

MODELOS ESTADÍSTICOS PARA EL ANÁLISIS FUNCIONAL DE LOS PROCESOS DE GRAMATICALIZACIÓN EN LAS CONSTRUCCIONES PROGRESIVAS DEL INGLÉS¹

STATISTICAL MODELS FOR THE FUNCTIONAL ANALYSIS OF GRAMMATICALIZATION PHENOMENA IN THE ENGLISH PROGRESSIVE CONSTRUCTION

CARLOS VAN ARKEL-SIMÓN²

Universidad Nacional de Educación a Distancia (UNED)

cvanarkel@la-laguna.uned.es

RESUMEN

En este artículo se analizan las relaciones cuantitativas existentes entre las construcciones progresivas del inglés y diferentes patrones funcionales, cuantificables a partir del conjunto de corpus *York-Toronto-Helsinki-Penn*, en aras de ilustrar los procesos de gramaticalización que experimentan estas construcciones. Se elabora un análisis estadístico que prueba la existencia de relaciones significativas entre los aumentos en las frecuencias de aparición normalizadas de las construcciones progresivas y la evolución de cuatro dimensiones funcionales del modelo multifactorial, así como entre los aumentos en el indicador estimado de los procesos de gramaticalización y tres de dichas dimensiones, en ambos casos, en función del género textual. Asimismo, se describen las principales tendencias de las formas progresivas totales y cinco de sus variantes sintácticas en función de las dimensiones relevantes del modelo multifactorial, así como el período de la lengua.

¹ El contenido de este artículo ha sido elaborado a partir del manuscrito de mi tesis doctoral *El cambio gramatical en las construcciones progresivas del inglés* que se encuentra pendiente de defensa pública.

² Doctorando de la Escuela Internacional de Doctorado y Profesor-Tutor de la UNED.

PALABRAS CLAVE: lingüística histórica inglesa; gramaticalización; lingüística del corpus; metodología cuantitativa; análisis multidimensional

ABSTRACT

In this paper we examine the quantitative relations between English progressive constructions and different functional patterns, measurable on the family of corpora *York-Toronto-Helsinki-Penn*, in order to illustrate the grammaticalization processes undergone by these constructions. We conduct a statistical analysis that demonstrates significant relationships between the increments in the standardized frequencies of progressive constructions and the changing patterns of four functional dimensions from the multidimensional model, and between the increase in the estimated index of grammaticalization and three of these dimensions, in both cases, according to textual genre. We also describe the main tendencies of the total number of progressive forms and five of its syntactic variants related to the relevant dimensions of the multidimensional model and the historical period.

KEYWORDS: English Historical Linguistics; Grammaticalization; Corpus Linguistics; quantitative methods; Multidimensional Analysis

1. INTRODUCCIÓN

Las principales hipótesis sobre la evolución de las construcciones progresivas del inglés indican que las variaciones en la distribución funcional de las formas progresivas permiten explicar gran parte de los procesos de cambio gramatical que ha experimentado la construcción (Smith 2005: 32-36). Estos procesos de cambio lingüístico se identifican, tradicionalmente, con procesos de gramaticalización (Fisher 1992: 254; Traugott 1992: 188). En la gramaticalización, la premisa básica es que el proceso está asociado a un incremento en la frecuencia de aparición de las estructuras que la experimentan, ya que la gramaticalización se acompaña de diseminación y expansión contextual.

Brinton y Traugott (2005: 110) mantienen que la expansión contextual se relaciona con un aumento de la frecuencia de aparición de las clases y de los tipos. El aumento en la frecuencia de aparición se convierte en un acelerador del propio proceso de gramaticalización. Es decir, el aumento de la ratio clase-tipo indica que ciertos elementos léxicos presentan mayor frecuencia de aparición, lo que posibilita la gramaticalización, al tiempo que es un indicativo de ella (Bybee 2003; Hopper y Traugott 2003: 126-130). Por tanto, en estos procesos, se asume que los aumentos en las relaciones clase-tipo son concomitantes con la aparición de las diferentes formas progresivas en un número mayor de contextos semánticos.

Los estudios más recientes sobre las construcciones progresivas del inglés (Kranich 2010; Núñez-Pertejo 2004; Smitterberg 2005) presentan análisis cuantitativos correspondientes a períodos particulares de la lengua (inglés moderno temprano e inglés moderno tardío) y no del inglés en su conjunto.

Para ampliar estos análisis a todos los períodos de la lengua, deben justificarse unos criterios de representatividad estándar para el conjunto de corpus seleccionado. En base a estos parámetros, el análisis cuantitativo del conjunto de corpus *York-Toronto-Helsinki-Penn* permite extraer las frecuencias de aparición de las construcciones objeto de estudio, probar la existencia de relaciones entre la evolución de determinadas dimensiones funcionales cuantificables y los desarrollos de las formas progresivas, así como entre los patrones que muestran estas dimensiones y los cambios del indicador estimado de los procesos de gramaticalización.

Por tanto, este estudio pretende ilustrar cuantitativamente los procesos de gramaticalización que experimentan las formas progresivas a lo largo de su evolución, tomando como muestra diacrónica representativa del inglés al conjunto de corpus *York-Toronto-Helsinki-Penn*.

2. METODOLOGÍA DEL ESTUDIO

El cumplimiento de una serie de criterios metodológicos, debidamente justificados, resulta sin duda aconsejable para conducir un estudio de estas características. No en vano, en Lingüística del Corpus se requiere canónicamente el cumplimiento del criterio de representatividad de la muestra como espécimen de una variedad de la lengua particular (McEnery y Wilson 2001: 29-30). En cualquier caso, también es conveniente justificar una serie de criterios metodológicos adicionales, adaptados a las particularidades de la Lingüística Histórica.

El primer criterio consiste, obviamente, en la representatividad lingüística de los corpus en función de cada período. El segundo criterio reside en la posibilidad de establecer comparaciones cuantitativas entre ellos (Leech, Hundt, Mair y Smith 2009: 27-31). Esto permite, en gran parte, establecer las condiciones para una posible extrapolación de los resultados obtenidos. Debe seleccionarse, por tanto, un corpus representativo de cada período en el que el conteo total de palabras sea prácticamente idéntico al del resto de los corpus.

El tercer criterio consiste en aplicar unas técnicas estadísticas que resulten congruentes con el carácter de los datos, así como con el tamaño de las muestras de los que se extraen. En otras palabras, las técnicas estadísticas seleccionadas deben ajustarse a unos modelos que respeten la naturaleza lingüística del objeto de estudio, pues sólo de este modo es posible alcanzar la suficiente significación como para formular hipótesis generales sobre la evolución de la lengua.

2.1. Material

El conjunto *York-Toronto-Helsinki-Penn* respeta el primer y segundo criterios metodológicos antes expuestos. Se trata de un conjunto de corpus históricos equiparables que cubre todos los períodos de la lengua, desde el inglés antiguo hasta el inglés moderno tardío. El criterio de inclusión de los textos compilados en este conjunto de corpus se basa en el principio de variación socio-histórica, por lo que se estima que sus muestras son representativas

de la lengua escrita de cada período. La periodización es de capital importancia en su compilación, así como la variación dialectal, el género textual, el registro empleado por el autor de cada texto y la variación sociolingüística.

La estructura particular de este grupo de corpus permite establecer comparaciones significativas entre períodos del inglés, pues se ha seleccionado una colección de muestras que no se limita a una variedad geográfica, registro, estrato social o género textual específico. La diversidad textual y dialectal del conjunto contribuye a la consideración de este corpus como plenamente representativo de los períodos del inglés.

El *York-Toronto-Helsinki Parsed Corpus of Old English Prose* (YCOE) constituye corpus un representativo del inglés antiguo (IA) que permite la descripción lingüística de este período de la lengua (Taylor, Warner, Pintzuk y Beths 2003). El YCOE abarca el período ca.450 A.D.-1100 A.D. Se trata de un corpus de 1,5 millones de palabras anotadas gramatical y sintácticamente, compiladas de textos en prosa del *Toronto Dictionary of Old English Corpus*, siguiendo las convenciones del *Helsinki Corpus* (HC). En cuanto a la representatividad del período, cabe destacar que el dialecto del IA preponderante en los textos de los que se obtiene la compilación es el sajón occidental, con algunos textos escritos en los dialectos de Kent, Mercia, así como el ánglico.

El *Penn Parsed Corpus of Middle English*, segunda edición (PPCME2), tercer lanzamiento, constituye un corpus representativo del inglés medio (IM) que permite el estudio de este período de la lengua (Kroch y Taylor 2000). El PPCME2 consta de 1,2 millones de palabras anotadas gramatical y sintácticamente, compiladas, con algunas modificaciones, de textos basados en la sección diacrónica del HC. El PPCME2 cubre el período 1150 A.D.-1500 A.D y muestrea las cinco áreas dialectales del IM: *East Midland, West Midland, Northern, Southern y Kentish*.

Por último, el *Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME), segundo lanzamiento, es un corpus representativo del inglés moderno temprano (IMT) que asimismo parte de los criterios de compilación y estructura del HC para su construcción (Kroch, Santorini y Diertani 2004). El PPCEME contiene 1,7 millones de palabras anotadas gramatical y sintácticamente. La muestra para este período abarca del 1500 A.D. al 1710 A.D.

El *Penn-Helsinki Parsed Corpus of Modern British English* (PPCMBE), primera edición, primer lanzamiento, es un corpus representativo del inglés moderno tardío (IModT) que contiene 1 millón de palabras anotadas gramatical y sintácticamente, del mismo modo que el resto de corpus del grupo (Kroch, Santorini y Diertani 2010). El PPCMBE abarca un período de 210 años, concretamente de 1700 A.D. a 1914 A.D. Este corpus se compiló manteniendo las convenciones del PPCEME, segundo lanzamiento, puesto que el HC no consta de una sección correspondiente al inglés moderno tardío.

2.2. Tipología de géneros textuales

A partir de la estructura del conjunto de corpus seleccionado y teniendo en consideración los dos primeros criterios metodológicos antes establecidos, es decir, la representativi-

dad lingüística y la comparabilidad inter-corpus, se procedió a la reclasificación de la tipología textual contenida en el conjunto *York-Toronto-Helsinki-Penn*. La reclasificación en 10 géneros textuales resultante es la que sigue:

- 1) Philosophy
- 2) History
- 3) Travelogue
- 4) Bible and Apocrypha
- 5) Homilies and Sermons
- 6) Biography, Diaries, Prefaces, Epilogues, Non-private and Private letters
- 7) Rules, Charters, Wills, Laws, Ecclesiastical Laws and Proceedings-trials
- 8) Medicine, Handbooks, Religious and Educational Treatises
- 9) Science, Astronomy, Geography
- 10) Fiction, Romances, Drama-comedies and Verse

La compatibilidad del tipo de texto para su inclusión en un género particular se justifica por los principios intrínsecos de compilación que presenta el conjunto de corpus. De esta forma, cada texto individual contiene una etiqueta que lo identifica como espécimen de un género concreto. Con todo, asumiendo que las diferencias extralingüísticas influyen en la distribución de los rasgos lingüísticos que presentan los textos, seleccioné el género textual que mejor define a cada texto en virtud de los siguientes parámetros:

- 1) El medio: La Filosofía, la Historia, los Cuadernos de viaje, la Biblia, la Biografía, los Tratados, la Ciencia y la Ficción son géneros que incluyen textos concebidos para ser fundamentalmente leídos. Por su parte, las Homilías y las Leyes comprenden textos que pueden ser declamados.
- 2) El destinatario: La Filosofía, la Historia, la Biblia, las Homilías, los Tratados, la Ciencia y la Ficción abarcan textos ideados para ser difundidos. Los Cuadernos de viaje, la Biografía y las Leyes contienen textos que no necesariamente son concebidos para ser publicados.
- 3) El propósito: Los Cuadernos de viaje, la Biografía y la Ficción comprenden textos con pasajes en los que domina la intención narrativa. La Filosofía, los Tratados y la Ciencia incluyen textos de carácter principalmente expositivo. En el resto de géneros (Historia, Biblia, Homilías y Leyes) se registra una mezcla de textos con ambos propósitos.

2.3. Técnicas de análisis estadístico

Para analizar los cambios en la distribución funcional de las estructuras progresivas se seleccionó el método estadístico multifactorial o multidimensional (MD). Este método, aplicado a la Lingüística del Corpus para el estudio de las estructuras gramaticales y léxicas en relación con los géneros textuales por Berber y Veirano (2014); Biber (1988); Biber y Finegan (1997); Geisler (2002) y Smitterberg (2005), consiste en la construcción de una se-

rie de dimensiones o factores funcionales a partir de características morfosintácticas presentes en los corpus analizados en función del género textual. En base a la evolución de estas dimensiones en los géneros, se pueden conducir análisis estadísticos de varianza o regresión para probar su relación con los desarrollos que experimentan diversas construcciones gramaticales.

Este método asume la existencia de una relación entre situaciones comunicativas y características morfosintácticas, pues las dimensiones o factores se construyen a partir de características lingüísticas que se definen por las frecuencias de aparición de determinadas estructuras en los diferentes géneros textuales. Esto permite obtener variables que son susceptibles de un análisis factorial. El análisis factorial da lugar a unas variables derivadas, también denominadas factores o dimensiones, sobre las que se pueden establecer modelos de inferencia que permiten probar, en última instancia, la existencia de relaciones entre las dimensiones y las estructuras objeto de estudio.

Los factores establecidos para aplicar esta técnica estadística son canónicamente seis. El cómputo de estos factores implica su correlación con una dimensión semántico-pragmática o funcional que presenta, en las dimensiones con rasgos de peso negativo, dos polos opuestos en su definición. En el diseño de este estudio, siguiendo especialmente los trabajos de Geisler (2002) y Smitterberg (2005), se construyeron para el análisis funcional de los procesos de gramaticalización que experimentan las construcciones progresivas del inglés cuatro dimensiones o factores que se corresponden con cuatro de las seis dimensiones definidas por Biber (1988) y Biber y Finegan (1997):

- 1) *Dimension 1: Involved versus informational production*
- 2) *Dimension 2: Narrative versus non-narrative concerns*
- 3) *Dimension 3: Elaborated reference versus situation-dependent reference*
- 4) *Dimension 5: Abstract versus non-abstract information*

Las dimensiones 1, 3 y 5 del modelo MD permiten distinguir géneros textuales con características prototípicamente conversacionales de géneros con rasgos estereotípicamente expositivos. La Dimensión 2 permite distinguir funciones narrativas no especializadas de funciones narrativas con propósitos especializados. Las dimensiones 2 y 5 no presentan rasgos morfosintácticos con peso negativo en su constitución, por lo que sólo el crecimiento de las mismas se considera indicativo de la función a la que representan. Por tanto, una disminución en los valores de las dimensiones 2 y 5 no puede equipararse con la presencia del polo funcional negativo.

Por su parte, para probar la existencia de relaciones entre la evolución de estas dimensiones funcionales y los desarrollos de las formas progresivas, en función del género textual y el período de la lengua, se estimó un modelo de análisis múltiple de covarianza (MANCOVA). En este modelo, se consideran variables explicativas o independientes al período de la lengua, el género textual y las cuatro dimensiones del modelo MD. El período y los géneros textuales constituyen variables cualitativas y las dimensiones son variables cuantitativas o covariables. Las variables dependientes o explicadas, es decir, el correspondiente

coeficiente representativo de las construcciones progresivas y la ratio clase-tipo léxica estimada son, por su parte, cuantitativas.

Posteriormente, para describir las tendencias relacionadas de las formas progresivas totales y sus subtipos sintácticos en función de las dimensiones y el período de la lengua, se estimó un modelo de regresión logística binomial. Para ello, se clasificó a la variable dependiente o explicada, es decir, las formas progresivas totales y sus subtipos sintácticos, como una variable cualitativa dicotómica. Esta última transformación permitió predecir las tendencias de la variable dependiente, que adopta dos valores discretos, esto es, uso frecuente o esporádico, en función de las variables independientes explicativas o predictorias, es decir, las dimensiones y los períodos de la lengua.

2.4. Extracción y tratamiento de las variables lingüísticas

La categorización de cada una de las variables como ocurrencias de un tipo concreto siguió el principio de máxima objetividad, que viene dado por la no intervención en el proceso de cuantificación de las variables. Para evitar interferencias en el proceso de conteo, se consideró, en primer lugar, construcción progresiva a toda construcción que contuviese una estructura de gerundio, partícula *-ing* adjetival o participio presente en el conjunto de corpus. La normalización posterior de las frecuencias de aparición obtenidas se expresa por los coeficientes-M y -V.

El primero de estos coeficientes, el *Coefficiente-M*, se utiliza cuando la forma progresiva se considera un rasgo lingüístico autónomo (Smutterberg 2005: 40). Esto implica que la frecuencia de las formas progresivas (N_{PROG}) se normaliza en función del número total de palabras del corpus (N_{WORD}). Para el cálculo de este coeficiente se normaliza la frecuencia por cada 100.000 palabras:

$$M = N_{\text{PROG}} / N_{\text{WORD}} \times 100.000$$

El *Coefficiente-V* relaciona el número de formas progresivas (N_{PROG}) con el número de sintagmas verbales (N_{VERB}) sin excluir las formas verbales que no pueden ocurrir en la forma progresiva, lo que excluye el factor subjetivo que implica la necesaria categorización de ocurrencias (Smutterberg 2005: 44):

$$V = N_{\text{PROG}} / N_{\text{VERB}} \times 10.000$$

Además del número total de formas progresivas, se extrajeron 15 variantes sintácticas de estas construcciones: (1) *Progressive forms in subjunctive mood*, (2) *in noun phrases predicate*, (3) *in noun phrases adjuncts*, (4) *in noun phrases subject non-nominative*, (5) *in noun phrases appositive*, (6) *in adjective phrases predicate*, (7) *in adjective phrases appositive*, (8) *in reduced relatives*, (9) *in prepositional phrases appositive or parenthetical*, (10) *added to non-verbal predicates*, (11) *in a generic predicate*, (12) *added to an appositive or parenthetical constituent*, (13) *added to adjuncts*, (14) *in participial absolutes* y (15) *in participial phrases*.

Por su parte, siguiendo a Geisler (2002: 250-251), para el cálculo de la Dimensión 1 del modelo MD, se extrajeron 10 variables lingüísticas, según el género textual y período de la lengua: (1) *Present tense verbs*, (2) *Personal pronouns*, (3) *It pronouns*, (4) *Main verb be*, (5) *Contractions*, (6) *Wh-questions*, (7) *Average word length*, (8) *Nouns*, (9) *Prepositions* y (10) *Attributive adjectives*. Las cuatro últimas son de peso negativo en el cálculo de esta dimensión.

Para el cálculo de la Dimensión 2, se obtuvieron 2 variables correspondientes a cada género textual y período: (1) *Past tense verbs* y (2) *Perfect aspect verbs*.

Para el cálculo de la Dimensión 3, se extrajeron 4 variables, en función del género y el período: (1) *Wh-relatives*, (2) *Time adverbials*, (3) *Place adverbials* y (4) *Other adverbs*. Las tres últimas variables son de peso negativo en el cálculo de esta dimensión.

Para el cálculo de la Dimensión 5, se obtuvieron 2 variables según el género textual y el período de la lengua: (1) *Adverbial subordinators* y (2) *By passives*.

Cada una de las variables gramaticales del modelo MD fue normalizada por 1000 palabras. Se procedió a esta normalización para cada dimensión según los géneros textuales y el período de la lengua, salvo la correspondiente a la variable *Average word length*, cuyo particular cálculo no requiere de dicha normalización, lo que resulta en un total de 17 variables normalizadas.

Posteriormente, se estandarizaron todas las variables obtenidas para construir modelo MD (18), las normalizadas y la no normalizadas, a una media de 0.0 y a una desviación estándar de 1.0 por género textual, para luego realizar el sumatorio correspondiente a la computerización final de los factores. Este sumatorio incluye la extracción de los factores negativos para las dimensiones que los contienen (1 y 3).

Con respecto a la ratio clase-tipo léxica, se dividió el número distintivo de tipos léxicos progresivos estimados en cada uno de los corpus (47 para el IA; 142 para el IM; 3899 para IMT y 2241 para IModT), según el género textual y período de la lengua, entre el número total de ocurrencias progresivas extraídas de cada corpus, normalizando por 100 palabras.

3. RESULTADOS OBTENIDOS

3.1. Significación ANOVA del modelo multidimensional (MD)

Para demostrar la significación estadística de la variación de las dimensiones calculadas en el conjunto de corpus (variable dependiente métrica o variable respuesta), se aplicó el test de diferencia honestamente significativa (DHS) de Tukey con las puntuaciones por dimensión en función del género textual y el período histórico (variables independientes cualitativas o factores), aplicando un análisis de varianza (ANOVA) e incluyendo los subsegmentos homogéneos para cada una de las dimensiones según el género y el período.

El modelo estadístico MD demostró su validez al probarse la existencia de una relación estadísticamente significativa entre la varianza de tres de las dimensiones construidas

(variable dependiente métrica o variable respuesta) y los períodos de la lengua (variable independiente no métrica o factor), así como entre la varianza de todas las dimensiones y los géneros textuales propuestos (variable independiente no métrica o factor). Por tanto, los valores no métricos de los factores, es decir, los períodos de la lengua y los géneros textuales, determinan una serie de grupos en la variable respuesta, esto es, en determinadas dimensiones del modelo MD.

En primer lugar, la Dimensión 1 del modelo MD, correspondiente a la producción conversacional frente a la producción informativa, muestra un claro patrón de decrecimiento.

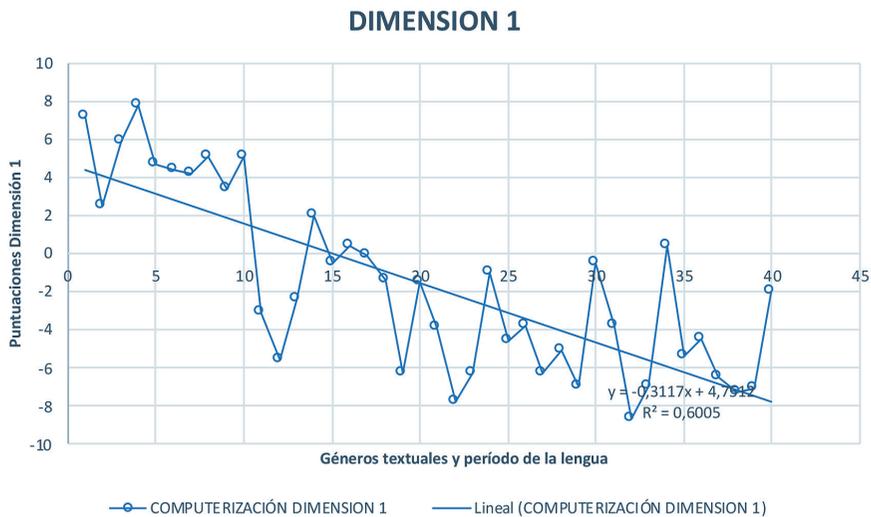


FIGURA 1.

Patrón de decrecimiento de la Dimensión 1 del modelo MD por género textual y período de la lengua

El gráfico de dispersión de la Figura 1 indica la evolución que experimenta la Dimensión 1 a lo largo del tiempo, especificada según los 10 géneros textuales. En el eje de abscisas, se incluyen los 10 géneros textuales por cada período diacrónico. En el eje de ordenadas, se indican los valores métricos que adopta la Dimensión 1. Como puede apreciarse, se trata de un patrón de seguimiento de una tendencia, siguiendo la tipología de contrastes establecida por Leech et al. (2009: 43). Por tanto, la producción de carácter conversacional disminuye con el paso del tiempo para dar lugar a patrones cuya función se corresponde a la producción expositiva o informativa.

Las pruebas *post-hoc* aplicadas, esto es, DHS de Tukey y clasificación mediante subconjuntos homogéneos, indican que el inglés moderno tardío (IModT) y el inglés moderno temprano (IMT) expresan esta tendencia con mayor claridad, formando un subconjunto homogé-

neo propio. Por su parte, el período del inglés medio (IM) constituye un estadio con valores intermedios que, a su vez, forma otro subconjunto homogéneo. Por último, el período del inglés antiguo (IA) presenta la mayor producción de tipo conversacional, formando también su propio subconjunto homogéneo. Esta variación de la Dimensión 1 en función del paso del tiempo es estadísticamente significativa con un valor $F_{1,32} = 117,743$ y un valor $p < 0,001$.

Los géneros textuales presentan un patrón en su evolución que configura dos polos a lo largo de la Dimensión 1. El primero, constituido por la Historia (género 2) y la Ciencia (género 9), que resultan los máximos exponentes de la producción informativa o expositiva y, la Biblia (género 4) y la Ficción (género 10), en el otro extremo, como máximos exponentes de la producción de tipo conversacional. La significación estadística de esta variación de la Dimensión 1 en función de los géneros textuales se sustenta en un valor $F_{1,32} = 9,228$ y un valor $p < 0,001$.

La Dimensión 3 del modelo MD, correspondiente a los usos con referencia específica frente a los usos dependientes de la situación, muestra un patrón de crecimiento. El gráfico de dispersión de la Figura 2, que se muestra a continuación, refleja la variación que experimenta la Dimensión 3 en función de los períodos de la lengua, especificada también según los 10 géneros textuales. Como puede apreciarse, la Dimensión 3 aumenta sus valores con el transcurso del tiempo. Se trata, asimismo, de un patrón de seguimiento de una tendencia, según la tipología de contrastes establecida por Leech et al. (2009: 43). Es decir, la referencia específica en el uso de la lengua aumenta con el paso tiempo y disminuyen, consecuentemente, los patrones dependientes de la situación.

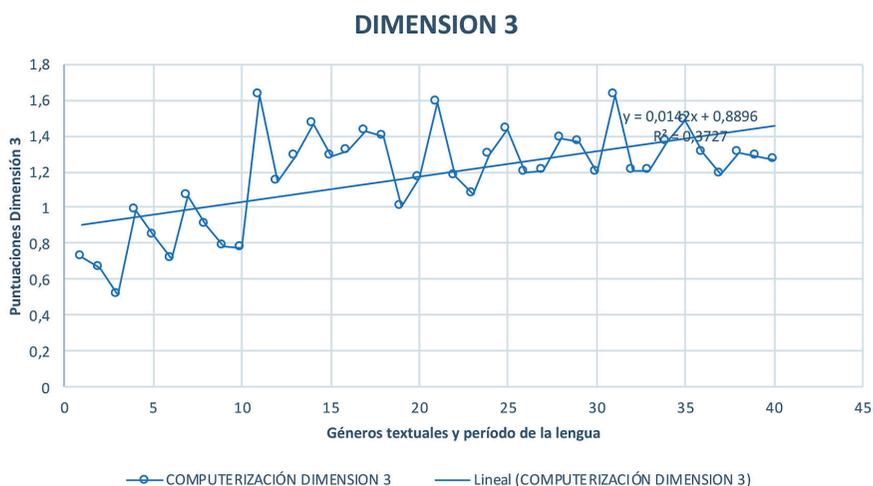


FIGURA 2.

**Patrón de crecimiento de la Dimensión 3 del modelo MD
por género textual y período de la lengua**

Tras las pruebas *post-hoc* aplicadas, se observa que los períodos del IModT, IM e IMT muestran esta tendencia con mayor claridad, formando un subconjunto homogéneo propio. Por su parte, el período del IA, cuyas funciones presentan patrones con mayor dependencia de la situación, forma su propio subconjunto homogéneo. Esta variación de la Dimensión 3 en función del paso del tiempo es estadísticamente significativa con un valor $F_{1,32} = 44,586$ y un valor $p < 0,001$.

Los géneros textuales describen un patrón a lo largo de la Dimensión 3 que presenta también dos polos claramente marcados. El primero está constituido por los Cuadernos de viaje (género 3) y la Historia, como máximos exponentes de los patrones dependientes de la situación y, en el otro extremo, la Filosofía (género 1), como máximo exponente de las funciones con referencia específica. La significación estadística de esta variación en función de los géneros textuales se sustenta en un valor $F_{1,32} = 3,684$ y un valor $p < 0,004$.

La Dimensión 5 del modelo, correspondiente a los usos que conllevan información abstracta frente a los usos que contienen información no abstracta, muestra un patrón de crecimiento con puntos de claro cambio.

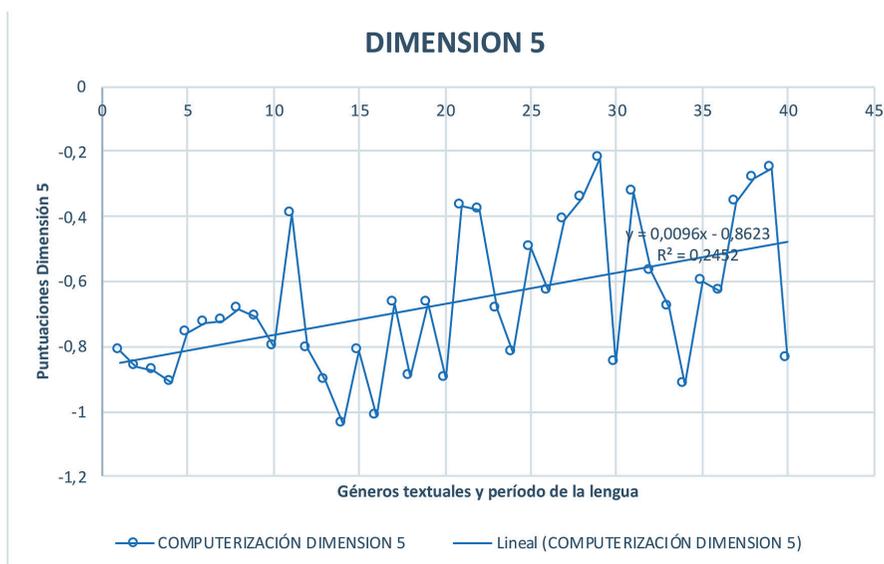


FIGURA 3.

Patrón de crecimiento de la Dimensión 5 del modelo MD por género textual y período de la lengua

El gráfico de dispersión de la Figura 3 indica un patrón de diferentes tasas de variación, según la tipología de contrastes propuesta por Leech et al. (2009: 43). En cualquier caso, se aprecia una tendencia de crecimiento. En otras palabras, los patrones abstractos tienden

a aumentar con el paso del tiempo, aunque esta función no llega a estar marcada en el conjunto de corpus.

Los períodos del IMT y el IModT presentan esta tendencia con mayor claridad, por lo que contienen expresiones con funciones más abstractas, formando un subconjunto homogéneo. Por su parte, los períodos del IM y el IA, los cuales forman su propio subconjunto homogéneo, contienen patrones con funciones abstractas menos marcadas. Esta variación de la Dimensión 5 en función del paso del tiempo es estadísticamente significativa con un valor $F_{1,32} = 14,813$ y un valor $p < 0,001$.

Los géneros textuales muestran un patrón a lo largo de la Dimensión 5 que presenta también dos polos dentro de los usos con información abstracta. El primero, constituido por la Biblia y la Ficción, como máximos exponentes de los usos con información no abstracta, los cuales forman un subconjunto homogéneo y, en el otro extremo, la Ciencia y las Leyes (género 7), como máximos exponentes de los usos con información abstracta. La significación estadística de esta variación en función de los géneros textuales se sustenta en un valor $F_{1,32} = 6,563$ y un valor $p < 0,001$.

La Dimensión 2 del modelo, correspondiente al propósito narrativo frente al propósito no narrativo, muestra un patrón de estancamiento con claros puntos de cambio.

La variación de la Dimensión 2 en función del paso del tiempo no es significativa con un valor $F_{1,32} = 2,311$ y un valor $p < 0,099$.

Los géneros textuales configuran un patrón a lo largo de la Dimensión 2 que presenta dos polos también claramente marcados. El primero, constituido por los Tratados (género 8) y la Filosofía, como máximos exponentes de la función narrativa especializada y, la Historia, la Biblia y la Ficción, como máximos exponentes de los propósitos narrativos no especializados. La significación estadística de esta variación en función de los géneros textuales se sustenta en un valor $F_{1,32} = 12,539$ y un valor $p < 0,001$.

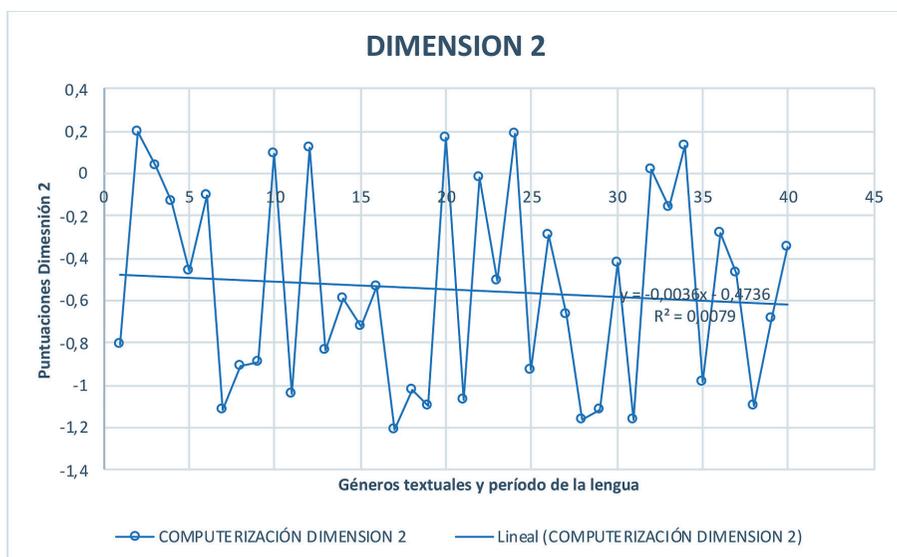


FIGURA 4.

Patrón de estancamiento de la Dimensión 2 del modelo MD por género textual y período de la lengua

El gráfico de dispersión de la Figura 4 indica un patrón de diferentes tasas de variación, siguiendo la clasificación establecida por Leech et al. (2009: 43). No obstante, se aprecia un ligero patrón de decrecimiento, el cual refleja la evolución de la Dimensión 2 en función de los géneros textuales. Es decir, los usos con propósito narrativo no especializado disminuyen en función de los géneros. En cualquier caso, la función narrativa no especializada no está marcada en el conjunto de corpus analizado.

3.2. Significación MANCOVA de las relaciones entre las variables del estudio

A la luz de la viabilidad del modelo MD, se procedió a probar la existencia de una relación de covarianza múltiple (MANCOVA) entre el Coeficiente-M de las frecuencias de aparición las formas progresivas, es decir, el coeficiente de las estructuras progresivas consideradas como rasgo lingüístico autónomo (variable dependiente métrica); el Coeficiente-V, esto es, el coeficiente de las construcciones progresivas en relación al número de sintagmas verbales (variable dependiente métrica); la ratio clase-tipo léxica estimada de las formas progresivas (variable dependiente métrica) y cada una de las dimensiones calculadas (variable independiente métrica), en función del paso del tiempo y de los géneros textuales (variables independientes no métricas).

En primer lugar, existe una relación de covarianza estadísticamente significativa entre el Coeficiente-M y la interacción de la Dimensión 1 con los géneros textuales en virtud de

un valor $F_{1,32}=3,107$ y un valor $p<0,021$. Por su parte, el período de la lengua no influye en esta relación de la Dimensión 1 con el Coeficiente-M. Por tanto, la evolución de la variable dependiente métrica (Coeficiente-M) se relaciona de forma significativa con la interacción de la covariable explicativa métrica (Dimensión 1) y la variable explicativa categórica o factor (género textual). En otras palabras, dependiendo del género textual, la relación entre la Dimensión 1 y el Coeficiente-M difiere. De esta forma, el aumento de la frecuencia de aparición de las construcciones progresivas, como rasgo lingüístico autónomo, se relaciona significativamente con la disminución de los usos con propósito conversacional. Esta relación inversa es más acusada en la Ficción que en la Biblia, es decir, los géneros textuales de características conversacionales marcadas en el conjunto de corpus (cf. 3.1.). Por su parte, la Ciencia y la Historia expresan esta relación de forma más leve, pues constituyen los géneros textuales marcados con respecto a la producción informativa o expositiva.

La Figura 5 muestra este patrón de seguimiento de una tendencia, según la tipología de contrastes establecida por Leech et al. (2009: 43). En el eje de abscisas, se muestran cuatro puntuaciones de la Dimensión 1 en relación con los valores del Coeficiente-M en cuatro géneros textuales, para los cuatro períodos de la lengua. En el eje de ordenadas se escalan los valores que adopta el Coeficiente-M.

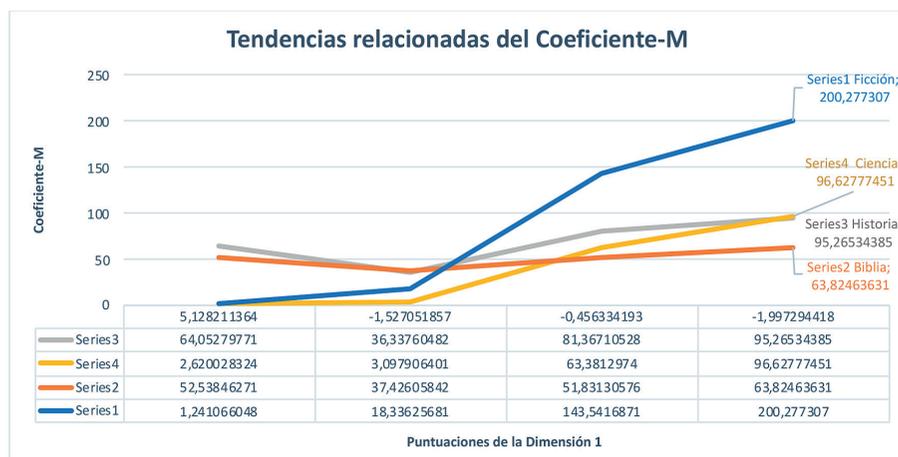


FIGURA 5.

Patrón de crecimiento del Coeficiente-M en función de la Dimensión 1 por género textual

Existe también una relación de covarianza estadísticamente significativa entre el Coeficiente-V y la interacción de la Dimensión 1 con los géneros textuales, en virtud de un valor $F_{1,32}=2,974$ y un valor $p<0,025$. El período de la lengua (variable explicativa categórica o factor) tampoco influye en la relación entre el Coeficiente-V y la interacción de la Dimensión 1 con los géneros textuales. Por tanto, la evolución de la variable dependiente métrica (Coeficiente-V) se relaciona significativamente con interacción de la covariable explicativa

métrica (Dimensión 1) y la variable explicativa categórica o factor (género textual), aunque esta relación depende, asimismo, del género textual. Como puede apreciarse, esta tendencia es menos acusada que la correspondiente al Coeficiente-M. No obstante, se registra el mismo patrón en los géneros textuales. Es decir, el Coeficiente-V se relaciona de forma significativa con la disminución de los usos conversacionales de forma más acusada en la Ficción que en la Biblia. Por su parte, la Ciencia y la Historia reflejan esta relación de una forma intermedia.

Se prueba también la existencia de una relación de covarianza estadísticamente significativa entre la ratio clase-tipo léxica estimada y la interacción de la Dimensión 1 con los géneros textuales, sustentada en unos valores $F_{1,32}=3,183$ y $p<0,019$. El período de la lengua (variable explicativa categórica o factor) tampoco influye en esta relación de la Dimensión 1 con la ratio clase-tipo léxica estimada, expresada en función de los géneros textuales. En otras palabras, el aumento del indicador estimado de los procesos de cambio gramatical (variable dependiente métrica) se relaciona de forma significativa con la disminución de los usos conversacionales (covariable explicativa métrica), aunque esta relación depende también de los géneros textuales (variable explicativa categórica o factor). Esta tendencia relacionada es más pronunciada que las correspondientes a los Coeficientes-M y -V, observándose el mismo tipo de relación con los géneros textuales: la Ficción muestra esta relación inversa de forma más acusada que la Biblia, mientras que la Historia y la Ciencia manifiestan esta relación de una manera intermedia.

La interacción de la Dimensión 3 con los géneros textuales no presenta una relación de covarianza estadísticamente significativa con el Coeficiente-M, el Coeficiente-V, ni la ratio clase-tipo léxica estimada. En otras palabras, la evolución de las variables dependientes o explicadas no difiere en función de la interacción de la Dimensión 3 con los géneros textuales (variables explicativas), con unos valores respectivos $F_{1,32}=0,944$ y $p<0,515$; $F_{1,32}=1,020$ y $p<0,463$; y $F_{1,32}=0,406$ y $p<0,914$.

Por su parte, existe una relación de covarianza estadísticamente significativa entre el Coeficiente-M, el Coeficiente-V, la ratio clase-tipo léxica estimada (variables dependientes métricas) y la interacción de la Dimensión 5 (variable independiente métrica) con los géneros textuales (variable independiente categórica o factor). Estas relaciones son significativas, respectivamente, con unos valores $F_{1,32}=3,408$ y $p<0,014$; $F_{1,32}=3,380$ y $p<0,015$; y $F_{1,32}=4,978$ y $p<0,002$. Por tanto, el aumento de la frecuencia de aparición de las construcciones progresivas, como rasgo lingüístico autónomo y en función del número de sintagmas verbales, así como el aumento en el indicador estimado de los procesos de gramaticalización se relacionan de forma significativa con el aumento de los usos con referencia abstracta, aunque esta función semántico-pragmática no llega a estar marcada en el conjunto de corpora.

El patrón observado en la Figura 6 se corresponde con el tipo de seguimiento de una tendencia, según la clasificación establecida por Leech et al. (2009: 43). En el eje de abscisas, se muestran cuatro puntuaciones de la Dimensión 5 en relación con los valores del Coeficiente-M en cuatro géneros textuales para los cuatro períodos de la lengua. En el eje de ordenadas, se indica el rango de valores que toma el Coeficiente-M. Como puede apreciarse, la relación directa entre el aumento del Coeficiente-M y aumento de los usos con referencia abstracta es más acusada en la Ficción que en la Biblia, los géneros de marcas

características conversacionales. Las Leyes y la Ciencia, esto es, los géneros de tipo expositivo, expresan esta relación de forma más leve.

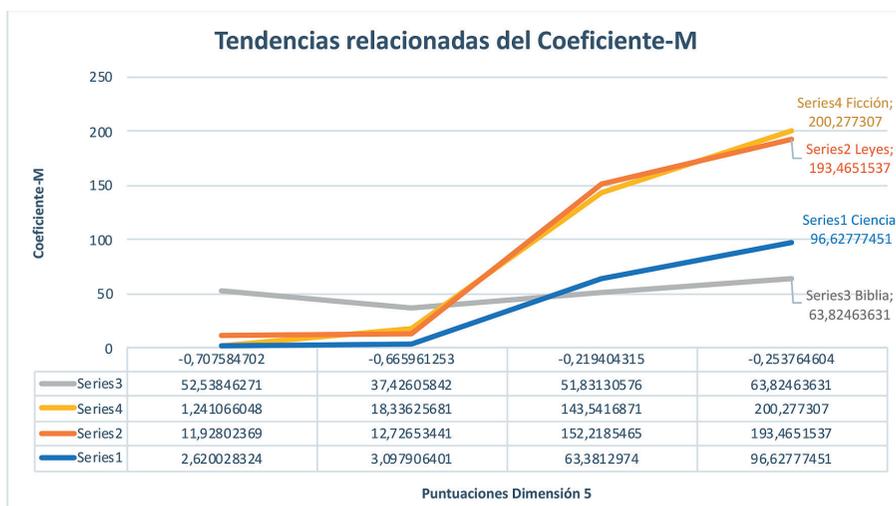


FIGURA 6.

Patrón de crecimiento del Coeficiente-M en función de la Dimensión 5 por género textual

Por su parte, la relación directa entre el Coeficiente-V y el aumento de los usos con referencia abstracta es asimismo más acusada en la Ficción que en la Biblia. Las Leyes y la Ciencia manifiestan también esta relación de un modo intermedio. No obstante, esta tendencia relacionada del Coeficiente-V y la interacción de la Dimensión 5 con los géneros textuales es menos acusada que la correspondiente al Coeficiente-M.

La relación de la ratio clase-tipo léxica estimada y la interacción de la Dimensión 5 con los géneros textuales es la más acusada de las tres tendencias descritas. Esta relación directa presenta un patrón en los géneros similar a los anteriores: es más marcada en la Ficción que en la Biblia. En cualquier caso, las Leyes expresan esta relación de forma más clara que la Ficción y la Ciencia, géneros que manifiestan esta relación de una manera intermedia.

Por último, existe una relación de covarianza estadísticamente significativa entre el Coeficiente-M, el Coeficiente-V y la interacción de la Dimensión 2 con los géneros textuales, con unos valores respectivos $F_{1,32}=2,871$ y $p<0,029$, y $F_{1,32}=3,116$ y $p<0,021$. La covarianza entre los aumentos en la ratio clase-tipo léxica estimada y la interacción de la Dimensión 2 con los géneros textuales no es significativa con un valor $F_{1,32}=1,067$ y un valor $p<0,433$. Por tanto, el aumento en la frecuencia de aparición de las formas progresivas, como rasgo lingüístico autónomo y en relación al número de sintagmas verbales, se relaciona de forma significativa con el descenso de los usos narrativos no especializados en función de los géneros textuales. En cualquier caso, la función narrativa no especializada no está marcada en el conjunto de corpus.

El patrón observado en la Figura 7, que se muestra a continuación, se corresponde con el tipo de seguimiento de una tendencia, según la clasificación establecida por Leech et al. (2009: 43). Este patrón refleja el aumento del Coeficiente-M en función de la interacción de la Dimensión 2 con los géneros textuales. Como puede apreciarse, la relación inversa entre el aumento del Coeficiente-M y la disminución de los usos narrativos no especializados es más acusada en la Ficción que en la Biblia. La Filosofía es el género que presenta esta relación de forma menos marcada, mientras que los Tratados encabezan la expresión de esta relación. Por su parte, la Historia, junto con la Biblia, expresan esta tendencia relacionada de una forma intermedia.

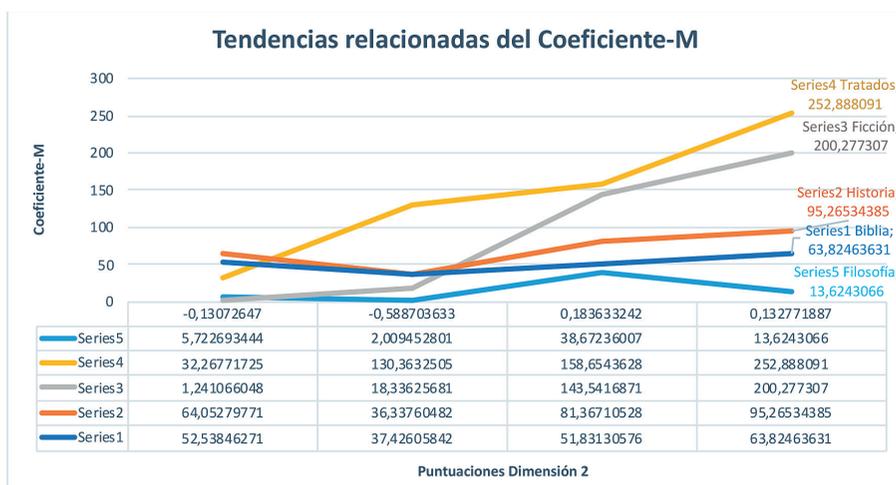


FIGURA 7.

Patrón de crecimiento del coeficiente-M en función de la Dimensión 2 por género textual

La relación inversa entre el aumento del Coeficiente-V y la interacción de la disminución de los usos narrativos no especializados con los géneros textuales es más acusada que la correspondiente al Coeficiente-M, observándose el mismo patrón en los géneros: es más marcada en la Ficción que en la Biblia, mientras que la Filosofía es el género que presenta esta relación de forma menos intensa. Por su parte, los Tratados encabezan la expresión esta relación y, la Historia, junto con la Biblia, manifiestan esta relación de un modo intermedio.

3.3. Significación del modelo logístico de regresión para la descripción de las tendencias relacionadas de las formas progresivas

Para analizar la distribución de las formas progresivas totales y sus variantes sintácticas en función de las dimensiones y los períodos de la lengua, se estimó un modelo estadístico de regresión logística binomial, dado el escaso número de ocurrencias registradas en algu-

nas de estas variantes en el grupo de corpus. En primer lugar, se categorizaron las ocurrencias de las formas progresivas según el siguiente binomio:

1. ≥ 100 ocurrencias: uso frecuente
2. < 100 ocurrencias: uso esporádico

Posteriormente, se seleccionó el estadístico de Wald por pasos hacia atrás para la selección de las variantes significativas en el modelo en función de las dimensiones y el período de la lengua, con un valor 0,05 de entrada y 0,10 de salida.

El modelo logístico estimado resultó significativo para las formas progresivas totales y 5 de sus variantes sintácticas de un total de 15 (cf. 2.4.). El modelo explica la probabilidad de encontrar patrones de uso frecuente o esporádico según los valores del conjunto de variables predictorias. Por tanto, según el caso, los valores conocidos de las variables predictorias (dimensiones y períodos de la lengua) determinan la categoría de la variable dependiente (uso frecuente o esporádico) con la que se corresponden en términos probabilísticos. De este modo, podemos asociar cuáles de las características de las variables predictorias contribuyen, en mayor medida, a formar parte de uno de los dos grupos de decisión.

En primer lugar, el número total de ocurrencias de las formas progresivas presenta un patrón de *uso frecuente* que se explica con un 85% de probabilidad en función de la Dimensión 3 con un valor $p=0,016$. En otras palabras, dado un valor conocido de la Dimensión 3 (referencia específica vs. dependiente de la situación), se predice correctamente, en un 85% de los casos, que las formas progresivas totales ocurren de forma frecuente.

La tendencia relacionada de las ocurrencias en frases de participio se ilustra en la Figura 9. Estas formas presentan un patrón de *uso frecuente* con una probabilidad de acierto del 75% en función de las dimensiones 1 y 2, con unos valores respectivos $p=0,014$ y $p=0,022$. Es decir, dados unos valores conocidos de las dimensiones 1 y 2 (cf. 3.1.), se predice correctamente un patrón de uso frecuente de las formas progresivas en frases de participio en un 75% de los casos.

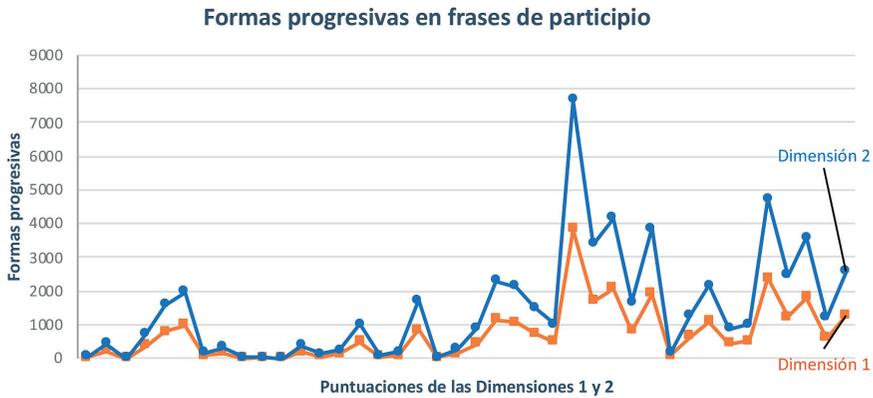


FIGURA 9.

Patrón de uso frecuente de las formas progresivas en frases de participio en función de las Dimensiones 1 y 2

Las ocurrencias en participios absolutos presentan un patrón de *uso esporádico* con una probabilidad de acierto del 85% en función del período de la lengua y la Dimensión 3, con unos valores respectivos $p=0,081$ y $p=0,053$. Esta tendencia relacionada se ilustra en la Figura 10.



FIGURA 10.

Patrón de uso esporádico de las formas progresivas en participios absolutos en función de la Dimensión 3 y el período de la lengua

Por último, las ocurrencias en modo subjuntivo presentan un patrón de *uso esporádico* explicado con un 80% de probabilidad de acierto en función de la Dimensión 1 (producción conversacional vs. producción expositiva) con un valor $p=0,001$. Las ocurrencias en sujetos no nominales presentan un patrón de *uso esporádico* que se explica con una probabilidad del 80% de acierto en función de la Dimensión 5 (información abstracta vs. no abstracta) con un valor $p=0,011$. Finalmente, las ocurrencias en cláusulas de relativo incompletas presentan un patrón de *uso esporádico* con una probabilidad de acierto del 80% en función de la Dimensión 1 con un valor $p=0,003$.

4. DISCUSIÓN

Los resultados presentados en este estudio han probado la existencia de una relación significativa entre los Coeficientes-M y -V y 3 de las dimensiones del modelo multidimensional en función de su interacción con los géneros textuales. En el caso de las dimensiones 1 y 5, la significación de esta relación es más acusada para Coeficiente-M, por lo que las formas progresivas como rasgo lingüístico autónomo manifiestan esta relación de forma más clara en los citados casos. Por tanto, se ha probado la existencia de una relación significativa, que depende del género textual, entre los aumentos de las formas progresivas, totales y verbales, y la disminución de los usos conversacionales (Dimensión 1), la disminución de los usos con propósito narrativo no especializado (Dimensión 2), así como el aumento de los usos que expresan información abstracta (Dimensión 5).

Por una parte, los aumentos en las formas progresivas se relacionan con la disminución de los usos conversacionales de forma marcada en la Ficción y de forma menos acusada en la Ciencia. Un patrón que sugiere una distinción entre géneros con características conversacionales frente a géneros de corte expositivo. Por otra parte, los aumentos en las formas progresivas se relacionan con la disminución de los usos narrativos no especializados de forma marcada también en la Ficción y de forma leve en la Filosofía, lo cual refleja la misma distinción entre tipos discursivos. Por su parte, el aumento en las formas progresivas se relaciona con el aumento en los usos que expresan información abstracta de forma acusada asimismo en la Ficción y de forma más leve en la Ciencia.

El incremento del indicador estimado de los procesos de gramaticalización se relaciona con las dimensiones siguiendo el mismo patrón, con la salvedad de la Dimensión 2 (cf. 3.2.). En cualquier caso, se observa una tendencia constante de aumento en la evolución de esta ratio clase-tipo, con un claro incremento en el período correspondiente al inglés moderno temprano (IMT), donde se espera un marcado proceso de gramaticalización, según la literatura especializada (Kranich 2010; Núñez-Pertejo 2004; Smith 2005; Smitterberg 2005).

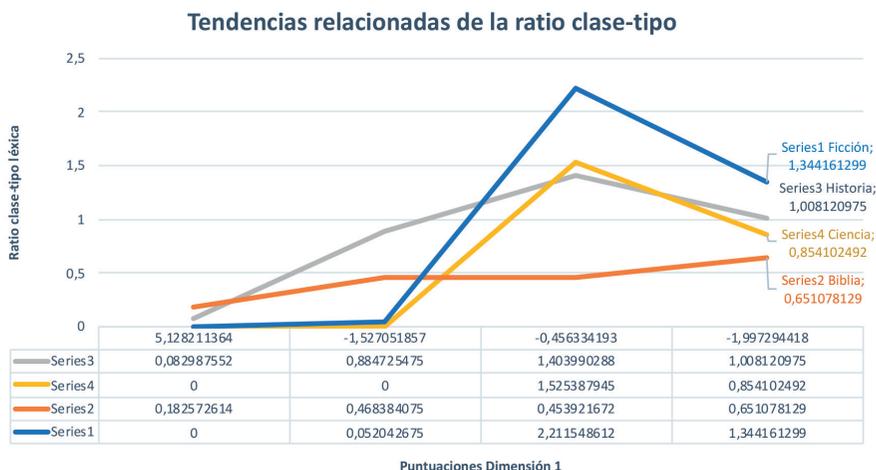


FIGURA II.

Patrón de crecimiento de la ratio clase-tipo léxica estimada en función de la Dimensión 1 por género textual

La evolución de la lengua inglesa tiende a reflejar patrones de menores rasgos conversacionales en los géneros textuales, y las construcciones progresivas se relacionan positivamente con esta tendencia general (cf. 3.2.). El inglés moderno temprano (IMT) y el inglés moderno tardío (IModT) manifiestan esta tendencia con mayor claridad, mientras que los períodos del inglés medio (IM) y el inglés antiguo (IA) muestran una menor producción de tipo expositivo. El hecho de que estas tendencias sean más acusadas en los géneros típicamente conversacionales que en los géneros expositivos sugiere una confluencia hacia los patrones especializados conforme evoluciona la lengua. Asimismo, los procesos de gramaticalización se asocian con esta tendencia, especialmente en los períodos en los que el indicador estimado de los procesos de gramaticalización es mayor.

Por su parte, las formas progresivas en frases en modo subjuntivo, en sujetos no nominales, en cláusulas de relativo incompletas y en participios absolutos presentan un patrón de uso esporádico que se explica, respectivamente, en función de los valores de las Dimensiones 1, 5, 1, y la Dimensión 3 y el período de la lengua (cf. 3.3.). En otras palabras, conocido el valor de la disminución de los usos prototípicamente conversacionales, así como el período de la lengua, según el caso, puede predecirse un patrón de uso esporádico en estas variantes sintácticas progresivas.

Las formas progresivas en frases de participio y las formas progresivas totales presentan un patrón de uso frecuente que se explica en función de los valores que adoptan las Dimensiones 1 y 2, y la Dimensión 3, respectivamente. Es decir, conocidos los valores de la disminución en las funciones de tipo conversacional y los propósitos narrativos no especializados, según el caso, puede predecirse un patrón de uso frecuente en estos tipos progresivos.

Finalmente, estos cambios relativos en la distribución de las variantes sintácticas de las formas progresivas, explicados en función de la evolución de los citados patrones funcionales, podrían asociarse a la paulatina adquisición de una mayor rigidez morfosintáctica, característica de la gramaticalización, tal y como señala la literatura al respecto (Lehmann 2002; Traugott 2003; Traugott y Heine 1991).

5. CONCLUSIONES

Los cambios funcionales descritos en este estudio, dada su relación, tanto con los aumentos de las ocurrencias progresivas como con el aumento de la ratio clase-tipo estimada, parecen influir en los fenómenos de gramaticalización que experimentan estas construcciones. Por su parte, la distribución relativa de las formas progresivas totales y sus variantes sintácticas, explicada en función de estos mismos patrones funcionales, ilustra tentativamente la forma en la que las variantes estructurales progresivas se propenden en los procesos de gramaticalización.

El tamaño de las muestras representativas de cada período, contenidas en el conjunto de corpus *York-Toronto-Helsinki-Penn*, puede considerarse lo suficientemente amplio como para formular esta tesis tentativa sobre las relaciones entre las tendencias funcionales del modelo multidimensional y la distribución relativa de las construcciones progresivas en la gramaticalización.

Con todo, los resultados presentados en esta investigación son indicativos de los principales patrones que se relacionan con los procesos de gramaticalización hasta donde lo permite el desarrollo actual de corpus históricos, así como de técnicas estadísticas adecuadas para el análisis funcional, siguiendo las propuestas de Heine y Kuteva (2005: 45).

REFERENCIAS BIBLIOGRÁFICAS

- BERBER, TONY Y MARCIA VEIRANO (2014): *Multi-Dimensional Analysis, 25 years on. A Tribute to Douglas Biber*, Amsterdam/Philadelphia, John Benjamins.
- BIBER, DOUGLAS (1988): *Variation across Speech and Writing*, Cambridge, Cambridge University Press.
- BIBER, DOUGLAS Y EDWARD FINEGAN (1997): «Diachronic Relations Among Speech-Based and Written Registers in English», en Terttu Nevalainen and Leena Kahlas-Tarkka (eds.), *To Explain the Present. Studies in the Changing English Language in Honour of Matti Rissanen*, Helsinki, Société Néophilologique, pp. 253-275.
- BRINTON, LAUREL J. Y ELIZABETH CLOSS TRAUOGOTT (2005): *Lexicalization and Language Change*, Cambridge, Cambridge University Press.
- BYBEE, JOAN (2003): «Mechanisms of Change in Grammaticalization. The Role of Frequency», en Brian Joseph y Richard Janda (eds.), *The Handbook of Historical Linguistics*, Oxford, Blackwell, pp. 602-623.
- FISHER, OLGA (1992): «Syntax», en Norman Blake (ed.), *Vol. 2 of The Cambridge History of the English Language*, Cambridge, Cambridge University Press, pp. 207-408.
- GEISLER, CHRISTER (2002): «Investigating Register Variation in Nineteenth-Century English», en

- Randy Reppen, Susan M. Fitzmaurice y Douglas Biber (eds.), *Using Corpora to Explore Linguistic Variation*, Amsterdam/Philadelphia, John Benjamins, pp. 249-271.
- HEINE, BERND Y TANIA KUTEVA (2005): *Language Contact and Grammatical Change*, Cambridge, Cambridge University Press.
- HOPPER, PAUL Y ELIZABETH CLOSS TRAUOGOTT (2003): *Grammaticalization*, Cambridge, Cambridge University Press.
- KRANICH, SVENJA (2010): *The Progressive in Modern English: A Corpus-Based Study of Grammaticalization and Related Changes*, Amsterdam/New York, Rodolpi.
- KROCH, ANTHONY Y ANN TAYLOR (2000): *Penn-Helsinki Parsed Corpus of Middle English (PPC-ME2)*, 2ª edición, tercer lanzamiento, Departamento de Lingüística, Universidad de Pensilvania. CD-ROM.
- KROCH, ANTHONY, BEATRICE SANTORINI Y ARIEL DIERTANI (2004): *Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)*, 1ª edición, segundo lanzamiento, Departamento de Lingüística, Universidad de Pensilvania. CD-ROM.
- KROCH, ANTHONY, BEATRICE SANTORINI Y ARIEL DIERTANI (2010): *The Penn Parsed Corpus of Modern British English (PPCMBE)*, 1ª edición, primer lanzamiento, Departamento de Lingüística, Universidad de Pensilvania. CD-ROM.
- LEECH, GEOFFREY, MARIANNE HUNDT, CRHISTIAN MAIR Y NICHOLAS SMITH (2009): *Change in Contemporary English. A Grammatical Study*, Cambridge, Cambridge University Press.
- LEHMANN, CHRISTIAN (2002): *Thoughts on Grammaticalization*, 2ª ed. revisada, Erfurt, Seminar für Sprachwissenschaft der Universität.
- MCENERY, TONI Y ANDREW WILSON (2001): *Corpus Linguistics*, Edinburgh, Edinburgh University Press.
- NÚÑEZ-PERTEJO, PALOMA (2004): *The Progressive in the History of English with Special Reference to the Early Modern English Period: A Corpus-Based Study*, Munich, LINCOM.
- SMITH, NICHOLAS I. (2005): *A Corpus-Based Investigation of Recent Change in the Use of the Progressive in British English*. Tesis doctoral, Universidad de Lancaster.
- SMITTERBERG, ERIK (2005): *The Progressive in 19th-Century English: A Process of Integration*, Amsterdam/New York, Rodolpi.
- TAYLOR, ANN, ANTHONY WARNER, SUSAN PINTZUK Y FRANK BETHS (2003): *The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)*, 1ª edición, Departamento de Lingüística, Universidad de York, Oxford Text Archive. Consultado el 4 de Abril de 2018. Disponible en: <http://www-users.york.ac.uk/~lang22/YcoeHome1.htm>
- TRAUGOTT, ELIZABETH CLOSS (1992): «Syntax», en Richard Hogg (ed.), *Vol. 1 of The Cambridge History of the English Language*, Cambridge, Cambridge University Press, pp. 168-289.
- TRAUGOTT, ELIZABETH CLOSS (2003): «Constructions in Grammaticalization», en Brian Joseph y Richard Janda (eds.), *The Handbook of Historical Linguistics*, Oxford, Blackwell, pp. 624-647.
- TRAUGOTT, ELISABETH CLOSS Y BERN HEINE (1991): *Approaches to Grammaticalization. Vol. 1. Focus on Theoretical and Methodological Issues*, Amsterdam/Philadelphia, John Benjamins.