

MARCOS GARCÍA SALIDO
VICTORIA VÁZQUEZ ROZAS

Scriptum Digital
1 (2012), pp. 67-84

LOS CORPUS DIACRÓNICOS COMO INSTRUMENTO PARA EL ESTUDIO DEL ORIGEN Y DISTRIBUCIÓN DE LA CONCORDANCIA DE OBJETO EN ESPAÑOL¹

MARCOS GARCÍA SALIDO (*Universidade de Santiago de Compostela*)
VICTORIA VÁZQUEZ ROZAS (*Universidade de Santiago de Compostela*)

RESUMEN:

Partimos de una investigación previa sobre la génesis y el desarrollo histórico de la concordancia de objeto en español que pone de relieve las ventajas de un enfoque basado en corpus frente a otras alternativas de análisis. Teniendo en cuenta los corpus diacrónicos de que dispone el español, discutiremos hasta qué punto resulta factible recuperar los datos precisos para un estudio de estas características y qué tipo de información (etiquetado, analizado, etc.) sería necesaria para facilitar tal tarea. Finalmente, se verá en qué medida se manifiesta la preferencia de ciertas funciones sintácticas por una codificación y un estatus informativo determinados en el corpus seleccionado y se planteará si los géneros discursivos documentados a lo largo de la historia del español son representativos en igual medida de dicha preferencia.

PALABRAS CLAVE: diacronía, anotación de corpus, concordancia de objeto

ABSTRACT:

The article is based on previous research on the genesis and diffusion of object agreement in Spanish, which highlights the advantages of a corpus-driven approach as contrasted with other lines of analysis. We take into account the Spanish historical corpora available and discuss the feasibility of retrieving the necessary data for this type of study. Furthermore, we put forward some proposals about what kind of information (tagging, parsing, etc.) could make the task easier. Lastly, in the light of corpus data, we can see to what extent certain syntactic functions favour particular coding devices and information status, and consider if the various discourse genres registered in the history of the Spanish language are equally representative of these preferences.

KEY WORDS: Diachrony, Corpus annotation, Object agreement

1. INTRODUCCIÓN

En español, la aparición en la misma cláusula de un constituyente sintáctico en función de objeto directo o indirecto y un afijo del tipo *le, les, la, lo*, etc., se ha venido equiparando desde hace ya algunos años con la relación de concordancia entre un constituyente sujeto y la desinencia verbal de persona correspondiente (cfr. Llorente y Mondéjar 1972; Bossong 1980; García-Miguel 1991; Franco 2000). Para trazar la diacronía de este patrón se ha acudido a la hipótesis formulada por Givón (1976), según la cual su origen radica en la construcción denominada *topic-shift construction* —una estructura con tema dislocado retomado anafóricamente mediante un pronombre— (cfr. Silva-Corvalán 1984)².

¹ Este estudio se ha desarrollado dentro del proyecto de investigación ESLORA, que cuenta con financiación del Ministerio de Ciencia e Innovación (FFI2010-17417). Los autores quieren hacer constar su agradecimiento a los dos revisores anónimos de la revista *Scriptum Digital* por sus sugerencias de corrección y mejora.

² Mencionamos en particular la propuesta de Givón y su adaptación al caso particular del español por parte de Silva-Corvalán por el interés que reviste su ambicioso intento de explicar de forma unitaria el origen de la concordancia de objeto y por los problemas que ello acarrea. No podemos dejar de reconocer la multitud de estudios que se han encargado de la duplicación de objetos en español (la consideren

Aunque ha alcanzado un grado de aceptación considerable, la hipótesis de Givón plantea algunos problemas. En primer lugar, para que los elementos anafóricos que retoman el constituyente dislocado se reanalicen como marcas de concordancia, Givón defiende que este tipo de construcciones ha de interpretarse como básico, al usarse fuera de los contextos a los que inicialmente estaba asociado. Pero para el español tal análisis es cuestionable, ya que en corpus actuales la construcción con objetos dislocados es minoritaria dentro del conjunto de cláusulas con objetos directos e indirectos y su uso obedece a funciones discursivas específicas (cfr. Hidalgo Downing 2003: § 5.3). Por otro lado, existen objetos concordados difícilmente relacionables con temas dislocados —objetos pospuestos seguidos de algún otro constituyente, por ejemplo— y la concordancia de objeto no está igualmente extendida en todos los tipos de objeto que cuentan con un afijo como posible codificación, un resultado inesperado si se asume que estas formas, a partir de un cierto momento, se reanalizan como marcas de concordancia de objetos plenos.

En un trabajo previo (Vázquez Rozas y García Salido 2012), en lugar de basarnos en la idea de Givón, partimos de la hipótesis de que la concordancia es simplemente el resultado de la fijación de morfemas de persona como integrantes de la morfología verbal —que lleva a la obligatorización en unos casos y, en otros, a la pérdida de distribución complementaria con objetos sintácticos—. Esta fijación es producto de la asociación frecuente entre determinados predicados y ciertas codificaciones de objeto³. A tal hipótesis nos llevaron varias observaciones.

En primer lugar, la desigual extensión de la concordancia de objeto se corresponde con ciertas preferencias en cuanto al tipo de unidad que funciona como tal. Así, los pronombres personales objeto concuerdan obligatoriamente con el predicado (1) y, fuera de este conjunto de formas, los objetos indirectos (2) concuerdan con más regularidad que los directos, cuya concordancia, en muchas variedades del español, está limitada a los casos de tematización. Esta asimetría se corresponde con la tendencia del español contemporáneo a codificar objetos indirectos y objetos de primera y segunda persona mediante afijos y objetos directos mediante constituyentes sintácticos —frases o cláusulas—, cosa que pudimos comprobar consultando la *Base de Datos Sintácticos (BDS)*⁴ en una aproximación inicial al problema.

(1)

- a. *A mí me gusta mucho cocinar (CREA) / *A mí gusta mucho cocinar*
- b. [...] este mellizo dizigótico no fue creado por el mismo padre que *me creó a mí (CREA) / *creó a mí*

(2)

- a. [...] se *le* concedió asilo político *a su actual esposo (CREA)*

concordancia o no) desde una perspectiva diacrónica (Riiho 1988a y b; Rini 1990, 1991; entre otros), aunque sea imposible reseñarlos y discutirlos aquí.

³ Una idea similar aparece esbozada en Bybee (1985: 39). Ariel (2000) también sostiene que la concordancia en muchos casos es producto de la fijación como parte de la morfología verbal de unidades pronominales en principio relativamente autónomas con respecto al verbo, pero no considera la frecuencia de uso un factor relevante en dicho proceso.

⁴ Vid. www.bds.usc.es.

- b. [...] por vez primera en aquella plaza se concedieron orejas a un diestro (CREA)
 (3)
 a. [...] -y piensa que *la última batalla* la tuve ayer [...] (CREA)
 b. Los socialistas pueden perder *la batalla* por Madrid (CREA)

	Afijo	Forma independiente	Objeto concordado
OD	13333 (21%)	49131 (77.4%)	1005 (1.6%)
OI	9727 (78.6%)	1085 (8.8%)	1554 (12.6%)

Figura 1. Codificación de objeto directo (OD) y objeto indirecto (OI) en español contemporáneo (BDS)

	Afijo	Pronombre concordado
OD	6749 (91.7%)	559 (8.3%)
OI	3505 (95%)	173 (5%)

Figura 2. Codificación de objetos de primera y segunda persona en español contemporáneo (BDS)

Tales hallazgos son indicativos de diferencias en cuanto a la accesibilidad de los referentes de los distintos tipos de objeto, pues como se ha venido observando cuanto más fácil sea el acceso del destinatario a la representación mental de una entidad determinada, tanto menos compleja será la expresión lingüística que necesite el hablante para referirla (cfr. Givón 1983 «*topic continuity*»; Ariel 1990). Se corresponden también con las previsiones defendidas por el modelo de la Estructura Argumental Preferida (cfr. Du Bois *et al.* 2003), que refleja el hecho de que en las lenguas naturales existen ciertas preferencias en cuanto a la distribución de información a través de diferentes funciones sintácticas: ciertas funciones están asociadas de manera recurrente a referentes altamente accesibles, mientras que con otras se produce la situación contraria.

Ahora bien, para confirmar la validez general de esta asociación regular entre ciertos objetos y un determinado tipo de codificación lingüística en la diacronía del español es necesario acudir a corpus que faciliten los datos pertinentes. La información necesaria para el estudio requiere, por una parte, la identificación de la estructura funcional de las cláusulas que componen el corpus o la muestra textual que sirva como base empírica, y en particular la identificación de todos los objetos directos e indirectos, tanto plenos como afijales/clíticos, incluidos en los textos; y por otra parte, la caracterización morfológica de los objetos directos e indirectos, determinada por el grado de accesibilidad de los correspondientes referentes.

Además, para sustentar la propuesta es preciso analizar los datos no en tanto «tipos» (*types*) sino como ejemplares concretos (*tokens*), es decir, atendiendo a la frecuencia de uso registrada en el corpus, pues asumimos que la concordancia de objeto en español es resultado de un proceso de fijación o rutinización de ciertas construcciones

muy frecuentes: en los contextos en que los objetos son recurrentemente codificados mediante clítico (indirectos, 1.^a y 2.^a personas), el clítico acaba por «morfologizarse» como afijo verbal, y cuando con este tipo de objetos se hace necesario un constituyente sintáctico (forma plena) no se prescinde del afijo/clítico sino que ambos elementos aparecen y la distribución complementaria entre el afijo/clítico y el objeto pleno desaparece⁵.

Así pues, para llevar a cabo el estudio propuesto con una base empírica adecuada, necesitábamos, o bien un corpus histórico suficientemente amplio y representativo de las diferentes épocas y con un cierto nivel de anotación morfológica y sintáctica, o bien un corpus «crudo» y los programas o herramientas precisas para realizar de forma automática o semiautomática los análisis gramaticales requeridos, o, como tercera opción, un corpus crudo anotado manualmente.

Finalmente, adoptamos la última alternativa y los datos confirmaron que las tendencias observadas para el español contemporáneo en cuanto a la codificación de objetos se observaban también en textos de épocas anteriores, como puede comprobarse en los siguientes gráficos.

⁵ En aquellos contextos en que se usan pronombres independientes de primera y segunda persona, la presencia de un afijo que los duplica puede resultar aparentemente contradictoria con la relación que se señala más arriba entre la codificación y la accesibilidad de un referente, como señala uno de los revisores de este trabajo. Sin embargo, la propia obligatorización de una copia afijal en este tipo de contextos (es decir, lo que entendemos que da lugar a la concordancia de objeto) impide el uso de una estructura menos compleja del tipo **vi a ti*. Las alternativas gramaticales son *te vi* y *te vi a ti*, en caso de que se necesite una forma independiente (porque esté focalizada, tematizada, coordinada, etc.). Podría defenderse que en estos casos la gramaticalización de un determinado patrón veta una solución que, en principio, se ajustaría a la correspondencia funcional entre codificación y accesibilidad. Tendríamos así un conflicto entre un patrón gramaticalizado y otro motivado funcionalmente del que sale vencedor el primero. Con todo, si comparamos las posibles soluciones de primera y segunda persona con las de tercera, no está tan claro que la obligatoriedad del afijo en las primeras produzca estructuras de mayor complejidad. Así, en *vio a los niños* tenemos una expresión más compleja que en *te vio a ti*, *me vio a mí*. Desde el punto de vista fónico, es evidente el mayor cuerpo de la primera. Esta es además más compleja desde el punto de vista del contenido ya que aporta información léxica ausente en las alternativas pronominales. Por último, la copia afijal en estas no es informativa, ya que no aporta ningún contenido que esté ausente en los pronombres independientes.

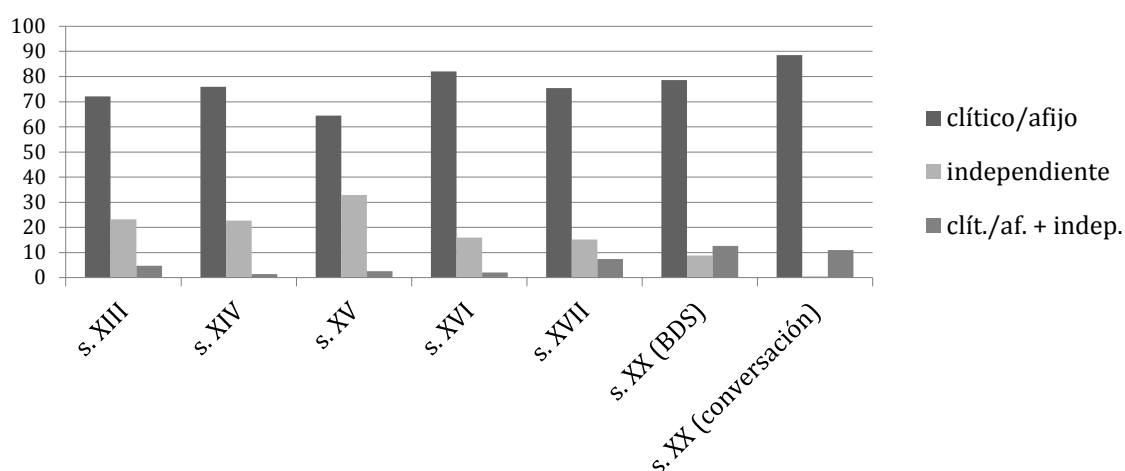


Figura 3. Diacronía de la codificación de OI

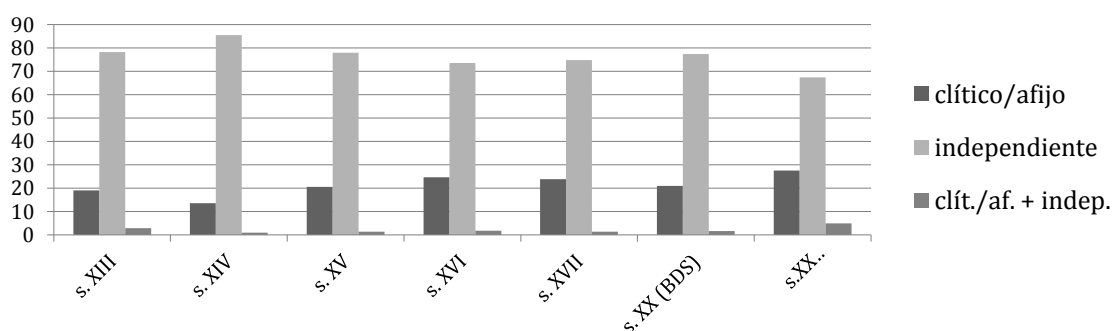


Figura 4. Diacronía de la codificación de OD

El objetivo del presente trabajo es pasar revista a las posibilidades de extracción de información que ofrecen los corpus diacrónicos del español para un estudio como el descrito y a la adecuación de diferentes tipos de textos históricos que han llegado hasta nosotros para estudiar tendencias de carácter supuestamente general en cuanto a la accesibilidad de los diferentes argumentos del predicado.

2. REVISIÓN DE CORPUS DIACRÓNICOS DISPONIBLES PARA EL ESPAÑOL

Existen actualmente varios corpus históricos en formato electrónico de los que obtener datos para estudiar cuantitativa y cualitativamente un hecho lingüístico como el que aquí nos ocupa. Otro asunto es si dichos corpus proporcionan realmente la base empírica más

adecuada para el análisis y si presentan las condiciones para que la información precisa pueda extraerse por medios automáticos o semiautomáticos.

Por lo que respecta a la primera cuestión, cualquier investigación histórica que aborde el origen y desarrollo de un cambio gramatical se ve obligada a aceptar los textos escritos como fuente legítima de datos, sin que ello impida reconocer la primacía del uso oral como motor del cambio diacrónico (cfr. Taavitsainen y Fitzmaurice 2007: 18 sigs.; Enrique-Arias 2009b).

Además de las limitaciones inherentes al trabajo con textos escritos, que es forzoso asumir, se han planteado otros inconvenientes de los corpus históricos disponibles en formato digital. Nieuwenhuijsen (2009: 371), que basa su análisis de algunos pronombrs personales en el *CORDE*, alude a las consecuencias del sistema adoptado para datar los textos en aquellos casos en que la fecha de composición de la obra no coincide con la de la copia que se reproduce en el corpus⁶.

Una dificultad añadida, conectada con la anterior, es la imposibilidad de cotejar los diferentes manuscritos de un texto, lo que puede llevar a una visión distorsionada de la evolución diacrónica de los fenómenos estudiados, como muestra Nieuwenhuijsen (2009: 374-376) mediante su análisis de las formas pronominales en tres versiones del *Libro del caballero Zifar*. No obstante, tal dificultad queda salvada en algunas colecciones —en el panorama hispánico, el corpus paralelo *Biblia medieval (BM)*—.

A las dificultades que pueden achacarse a la disponibilidad de textos para la elaboración de corpus diacrónicos hay que añadir los problemas a los que se enfrenta el investigador cuando pretende recuperar determinada información a partir de dichos corpus.

En cuanto al formato de los textos, el respeto por las opciones gráficas de los diferentes códices a través de la transcripción paleográfica resulta idóneo para investigaciones de carácter fonético-fonológico, pero se convierte en un obstáculo cuando se trata de abordar estudios en los niveles léxico, morfológico o sintáctico, como reconoce Enrique-Arias (2009c: 275) al afirmar que la variedad gráfica que se ofrece en un repertorio como la *BM* «entorpece las búsquedas automáticas en el interfaz de consulta del corpus».

La observación de Enrique-Arias enlaza con la segunda cuestión que planteábamos al inicio de este apartado: si los corpus electrónicos disponibles facilitan la obtención de la información necesaria para un estudio sintáctico sobre el desarrollo diacrónico de la concordancia de objeto en español (o para otras posibles investigaciones de carácter gramatical).

Las herramientas de recuperación de información requieren una estandarización previa de los textos cuyas implicaciones suscitan discusión. La unificación de variantes gráficas y la delimitación de unidades textuales suelen mencionarse entre las tareas ineludibles para facilitar la extracción de datos de interés lingüístico (cfr. Enrique-Arias

⁶ Nieuwenhuijsen (2009: 371) justifica su datación diferente de la ofrecida en el *CORDE* en dos obras, *Facienda* y *Calila*, ya que las versiones incluidas en dicho corpus corresponden a épocas bastante posteriores a las fechas de creación.

2009c: 272 y 275)⁷, pero ello no impide hacer hincapié en la trascendencia de las decisiones de los editores, cuyos criterios deben quedar claramente expuestos para consideración de los usuarios de los corpus (*vid.* Taavitsainen y Fitzmaurice 2007: 21-22 y 31 n. 17). En general se prefieren aquellas opciones de edición que conllevan una menor intervención del transcriptor, intervención que en cualquier caso siempre arrastra una dosis de subjetividad:

The process of editing involves decisions between alternatives from blurred and indistinct letter forms that might invite different interpretations to expanding abbreviations, marking corrupt forms and possible emendations [...]. Punctuating a medieval text is always a problem as medieval practices were very different from ours, and imposing modern ideas and practices on medieval texts is a form of interpretation. For instance, the notion of *verbatim* quotations is a modern concept, and sentence boundaries can be ambiguous (Taavitsainen y Fitzmaurice 2007: 21-22)⁸.

Otros niveles de anotación lingüística, como «la introducción, al menos parcial, de lematización e información gramatical» (Enrique-Arias 2009c: 275), son aún un desiderátum más que una realidad en la mayor parte de los casos. Y cuando se han incorporado, siquiera en parte, como ocurre en el *Corpus del español* (CE), se advierte que «para los textos medievales del corpus la codificación léxica y gramatical en la base de datos dista de ser perfecta» (*ibíd.*). Nieuwenhuijsen (2009: 376-378), por su parte, se refiere a la dificultad para localizar con procedimientos automáticos las unidades dependientes o formas ligadas y, tras comparar las opciones de búsqueda que ofrecen el *CORDE* y el *CE*, concluye que ambas alternativas son igualmente ineficaces para recuperar las formas pronominales enclíticas.

Por lo que se refiere al análisis de la concordancia de objeto, difícilmente podríamos disponer de un corpus histórico analizado sintácticamente si, como vemos, aún no está resuelto el nivel básico de anotación morfológica. La situación actual en cuanto a las posibilidades de extracción de datos lingüísticos en corpus electrónicos es sintetizada por Enrique-Arias (2009c: 276): «Tanto si se trata de textos lematizados como si no, la metodología de la lingüística a base de corpus gravita hacia la búsqueda y observación de marcadores explícitos considerando un número limitado de opciones».

Existen actualmente o están en proceso de desarrollo algunas aplicaciones informáticas que facilitan el análisis de fenómenos gramaticales en corpus electrónicos. Sin embargo, los programas mencionados en las publicaciones más recientes (Blake y Lee 2009; Nieuwenhuijsen 2009) se centran en el manejo de elementos léxicos particulares, o como mucho en variantes de unidades léxicas (los «marcadores explícitos» a los que alude Enrique-Arias 2009c), pero no logran un análisis textual con el nivel de abstracción

⁷ «[...] está previsto mejorar la versatilidad del corpus *Biblia medieval* mediante la adición de una versión con ortografía y puntuación normalizadas [...]» (*ibíd.*: 275).

⁸ Añaden una nota (*ibíd.*: 32, n. 18) con una referencia a Blockley (2001) sobre las consecuencias de la edición textual en la interpretación de las relaciones interclausulares (hipotaxis frente a parataxis).

requerido para identificar el esquema sintáctico de las cláusulas y reconocer los objetos directos e indirectos en sus varias realizaciones categoriales.

3. EL RECURSO A LA ANOTACIÓN MANUAL

El panorama que acabamos de resumir no deja otra alternativa, para el estudio que pretendíamos, que la anotación manual.

A pesar de que las publicaciones en lingüística de corpus no suelen prestar mucha atención a los métodos de anotación manual, el hecho es que esta técnica es la única posibilidad en los análisis que no parten de unos marcadores formales específicos y adoptan una perspectiva que va de la función a la forma. Los estudios de pragmática, tanto histórica como actual, entran con frecuencia en esta categoría, de ahí que las referencias a la etiquetación manual sean más comunes y más detalladas en el nivel pragmático-discursivo (cfr. Tottie y Hoffmann 2006; Kohonen 2007; Traugott y Pintzuk 2008, entre otros)⁹. Pero también requieren anotación a mano las investigaciones de carácter sintáctico que pretenden trascender las construcciones particulares de determinados elementos léxicos. Abundan en la bibliografía reciente los estudios de corpus que analizan con gran precisión el comportamiento sintagmático de unidades concretas (por ejemplo, varios de los incluidos en Enrique-Arias 2009a: Schulte 2009; Blake y Lee 2009, Alba-Salas 2009, Nieuwenhuijsen 2009, Pato 2009, etc.), y sin duda todos ellos se benefician de las herramientas que permiten extraer de forma más o menos automática la información textual necesaria.

Pero hay que señalar también los inconvenientes de un método de obtención de datos excesivamente dependiente de las formas léxicas particulares. El efecto de no tener en cuenta la estructura funcional de las unidades sintácticas puede verse en la ilustración que ofrece Davies (2009: 153-154) de las ventajas que el *CE* presenta sobre el *CORDE* para investigar hechos sintácticos. Dice Davies que para poder estudiar «the causative construction, which is composed of a form of *hacer* followed by an infinitive (*fizo llamar, haze venir, fizieron escriuir*, etc.)» (*ibid.*: 153), basta con pedir en el *CE* la secuencia «*[hacer] [vr*]*» (*ibid.*: 154), sin percatarse de otras posibilidades secuenciales perfectamente posibles como realizaciones de la construcción causativa *hacer* + infinitivo: *hágale saber; no lo hizo así constar; le hizo también cortar las orejas; e hizo también pasar nadadores de la otra parte...* La construcción sintáctica *hacer* + infinitivo es una estructura relacional abstracta que no se identifica con una secuencia significativa particular (*[hacer]*

⁹ Otra opción es reformular la *research question* de tal modo que se adapte a un enfoque forma → función, como hacen por ejemplo Taavitsainen y Jucker (2007), que analizan ciertos actos de habla basando sus búsquedas en elementos léxicos concretos identificados en un cierto campo semántico. Ello no impide que en otro trabajo del mismo volumen Taavitsainen y Fitzmaurice expresen sus reticencias sobre la validez del método: «corpus-based methods can be barriers to the investigation of pragmatic phenomena like conversational implicatures, which are neither routinely nor conventionally realized in lexical grammatical expressions. In these cases, corpus methodology encourages the identification of the function with a linguistic expression; such reliance can severely limit findings and thus jeopardize the validity of the study» (Taavitsainen y Fitzmaurice 2007: 18).

[*vr**]), sino que se manifiesta materialmente a través de variados y no siempre previsibles usos concretos.

A la vista de lo expuesto, cabe plantearse si realmente la lingüística basada en el análisis de los corpus electrónicos disponibles constituye en la práctica un avance tan positivo como suele presuponerse, y si la «revolución instrumental» (Rojo 2010a: 1153) que en términos generales supone la introducción de los nuevos métodos analíticos alcanza también a la sintaxis (y a la pragmática). La respuesta es afirmativa, pues incluso cuando no es factible la búsqueda (semi)automática y se requiere anotación manual, los corpus electrónicos resultan útiles al investigador al incorporar un sistema de codificación o «anotación no lingüística» (Rojo 2010b: 18) que clasifica los textos con indicaciones sobre género o tipo textual, área geográfica, autor, obra, año de edición, etc., y permite búsquedas independientes¹⁰. Nieuwenhuijsen (2009: 372) destaca la utilidad del *CORDE*, que permite acceder a la información contextual precisa para un análisis cualitativo minucioso al dar la opción de seleccionar todos los ejemplos de una obra determinada. Fuera del ámbito hispánico, por ejemplo Kohlen (2007), señala asimismo las ventajas de usar una base de datos textual con codificación por géneros en un estudio de tipo cualitativo que por su enfoque *función* → *forma* excluye la posibilidad de un análisis automático del corpus.

Con todo, a falta de una anotación sintáctica de la estructura funcional de todas las cláusulas contenidas en un corpus, no queda otra alternativa que un análisis manual, pues sin anotación sintáctica no es posible extraer información acerca de las distintas codificaciones que reciben los objetos ni datos relativos a su frecuencia.

En resumidas cuentas, tenemos que decir con Schulte (2009: 169) que «[t]he main reason to use non-annotated corpora is the lack of availability of sufficiently large annotated corpora»¹¹.

4. SOBRE LA MUESTRA EMPLEADA

Para el estudio descrito en la introducción, se pretendía recopilar datos de un período que abarcara desde documentaciones tempranas del castellano hasta la época en que la gramaticalización de clíticos en afijos alcanza una cierta consolidación (ss. XVI-XVII) y comparar esos datos con los del español actual. En los siguientes párrafos se describe la

¹⁰ Rojo destaca también la ventaja del *CORDE* en este aspecto, frente a las limitaciones del *CE*: «el CE no permite hacer consultas con respuestas distribuidas por países. Ni por tipos de texto. Ni por áreas temáticas, autores u obras. No todos estos rasgos tienen la misma importancia, por supuesto, pero es difícil evitar la sensación de que un sistema de consultas que no permita seleccionar los ejemplos de una palabra que aparecen en documentos publicados en un cierto período históricamente relevante (no por siglos) o en un país determinado o en textos de un cierto tipo resulta ser un sistema que no responde a las necesidades de la mayor parte de quienes investigan sobre el español. Por supuesto, el problema no está solo en la aplicación de consulta, sino también en el sistema de codificación utilizado o bien en la interacción de ambos» (Rojo 2010b: 21).

¹¹ Esta situación no es exclusiva del español. Con respecto al inglés, Smith *et al.* (2008: 164) señalan que «for many research questions the level of annotation provided in existing corpora is inadequate, and linguists therefore often need to categorize query results to their own classification criteria».

muestra recopilada para tal fin, que no está libre de problemas. Los principales derivan de las limitaciones que en cuanto a la elaboración de corpus diacrónicos y a la recuperación de datos a partir de ellos se comentan en los apartados anteriores.

Para el español actual, al lado de datos extraídos de un corpus constituido mayoritariamente por textos literarios y periodísticos, pudimos contar con otros procedentes del discurso conversacional (*Conversaciones*). Esta es la actividad discursiva más representativa de todas, pues la cultivan todos los hablantes (a diferencia del discurso literario) y, en general, le dedican más tiempo que a las demás.

Para etapas más antiguas evidentemente es imposible obtener datos conversacionales. Para la primera época representada en la muestra (ss. XIII y XIV) contábamos ya con datos obtenidos a partir de recuentos hechos para trabajos anteriores y procedentes de ejemplos fichados para un antiguo proyecto desarrollado en el Departamento de Lengua española de la USC.

Quedaba por cubrir el período que va desde el s. XV al XVII. Para ello buscamos tres textos en prosa para cada uno de los tres siglos, intentando que procedieran del principio, medio y final de cada siglo. De cada texto se analizaron fragmentos de 10.000 palabras, con lo que los siglos XV, XVI y XVII cuentan con muestras de aproximadamente 30.000 palabras cada uno. Además tanto para calcular con relativa comodidad que cada período estuviera representado por fragmentos de extensión equivalente como para poder incluir los ejemplos en una base de datos y anotarlos con relativa facilidad, se buscaba que los textos estuvieran en formato electrónico.

A la muestra puede achacársele, en cuanto a los textos más antiguos incluidos, la distancia existente entre la fecha de composición y la de los manuscritos disponibles en las ediciones que han llegado a nosotros, con lo cual no se puede saber con certeza qué rasgos del texto son propios de la época que en teoría representa y cuáles provienen de la intervención de copistas posteriores —problema sobre el que alertan Fernández-Ordóñez (1993: 69 y, especialmente, 2006: 1790 y ss.) o Nieuwenhuijsen (2009: 374-375), por ejemplo—. Esto se debe a que para los siglos XIII a XV se han aprovechado bases de datos ya existentes que ofrecen un volumen considerable de datos sin necesidad de acudir a recuentos manuales, con el tiempo y el trabajo que ello requeriría. Tales recursos fueron elaborados hace aproximadamente dos décadas a partir de ediciones de obras literarias disponibles en la época. Se trata, pues, de bases de datos cerradas sobre cuya composición ya no se puede decidir.

De otra parte, al buscar textos disponibles en formato electrónico para documentar épocas más recientes, se han incluido ediciones de calidad variable.

Con todo, las tendencias que se pretendía documentar son relativamente marcadas y, hasta cierto punto, estables diacrónicamente, de manera que la existencia de alteraciones significativas de los textos a este respecto habría de deberse a cambios sistemáticos en la codificación de objetos —formas independientes en lugar de clíticos o afijos— por parte de copistas y editores, lo cual parece improbable.

OI			OD		
clít./af.	f. indep.	clít. + f. indep.	clít. / af.	f. indep.	clít. + f. indep.

s. XIII		1044 (72.1%)	336 (23.2%)	67 (4.7%)	1427 (19%)	5915 (78.3%)	207 (2.7%)
s. XIV		1738 (75.9%)	519 (22.7%)	33 (1.4%)	970 (13.5%)	6145 (85.6%)	66 (0.9%)
s. XV		302 (64.5%)	154 (32.9%)	12 (2.6%)	424 (20.6%)	1602 (78%)	28 (1.4%)
s. XVI		579 (82%)	112 (15.9%)	15 (2.1%)	628 (24.7%)	1870 (73.6%)	43 (1.7%)
s. XVII		466 (75.4%)	106 (15.2%)	46 (7.4%)	563 (23.8%)	1771 (74.9%)	31 (1.3%)
s. XX	<i>BDS</i>	9727 (78.6%)	1085 (8.8%)	1554 (12.6%)	13333 (21%)	49131 (77.4%)	1005 (1.6%)
	<i>Conversaciones</i>	361 (88.5%)	2 (0.5%)	45 (11%)	302 (27.6%)	740 (67.5%)	54 (4.9%)

Figura 5. Codificación de objetos a lo largo de la historia del español

Como se puede comprobar en los datos obtenidos a partir de la muestra recopilada, expuestos en la figura 5, las tres codificaciones distinguidas se asocian con cada tipo de objeto de manera muy similar en todos los períodos analizados. Los referentes de objetos indirectos se codifican preferentemente mediante clíticos, lo que supone un indicio de su alta accesibilidad. Los objetos directos, por su parte, se asocian mayoritariamente con formas independientes, con lo que cabe suponer que están asociados a referentes de accesibilidad más baja con mayor frecuencia que los primeros. Los objetos doblados son minoritarios en ambos casos, si bien se observa un incremento de indirectos concordados conforme avanzamos en el tiempo.

En una parte de la muestra, sin embargo, la preferencia de los objetos indirectos por codificaciones asociadas a referentes de alta accesibilidad es menos clara que en el resto: la que corresponde al s. xv. Aquí, los objetos indirectos clíticos no llegan al 65%, mientras que en el resto no bajan del 70%. Con respecto a los objetos indirectos no clíticos, su frecuencia en los textos correspondientes a este siglo se sitúa 10 puntos porcentuales por encima de las frecuencias más altas del resto de la muestra.

En la siguiente tabla se presentan los mismos datos pero esta vez desglosados por sus textos de procedencia en lugar de por siglos.

	OI			OD		
	clít. / af.	f. indep.	clít. + f. indep.	clít. / af.	f. indep.	clít. + f. indep.
<i>Cid</i> (1. ^{er} <i>Cantar</i>)	161 (72.9%)	39 (17.6%)	21 (9.5%)	201 (26.5%)	497 (65.5%)	60 (8%)
<i>PCG</i> (1. ^a parte)	883 (72%)	297 (24.2%)	46 (3.8%)	1226 (18%)	5418 (80%)	147 (2%)
<i>CL</i>	1738 (75.9%)	519 (22.7%)	33 (1.4%)	970 (13.5%)	6145 (85.6%)	66 (0.9%)
<i>DA</i>	86 (53.4%)	68 (42.2%)	7 (4.3%)	99 (19.4%)	404 (79.2%)	7 (1.4%)
<i>Corbacho</i>	101 (66%)	47 (30.7%)	5 (3.3%)	118 (17.3%)	555 (81.3%)	10 (1.4%)

<i>Celestina</i>	115 (74.7%)	39 (25.3%)	0	207 (24%)	643 (74.7%)	11 (1.3%)
<i>Amadís</i>	220 (82%)	46 (17.2%)	2 (0.8%)	226 (28.1%)	566 (70.3%)	13 (1.6%)
Lazarillo	205 (81.7%)	40 (16%)	6 (2.3%)	206 (24.6%)	622 (74.2%)	10 (1.2%)
<i>Moradas</i>	154 (82.4%)	26 (13.9%)	7 (3.7%)	196 (21.8%)	682 (76%)	20 (2.2%)
<i>Quijote</i>	116 (63.4%)	46 (25.1%)	21 (11.5%)	160 (21.4%)	568 (76%)	20 (2.6%)
<i>Dorotea</i>	167 (80.3%)	32 (15.4%)	9 (4.3%)	236 (27.9%)	606 (71.6%)	4 (0.5%)
<i>NE</i>	183 (80.6%)	28 (12.3%)	16 (7.1%)	167 (21.7%)	597 (77.3%)	7 (1%)
<i>BDS</i>	9727 (78.6%)	1085 (8.8%)	1554 (12.6%)	13333 (21%)	49131 (77.4%)	1005 (1.6%)
<i>Conversaciones</i>	361 (88.5%)	2 (0.5%)	45 (11%)	302 (27.6%)	740 (67.5%)	54 (4.9%)

Figura 6. Codificación de objetos. Distribución por textos

Entre los textos correspondientes al siglo xv, dos de ellos muestran diferencias con respecto al resto de la muestra: el *Corbacho* y los procedentes de *Documentos andaluces (DA)*—una colección de documentos notariales en su mayoría—. En ambos casos, la frecuencia de objetos indirectos clíticos es menor que en el conjunto de la muestra y la de objetos indirectos independientes es mayor, pero esta diferencia es especialmente acusada en el caso de los textos notariales.

Los dos ejemplos siguientes pueden resultar ilustrativos acerca de las diferencias entre textos notariales y otros tipos. En primer lugar, abundan las aposiciones a pronombres, probablemente para asegurar la identificación inequívoca de los firmantes de cada documento y del papel de cada uno de ellos en las relaciones contractuales a las que se comprometen. En (4) puede verse que una aposición a un pronombre de segunda persona se repite parcialmente con otro pronombre correferencial a solo una cláusula de distancia.

- (4) Sepan qu<<a>>ntos esta c<<a>><rt>a viere<n> Com<m>o yo estea<n> sanch<e>s' Recabdador mayor Por n<uest>ro Sen<n>or el Rey enla muy noble c'ibdat de Cordoua & de todo su ob<i>><s>pado este an<n>o enq<ue> estamos' d<e>la ffecha desta c<<a>><rt>a ves'jno q<ue> so en<e>sta di<c>ha c'ibdat enla collac'io<n> de s'antiago oto<r>go avos' anto<n> gom<e>s' co<n>tador mayor de n<uest>ro s'en<n>or el Rey q<ue> s'od<e>s' absente bje<n> as'y com<m>o sy fuesed<e>s' prese<n>te E digo q<ue> en<e>ste di<c>h<<o>> dia d<e>la f<e>c<e>ha desta carta yo co<n>pre vnas' casas' q<ue> s'on en<e>sta di<c>ha c'ibdat enla collac'io<n> de s'antiago q<ue> s'e tien<e> co<n> casas' de vos' el di<c>h<<o>> anto<n> gom<e>s' & co<n> casas' de gonc'alo s'anch<e>s' de valenc'uela (DA, 1406).

Tanto en el ejemplo anterior (4) como en el siguiente (5) encontramos predicados triactanciales en los que tanto el complemento directo como el indirecto están asociados a constituyentes independientes. En ambos casos, los objetos indirectos son estructuras apositivas en las que intervienen un pronombre o un nombre propio, o la combinación de ambos, además de frases con contenido descriptivo.

- (5) [...] p<ar>a q<ue> pueda por mj & en mj nonbre dar & entregar A joh<a>n ferr<<a>>n<de>s' delos rrios c<r><<i>>ado del alcall<e> m<art><<i>>n> ferr<<a>>n<de>s' c'eron ves'jno desta c'ibdat ala collac'io<n> de santa m<ar>j<<a>> magdalena la tene<n>c'ia & possessio<n> de vnas' casas' alfondeg<u><<i>>llas enq<ue> ay dos puertas' con sus sobrados' [...] (DA, 1408).

Este tipo de codificación, atípica en general para los objetos indirectos, se da con cierta regularidad en este tipo de textos tal como indican los datos presentados en la figura 6. Aun dejando de lado el caso de los objetos indirectos, puede comprobarse que los argumentos de los predicados presentan en los documentos notariales codificaciones de una complejidad insólita en cualquier otro texto. Probablemente este tipo de acotaciones pueda verse como un rasgo característico de una cierta tradición discursiva (cfr. Kabatek 2004, por ejemplo) y, por tanto, ajena a los textos que se encuadran fuera de esta tradición.

Lo anterior sugiere que en los textos legales la relación entre accesibilidad referencial y codificación lingüística no opera de la misma manera que en otros géneros textuales. Sin embargo, en el momento de recopilar los datos no tuvimos en cuenta posibles diferencias en el fenómeno estudiado debidas a la variación en cuanto al género textual. Para afirmar con mayor seguridad que los textos legales son hasta cierto punto excepcionales en lo que toca a las tendencias en cuanto a la codificación de objetos hubiera sido necesario ampliar la muestra para hacer posible una comparación entre textos legales y de otro tipo, con la dificultad que ello supone tanto por la disponibilidad de este tipo de textos para todas las épocas estudiadas como por el aumento del trabajo que para el investigador supondría analizar manualmente una muestra aun mayor.

De lo anterior se colige que la existencia de materiales que permitan extraer automáticamente de textos diacrónicos información de tipo morfológico y sintáctico hubiera permitido incluir textos de diferentes géneros para los diferentes períodos estudiados sin añadir un esfuerzo considerable para los investigadores. De este modo, se hubieran podido alcanzar conclusiones con respecto a la influencia del factor género más fundamentadas que la simple intuición que se apunta en los párrafos anteriores.

5. CONCLUSIONES

Este artículo pretendía valorar las posibilidades que los corpus diacrónicos existentes en la actualidad ofrecen al investigador para abordar un estudio del origen de la concordancia de objeto en español. Se partía de un trabajo previo (Vázquez Rozas y García Salido 2012) en el que se defendía que este fenómeno era consecuencia de la

fijación de tendencias mayoritarias en la codificación de objetos, ya que los tipos de objeto que se expresan mayoritariamente mediante clíticos/afijos (indirectos, expresados mediante marcas de persona) son los mismos que o bien no experimentan restricciones en cuanto a la duplicación o la presentan obligatoriamente. Probar esta idea requiere un estudio cuantitativo con datos tanto de tipo sintáctico (distinción entre objetos directos e indirectos) como morfológico (objeto clítico/afijal frente a objeto independiente). Aquí se ha reflexionado sobre qué recursos tiene el investigador para obtener este tipo de datos y los procedimientos que ha de seguir. Las conclusiones de esta reflexión se exponen a continuación.

En primer lugar, al compilar la muestra, el investigador ha de ser consciente de que algunos de los géneros textuales representados en los corpus diacrónicos disponibles para el español pueden presentar peculiaridades informativas con respecto a las variedades discursivas utilizadas mayoritariamente —fundamentalmente, la conversación—. No obstante, para asegurarse del carácter excepcional de fenómenos observados en determinados textos, sería conveniente poder aislar este factor con respecto a otros (geográfico, temporal, etc.).

En segundo lugar, para rentabilizar las técnicas de la lingüística de corpus en el análisis de las construcciones complejas es preciso enriquecer la anotación sintáctica de los textos electrónicos. Se ha constatado que el nivel de anotación de los corpus diacrónicos disponibles para el español es insuficiente para llevar a cabo un estudio como el descrito prescindiendo de la anotación manual. Lamentablemente, hasta donde sabemos, el único recurso para el español que permitiría recuperar automáticamente los datos necesarios para una investigación como la descrita es el corpus ANCORA¹², que en cualquier caso, recoge únicamente datos de español contemporáneo y cuyo tamaño plantea dudas sobre la viabilidad de un estudio cuantitativo. En fase de elaboración existen dos recursos similares: CSA, dentro del proyecto DRASAE¹³, de nuevo sobre español contemporáneo, y el corpus del proyecto ISWOC¹⁴, que incluirá, esta vez sí, muestras de la diacronía del español, entre otras lenguas.

Mientras recursos como el último mencionado no estén disponibles para la comunidad científica, lo adecuado sería conjugar la anotación manual con la información que proporcionan los recursos (semi)automáticos de análisis de corpus (programas de extracción de concordancias, de búsqueda de colocaciones, lematización, etc.) y la anotación existente (datos cronológicos, de autoría, de género, etc.). Los corpus disponibles en línea también podrían permitir la inclusión de anotaciones parciales realizadas para investigaciones específicas (cfr. las posibilidades que ofrece el *BNC* en cuanto a «*in-tool-annotation*» destacada en Smith *et al.* 2008). Ahora bien, tal recurso solo sería realmente útil si se garantiza que los criterios de dichas anotaciones se detallan adecuadamente para hacer posible su reutilización.

¹² <http://clic.ub.edu/corpus/ancora>

¹³ <http://gramatica.usc.es/proyectos/drasae/>

¹⁴ <http://www.hf.uio.no/ilos/english/research/projects/iswoc/>

FUENTES BIBLIOGRÁFICAS

- Amadís* = FONDOW, Steven R. (ed.) (1999): *The Electronic Text and Concordances of the «Amadís de Gaula», 1539 Seville Edition*. Madison: HSMS.
- ANCORA = *Corpus AnCora*. <http://clic.ub.edu/corpus/ancora>.
- BDS = *Base de datos sintácticos*. <http://www.bds.usc.es>.
- BM = Enrique-Arias, Andrés (dir.) (2008): *Biblia Medieval*. <http://www.bibliamedieval.es>.
- BNC = *The British National Corpus*. <http://www.natcorp.ox.ac.uk/>.
- CE = DAVIES, Mark (2003): *Corpus del español*. <http://www.corpusdelespanol.org/>.
- Celestina* = CORFIS, Ivy A. y John O'NEILL (eds.) (1997): *Early «Celestina» Electronic Texts and Concordances*. Madison: HSMS.
- Cid* = ANÓNIMO (2000 [1195-1207]): *Cantar de Mio Cid*. Barcelona: Crítica [Edición de Alberto Montaner].
- CL = DON JUAN MANUEL (1983 [1335]): *Conde Lucanor*. Madrid: Gredos [Edición de José Manuel Bleca].
- Conversaciones* = BRIZ, Antonio y Grupo VAL.ES.CO (2002): *Corpus de conversaciones coloquiales*. Madrid: Arco/Libros.
- Corbacho* = MARTÍNEZ DE TOLEDO, Alfonso (1901 [1438]): *Arcipreste de Talavera o Corbacho*. Madrid: Sociedad de Bibliófilos Españoles [Edición electrónica (2004), Alicante: Biblioteca virtual Miguel de Cervantes].
- CORDE = Real Academia española: *Corpus diacrónico del español*. <http://corpus.rae.es/cordenet.html>.
- CREA = Real Academia española: *Corpus de referencia del español actual*. <http://corpus.rae.es/creanet.html>.
- DA = KAUFFELD, Cynthia (ed.) (1999): *Electronic texts and concordances of Andalusian documents (1324-1500) selected from the collection of The Hispanic Society of America*. Madison: HSMS.
- Dorotea* = VEGA, Lope de (1968 [1632]): *La Dorotea*. Madrid: Castalia [Edición de Edwin S. Morby].
- Lazarillo* = ANÓNIMO (2004 [1554]): *Vida de Lazarillo de Tormes y sus fortunas y adversidades*. Alicante: Biblioteca Virtual Miguel de Cervantes.
- Moradas* = SANTA TERESA DE JESÚS (2000 [1577]): *Las Moradas*. Alicante: Biblioteca Virtual Miguel de Cervantes.
- NE = CASTILLO SOLÓRZANO, Alonso de (1906 [1632]): *Teresa de Manzanares o La niña de los embustes*. Madrid: Viuda de Rico.
- PCG = ALFONSO X (1979 [1289]): *Primera Crónica General*. Madrid: Gredos [Edición de Ramón Menéndez Pidal].
- Quijote* = CERVANTES, Miguel de (1998 [1615]): *Segunda parte del ingenioso hidalgo don Quijote de la Mancha*. Barcelona: Instituto Cervantes/Crítica [Edición de Francisco Rico].

REFERENCIAS BIBLIOGRÁFICAS

- ALBA-SALAS, Josep (2009): «Las estructuras tipo *meter miedo* en diacronía: más detalles sobre la evolución histórica de las colocaciones causativas», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 343-364.
- ARIEL, Mira (1990): *Accessing Noun-Phrase Antecedents*. Londres: Routledge.

- ARIEL, Mira (2000): «The Development of Person Agreement Markers: from Pronouns to Higher Accessibility Markers», en Michael Barlow y Suzanne Kemmer (eds.), *Usage-based Models of Language*. Stanford: CSLI, pp. 197-260.
- BLAKE, Robert y Gina LEE (2009): «MOCA: análisis de corpus en línea en un contexto histórico», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 327-341.
- BLOCKLEY, Mary (2001): *Aspects of Old English Poetic Syntax. Where Clauses Begin*. Urbana: University of Illinois Press.
- BOSSONG, Georg (1980): «Aktantenfunktionen in Romanischen Verbalsystem», *Zeitschrift für romanische Philologie*, 96, pp. 1-22.
- BYBEE, Joan L (1985): «Diagrammatic Iconicity in Stem - Inflection Relations», en John Haiman (ed.), *Iconicity in syntax*. Amsterdam: John Benjamins, pp. 11-47.
- DAVIES, Mark (2009): «Creating Useful Historical Corpora: A Comparison of *CORDE*, the *Corpus del español*, and the *Corpus do Português*», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 137-166.
- DU BOIS, John W., Lorraine E. KUMPF y William J. ASHBY (eds.) (2003): *Preferred Argument Structure. Grammar as Architecture for Function*. Amsterdam: John Benjamins.
- ENRIQUE-ARIAS, Andrés (ed.) (2009a): *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- ENRIQUE-ARIAS, Andrés (2009b): «Introducción», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 11-21.
- ENRIQUE-ARIAS, Andrés (2009c): «Ventajas e inconvenientes del uso de *Biblia medieval* (un corpus paralelo y alineado de textos bíblicos) para la investigación en lingüística histórica del español», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 269-283.
- FERNÁNDEZ-ORDÓÑEZ, Inés (1993): «Leísmo, laísmo y loísmo: estado de la cuestión», en Olga Fernández Soriano (ed.), *Los pronombres átonos*. Madrid: Taurus, pp. 63-96.
- FERNÁNDEZ-ORDÓÑEZ, Inés (2006): «La historiografía medieval como fuente de datos lingüísticos. Tradiciones consolidadas y rupturas necesarias», en José Jesús Bustos Tovar y José Luis Girón Alconchel (eds.), *Actas del VI Congreso Internacional de Historia de la Lengua Española*, vol II. Madrid: Arco/Libros, pp. 1779-1807.
- FRANCO, Jon (2000): «Agreement as a Continuum: The Case of Spanish Pronominal Clitics», en Frits Beukema y Den Dikken (eds.), *Clitic Phenomena in European Languages*. Amsterdam: John Benjamins, pp. 147-190.
- GARCÍA-MIGUEL, José María (1991): «La duplicación de complemento directo e indirecto como concordancia», *Verba*, 18, pp. 375-410.
- GIVÓN, Talmy (1976): «Topic, Pronoun and Grammatical Agreement», en Charles N. Li (ed.), *Subject and Topic*. New York: Academic Press, pp. 149-188.
- GIVÓN, Talmy (1983): «Topic Continuity in Discourse: an Introduction», en Talmy Givón (ed.), *Topic Continuity in Discourse. A Quantitative Cross-language Study*. Amsterdam: John Benjamins, pp. 5-41.
- HIDALGO DOWNING, Raquel (2003): *La tematización en el español hablado*. Madrid: Gredos.
- KABATEK, Johannes (2004): «Tradiciones discursivas jurídicas y elaboración lingüística en la España

- medieval», *Cahiers de Linguistique Hispanique Médiévale*, 28, pp. 249-261.
- KOHNE, Thomas (2007): «Text Types and the Methodology of Diachronic Speech Act Analysis», en Susan Fitzmaurice e Irma Taavitsainen (eds.), *Methodological Issues in Historical Pragmatics*. Berlin: Mouton de Gruyter, pp. 139-166.
- NIEUWENHUIJSEN, Dorien (2009): «El rastreo del desarrollo de algunos pronombres personales en español: (im)posibilidades de los corpus diacrónicos digitales», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 365-384.
- LLORENTE, Antonio y José MONDÉJAR (1972): «La conjugación objetiva en las lenguas románicas», *Prohemio*, 3, 1, pp. 5-27.
- PATO, Enrique (2009): «Notas aclaratorias sobre la historia del indefinido *alguien*: una aplicación directa del uso de corpus diacrónicos», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 401-416.
- RIIHO, Timo (1988a): «Sobre el carácter fortuito de la redundancia pronominal en el español antiguo», en Manuel Ariza, Álvaro Salvador y Antonio Viudas (eds.), *Actas del I Congreso Internacional de Historia de la Lengua*, vol. I. Madrid: Arco/Libros, pp. 583-591.
- RIIHO, Timo (1988b): *La redundancia pronominal en el iberorromance medieval*. Tübingen: Niemeyer.
- RINI, Joel (1990): «Dating the Grammaticalization of the Spanish Clitic Pronoun», *Zeitschrift für romanische Philologie*, 106, pp. 354-370.
- RINI, Joel (1991): «The Redundant Indirect Object Constructions in Spanish: A New Perspective», *Romance Philology*, 45, 2, pp. 269-286.
- ROJO, Guillermo (2010a): «Aguja de navegar corpus», en Víctor M. Castel y Liliana Cubo de Severino (eds.), *La renovación de la palabra en el bicentenario de la Argentina. Los colores de la mirada lingüística*. Mendoza: Editorial FFyL, UNCuyo, pp. 1151-1163.
- ROJO, Guillermo (2010b): «Sobre codificación y explotación de corpus textuales: Otra comparación del *Corpus del Español* con el *CORDE* y el *CREA*», *Lingüística*, 24, pp. 11-50.
- SCHULTE, Kim (2009): «Using non-annotated diachronic corpora: benefits, methods and limitations», en Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 167-182.
- SILVA-CORVALÁN, Carmen (1984): «Semantic and Pragmatic Factors in Syntactic Change», en Jacek Fisiak (ed.), *Historical Syntax*. Berlin: Mouton, pp. 555-573.
- SMITH, Nicholas, Sebastian HOFFMANN y Paul RAYSON (2008): «Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations», *Literary and Linguistic Computing*, 23, 2, pp. 163-180.
- TAAVITSAINEN, Irma y Susan FITZMAURICE (2007): «Historical pragmatics: What it is and How to Do it», en Susan M. Fitzmaurice e Irma Taavitsainen (eds.), *Methodological Issues in Historical Pragmatics*. Berlin: Mouton de Gruyter, pp. 11-36.
- TAAVITSAINEN, Irma y Andreas H. JUCKER (2007): «Speech Act Verbs and Speech Acts in the History of English», en Susan M. Fitzmaurice e Irma Taavitsainen (eds.), *Methodological Issues in Historical Pragmatics*. Berlin: Mouton de Gruyter, pp. 107-138.
- TOTTIE, Gunnel y Sebastian HOFFMANN (2006): «Tag Questions in British and American English», *Journal of English Linguistics*, 34, 4, pp. 883-911.
- TRUGOTT, Elizabeth Closs y Susan PINTZUK (2008): «Coding the York-Toronto-Helsinki Parsed Corpus of Old English Prose to investigate the Syntax-Pragmatics Interface», en Susan M. Fitzmaurice y Donka Minkova (eds.), *Studies in the History of the English Language IV*.

Empirical and Analytical Advances in the Study of English Language Change. Berlin/New York: Mouton de Gruyter, pp. 61-80.

VÁZQUEZ ROZAS, Victoria y Marcos GARCÍA SALIDO (2012): «A Discourse-Based Analysis of Object Clitic Doubling in Spanish», en Tine Breban, Lieselotte Brems, Kristin Davidse y Tanja Mortelmans (eds.), *Grammaticalization and Language Change: New Reflections*. Amsterdam: John Benjamins, pp. 269-296.