

«DELOS SOS OIOS TAN FUERTE MIENTRE LORANDO»: UN CAMBIO SINTÁCTICO EXAMINADO A TRAVÉS DE UN CORPUS DIGITALIZADO

ROBERT J. BLAKE (*University of California, Davis*)
CARLOS SÁNCHEZ LANCIS¹ (*Universitat Autònoma de Barcelona*)

RESUMEN:

En este trabajo se aplica la noción del *continuo sociolingüístico* al plano diacrónico del castellano para estudiar el origen y la evolución de un cambio sintáctico a través de los documentos notariales del monasterio de Sahagún (siglos IX al XII). Se trata de un conjunto de textos medievales no literarios con escritura latinizante y parcialmente logográfica en transición hacia las reformas carolingias, en los que se demuestra que los escribas mezclaban el latín con el romance hablado innovador. Gracias al estudio de estos documentos mediante la herramienta de análisis digital MOCA (*Multimedia Online Corpus Analysis*), se ha podido establecer, por un lado, una clara relación entre diversas estructuras gramaticales distintas formadas por la secuencia *demonstrativo/artículo + posesivo + sustantivo* ya desde época muy temprana; por otro, constatar la presencia de la construcción sintáctica *artículo + posesivo + sustantivo* incluso en los textos notariales, a pesar de estar escritos aparentemente en latín.

PALABRAS CLAVE: continuo sociolingüístico, corpus digitalizado del castellano antiguo, documentos notariales, cambio sintáctico, construcción posesiva

ABSTRACT:

In this study, we extend the idea of a synchronic *sociolinguistic continuum* to a diachronic framework in order to examine the origin and evolution of a particular Spanish syntactic change –the possessive construction *article + possessive adjective + noun*, for example *los sos oios*– with data extracted from a digitalized corpus of early medieval notarial documents from the Sahagun Monastery on the border of Castile and Leon (9th to 12th centuries). These texts are written in a mixed or Latinized style that employs both Latin morphology and lexis along with other forms rendered in completely reformed, phonetic spellings patterned after the spoken Romance vernacular at the time. We used a digital corpus tool called MOCA (*Multimedia Online Corpus Analysis*) to follow the evolution of this possessive construction over time. This methodology allowed us to establish a linguistic continuum between Romance syntax in documents written both before and after the adoption of Carolingian orthographic reforms.

KEY WORDS: Sociolinguistic continuum, Digital medieval Spanish corpus, Notarial documents, Syntactic change, Possessive construction

1. EL CONTINUO LINGÜÍSTICO Y LOS CORPUS DIGITALIZADOS

La evolución fonológica de una lengua a través del tiempo es un tema fundamental para comprender el campo de la lingüística, a pesar de que no se revelaron con claridad los mecanismos del cambio lingüístico hasta el brillante estudio de Weinreich, Labov y Herzog en 1968. Antes de esta fecha, los principales modelos teóricos no ofrecían una explicación

¹ La presente investigación ha sido parcialmente financiada con una ayuda del MEC (nº de ref. PR2009-0144, Modalidad A) dentro del Programa Nacional de Movilidad de Recursos Humanos de Investigación para «Estancias de movilidad de profesores e investigadores séniores en centros extranjeros de enseñanza superior o investigación», del MiCInn (nº de ref. FFI2011-29440-C03-01) y de la CIRIT del Comissionat per Universitats i Recerca de la Generalitat de Catalunya (nº de ref. 2009SGR 1067).

satisfactoria del cambio. Por ejemplo, Ferdinand de Saussure, padre de la lingüística moderna, apenas trató el tema del cambio, ya que sus nociones de *langue* y *parole* captaban más bien un contrato social donde la *langue* (entiéndase por la sociedad) inhibía y contrarrestaba cualquier desviación idiosincrásica para que el sistema lingüístico no se desvirtuara a causa de la *parole*, o sea, como resultado de las peculiaridades e innovaciones del uso individual. En este marco estático no había lugar para explicar el proceso del cambio lingüístico. Por otra parte, los modelos de Chomsky tampoco permitían dar cuenta del cambio lingüístico porque perseguían descubrir la *Gramática Universal*, un concepto más bien psicológico, no social, donde el uso no tenía papel alguno. Por fin, Weinreich, Labov y Herzog (1968) nos dieron la clave del cambio lingüístico revelando que las innovaciones del individuo, presentes en toda lengua natural, promueven el cambio de manera que se vaya adoptando por una mayoría de hablantes o, por lo menos, un grupo suficientemente preponderante y prestigioso en la sociedad. Así, durante la realización de un cambio la variación procede por medio de un continuo lingüístico que depende de factores estilísticos, como el grado de formalidad o informalidad, y de otros factores sociales, como la clase, el sexo, o la educación del hablante. Todo hablante participa en la vida de una variante, en su promoción o su supresión, pero nadie la emplea todo el tiempo ni en todos los ámbitos sociales hasta que finaliza el cambio lingüístico por aclamación de todos. Por lo tanto, el continuo se deriva de las diferentes frecuencias de uso según unas normas que rigen un determinado contexto social durante un largo período de variación y de aceptación social.

A primera vista, el continuo parece presentar un cuadro de *variación libre*, un término inventado por los estructuralistas, pero las variantes no ocurren al azar sino que están condicionadas por una serie de normas de uso. Una vez que una variante en particular es aceptada por todos los hablantes en todos los contextos posibles, se puede hablar de cambio lingüístico y de un nuevo uso obligatorio, ya que el rasgo en cuestión deja de ser una variante. Sin embargo, antes de llegar a este nuevo *acuerdo* general, existe un continuo de variación donde la nueva realización se emplea según unas normas de uso vigentes en una comunidad de habla. Descubrir el inicio de un cambio lingüístico resulta ser muy difícil en el contexto sincrónico y casi imposible en el contexto diacrónico. Pero el proceso del cambio lingüístico parece ser lo mismo en toda época y, por ello, muchos investigadores han recurrido a la teoría sociolingüística para guiar sus estudios históricos (p. ej., Koch y Oesterreicher 2007 [1990], Gimeno 1995, Penny 2004 [2000], Wright 2002, Tuten 2003, Emiliano 2005, Penny 2006, Blake y Lee 2009, Velázquez-Mendoza y Blake 2010).

Aunque los mecanismos del cambio expuestos por la sociolingüística son los mismos en el estudio sincrónico de la lengua que en el marco histórico, los métodos no lo son. La diacronía se trabaja dentro del marco filológico; es decir, se basa en documentos escritos y no en recuentos estadísticos sacados de unas muestras del habla. La escritura aporta sus propias normas que resisten, en gran parte, las innovaciones encontradas en la lengua vernácula a fin de que se mantenga la mayor comprensión posible entre los miembros de una comunidad de habla que se extiende a través del tiempo y el espacio. Koch y Oesterreicher (2007 [1990]: 35) captan esta noción en la diacronía de una lengua al destacar los factores de lo hablado, la *lengua de la inmediatez*, en conjunto con lo escrito, la *lengua de la distancia*. Por lo tanto, se espera hallar en el continuo escrito a través del tiempo muchas inconsistencias que resultan de inserciones de formas procedentes de lo hablado, al

lado de otras formas petrificadas en el tiempo que han caído o que van cayendo en desuso en el habla.

Ahora bien, los que sabían leer en tiempos pasados se percataban de esta situación, tan confusa para el lector moderno que no conoce más que los resultados de unos cambios lingüísticos totalmente acabados, y acertaban con la lectura correcta en voz alta. Así es como se tienen que entender, por ejemplo, los documentos medievales de la *Colección diplomática del monasterio de Sahagún (CDMSahagún)* de los siglos IX al XII², donde los escribas empleaban una escritura latinizante y parcialmente logográfica en transición hacia las reformas carolingias, que establecieron unas equivalencias fonéticas en España a partir del siglo XIII, como nos ha explicado Wright (1989 [1982], 1993, 2002) en repetidas ocasiones. Para leer estos documentos latinizantes antes del siglo XIII, se exige reconocer cómo los escribas mezclaban lo petrificado (el latín) con lo innovador (el romance hablado) con una flexibilidad natural. Esta mezcla nos resulta rara hoy día si se cree equivocadamente que aquella escritura tenía un valor puramente fonético. De ser así no habría hecho falta la reforma ortográfica que luego se implementó en el siglo XIII. Estas prácticas ortográficas antes de la reforma tampoco implicaban la existencia de dos lenguas habladas en una misma sociedad, o sea, la diglosia, sino que reflejaban un criterio muy flexible para lo que se entendía por *escribir*, siendo el *escribir latín* un concepto muy abarcador en el Alto Medioevo (Wright 2002, Penny 2003, Emiliano 2005, Wright 2009). Desde luego, ciertos vocablos latinos les seguían siendo inteligibles a los letrados de la época en su forma escrita petrificada a pesar de no ser usados en la lengua vernácula del momento (Wright 2002: 32).

Pese a todos los esfuerzos y todo el avance que ha supuesto el gran desarrollo que han experimentado los estudios de sintaxis histórica últimamente, todavía quedan muchos puntos oscuros en las explicaciones gramaticales que se han dado de ciertos cambios. Esto es debido, por una parte, al hecho de que estas investigaciones se han dedicado preferentemente al estudio de la lengua literaria, lengua mucho más cuidada que la lengua oral y mucho más opaca para reflejar el cambio gramatical en el momento en que éste se produce, sin tener nada en cuenta la lengua no literaria, mucho menos sujeta a normas de retórica y de escritura de su tiempo (*vid.* Lapesa 1981); por otra parte, aunque durante los últimos años estamos asistiendo a un considerable incremento del número de herramientas informáticas que proporcionan corpus formados por bases de datos textuales que facilitan el acceso a un gran número de datos lingüísticos diacrónicos, entre los que destacan el *Corpus diacrónico del español (CORDE)* de la Real Academia Española y el *Corpus del español* del Dr. Mark Davies, la lengua de los documentos medievales, sobre todo de los primeros siglos, sigue sin estar suficientemente representada.

El objetivo principal del presente trabajo es demostrar las posibilidades de investigación que ofrece un determinado tipo de herramienta informática de análisis de corpus digitalizados para el estudio del cambio gramatical. Esta metodología se está aplicando en estos momentos en la Universidad de California (UC Davis) con la finalidad de poder encontrar indicios, en textos no literarios nunca antes estudiados desde este punto de

² Se ha elegido la *Colección diplomática del monasterio de Sahagún* debido al gran valor lingüístico que tiene, ya que este monasterio, aunque situado en León, tuvo una gran importancia durante la Edad Media no sólo por su ubicación, sino por la gran extensión de tierras que poseía, las cuales iban desde Tierra de Campos hasta Liébana y Segovia, por lo que su influencia fue muy grande. Debido a ello, los datos que ofrece no sólo se limitan al dialecto leonés, sino que también sirven para reflejar el romance castellano.

vista, del origen de ciertos cambios sintácticos. Como ejemplo de explotación de corpus digitalizados se va a llevar a cabo un análisis no exhaustivo de la existencia de una construcción sintáctica (*artículo + posesivo + sustantivo*; p. ej., «delos sos oios...») que marcará la configuración y las particularidades del castellano medieval (*vid.* principalmente Lapesa 1971 y Company 2009)³. El corpus seleccionado está compuesto actualmente por los documentos medievales transcritos paleográficamente de los siglos IX al XII del monasterio de Sahagún (*CDMSahagún*), que linda con Castilla, los cuales han sido introducidos en una base de datos textual, que permite ser consultada y analizada a través de una herramienta de análisis digital denominada *MOCA (Multimedia Online Corpus Analysis)* (Blake y Lee 2009).

Dadas las características de esta clase de textos no literarios, de tipo documental principalmente, los cuales están escritos aparentemente en un bajo latín (Bastardas 1953), es decir, en escritura logográfica latino-romance de la Alta Edad Media (Wright 1992 y 2002, Emiliano 2005), pero que en realidad, por los indicios que se tienen, en cierta forma recubren un romance castellano, nuestro propósito es constatar la existencia de algunas de las características morfológicas y sintácticas de esta estructura, propia del castellano medieval, que contiene el demostrativo/artículo y el posesivo junto a un sustantivo. Para ello se va a emplear la herramienta en línea *MOCA*, de gran versatilidad y potencia en relación a otros bancos de datos textuales diacrónicos, la cual nos va a permitir afinar mucho más los análisis realizados en las distintas gramáticas históricas en relación a esta construcción sintáctica que se halla durante los primeros siglos del romance castellano⁴, pero que curiosamente no aparece reflejada del mismo modo en los diferentes tipos de textos. Hay que tener en cuenta que ciertos textos no literarios (colecciones documentales, cartularios, becerros, etc.), que normalmente han sido desechados del estudio histórico de la lengua española por ser considerados exclusivamente latinos, pueden aportar datos muy valiosos que nos muestren en qué manera se gestó durante un largo periodo nuestro idioma, dada la transición entre la tradición escrita visigótica y las nuevas reformas carolingias apenas iniciadas en el siglo XI.

2. PRINCIPALES POSIBILIDADES DE INVESTIGACIÓN CON LA HERRAMIENTA *MOCA* (ANÁLISIS DE CORPUS EN LÍNEA MULTIMEDIA)

La herramienta de análisis de corpus multimedia en línea denominada *MOCA* posee, respecto a otros programas de concordancia existentes, una serie de ventajas, que permiten grandes posibilidades de cara a la realización de una investigación como, por ejemplo, la que nos ocupa en el presente estudio. En la siguiente figura aparece la pantalla de acceso a esta herramienta:

³ Para un exhaustivo estudio diacrónico de este cambio sintáctico del español, objetivo no contemplado en nuestro análisis, remitimos al completísimo trabajo de Company (2009).

⁴ La primera vez que se documenta esta construcción es en las *Glosas emilianenses*, aproximadamente de mediados o finales del siglo XI, en la siguiente frase: «Facanos Deus omnipotes tal serbitjo fere ke delante *ela sua face* gaudioso segamus».

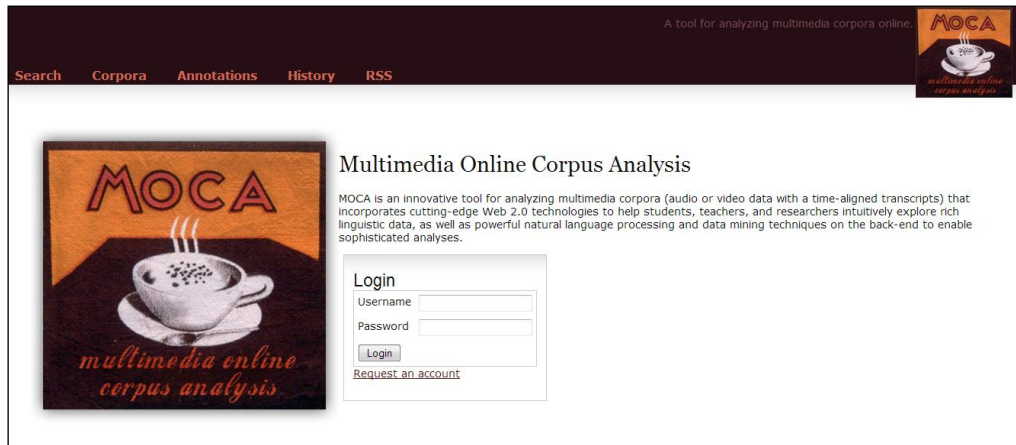


Figura 1. Pantalla de acceso a la herramienta *MOCA*

La primera de estas posibilidades es que cada investigador puede formar su propio corpus individual añadiendo documentos a la base de datos textual, mediante la transcripción directa del documento o a través de la simple copia de otro texto que ha sido previamente digitalizado (en el presente caso, los documentos medievales transcritos paleográficamente de los siglos IX al XII del monasterio de Sahagún). Para ello, debe acceder a esta función a través de la pestaña *Corpora* que aparece en la parte superior de la pantalla anterior, situada junto a la de búsqueda (*Search*), anotaciones (*Annotations*) e historial de búsquedas (*History*). Si se opta por diseñar el propio corpus, existe la posibilidad de establecer una combinación o compaginación dentro de la misma base de datos de los documentos introducidos con otros archivos de imagen, audio o incluso vídeo, lo cual puede resultar de gran utilidad sobre todo cuando se trabaja, como es nuestro caso, con documentos notariales en los que una interpretación equivocada en su transcripción puede llevar a conclusiones erróneas. De ahí la importancia para cualquier investigador de disponer siempre de la posibilidad de realizar una consulta del manuscrito original para poder comprobar la existencia o no de una determinada forma gráfica, morfológica o una construcción sintáctica. Asimismo, también se puede asociar a cada texto una serie de variables extralingüísticas de gran utilidad como son su fecha de redacción, su origen geográfico, el nombre específico con que se conoce el documento, el nombre del escriba, el tipo al que pertenece su contenido (si se trata de una venta, una donación, etc.), el siglo y el cuarto de siglo al que corresponde, u otras categorías que el investigador considere oportunas, ya que no se trata en ningún momento de una base de datos cerrada. Luego la herramienta *MOCA* permite hacer búsquedas y correlaciones según estas variables. Todos estos aspectos se pueden observar en la Figura 2:

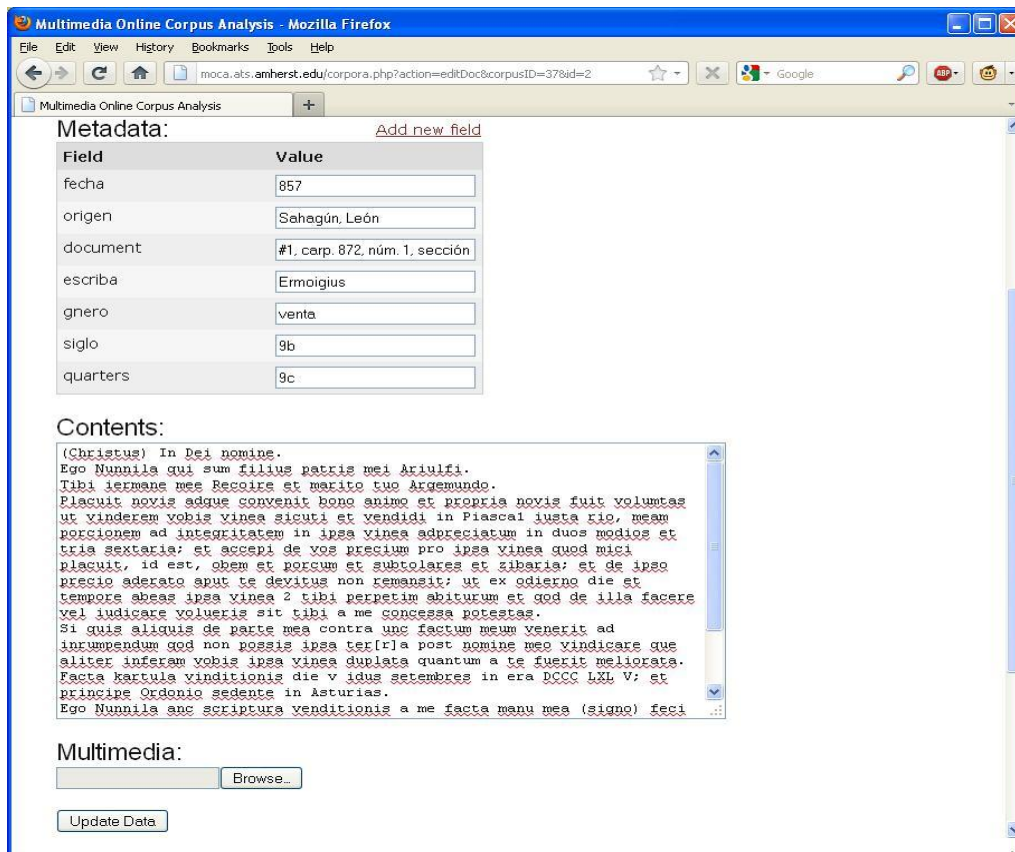


Figura 2. Entrada de datos al sistema MOCA

Por otra parte, en esta misma pantalla titulada *Corpora* aparecen los distintos corpus a los que se puede acceder, ya sean propios o de uso general, pudiendo consultar sus características (*stats*) antes de realizar cualquier consulta, como se puede observar en la siguiente figura por lo que respecta al corpus analizado en el presente estudio:

The screenshot shows the 'Corpora Stats' page for 'español antiguo'. The page has a navigation bar with 'Search', 'Corpora', 'Annotations', 'History', and 'RSS'. The main content includes a description of the corpus and an overview table.

Corpora Stats: "español antiguo"

[edit](#)

Description
Este corpus cuenta con documentos del monasterio de Sahagún (León) y la catedral de León desde los primeros momentos hasta la transición al romance, o sea, desde los siglos VIII en adelante. El corpus tiene interés histórico en cuanto a la transición y evolución del latín tardío al romance.
Documents were last updated on 06/28/11 at 17:06:53.

Overview

Item	Count
Documents	1434
RSS Feeds	0
Words	501199
Distinct Words	35068

Top Words

Word	Count
et	29052
in	17009
de	16143
cum	5427
conf	4974
ad	3974
ego	3542
uel	3098
a	3036
que	3002

Figura 3. Características del corpus «español antiguo»

Así, según los datos anteriores, estamos ante un corpus compuesto por 1434 documentos, 501199 ocurrencias, de las cuales 35068 son formas o palabras diferentes. Como información complementaria, se presentan en una tabla los diez vocablos con mayor frecuencia de uso, colocados en orden descendente. Lógicamente, dadas las características de estos textos, la conjunción copulativa *et*, con 29052 ocurrencias, se halla como la forma más utilizada, seguida de las preposiciones *in* (17009) y *de* (16143), y destaca como la séptima más empleada significativamente el pronombre personal *ego* con 3542 apariciones. A pesar de que las palabras más gramaticales son las que acostumbran a encontrarse con mayor frecuencia, la presencia del pronombre personal de primera persona *ego* en esta lista indica con gran claridad la tipología textual del corpus.

Cuando se accede a esta herramienta, la primera pantalla que aparece por defecto es la de selección de corpus para la realización de búsquedas (pestaña *Search*). El investigador puede elegir uno o más corpus simultáneamente y agruparlos según sus intereses. Una vez seleccionado, se presentan las dos opciones que se pueden realizar: analizar (*analyze*) o buscar (*search*). En el primer caso, se puede optar por hacer una búsqueda en el corpus entero de una serie de vocablos que aparecen en un número concreto de combinaciones o colocaciones, formadas por 1-6 o una cantidad indeterminada de palabras, que se hallan en un número igual, inferior o superior a *x* veces, pudiendo ordenarse los resultados por orden alfabético o frecuencia de aparición y visualizarse además con una gráfica de tipo nube o relacional. Asimismo, los resultados obtenidos, que se limitan a recoger en un listado las combinaciones posibles junto a su frecuencia, pueden ser contrastados con el vocabulario que compone el corpus respecto a cualquiera de las variables extralingüísticas introducidas. Además, a partir de todas estas colocaciones se pueden visualizar los contextos en los que aparecen. Toda esta serie de posibilidades se puede ver en la siguiente figura:

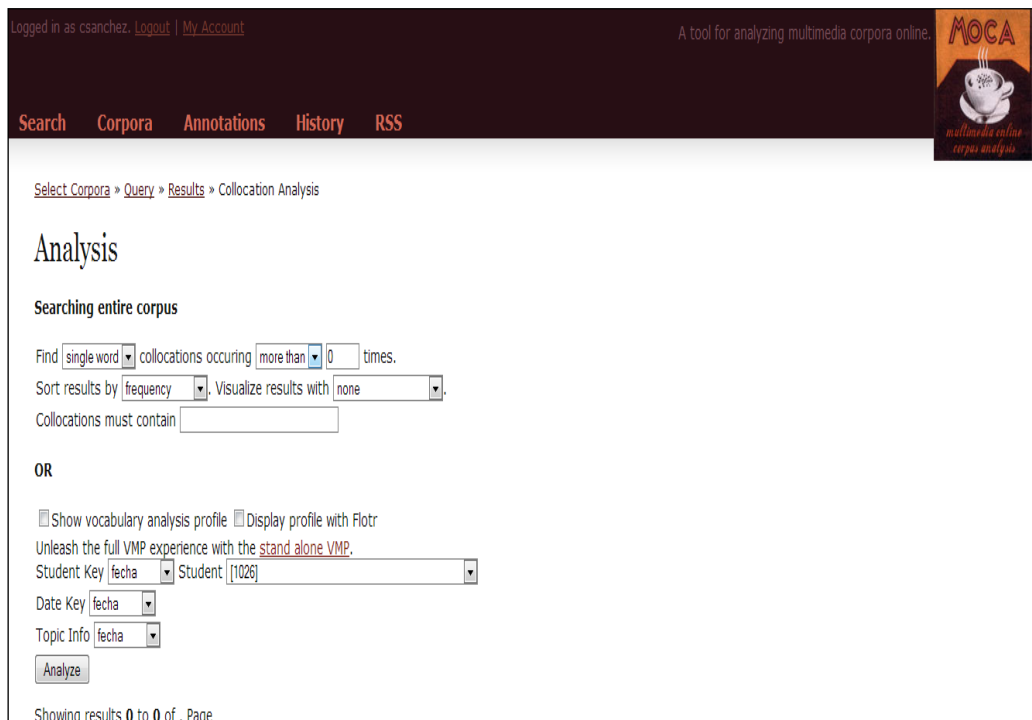


Figura 4. Posibilidades de análisis del corpus

En el segundo caso, el de búsqueda en el corpus (*Corpus Search*), simplemente se introduce la opción de localizar una o más palabras en los textos, e incluso formas lematizadas, restringidas o no por cualquiera de las diferentes variables extralingüísticas tenidas en cuenta. El resultado obtenido, a diferencia del caso anterior, es directamente el de las concordancias o fragmentos en los que se hallan los elementos lingüísticos elegidos, los cuales pueden ser ordenados de distinto modo y representados además, según todas estas variables, mediante un gráfico. Algunas de estas opciones se pueden observar en la Figura 5:

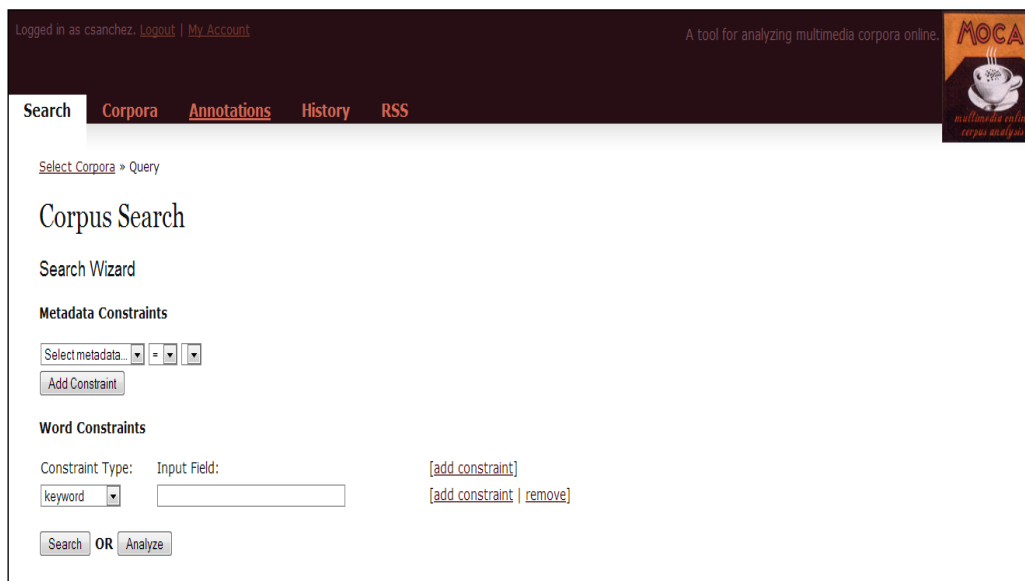


Figura 5. Posibilidades de búsqueda en el corpus

Así, una vez realizada la introducción de todos estos datos, en este caso referidos a los documentos transcritos paleográficamente del monasterio de Sahagún de los siglos IX al XII (*CDMSahagún*), por ejemplo, se puede hacer la búsqueda de todas las ocurrencias de la secuencia sintáctica *hunc meum factum*, en donde aparecen un determinante demostrativo, un adjetivo posesivo y un sustantivo, en la segunda mitad del siglo X, como se constata en la Figura 6:

The screenshot shows a web browser window titled "Multimedia Online Corpus Analysis - Mozilla Firefox". The address bar shows the URL "moca.ats.amherst.edu/search.php?action=results". The page has a navigation menu with "Search", "Corpora", "Annotations", "History", and "RSS". The main content area displays "Search Results for 'hunc meum factum'".

On the right side, there are two panels: "Toggle Metadata" with checkboxes for "fecha", "origen", "document", "escriba", "gnero", "siglo", and "quarters", all of which are checked; and "Annotation Schema Selector" with a dropdown menu set to "Select Schema".

The search results are displayed in a table-like format with columns for metadata and text snippets. The first result is:

+	view	edit	fecha	origen	document	escriba	gnero	siglo	quarters
			950-967 (véase nota)	Sahagún, León	#126, carp. 872, núm. 17, sección de clero, Sahagún; Becerro, fol. 59r, Sahagún	Servusdei	donación	10b	10b

The text snippet for this result is: "hominis stare nos faciat sane quod minime credo qui contra hunc meum factum ad disrumpendum venerit vel venerimus tam ex propinquis".

Other results in the list include entries from Sahagún, León, with document numbers #132, #141, and #146, and various dates and origins.

Figura 6. Lista de ejemplos según la búsqueda

En este caso, la lista de ejemplos obtenida de la combinación gramatical *hunc meum factum* no sólo aparece por defecto ordenada cronológicamente, sino que también presenta visualmente superpuestas todas las variables extralingüísticas comentadas más arriba, lo cual permite observar a primera vista si realmente existe una relación entre esta construcción y el origen determinado del documento, el tipo al que pertenece o el escriba que lo redactó, como se puede apreciar en la Figura 7:

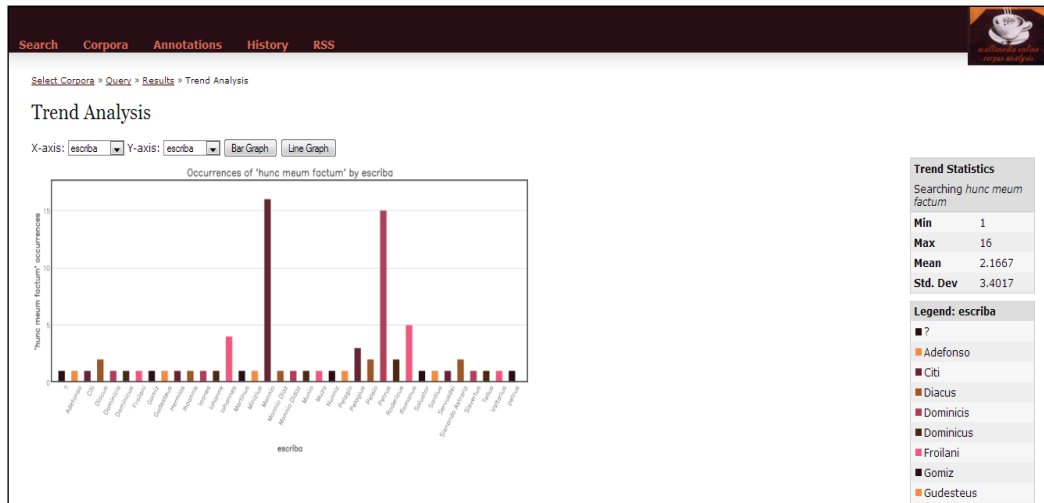


Figura 7. Representación gráfica del uso de esta construcción según los escribas

A partir de aquí se puede hacer, por tanto, una selección de datos únicamente en función de cualquier variable extralingüística según nuestra intuición investigadora, con el fin de confirmar o no su vinculación a una secuencia sintáctica concreta.

Además, el contexto en el que se incluye(n) la(s) forma(s) buscada(s) puede irse ampliando progresivamente, según las necesidades de información del propio investigador, a partir del signo + que aparece a la izquierda del ejemplo, hasta llegar a la visualización de la totalidad del documento (opción *view*), el cual incluso se puede llegar a editar (opción *edit*) a medida que se van realizando las diferentes consultas, si se desea completar alguna de las variables introducidas o corregir algún dato concreto, como se puede comprobar en la Figura 8:

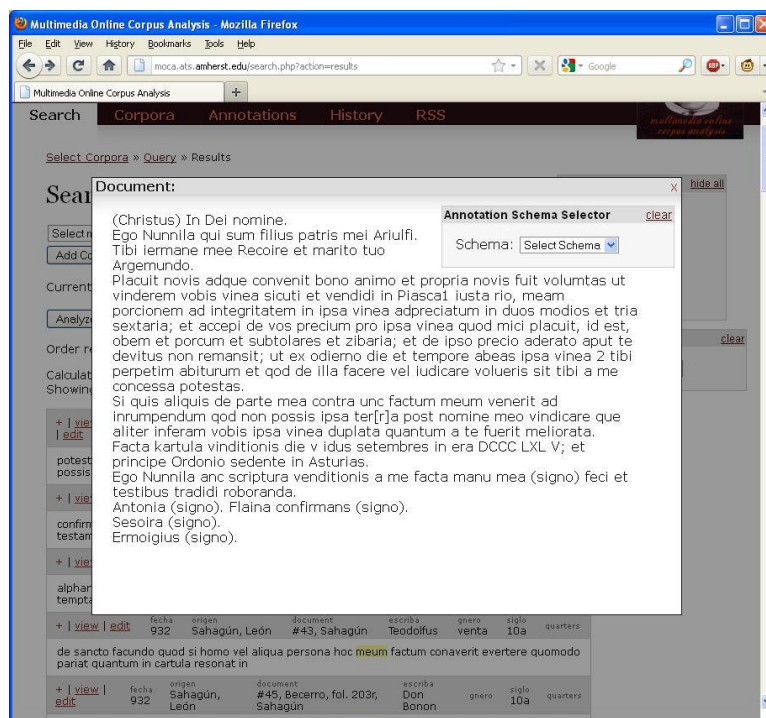


Figura 8. Búsqueda de ejemplos en su contexto

Hay que tener muy en cuenta que las listas de datos obtenidas con el empleo de herramientas informáticas no siempre son suficientes para poder realizar el análisis diacrónico de una determinada estructura gramatical, ya que tan importante es su presencia como su no aparición o contraste con otra clase de construcciones de igual o similar valor y función en un mismo documento. Por ello, la posibilidad de ir al texto completo siempre que ello sea necesario se muestra como una función que mejora y garantiza en gran manera el desarrollo de cualquier investigación.

Como ya se ha visto, del mismo modo que sucede con cualquier programa de concordancias, *MOCA* le permite al investigador realizar la búsqueda de cada ejemplo de una determinada palabra o una combinación sintáctica concreta (o sea, una colocación lingüística) en conjunto con su contexto original a fin de sacar sus estadísticas de uso. Sin embargo, *MOCA* posee otras ventajas que no tiene la mayoría de estos programas. Al tratarse de una colocación sintáctica, por una parte, también ofrece el denominado *cálculo de z* (*Z-Score*) de cada ejemplo, cifra que expresa la representatividad de esa estructura sintáctica en el corpus; y, por otra, además proporciona el cálculo de la *información mutua* (*Mutual Information*), la cual refleja las posibles interrelaciones existentes entre dos o más formas lingüísticas. En otras palabras, tanto el *cálculo de z* como el de la *información mutua* vienen siendo un índice mismo de la noción de *colocación*, aspecto fundamental a considerar cuando se trabaja sobre todo con documentos notariales. Estas funciones aparecen recogidas en la Figura 9:

The screenshot shows the MOCA web interface for a search query. The search criteria are: 'three word' collocations occurring 'more than' 0 times, sorted by 'frequency', and visualized with 'none'. The search term is 'meum factum'. The results table is as follows:

Collocation	Count	Mutual Information	Z-Score
hoc meum factum	94	0.0000	5.3603
hunc meum factum	88	0.0000	4.9945
meum factum uenerit	32	0.0000	1.5809
meum factum ad	24	0.0000	1.0933
meum factum sit	17	0.0000	0.6666
meum factum conauerit	12	0.0000	0.3618
meum factum infringere	9	0.0000	0.1789
meum factum semper	6	0.0000	-0.0039
meum factum rumpere	6	0.0000	-0.0039
meum factum irrumpere	4	0.0000	-0.1258
meum factum conf	4	0.0000	-0.1258

Figura 9. Cálculo de z y de la información mutua

Asimismo se puede representar de forma totalmente visual estos cálculos de frecuencia y distribución a través de un gráfico de tipo *nube* (*Tag Cloud*) o *relacional* (*Relationship Graph*). Esta manera de presentación nos permite ver a simple vista, por ejemplo, en función del tamaño creciente de las letras utilizado, cuáles son las combinaciones más habituales de una serie de vocablos en los textos analizados, como se puede observar muy claramente en la Figura 10:

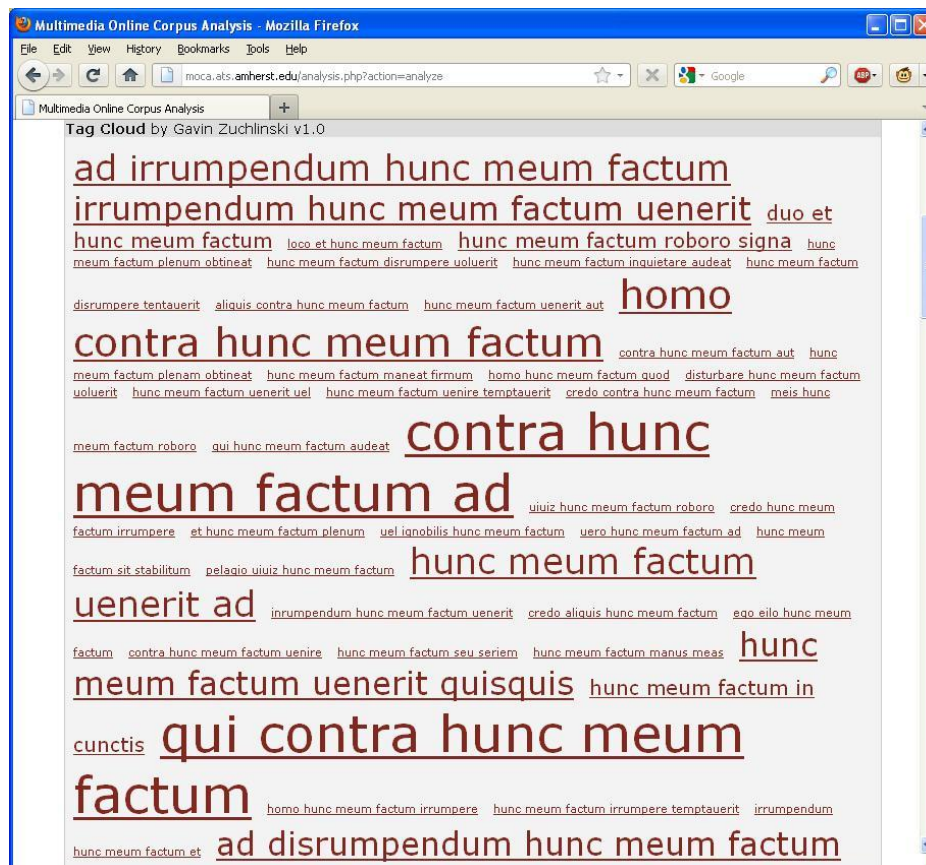


Figura 10. Representación visual de frecuencias según el esquema *nube*

Y también es posible estudiar qué relaciones establecen las palabras objeto de la búsqueda, que se hallan en una determinada construcción sintáctica, con el resto de vocablos del propio texto o del conjunto de documentos. De esta forma el investigador puede comprobar, con una simple ojeada, mediante el sistema de representación de gráfico relacional, la variedad total de estructuras en las que se encuentran los diferentes elementos lingüísticos que componen la combinación analizada, teniendo la posibilidad, en todo momento, de formar un nuevo gráfico relacional a partir de una de estas nuevas secuencias. Además, aparecen a continuación en orden de frecuencia todas estas construcciones, por lo que se le facilita al estudioso el acceso inmediato al conjunto de ejemplos de estas estructuras. De esta manera, una herramienta como *MOCA* posee unas técnicas de descubrimiento que permiten la observación de patrones sintácticos que de otro modo difícilmente se podrían constatar. Este sistema de visualización se encuentra representado en la Figura 11:

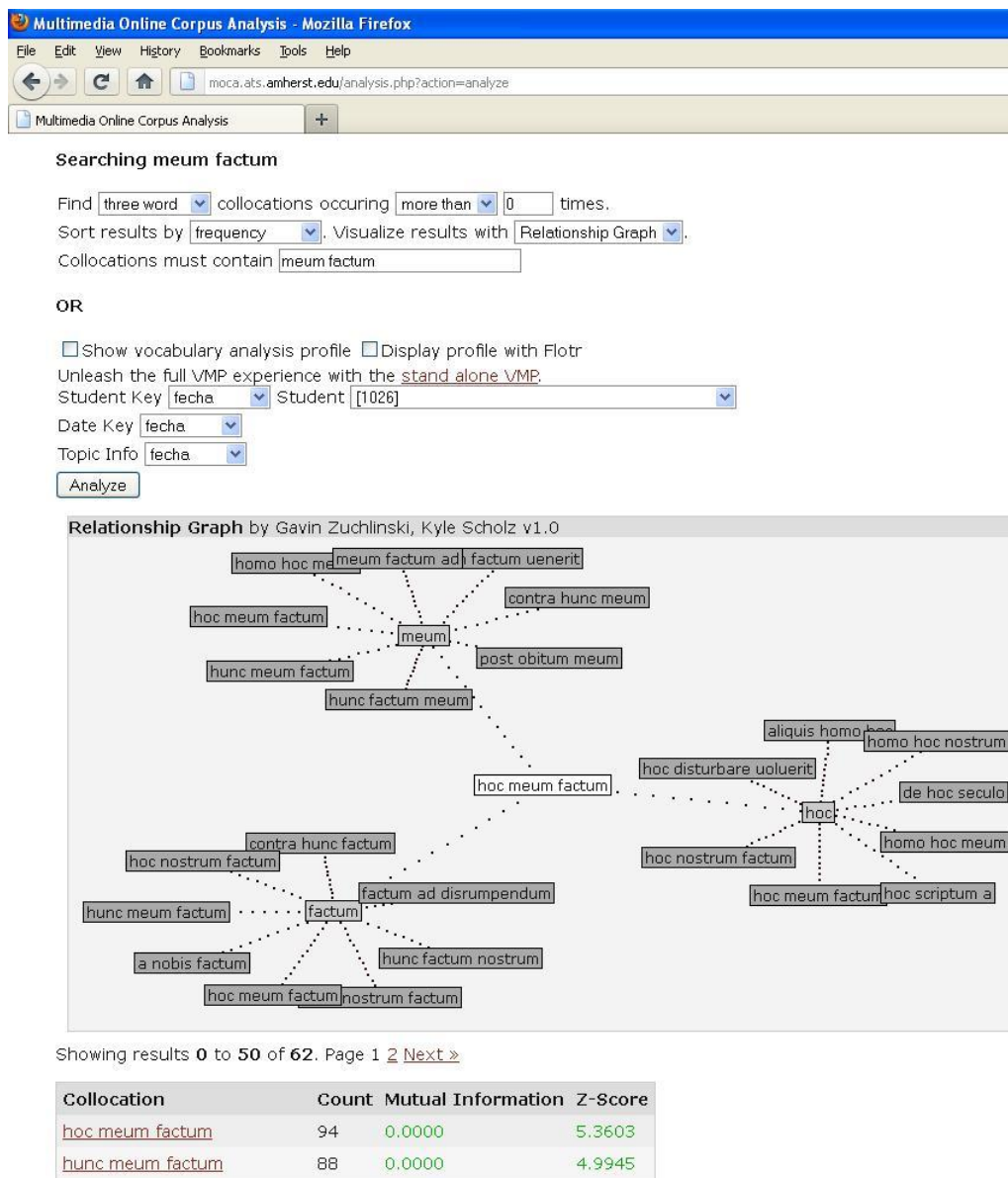


Figura 11. Representación visual de frecuencias según el esquema de *gráfico relacional*

Asimismo, se debe señalar que si los textos objeto del presente análisis estuvieran escritos con una ortografía reformada (o sea, como sucede con todo manuscrito después del siglo XII) y netamente castellana, MOCA podría llevar a cabo búsquedas a partir de las categorías gramaticales (p. ej., sustantivo, pronombre, verbo, adjetivo, adverbio, preposición, etc.) gracias a la existencia de un *lematizador* automático. Sin embargo, debido a la flexibilidad de los patrones escritos del Alto Medioevo (Emiliano 2005), sería difícil, si no imposible, diseñar el correspondiente lematizador para todo documento escrito antes del siglo XIII.

Finalmente, se debe comentar la posibilidad que posee el investigador de acceder a todo el historial de consultas realizadas en diferentes días y horas a través de la pestaña *History*. De esta forma, se puede ver fácilmente qué búsquedas concretas se han hecho y en

qué orden, lo cual constituye una gran ayuda en todo proceso de investigación que requiere un trabajo de varios días, con el fin de recuperar con cierta facilidad el hilo del análisis y poder así evitar repeticiones de análisis más o menos infructuosos.

3. UN EJEMPLO DE EXPLOTACIÓN DE UN CORPUS DIGITALIZADO: LA CONSTRUCCIÓN *DEMOSTRATIVO/ARTÍCULO + POSESIVO + SUSTANTIVO*

Como ejemplo del análisis propuesto, en el corpus documental estudiado, a pesar de que son muy poco frecuentes las combinaciones de demostrativo más posesivo seguido de un sustantivo, éstas presentan una interesante variación, como sucede con la construcción en donde aparece el posesivo singular de primera persona *meus, mea, meum* con los diferentes demostrativos o pronombres de identidad que encontramos en estos documentos (tanto con *illam, hunc, istum, ipsum* y sus otras variantes morfológicas). Así, en la combinación en donde se halla el posesivo *meus*, solo se documentan los siguientes datos en todo el corpus:

- (1) [...] et *iste meus scriptus plenus* abeat roborem facta series testamenti ipsas nonas decembris era millesima x^a regnante ranimiro prolis sanctioni [...]. (MOCA: 972, CDMSahagún)⁵.
- (2) [...] et *iste meus scriptus* maneat inconvulsus facta scriptura vii^o kalendas februarias era [...]. (MOCA: 973, CDMSahagún).
- (3) [...] et *isto meus scriptus* facta series testamenti notum die viiii kalendas decembris era [...]. (MOCA: 990, CDMSahagún).
- (4) [...] ea uidelicet racione ut *ipse meus uir* indutus monastico habitu in claustrum sancti facundi [...]. (MOCA: 1115, CDMSahagún).
- (5) [...] ut *ipse meus uir* acciperet ipsum monasterium sancti saluatoris cum omnibus suis adiunctionibus [...]. (MOCA: 1115, CDMSahagún).

En el caso de la variante *meum*, que aparece en un total de 484 ocurrencias en el corpus, a pesar de que la combinación mayoritaria es su posposición (*factum meum* 74 casos, *obitum meum* 64 casos, *corpus meum* 13 casos, etc.), como podría esperarse en un texto latino, sin embargo, la anteposición del posesivo al sustantivo es muy frecuente (*meum factum* en 188 veces, que equivale a un 39% del total de casos):

Collocation	Count
meum factum	188
hoc meum	127
hunc meum	97
factum meum	74
obitum meum	64
uero meum	17

⁵ El número que aparece en los ejemplos dentro del paréntesis corresponde siempre al año del documento.

corpus meum	13
meum donum	8
meum testamentum	7
meum obitum	7
ouitum meum	6
meum scriptum	5

Tabla 1. Combinación de la variante del posesivo *meum*

Además, es muy significativo que se combine también con el demostrativo latino de primer grado antepuesto seguido de un sustantivo (*hoc meum* en 127 casos y *hunc meum* en 97), valores que representarían el 46% del total de casos recogidos:

Collocation	Count
meum factum	188
hoc meum	127
hunc meum	97

Tabla 2. Frecuencia de la combinación de *meum*

Por otra parte, si tenemos en cuenta el tipo de nombre con el que se combinan el demostrativo y el posesivo en estos documentos, se observa que no se produce una gran variedad, ya que predomina básicamente el sustantivo *factum* (94 casos), seguido de unos pocos ejemplos de *donum* (7 casos), *testamentum* (6 casos), *regale* (como adjetivo) y *mandatum* (3 casos), *pactum* y *uoluntarium* (2 casos) y, finalmente, *scriptum*, *decretum* y *regium* (1 caso):

Collocation	Count
hoc meum factum	94
hoc meum donum	7
hoc meum testamentum	6
hoc meum regale	3
hoc meum mandatum	3
hoc meum pactum	2
hoc meum uoluntarium	2
hoc meum scriptum	1
hoc meum decretum	1
hoc meum regium	1

Tabla 3. Combinación de la variante *meum* con sustantivo

Se trata, como se puede comprobar, de un tipo de nombre relacionado semánticamente con el lenguaje propio de los actos administrativos legales en donde se produce una donación, una venta, etc., por lo que *factum* es obviamente el elemento más frecuente. Algunos ejemplos en donde aparecen estos sustantivos se recogen a continuación:

- (6) quod si aliquis homo *hoc meum testamentum* infringere temptaverit a deo sit excommunicatus per secula [...]. (MOCA: 921, CDMSahagún).
- (7) quod si homo vel aliqua persona *hoc meum factum* conaverit evertere quomodo pariat quantum in cartula resonat [...]. (MOCA: 932, CDMSahagún).
- (8) si quis igitur homo ex regali dignitate uel ex seruili conditione uel cuiuscumque generis homo *hoc meum regale uotum* aliquo modo infringere temptauerit sit extraneus a corpore christi et sanguine [...]. (MOCA: 1130, CDMSahagún).
- (9) Si quis uero *hoc meum factum* infringere temptauerit sit maledictus et excommunicatus; et pectet regie parti mille morabetinos Et *hoc meum pactum* semper maneat firmum [...]. (MOCA: 1153, CDMSahagún).
- (10) Si quis uero contra *hoc meum mandatum* in aliquo fecerit centum aureos regie parti in coto pectabit et monasterio Sancti Facundi dampnum quod intulerit duplicatum in coto pariter restituet [...]. (MOCA: 1181, CDMSahagún).
- (11) Si quis autem de parentela mea uel cuiuslibet generis fuerit qui *hoc meum donum* irritum facere uel mutare uoluerit sit maledictus et excommunicatus; (MOCA: 1185, CDMSahagún).

Un primer problema que se nos plantea es la consideración de si nos encontramos únicamente ante un lenguaje formulario, en donde la construcción sintáctica así como la existencia de una posible variación quedarán totalmente limitadas, por lo que los datos hallados de ninguna manera pudieran reflejar la lengua oral. Ciertamente, la repetición constante de un mismo tipo de sustantivo (*factum*) que se combina con una misma clase de demostrativo (*hoc*) y de posesivo (*meum*), nos puede llevar a pensar que estamos ante una pura fórmula que ya aparece en los primeros documentos y que se mantiene hasta las últimas fechas del corpus. Por ello, no es de extrañar que el demostrativo empleado sea el más cercano al latín clásico (*hic, haec, hoc*), el cual se sustituyó ya en latín tardío o vulgar por el de segundo grado (*iste, ista, istud*). Sin embargo, si realmente nos encontráramos ante un texto puramente latino escrito por alguien que domina (habla y escribe) la lengua latina, no debería ser posible hallar variación morfosintáctica en este demostrativo en construcciones en donde aparecieran los mismos elementos combinados, lo cual sucede, como se observa en la Tabla 4:

Collocation	Count
hunc meum factum	88
hunc meum scriptum	3
hunc meum testamentum	1
hunc meum textum	1
hunc meum decretum	1

Tabla 4. Combinación de *hunc meum* con sustantivo

Como se puede ver, la sustitución de la forma *hoc* (nominativo o acusativo singular neutro) por *hunc* (acusativo singular masculino) acompañando a sustantivos de género neutro ya desde los primeros documentos, indica el reflejo en la lengua escrita de la desgramaticalización del neutro en la evolución del latín a las lenguas romances. Además, se produce con los mismos sustantivos que en el caso anterior y en proporciones muy similares:

- (12) [...] abzuleman fredinandiz confirmat; sisebutu petriz confirmat ego taion *hunc meum testamentum* quem fieri iussi cum propria manu roboravi signo [...]. (MOCA: 921, CDMSahagún).
- (13) [...] hominis stare nos faciat sane quod minime credo qui contra *hunc meum factum* ad dirumpendum venerit vel venerimus tam ex propinquis [...]. (MOCA: 950-967, CDMSahagún).
- (14) si quis vero *hunc meum textum* scripture ausu temerario irrumpere conaverit vel ex his rebus auferre alienare [...]. (MOCA: 956, CDMSahagún).
- (15) quod absit contra *hunc meum decretum* ad inrumpendum uenerit aut in alia parte usurpare uoluerit siue rex [...]. (MOCA: 1020, CDMSahagún).
- (16) [...] et hec scriptura maneat firma facta carta vi^o kalendas februaryi era i cviii^a regnante rege adefonso in legione ego auolualiti *hunc meum scriptum* roboro signum pro testes petro diaco citi testes monnio notuit [...]. (MOCA: 1071, CDMSahagún).
- (17) quod si aliquis homo de gens mea uel de extranea ad irrumpendum *hunc meum factum* uenerit uel uenero quisquis fuerit qui talia comiserit inprimis sit excommunicatus [...]. (MOCA: 1115, CDMSahagún).

Un segundo problema tiene que ver con la variación del tipo de demostrativo que se halla en esta clase de construcción más o menos formularia. Si solo es posible la presencia de un determinado adjetivo demostrativo en todos los documentos, realmente estaríamos ante una fijación formularia alejada de la lengua romance oral de la época, dada la distancia existente entre el demostrativo latino de primer grado (*hic, haec, hoc*) y el correspondiente demostrativo romance de primer grado que vino a sustituirlo (*este, esta, esto*). Los ejemplos en donde el posesivo *meus* (en sus variantes *meus* y *meum*) es precedido por otro tipo de demostrativo son muy escasos en nuestro corpus y de fechas más o menos tardías:

- (18) ut tam eglesias quam villas vel quantum in *isto meum scriptum* adnotare disposui donationem recessvindi abbati vel sanctorum facundi [...]. (MOCA: 938, CDMSahagún).
- (19) et *iste meus scriptus plenus* abeat roborem facta series testamenti ipsas nonas [...]. (MOCA: 972, CDMSahagún).
- (20) et *iste meus scriptus* maneat inconvulsus facta scriptura vii^o kalendas februaryas era [...]. (MOCA: 973, CDMSahagún).
- (21) [...] tum relinquis ad filis tuis et si alicus omo *istu meum factum* derumpere quesierit uel suocita persona asertoris uenerit [...]. (MOCA: 1046, CDMSahagún).
- (22) et si aliquis homo uiuente super terram qui ad *istum meum factum* ad inrumpendum uenerit uel uenerimus tam ego quam [...]. (MOCA: [1124], CDMSahagún).

- (23) Si quis uero ex meo uel alieno genere *istud meum donum* infringere uoluerit sit a Deo maledictus et excommunicatus [...]. (*MOCA*: 1168, *CDMSahagún*).

Se trata de apenas 6 casos en todo el corpus, pero no por ello dejan de ser muy significativos, ya que se combinan con la misma clase de sustantivos que antes, pero aquí el demostrativo presenta una mayor variación genérica y casual. Por otra parte, la presencia de este demostrativo latino etimológicamente de segundo grado (*iste, ista, istud*) en contextos en donde aparece usualmente un demostrativo latino de primer grado (*hic, haec, hoc*), no deja lugar a dudas del reflejo en la lengua escrita latina de una lengua oral ¿romance? que ha realizado una reformulación del sistema de los pronombres demostrativos. A ello hay que añadir que también es posible encontrar, aunque solo sea un único ejemplo, el pronombre de identidad latino *ipse, ipsa, ipsum*, que evolucionará como pronombre de segundo grado romance (*ese, esa, eso*), combinado con un posesivo y un sustantivo, como se puede observar en (24):

- (24) [...] in alio loco in villa que dicunt calzata *ipsos meos molinos* cum suis aqueductiles uel quicquid ad eum pertinet et *ipsum meum pomerium* qui ibidem est et post discessu namque meum omnia mea facultas ubique potueritis inuenire uel adplicare licitum abeatis [...]. (*MOCA*: 986, *CDMSahagún*).

Aparte del uso específico del pronombre *ipse*, con sus diferentes variaciones morfológicas, hay que destacar su combinación con un tipo de sustantivos distintos a los que aparecían en los ejemplos anteriores, de posible carácter formulario. En este caso, la estructura posesiva hace referencia a una serie de elementos físicos, posesiones, objetos, etc., que se van a donar o transmitir. Téngase en cuenta cómo la presencia de un sustantivo con un valor semántico diferente, más próximo al mundo real del hablante de la época, propicia la utilización de un demostrativo mucho más alejado de los usos canónicos latinos. Sin embargo, el empleo de este demostrativo en concreto, ¿realmente está expresando un demostrativo romance de segundo grado (*ese, esa, eso*), cuando en el resto de ejemplos la forma preferida es la de primer grado (*este, esta, esto*), aunque sea a través de su forma latina clásica (*hic, haec, hoc*)?

Ante estas dudas, parece lógico realizar un análisis en nuestro corpus de documentos de un determinado sustantivo que posea un valor concreto, no abstracto, que exprese un objeto físico y que tenga una frecuencia de uso muy elevada, como es el caso del sustantivo *casa* (*casa* en 77 ocasiones, *casam* en 11 ocurrencias y *casas* en 151 casos, con un total de 239 ejemplos). Los resultados obtenidos resultan muy interesantes y significativos, ya que se halla una gran variedad de combinaciones:

Collocation	Count
ipsa casa	12
illa casa	8
mea casa	6
alia casa	6

vestra casa	2
casa vestra	2
nostra casa	2
casa tota	2
casa mea	2
la casa	2
casa nostra	1
casa illa	1
ista casa	1

Tabla 5. Combinación del sustantivo *casa*

Collocation	Count
meam casam	2
illam casam	2
casam alia	1
ipsam casam	1
uestram casam	1
casam meam	1
talem casam	1

Tabla 6. Combinación del sustantivo *casam* en acusativo

Collocation	Count
suas casas	29
illas casas	10
meas casas	6
ipsas casas	4

Tabla 7. Combinación del sustantivo plural *casas*

Así, se constata en primer lugar que los demostrativos que preceden a este sustantivo son *ipsa(m/s)* (13 casos en singular y 4 en plural), *illa(m/s)* (10 casos en singular y 10 en plural) e *ista* (1 caso en singular), apareciendo en 2 casos la forma de artículo *la* también con el singular, pero en ninguna ocasión el demostrativo latino de primer grado *haec* o *hae*, como se puede comprobar en los siguientes ejemplos:

- (25) [...] adicimus quoque ad *ipsam casam* alia villa in ataula in fonte de ozva et [...].
(MOCA: 959, CDMSahagún).

- (26) ut sit cuncta que adnotata invenimus concessa et confirmata in *ipsa casa* de sanctum salvatoris post partem sanctorum facundi [...]. (MOCA: 960, CDMSahagún).
- (27) et ipsa diuisa fuit de filio meo iohannes et in pisorga dono tibi ipso solare ubi sedes ouecco bellitiz cum *ipsas casas* de suo ierno citi et [...]. (MOCA: 1077, CDMSahagún).
- (28) ut teneamus ipsas hereditates supradictas in uita nostra; post mortem uero nostram accepiatis ipsas hereditates et *illas casas* quas habemus in sancto facundo ad portam de illo monasterio cum omnibus [...]. (MOCA: 1101, CDMSahagún).
- (29) et libero ad *illa casa* de sancto saluatore et ad domno bernardo priore qui ibi per manum est de domno didaco abbate sancti facundi [...]. (MOCA: 1105, CDMSahagún).
- (30) et uno consensu uendidimus tibi unam kasam in hac uilla et fuit *ista casa* de arnulfo qui sine filiis ex hac uita discessit ideoque domus ipsa uenit in potestate domini diaci abbatis qui [...]. (MOCA: 1113, CDMSahagún).
- (31) [...] nec in uadimonio ponam Offero predictis martiribus Facundo et Primitiuo ad *illam casam* de coquina fratrum ut post obitum meum iure hereditario illa possideat [...]. (MOCA: 1181, CDMSahagún).
- (32) Domingo Pedrez Senior Pedro Pelaez en *la casa* de Uillada Merino: (MOCA: 1186, CDMSahagún).
- (33) Etsi non habemus potestate / de altra quincta dare, recebidos seamos en *la casa*, post obitum / nostrum. (MOCA: 1187, CDMSahagún).

Resulta muy interesante observar la convivencia en un mismo fragmento u oración de diferentes tipos de demostrativos para expresar en teoría relaciones deícticas muy semejantes, por no decir idénticas, de tal manera que es posible hallar formas más cercanas al romance (*ipsas hereditates* (28); *ista casa* (30)) junto a construcciones o usos más propios del latín (*illas casas* (28); *in hac uilla* (30)). Como antes se ha comentado, esta convivencia de uno y otro tipo es un buen indicador o un indicio de un continuo y de la coexistencia de dos tipos de lengua claramente distintas en una misma época, por lo que en cierto modo podríamos afirmar que esa necesidad de dar a este tipo de textos una apariencia latina, en muchas ocasiones se queda simplemente en eso, ya que con bastante frecuencia estamos asistiendo a la introducción de elementos gramaticales pertenecientes a un claro romance oral. Por otro lado, debe notarse que la mayoría de los ejemplos con el demostrativo *illa(m/s)* (12 del total) al igual que los del artículo *la* y el único caso de *ista* son del siglo XII, mientras que los ejemplos de *ipsa(m/s)* (7 del total) son del siglo X. Estos datos podrían hacernos pensar que como el artículo romance proviene de la gramaticalización del pronombre demostrativo latino de tercer grado (*ille, illa, illud*), el escriba simplemente se limita a pasar de una lengua oral romance a otra lengua escrita latina que conoce con mayor o menor rigor, relacionando dos elementos lingüísticos que ya lo estaban de forma etimológica. Y si en épocas anteriores la forma elegida mayoritariamente era *ipsa*, en cierto modo podría obedecer a un mayor desconocimiento de la lengua latina por parte del escriba que le llevaría a buscar la forma más semejante a la lengua oral de su tiempo.

En segundo lugar, en cuanto a la presencia única de un posesivo, ésta se da en contadas ocasiones y presenta variación posicional, predominando claramente la anteposición con 48 ocurrencias (*suas casas* en 29 ejemplos con un 60,40% del total de casos) frente a la posposición en 6 ocasiones:

- (34) et accepimus de uos uno orto iuxta *nostra casa* in uillapezennim in era i c xxxviii^a [...]. (MOCA: [1110], CDMSahagún).
- (35) et tali causa facio vobis cartam de terra mea propria quam habeo in loco que dicitur noantica ubi sunt *vestras casas*; et est determinata primo termino per *mea casa* et [...]. (MOCA: 962, CDMSahagún).
- (36) et post obitum nostrum ab integro uobis concedimus cum terris et uineis pratis et pascuis paludibus montes fontes molinarias piscarias cultum uel incultum corte cum *suas casas* et suas cupas et [...]. (MOCA: 1021, CDMSahagún).
- (37) Si uero frater ille pactum quod scripsimus mihi attendere recusauerit cum predicta *casa mea* et cum omnibus que memorauit libere [...]. (MOCA: 1181, CDMSahagún).
- (38) ego arella christi ancilla offero ad ipsos sanctos super *casa uestra* villare pro anima mea et terras et vineas quos habeo subtus villa mohore et super *casa uestra* villare; (MOCA: 994, CDMSahagún).
- (39) quod domino dignanter valebunt offerre in primis quoque offerimus in loco predictu ac martiribus superius fatis *casa nostra* ad integro uidelicet sancti andre in villa de mottarafi [...]. (MOCA: 959, CDMSahagún).

En principio, no parece existir una explicación sintáctica a partir de los ejemplos encontrados que dé cuenta de la preferencia por un tipo u otro de orden, por lo que resulta más coherente considerar que tal predilección obedezca al reflejo de una sintaxis oral romance que a un deseo de escribir correctamente la lengua latina. Finalmente, si solo se tienen en cuenta aquellos casos en que un demostrativo se combina con un posesivo y con el sustantivo *casa(m/s)*, descubrimos que se trata de una construcción que se halla en muy pocas ocasiones, pero que presenta unas ciertas peculiaridades, como se puede observar en los siguientes ejemplos:

- (40) [...] concedo ad perabendum uel seruiendum adicio aduc *illas meas casas* quem abeo in bustello cum suis terris uineis pratis padulibus montis fontis exitus uel regressus [...]. (MOCA: 1050, CDMSahagún).
- (41) et facio uobis kartam testamenti et priuilegii firmitatis de *illa mea casa tota integra* que est circa ecclesiam sancti tirsi [...]. (MOCA: 1123, CDMSahagún).
- (42) et ad ipsum monasterium sancti facundi et dono ibi *illas meas casas totas* et circa illas xi solares populatos; (MOCA: 1143, CDMSahagún).
- (43) ut semper possideat illa absque ulla contradicione monasterium domnis sanctis ad obsequium monachorum ibi degencium adicio eciam in uillauincencii *illas meas casas* que sunt iuxta ecclesiam sancte marie [...]. (MOCA: 1146, CDMSahagún).
- (44) ego Urracha spontanea uoluntate et equo animo offero Deo sanctisque martiribus Facundo atque Primitiuo pro remedio anime mee *illam meam casam* quam ex proprio censu una cum marito meo Petro de Furones iam defuncto lucrata sum [...]. (MOCA: 1181, CDMSahagún).
- (45) Matela pernominato in ripa de Aradoy vntre Castronouo et Merendela que conparamus de Petro Gonzaluez et de sua mulier Exemena Diaz et de suo irmano Gil Gonzaluez; *elas suas casas* ela media torre et duos solares circa que tenet se ela casa et suo orto que stetit ontre ela sua porta et rio; (MOCA: 1185, CDMSahagún).

De un total de 48 ejemplos con posesivo antepuesto al sustantivo, únicamente en 6 ocasiones (1 del siglo XI y 5 del siglo XII) esta combinación se encuentra precedida solo por un demostrativo latino de tercer grado (ejemplos (40)-(44)) o por la forma propia del artículo romance (*elas suas casas*, en (45)), a pesar de las distintas posibilidades que hemos visto anteriormente. Por otra parte, esta estructura puede llevar pospuesto un cuantificador, como sucede en (41) y (42) con *illa mea casa tota integra* e *illas meas casas totas*, respectivamente, construcciones completamente anómalas en romance. Sin embargo, resulta curiosa la presencia del cuantificador universal *totus* en este tipo de estructuras con posesivo, ya que se trata de la forma que ha triunfado en romance frente a la predominante en latín *omnis*:

- (46) et si uos quesieritis ambulare de uestros domos damus ad uobis foro et plazo que leuetis *totam uestram casam* ad caput de viiii^o dies qui leuetis uestras portas et uestros uscios et illas kasas et illas bardas non defacatisa; (MOCA: 1092, CDMSahagún).
- (47) si uero uenerit in particione de suos filios *tota ista mea hereditate* sicut in testamento resonat post mortem meam [...]. (MOCA: 1062, CDMSahagún).

En estos ejemplos, el cuantificador se puede unir directamente al sintagma formado por el posesivo y el sustantivo (*totam uestram casam*, en (46)), tal y como sucede en romance, o puede preceder a toda la construcción introducida también por un demostrativo (*tota ista mea hereditate*, en (47)), aunque no se trate del sustantivo *casa*.

Llegados a este punto, la pregunta que nos hacemos es si todas estas variantes estructurales en donde aparece combinado un demostrativo con un posesivo y un sustantivo tienen, en primer lugar, alguna relación entre sí, un común denominador que las caracterice y, en segundo lugar, si son el reflejo de la existencia de una nueva categoría gramatical, el artículo romance, que, entre otras posibilidades, podía sustituir al demostrativo en este tipo de construcciones y preceder al posesivo. En relación con la primera cuestión, parece lógico llegar a pensar que sí existe una relación entre esta clase de combinaciones y que los diferentes tipos de demostrativos que se encuentran obedecen más a razones de cambio en la escritura, fruto de las reformas ortográficas carolingias, que a una elección consciente de un determinado elemento. Por ello, resulta muy significativa no solo la convivencia de formas clásicas del demostrativo de primer grado (*hic, haec, hoc*) junto a otras del pronombre de identidad (*ipse, ipsa, ipsum*) y del demostrativo de segundo grado del latín reformulado como primer grado en romance (*iste, ista, istud*), sino también la posibilidad de preceder el mismo tipo de sintagmas nominales. Si se admite, por consiguiente, que no son tan distintas entre sí estas combinaciones, se puede llegar a pensar que en realidad se trata de una única estructura en la que diferentes escribas de distintas épocas con diferentes conocimientos de la lengua latina intentan expresar o reflejar la presencia en la lengua oral de una construcción sintáctica muy concreta (*demostrativo/artículo + posesivo + sustantivo*) y lo hacen de la mejor manera posible, que es buscando un correlato latino a una nueva categoría gramatical propia del romance oral. Como es de suponer que los escribas carecían de conocimientos de gramática histórica, lo que predomina en todos estos casos es la búsqueda de un elemento que pueda expresar la deixis en su forma más general. De ahí que sea posible encontrar fácilmente ejemplos de *illam, hunc, istum, ipsum* y sus otras variantes

morfológicas intercambiados en los mismos textos y contextos (*iste meus scriptus, hoc meum factum, ipse meus uir, ipsos meos molinos, illas meas casas, elas suas casas*).

Y respecto a la segunda cuestión, esto no significa que siempre que tengamos una construcción de esta clase se deba abandonar la lectura demostrativa o deíctica y tenga que ser sustituida por un artículo, ya que carecemos de pruebas suficientes que demuestren este hecho. Sin embargo, la variación de estas formas durante varios siglos en este tipo de textos así como su concreción a finales del siglo XII podrían hacernos ver, como ya señaló Lapesa (1961), que durante un tiempo la interpretación de un pronombre demostrativo podía hacerse a través del artículo, en tanto en cuanto expresaba la deíxis en forma neutra y no hacía referencia a un grado gramatical concreto. Por ello, no sería imposible considerar en algún contexto los ejemplos anteriores con valor más general (*el mi escrito, el mi hecho, el mi esposo, los mis molinos, las mis casas, las sus casas*), sin que el documento sufriera un cambio interpretativo sustancial. De este modo, la presencia de la estructura *artículo + posesivo + sustantivo* en romance ya desde las *Glosas emilianenses* en el siglo X y a partir de ese momento en los textos literarios, tendría un posible correlato en todo este conjunto de textos notariales, no literarios, considerados exclusivamente latinos hasta no hace mucho.

4. CONCLUSIÓN

Se sabe que la variación, siempre presente en el continuo sociolingüístico, precede a cualquier cambio en el sistema, lo cual viene siendo la conversión de una *regla variable* en una aplicación obligatoria en un contexto determinado. A lo largo de este estudio, se han buscado los antecedentes de la estructura medieval *artículo + posesivo + sustantivo*, que aparece en el verso tan conocido «*delos sos oios tan fuerte mientre lorando*», en un corpus de textos medievales no literarios, de tipo documental, de los siglos IX al XII del monasterio de Sahagún, mediante la herramienta informática de análisis de corpus digitalizados *MOCA (Multimedia Online Corpus Analysis)*. Así, del estudio de la combinación en el sintagma nominal de los posesivos en sus diversas variantes, al igual que de los demostrativos, en relación con el núcleo nominal, hemos podido constatar, gracias al empleo de la herramienta en línea *MOCA*, en primer lugar, una evolución a través de los textos latinizantes que nos ha permitido ver su aparición en la posición de determinante o de complemento del nombre de forma más frecuente a lo largo de los siglos; y se ha podido observar, en segundo lugar, el inicio de una relación entre estos dos elementos que puede dar indicios de la existencia, ya en este periodo, de una de las estructuras gramaticales más características del castellano medieval como es la formada por el artículo (procedente de la gramaticalización del demostrativo de tercer grado *ille*) + el posesivo en forma adjetiva + el sustantivo, dada la ausencia de concordancia casual así como la gran variabilidad de formas demostrativas latinas o romances presentes. Esta combinación sintáctica, marcada en términos de deíxis, también exclusiva y caracterizadora del romance medieval, que se desgramaticaliza a finales del siglo XV, pero que pasa a ser propia según las gramáticas del habla popular y rústica durante siglos, situándola en la actualidad como un rasgo dialectal en zonas de Asturias, León y Santander, que no se da en todas las lenguas románicas (se ha conservado en catalán e italiano actuales), sin embargo, se encuentra ya en las *Glosas* como primera

documentación, por lo que su estudio en un corpus de documentos medievales notariales de los siglos IX al XII como el que nos ocupa arroja luz sobre su origen y su temprana formación.

REFERENCIAS BIBLIOGRÁFICAS

- BASTARDAS PARERA, Juan (1953): *Particularidades sintácticas del latín medieval (cartularios españoles de los siglos VIII al XI)*. Barcelona: CSIC/Escuela de Filología.
- 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: Iberoamericana/Vervuert, pp. 327-342.
- COMPANY, Concepción (2009): «Artículo + posesivo + sustantivo y estructuras afines», en Concepción Company (dir.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal*, vol. 1. México D. F.: Universidad Nacional Autónoma de México/Fondo de Cultura Económica, pp. 761-880.
- CORDE = REAL ACADEMIA ESPAÑOLA: *Corpus diacrónico del español*. <http://corpus.rae.es/cordenet.html>.
- DAVIES, Mark: *Corpus del español*. <http://www.corpusdelespanol.org>.
- EMILIANO, Antonio (2005): «Representational Models vs. Operational Models of Literacy in Latin-Romance Legal Documents», en Roger Wright y Peter Ricketts (eds.): *Studies on Ibero-Romance Linguistics Dedicated to Ralph Penny*. Newark, Delaware: Juan de la Cuesta, pp. 17-58.
- GIMENO MENÉNDEZ, Francisco (1995): *Sociolingüística histórica (Siglos X-XII)*. Madrid: Visor Libros (*Biblioteca Filológica Hispana*, 25).
- KOCH, Peter y Wulf OESTERREICHER (2007 [1990]): *Lengua hablada en la romanía: español, francés, italiano*. Versión española revisada, actualizada y ampliada por los autores. Madrid: Gredos (*BRH, Estudios y Ensayos*, 448) [Traducción española por Araceli López].
- LAPESA, Rafael (1961): «Del demostrativo al artículo», *Nueva Revista de Filología Hispánica*, 15, pp. 23-44.
- LAPESA, Rafael (1971): «Sobre el artículo ante posesivo en castellano antiguo», en Eugenio Coseriu y Wolf-Dieter Stempel (eds.), *Sprache und Geschichte: Festschrift für Harri Meier zum 65*. München: Wilhelm Fink Verlag, pp. 277-296.
- LAPESA, Rafael (1981): *Historia de la lengua española*. Madrid: Gredos (*BRH, Manuales*, 45) [9ª edición corregida y aumentada].
- MOCA = PAYNE, Scott (adm.): *Multimedia Online Corpus Analysis*. <http://moca.ats.amherst.edu> [Consulta: 01/06/2010-31/08/2010].
- PENNY, Ralph (2003): «Ambigüedad grafemática: correspondencia entre fonemas y grafemas en los textos peninsulares anteriores al siglo XIII», en Hermógenes Perdiguero (ed.), *Lengua romance en textos latinos de la Edad Media: sobre los orígenes del castellano escrito*. Burgos: Universidad de Burgos/Instituto Castellano Leonés de la Lengua, pp. 221-228.
- PENNY, Ralph (2004 [2000]): *Variación y cambio en español*. Madrid: Gredos (*BRH, Estudios y Ensayos*, 438) [Traducción española por Juan Sánchez Méndez].
- PENNY, Ralph (2006): «What Did Sociolinguistics Ever Do for Language History? The Contribution of Sociolinguistic Theory to the Diachronic Study of Spanish», en Clare Mar-Molinero y Miranda Stewart (eds.), *Language Variation and Change: Historical and Contemporary Perspectives*. Amsterdam: John Benjamins, pp. 49-62 (Special issue of *Spanish in Context*, 3, 1).
- TUTEN, Donald N. (2003): *Koineization in Medieval Spanish*. Berlin: Mouton de Gruyter (*Contributions to the Sociology of Language*, 88).

- VELÁZQUEZ-MENDOZA, Omar y Robert BLAKE (2010): «Temprana marcación acusativa en el hispanorromance», *Romance Philology*, 64, pp. 285-294.
- WEINREICH, Uriel, William LABOV y Marvin I. HERZOG (1968): «Empirical Foundations for a Theory of Language Change», en Winfred Philipp Lehmann y Yakov Malkiel (eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press, pp. 95-195.
- WRIGHT, Roger (1989 [1982]): *Latín tardío y romance temprano en España y la Francia carolingia*. Madrid: Gredos (*BRH, Estudios y Ensayos*, 369) [Traducción española por Rosa Lalor].
- WRIGHT, Roger (1992): «La metalingüística del siglo XII español (y la *Chronica Adefonsi Imperatoris*)», en Manuel Ariza, Rafael Cano, José María Mendoza y Antonio Narbona (eds.), *Actas del II Congreso Internacional de Historia de la Lengua Española*, vol. II. Madrid: Pabellón de España, pp. 879-86.
- WRIGHT, Roger (1993): «Complex Monolingualism in Early Romance», en William J. Ashby, Marianne Mithun, Giorgio Perissinotto y Eduardo Raposo (eds.), *Linguistic Perspectives on the Romance Languages: Selected Papers from the 21st Linguistic Symposium on Romance Languages (LSRL XXI), Santa Barbara, California, 21-24 February 1991*. Amsterdam: John Benjamins, pp. 377-388 (*Amsterdam Studies in the Theory and History of Linguistic Science*, IV. *Current Issues in Linguistic Theory*, 103).
- WRIGHT, Roger (2002): *A Sociophilological Study of Late Latin*. Turnhout: Brepols (*Utrecht Studies in Medieval Literacy*, 10).
- WRIGHT, Roger (2009): «Bidialectalism or the Continuum? (In Four Different Millennia)», en Sanda Reinheimer Rîpeanu (ed.), *Studia Lingvistica in honorem Mariae Manoliu*. Bucharest: Editura Universitatii din Bucuresti, pp. 386-397.

FUENTES BIBLIOGRÁFICAS

- CDMSahagún1 = MÍNGUEZ FERNÁNDEZ, José María (1976): *Colección diplomática del monasterio de Sahagún (siglos IX-X)*, vol. I (857-999). León: Centro de Estudios e Investigación San Isidoro (*Fuentes y estudios de historia leonesa*, 17).
- CDMSahagún2 = HERRERO DE LA FUENTE, Marta (1988a): *Colección diplomática del monasterio de Sahagún (857-1230)*, vol. II (1000-1073). León: Centro de Estudios e Investigación San Isidoro (*Fuentes y estudios de historia leonesa*, 36).
- CDMSahagún3 = HERRERO DE LA FUENTE, Marta (1988b): *Colección diplomática del monasterio de Sahagún (857-1230)*, vol. III (1074-1109). León: Centro de Estudios e Investigación San Isidoro (*Fuentes y estudios de historia leonesa*, 37).
- CDMSahagún4 = FERNÁNDEZ FLÓREZ, José Antonio (1991): *Colección diplomática del monasterio de Sahagún (857-1300)*, vol. IV (1110-1199). León: Centro de Estudios e Investigación San Isidoro (*Fuentes y estudios de historia leonesa*, 38).