien simulando un suicidio’).

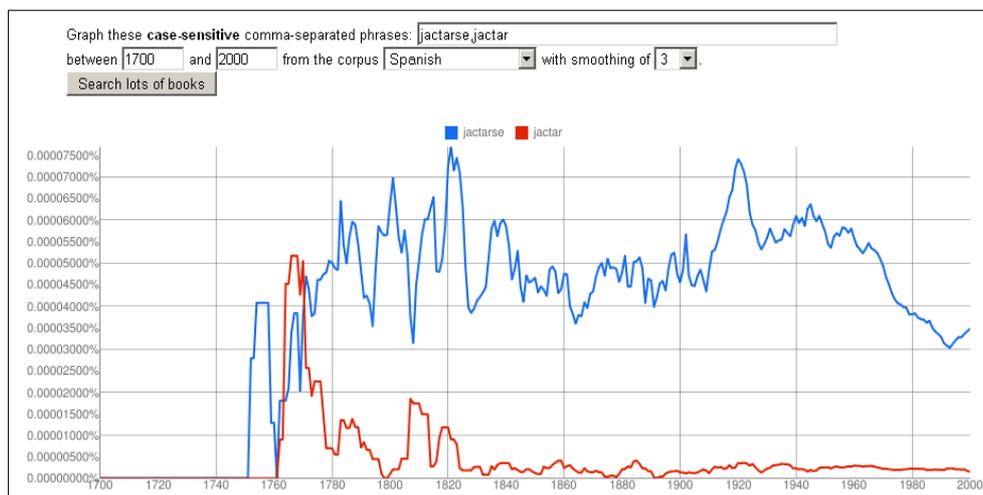


Figura. Gráfico ofrecido por Google Books N-gram Viewer (<http://ngrams.googlelabs.com>) que compara el uso de la forma *jactar* y *jactarse* desde el siglo XVIII hasta la actualidad.

2. REPRESENTACIÓN LEXICOGRÁFICA

Con el fin de observar la representación lexicográfica de los verbos pronominales estrictos, se tomaron tres diccionarios generales del español actual que tuvieran versión en CD-ROM y opciones de búsqueda avanzada: el *Diccionario de la lengua española* (DRAE), de la Real Academia Española; el *Diccionario de uso del español* (DUE), de María Moliner; y el *Diccionario de uso del español de América y España* (DUEAE), dirigido por Paz Battaner.

Se buscaron todos los lemas verbales que terminaran en *-se*, y se extrajeron manual o automáticamente. Los tres listados se compararon mediante un procedimiento también automático, y se obtuvo un listado de 1.002 lemas en total. La tabla 1 muestra el número de lemas pronominales de cada diccionario empleado en comparación con el total de lemas verbales. Se observa que constituyen una parte ínfima del lemarío verbal.

⁴ Ejemplo del Corpus del Español, de M. Davies (<http://www.corpusdelespanol.org>).

⁵ Para la noción de explotación frente a norma en el uso de la lengua, nos basamos en la *Theory of Norms and Exploitations* de Hanks (2004 y en prensa).

⁶ Tomado del Spanish Web Corpus, disponible en Sketch Engine (<http://the.sketchengine.co.uk>).

Tabla 1. Número de lemas pronominales en relación con el total de lemas verbales de los tres diccionarios empleados para el análisis.

| | Lemas verbales | Lemas verbales prnls. | % |
|--------------|----------------|-----------------------|------|
| DRAE | 12.017 | 760 | 0,70 |
| DUE | 10.928 | 587 | 0,59 |
| DUEAE | 6.653 | 335 | 0,55 |

En la tabla 2 se muestran los cruces entre obras.

Tabla 2. Número de lemas pronominales en el DRAE, el DUE y el DUEAE, y coincidencias de estos entre las tres obras.

| | n | % |
|---------------------------|--------------|--------------|
| DRAE | 760 | 75,85 |
| DUE | 587 | 58,58 |
| DUEAE | 335 | 33,43 |
| DRAE + DUE + DUEAE | 135 | 13,47 |
| DRAE + DUE + no DUEAE | 350 | 34,93 |
| DRAE + DUEAE + no DUE | 39 | 3,89 |
| DUE + DUEAE + no DRAE | 21 | 2,10 |
| DRAE + no DUE + no DUEAE | 236 | 23,55 |
| DUE + no DRAE + no DUEAE | 81 | 8,08 |
| DUEAE + no DRAE + no DUE | 140,00 | 13,97 |
| DRAE o DUE o DUEAE | 1.002 | 100 |

Puede observarse que las coincidencias mayores se observan entre DRAE y DUE. Por otro lado, es muy reducido el número de lemas coincidentes ($n = 135$, 13,47 %). En tercer lugar, un dato que no muestra la tabla es que, de los 1.002 lemas, 267 (26,65 %) tienen un equivalente no pronominal en alguna de las otras dos obras; por ejemplo, *jactarse* (DUEAE) se representa como *jactar* en DRAE y DUE.

3. HIPÓTESIS

La variación observada en las tres fuentes lexicográficas consultadas motiva nuestro estudio de corpus. Las sospechas acerca de los distintos criterios puramente

lexicográficos como origen de las diferencias tienen indicios evidentes: por ejemplo, el DRAE y el DUE contienen lemas anticuados (marcadas con la abreviatura *ant.*), mientras que el DUEAE solo refleja el uso actual. Pero nos preguntamos si tal diversidad es solo de origen lexicográfico, o podría haber diferencias lingüísticas reales que el corpus pudiera indicar. Se realizó una pequeña búsqueda manual preliminar en distintos corpus tanto sincrónicos como diacrónicos⁷, y se vio que muchos verbos no aparecían ni una vez en ninguno de ellos, o aparecían tan pocos casos que los datos no eran concluyentes. Se decidió entonces emplear un corpus más grande.

Teniendo en cuenta los datos buscados en los diccionarios y esta pequeña exploración manual, se formularon las dos siguientes hipótesis:

- Hipótesis 1: La gran variación en la representación lexicográfica es reflejo de variación en el uso. Existen verbos que no son realmente pronominales estrictos pese a estar representados de este modo en los diccionarios.
- Hipótesis 2: Verbos considerados estrictamente pronominales son más frecuentes en forma de participio o tienen concomitancias con la estructura *estar + participio*.

Se explica a continuación el análisis de corpus realizado para contestar estas dos hipótesis.

4. ANÁLISIS DE CORPUS

4.1. Descripción del corpus empleado

Se empleó un corpus de enagramas (de 1 a 5 enagramas) extraído de Google Books y puesto recientemente a libre disposición de la comunidad científica (Michel, Shen, Aiden, Veres, Gray, Google Books Team, et al., 2011). Este corpus está formado aproximadamente por 5 millones de libros publicados en varios idiomas desde 1500 hasta 2008 y posteriormente digitalizados, y las unidades léxicas que se muestran en los

⁷ Se buscaron los 23 primeros verbos del listado alfabético en el Corpus del Español (cit.); el Corpus Diacrónico del Español, CORDE (RAE), <http://corpus.rae.es/cordenet.html>; el Spanish Web Corpus (cit.); y el Corpus de Referencia del Español Actual, CREA (RAE), <http://corpus.rae.es/creanet.html>.

enigramas han de haber aparecido al menos 40 veces. La parte de lengua española consta de 45.000 millones de palabras, lo que convierte a este corpus en el más grande jamás compilado para este idioma⁸. En concreto, para nuestros objetivos, un período tan amplio (1500-2008) salva la diferencia de atención de los tres diccionarios en cuanto a datos históricos.

4.2. Metodología

El corpus de Google es en realidad un índice de enigramas, es decir, una serie de tablas con secuencias de hasta cinco palabras, con indicación de frecuencia de aparición en cada año. Este corpus no está lematizado, por lo tanto, los verbos tuvieron que buscarse por cada una de sus formas por medio de un procedimiento automatizado. Para obtener las formas de los lemas verbales se empleó la base de datos morfológica OSLIN (Janssen 2005). De los enigramas recuperados con las formas verbales, se contabilizaron como pronominales aquellos en los que aparecía un pronombre enclítico o en concordancia con la forma verbal y como no pronominales el resto de los enigramas, excepto las formas en participio, que se separaron en una clase distinta. Se ignoraron casos considerados ambiguos, como aquellos en los que existe una coincidencia formal con otra categoría gramatical (como el sustantivo en *persona, fuga, cautela, hernia*, etc.). Se ignoraron también las combinaciones con verbos de soporte, los casos en los que el verbo analizado aparecía en posición inicial en el enigma (excepto aquellos con enclíticos) y aquellos en los que aparecía un segundo verbo antes del verbo analizado. Por conveniencia práctica, se dividió el corpus en tres períodos: 1500-1870, 1871-1970 y 1971-2008, y los valores se expresaron en frecuencia relativa para compensar las diferencias en tamaño de las partes. El resultado del proceso es una matriz con los 1.002 verbos que registra su distribución de frecuencia en el tiempo en las clases *pronominal, no pronominal y participio*. En todos los casos, se calculó la significación estadística de las diferencias entre pronominal y no pronominal en el último período (test binomial), porque se trata de la franja que más interesa a nuestro estudio, y se ordenaron los verbos dividiendo su frecuencia como pronominales por su frecuencia como no pronominales.

⁸ Para lengua inglesa se están empezando a compilar corpus de grandes dimensiones, como el de Pomikálev, Rychly y Kilgarrif (2009). Por el momento, en español se dispone, entre otros, del Corpus del Español, de 100.000.000 de palabras; el CREA, de 160.000.000; y el CORDE, de 250.000.000.

5. RESULTADOS

Los resultados de los tres períodos en que se dividieron los datos se resumen en la tabla 3.

Tabla 3. Número y porcentaje de formas pronominales, no pronominales y participios de los tres períodos en que se dividió el análisis.

| Intervalos porcentuales | 1500-1870 | | | 1871-1970 | | | 1971-2008 | | |
|-----------------------------|-----------|-----------|-------|-----------|-----------|-------|-----------|-----------|-------|
| | Prnls. | No prnls. | Part. | Prnls. | No prnls. | Part. | Prnls. | No prnls. | Part. |
| 90-100 % | 68 | 100 | 26 | 105 | 148 | 37 | 116 | 153 | 37 |
| 80-89 % | 27 | 35 | 9 | 24 | 20 | 4 | 20 | 20 | 7 |
| 70-79 % | 19 | 17 | 12 | 22 | 17 | 11 | 16 | 18 | 9 |
| 60-69 % | 11 | 11 | 9 | 13 | 13 | 11 | 12 | 24 | 14 |
| 50-59 % | 21 | 19 | 12 | 15 | 17 | 14 | 8 | 11 | 13 |
| 40-49 % | 14 | 13 | 11 | 14 | 15 | 10 | 10 | 7 | 10 |
| 30-39 % | 10 | 7 | 11 | 11 | 12 | 12 | 25 | 13 | 11 |
| 20-29 % | 11 | 19 | 11 | 17 | 23 | 22 | 18 | 16 | 19 |
| 10-19 % | 18 | 19 | 17 | 16 | 22 | 28 | 16 | 20 | 21 |
| 0,1-9 % | 23 | 45 | 48 | 33 | 51 | 51 | 31 | 59 | 59 |
| 0,00 (sí OSLIN / No GBooks) | 611 | 548 | 667 | 563 | 495 | 633 | 561 | 492 | 633 |
| No OSLIN | 169 | 169 | 169 | 169 | 169 | 169 | 169 | 169 | 169 |

En ella se muestra el número de verbos que tienen formas pronominales, no pronominales o en participio en los rangos porcentuales indicados en la primera columna. Por ejemplo, en el primer período (1500-1870) hay 68 verbos que tienen un 90-100% de formas pronominales; hay 100 verbos que tienen un 90-100% de formas no pronominales; y hay 26 verbos que tienen un 90-100% de formas en participio. Así, los valores de cada celda de la misma fila no se refieren a los mismos verbos.

5.1. Porcentaje de verbos pronominales estrictos

En los tres períodos se observa que la mayoría de los verbos admiten con frecuencia usos no pronominales. Parece normal que, tal como se ha indicado en el apartado 2, haya un cierto porcentaje de desviación con respecto a la norma general por motivos expresivos o por errores. Pero más allá de un 20% (porcentaje que se toma como

umbral arbitrario) de casos de desviación por motivos expresivos (o errores de uso u otras circunstancias), se considera que no puede hablarse de forma tajante de «verbo pronominal estricto» (al menos si se contempla un corpus diacrónico de unos 500 años) en muchos de los casos objeto de estudio. Se muestran en la tabla 4 los mismos datos que los de la tabla 3, separando los datos en dos franjas porcentuales según este 20 % indicado.

Tabla 4. Número de verbos que aparecen en forma pronominal divididos en los rangos porcentuales que se indican en la primera columna. Se han dividido los porcentajes de la tabla 3 en dos bloques desiguales (80-100 % y 0,1-79 %) para mostrar que los verbos del primer bloque cumplen aproximadamente el rasgo de presentarse solo como pronominales (dejando de lado errores, variaciones expresivas y otras circunstancias del uso), mientras que los del segundo no.

| | 1500-1870 | | 1871-1970 | | 1971-2008 | |
|--------------------------------|-----------|----------|-----------|----------|-----------|----------|
| Intervalos porcentuales | n | % | n | % | n | % |
| 80-100 % | 95 | 9,48 | 129 | 12,87 | 136 | 13,57 |
| 0,1-79 % | 127 | 12,67 | 141 | 14,07 | 136 | 13,57 |
| 0,00 (sí OSLIN / No GBooks) | 611 | 60,98 | 563 | 56,19 | 561 | 55,99 |
| No OSLIN | 169 | 16,87 | 169 | 16,87 | 169 | 16,87 |

Puede observarse que, en los tres períodos históricos, el porcentaje de verbos mayoritariamente pronominales (80-100 %) es menor que aquellos que pueden encontrarse también en estructuras transitivas o intransitivas no pronominales (0,1-79 %). Verbos que se encuentran en la primera franja mencionada son *abstenerse*, *cerciorarse*, *entrometerse*, *esmerarse*, *obstinarse*, *prosternarse*, *quejarse*, *retrepase* y otros. En cambio, *aburguesarse*, *desdibujarse*, *fossilizarse*, *gramaticalizarse*, etc., se encuentran entre los casos con numerosas concordancias no pronominales, como en estas frases de Google Books:

La Revolución *aburguesó*, hasta cierto punto, la sociedad francesa.
 Estos dos universos *desdibujan* los límites de las distinciones culturales.
 ...aislamiento cultural de los españoles, que *fossilizó* la vida intelectual del país.
 Las lenguas codifican o *gramaticalizan* rasgos del contexto o evento del habla.

5.2. Porcentaje de verbos con usos en participio

En la tabla 3 se observa que existen verbos que los diccionarios categorizan como pronominales y que solo se han encontrado en su forma de participio, o que tienen numerosos casos de esta forma no personal. Son muy frecuentes los participios en casos como estos (tomados de Google Books):

En los delirios de mi mente *afiebrada*...
Una sonrisa maliciosa despuntó de sus labios *amoratados*.
Las capas *desclasadas* de nuestra sociedad...
Un gentío *endomingado* y devoto...

Resulta difícil o imposible averiguar la pronominalidad del verbo si este solamente se expresa como participio. Además, teniendo en cuenta especialmente los datos del período 1500-1870, si en esta franja temporal ya se encuentran casos de verbos que solamente se expresan en forma de participio, es razonable preguntarse si se trata de verbos defectivos, o simplemente de adjetivos con forma de participio, como es frecuente que ocurra⁹.

5.3. Verbos no encontrados en el corpus

Es llamativa la gran cantidad de verbos de los que no se ha hallado ninguna concurrencia en el corpus¹⁰. Si nos limitamos a la tabla 4, se observa que más de la mitad de los verbos, en los tres períodos, no se han encontrado en ninguna de sus formas (se dejan de lado los 169 casos que el conjugador empleado no tenía). Se trata de casos como *achaplinarse*, *amezquinarse*, *chimpilinearse*, *desamotinarse*, *encaratularse*, *salmuerarse*, *trasconejarse* y muchos otros. Consideramos este el resultado más relevante en cuanto al análisis metalexigráfico.

6. CONCLUSIONES Y TRABAJO FUTURO

En este estudio se ha realizado un análisis de corpus de los 1.002 verbos que, en DRAE, DUE o DUEAE, aparecen lematizados en su forma pronominal y son considerados, por

⁹ El DUEAE, como cualquier otro diccionario, recoge muchos lemas que son formas en participio, no asociadas por tanto a ningún lema verbal: *abuhardillado*, *alocado*, *disparatado*, *endemoniado*, etc.

¹⁰ Se recuerda que Google Books no ha incluido en su corpus ninguna forma que haya aparecido menos de 40 veces.

tanto, como estrictamente pronominales. Se había partido de un acercamiento previo a las tres obras y a un pequeño análisis manual de corpus, y ya se habían detectado divergencias de criterio, así como ausencia de datos empíricos. Los datos del análisis extraídos del corpus de enigramas de Google Books corroboran las intuiciones iniciales, formuladas concretamente en las dos hipótesis del apartado 3. Consideramos que:

- Se confirma la hipótesis 1, según la cual la gran variación en la representación lexicográfica es reflejo de variación en el uso. Muchos verbos no son realmente pronominales estrictos, sino que pueden entrar en construcciones no pronominales (transitivas o intransitivas), además del uso pronominal (generalmente intransitivo).
- Se confirma la hipótesis 2, según la cual verbos considerados estrictamente pronominales son más frecuentes en forma de participio o tienen concomitancias con la estructura *estar + participio*. Muchos verbos solo se emplean en participio en esta forma, de modo que incluso queda en interrogante su estatus de verbos. Otros aparecen tan frecuentemente en esta forma que resulta difícil confirmar su naturaleza estrictamente pronominal.

Los motivos por los cuales existen estas divergencias en la representación lexicográfica sobrepasan los límites de este estudio, pero pueden estar relacionados con la escasa documentación de que se ha dispuesto hasta hace poco y con la metodología básicamente introspectiva con que se han redactado los diccionarios también hasta hace poco.

Teniendo en cuenta los datos recogidos, se extraen las siguientes conclusiones:

- La especialización morfológica que implica un verbo pronominal estricto es menos frecuente que lo esperado. Como final de un proceso de gramaticalización que a veces puede durar siglos (véase el apartado 2 para ejemplos de variación diacrónica), es corriente que no fragüe en una forma exclusivamente con pronombre, sino que pueda desglosarse en otras variantes diatéticas, que el propio idioma ya contempla como un modo de explotación de los verbos. Si se tiene *aburguesarse*, será probable que pueda existir *aburguesar*, con más o menos frecuencia de uso.

- Es necesario que el lecionario de los diccionarios generales se ajuste a los datos empíricos, y no se alimente de datos que, a través de los siglos, han ido pasando de un diccionario a otro sin contrastarse con el uso real.
- El uso de corpus de grandes dimensiones arroja una nueva luz sobre el estudio de los fenómenos léxicos. En palabras de baja frecuencia, caso de muchos de los verbos estudiados, poder contar con un corpus que sobrepasa en gran medida las herramientas de que se disponía hasta ahora implica un cambio cualitativo de la investigación (aun basado en un aumento simplemente cuantitativo). La lexicología y uno de sus brazos aplicados, la lexicografía, deberían tener en cuenta la posibilidad de emplear estas herramientas de libre acceso.

Este trabajo no agota las posibilidades de análisis de los datos que se han presentado, y se considera que deben dejarse para estudios posteriores, entre otras, las siguientes cuestiones:

- Averiguar, mediante labor documental, el origen de los lemas verbales que no se han hallado en el corpus.
- Contrastar la semántica de estos verbos; por ejemplo, si los significados que aparecen en el diccionario son los que el corpus ofrece como más frecuentes.
- En el mismo sentido, indagar si las formas no pronominales de los verbos tienen un significado distinto de las pronominales, o son variantes diatéticas.

Bibliografía

- Battaner, P. (Dir.^a). (2003). *Diccionario de uso del español de América y España*. Barcelona: Spes Editorial.
- Hanks, P. (2004). The Syntagmatics of Metaphor and Idioms. *International Journal of Lexicography*, 17(3), 245-274.
- Hanks, P. (En prensa). *Lexical Analysis: Norms and Exploitations*. Massachusetts: MIT Press.
- Janssen, M. Open Source Lexical Information Network. *Third International Workshop on Generative Approaches to the Lexicon*, Geneva, Switzerland.
- Pomikálek, J., Rychlý, P., Kilgarriff, A. Scaling to Billion-plus Word Corpora. *Advances in Computational Linguistics*, 41, 3-13.

- Moliner, M. (2001-2002). *Diccionario de uso del español*, 2.^a ed. Madrid: Gredos.
Versión en CD-ROM.
- Lorente, M. (2010). Verbs pronominals de font lèxica: reflexivitat i reciprocitat inherents. En I. Creus; M. Puig; J. R. Veny (Eds.), *Actes del Quinzè Col·loqui Internacional de Llengua i Literatura Catalanes* (pp. 435-445). Barcelona: Publicacions de l'Abadia de Montserrat.
- Lorente, M. (en prensa). Verbs pronominals inherents: descripció i representació lexicogràfica. M. Á. Pradilla (Ed.), *Actes del III Col·loqui Internacional «La lingüística de Pompeu Fabra»*. Tarragona, desembre del 2008. Barcelona: Institut d'Estudis Catalans.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., Aiden, E. L. Quantitative Analysis of Culture Using Millions of Digitized Books (2011). *Science* 331(6014), 176-182.
- Otero, C. P. (1999). Pronombres reflexivos y recíprocos. En I. Bosque, V. Demonte, V. (Dir.), *Gramática descriptiva de la lengua española* (pp. 1427-1517). Madrid: Espasa.
- Real Academia Española (1726-1739). *Diccionario de autoridades*. Edición en línea en *Nuevo tesoro lexicográfico de la lengua española*: <http://buscon.rae.es/ntlle/SrvltGUILoginNtlle> (última consulta: 7.05.11).
- Real Academia Española (2003). *Diccionario de la lengua española*, 22.^a ed. Madrid: Espasa. Versión en CD-ROM.
- Real Academia Española (2009). Oraciones activas, pasivas, impersonales y medias. En *Nueva gramática de la lengua española* (pp. 3037-3112). Madrid: Espasa.
- Renau, I. (En preparación). *Representación de los usos pronominales en los verbos de un diccionario de aprendizaje de español como lengua extranjera*. Tesis doctoral. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona.
- Sánchez López, C. (2002). Las construcciones con *se*. Estado de la cuestión. En C. Sánchez López (Ed.), *Las construcciones con «se»* (pp. 13-163). Madrid: Visor Libros.