

## HACIA UNA NUEVA ARITMÉTICA DEL TEXTO: UNAS REFLEXIONES SOBRE EL MODELO MUTIDIMENSIONAL DE ANÁLISIS TEXTUAL<sup>1</sup>

JAVIER PÉREZ GUERRA  
DOLORES GONZÁLEZ ÁLVAREZ  
*Universidade de Vigo*

**RESUMEN.** *La elaboración de un modelo teórico construido sobre una base empírica que permita fundamentar cuestiones textuales como el estilo o el grado de “narratividad” o “dependencia del contexto” de un texto o género es la finalidad del llamado modelo multidimensional, expuesto primordialmente en Biber (1988).*

*En este trabajo, además de describir las fases fundamentales de dicho modelo, advertimos ciertas incorrecciones de índole tanto teórica como metodológica. Con respecto a las primeras, nos centramos en la composición del corpus y la selección de los rasgos lingüísticos no gramaticalizados. Desde una perspectiva más práctica, nos detenemos en los mecanismos globales de cuantificación de los rasgos lingüísticos, y apuntamos ciertas soluciones que pueden mejorar aspectos tales como la representatividad estadística de la propuesta multidimensional.*

**PALABRAS CLAVE.** *Análisis de discurso, análisis textual, género, análisis factorial, análisis multidimensional, lingüística informática.*

**ABSTRACT.** *The elaboration of a theoretical framework, built on an empirical basis which makes it possible to account for textual issues such as the style or the degree of “narrativeness” or “context-dependency” of a specific text or genre (text-type) is the main aim of the so-called multi-feature multi-dimensional model which is primarily expounded in Biber (1988).*

*In this paper, apart from describing the major stages of the multi-dimensional proposal, we will focus on some of its inconveniences of both a theoretical and a methodological nature. As far as the former are concerned, we will concentrate on the selection of the corpus of texts and the non-grammaticalised linguistic features. From a more practical perspective, we will explore the general mechanisms which are used to quantify the linguistic features, some specific solutions being suggested with the purpose of enhancing the statistical representativeness of the multi-dimensional model.*

**KEYWORDS.** *Discourse analysis, textual analysis, genre, factor analysis, multidimensional analysis, computational linguistics.*

## 1. INTRODUCCIÓN

Intentar elevar a la categoría de materias cuantificables cuestiones que hasta el momento habían quedado reservadas a la mirada de estudiosos de la retórica o el estilo, o incluso de analistas del discurso sorprende, desde el primer momento, por dos razones. En primer lugar, la pretensión es lo suficientemente compleja como para despertar las más justificadas reservas. En segundo lugar, la potencialidad de dicha técnica a la hora de ampliar los horizontes tradicionales de la lingüística del discurso es tan grande que muy pocos lingüistas podrán permanecer impasibles ante la tentación de abordar estos mecanismos cuantificadores.

La idea principal expuesta en el párrafo anterior es el objetivo de las investigaciones de Biber y más tarde Biber y Finegan<sup>2</sup>, así como de sus seguidores más inmediatos. Su análisis “multifactorial” o “multidimensional” no es más que la materialización de dicha propuesta, es decir, la plasmación en un número, esto es, en una categoría cuantificable y, en consecuencia, contrastable, de una serie de características textuales de un pasaje, de un grupo de pasajes, o de incluso un género estilístico determinado. Esa cuantificación servirá no sólo para establecer contrastes entre, de nuevo, pasajes, grupos de pasajes o géneros desde una perspectiva sincrónica, sino también para recorrer histórica o diacrónicamente la evolución de esas producciones lingüísticas. De hecho, los trabajos más conocidos de Biber & Finegan se centran en esta última posibilidad: el análisis diacrónico de las características funcionales de ciertos géneros estilísticos (*genres*) o, simplemente, taxonomías discursivas (*text types*). Véase, a modo de ilustraciones de trabajos sincrónicos y diacrónicos enmarcados dentro de esta metodología, Atkinson (1992), Biber (1988, 1995a, 1996), Biber, Conrad y Reppen (1998), Biber y Finegan (1989, 1992, 1997), Meurman-Solin (1993), Taavitsainen (1993, 1997) o nuestros (1998 y en prensa).

En este artículo nos proponemos una doble tarea. Por un lado, introduciremos el análisis multidimensional al lector novel, respondiendo así a la necesidad de conducir las explicaciones complejas y (debemos reconocer) confusas de Biber & Finegan al plano más simple. Por otro lado, nos detendremos en aquellos aspectos que, a nuestro juicio, desmerecen la teoría multidimensional e intentaremos aportar soluciones que, a la vez, no se escapen del modelo general y no introduzcan complejidad innecesaria. Ambas pretensiones nacen de un acto de fe que deseamos expresar desde estas primeras líneas: nuestra confianza absoluta no sólo en la necesidad sino también en la posibilidad de que la prototipicidad de un determinado “tipo” textual pueda ser determinada objetivamente mediante el análisis de rasgos cuantificables, o, en otras palabras, nuestra firme convicción de que la filosofía multidimensional es operativa y útil.

## 2. LA MECÁNICA DEL MODELO

En esta sección recorreremos las etapas fundamentales del modelo multidimensional aplicado, en particular, a partir de Biber (1988). Haremos referencia a los siguientes

estadios del modelo: la hipótesis de que las dimensiones funcionales de un texto pueden ser computadas mediante el análisis de rasgos cuantificables (§2.1), la estandarización estadística de dichos rasgos (§2.2), el agrupamiento de los resultados mediante técnicas estadísticas reputadas (§2.3) y la homologación de los resultados en torno al concepto central de dimensión (§2.4).

### 2.1. Punto de partida: concepto de “dimensión”

La idea central sobre la que gira todo el proceso es la necesidad de cuantificar la posición relativa de un género con respecto a una “dimensión” o “factor”. Ya que el concepto “dimensión” es central para llegar a un entendimiento pleno del modelo, detengámonos en dicha noción antes de introducirnos en los elementos más mecánicos de la teoría.

Por “dimensión”, Biber y Finegan entienden aquella perspectiva mediante la cual un género (o texto) determinado puede ganar una interpretación funcional concreta. Así, la dimensión “elaborado *versus* dependiente del contexto” nos permitirá concluir en qué medida un pasaje, texto o género es más o menos elaborado con respecto a otro, esto es, desde el ángulo opuesto, en qué proporción los referentes dispersos en el material textual a analizar están íntimamente ligados a circunstancias externas a la propia realidad textual. Ya que la pretensión que se esconde detrás del concepto “dimensión” es exclusivamente relacional, su denominación, en prácticamente la totalidad de los casos, contiene dos polos opuestos, enlazados convencionalmente mediante “*versus*”. A continuación listamos algunas de las dimensiones sobre las que más han investigado Biber y sus seguidores<sup>3</sup>:

- Dimensión 1: producción informativa *versus* producción participativa (*involved*)
- Dimensión 2: producción narrativa *versus* producción no narrativa
- Dimensión 3: producción elaborada *versus* producción dependiente del contexto
- Dimensión 4: producción abiertamente argumentativa *versus* producción no claramente argumentativa
- Dimensión 5: estilo impersonal *versus* estilo personal

La reformulación de estas dimensiones obedece a unos principios sociales, culturales, psicológicos, situacionales y, en general, funcionales –sobre esta interpretación (cuestionable) trataremos en §3.4 abajo–. La posición relativa de un texto o género en cada una de las dimensiones analizadas nos conducirá a la interpretación funcional (sincrónica o diacrónica) de dicho elemento.

Quedémonos, pues, con esta introducción del concepto “dimensión”, y dirijamos nuestras miradas al proceso técnico. La comprensión de dicho procedimiento desembochará en un mejor entendimiento de la noción anterior, a cuya vertiente estadística dedicamos la sección 2.4.

El Gráfico 1 muestra de modo esquemático todo el proceso descrito en las próximas secciones:

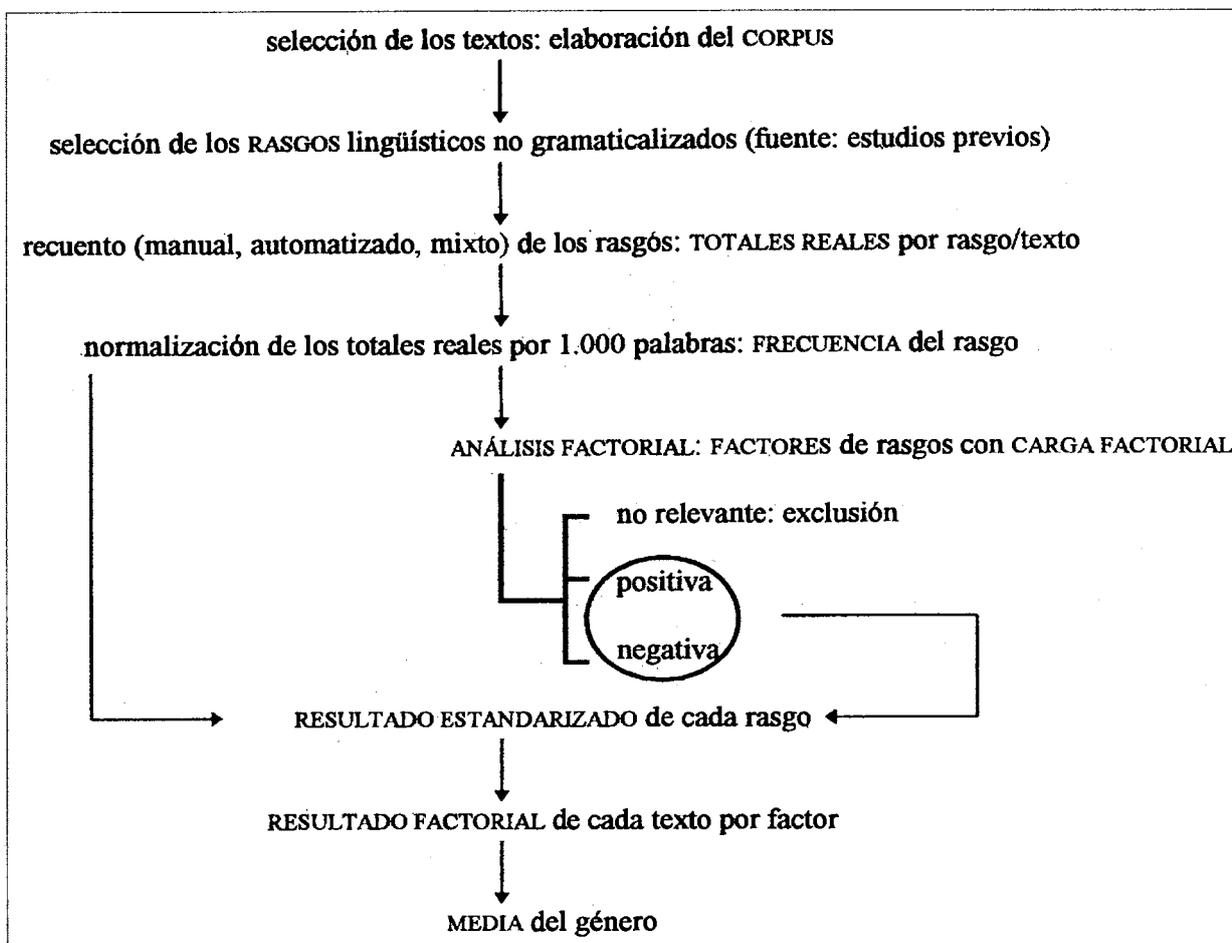


Gráfico 1. Esquema del modelo multidimensional de Biber & Finegan

## 2.2. La selección y recuento de los rasgos

Si todo el entramado multifactorial está basado en la necesidad de elevar a rango numérico ciertos elementos de un texto o género para, de esa manera, proceder a una interpretación funcional del material textual, el punto de partida del proceso deberá ser la selección de una serie de elementos que puedan ser determinados objetivamente, si es posible, mediante técnicas computacionales. Dichos elementos reciben el nombre de “rasgos” en el modelo de Biber & Finegan.

Casar la idea de que esos rasgos han de ser representativos del texto y a la vez la necesidad de que puedan ser computados y reducidos a una mera expresión numérica desemboca inevitablemente en la conclusión de que estos han de poseer naturaleza lingüística<sup>4</sup> no gramaticalizada. Por “no gramaticalizada”<sup>5</sup> nos referimos al requisito añadido de que la aparición de un rasgo lingüístico no puede estar condicionada por las estruc-

turas gramaticales invariables, esto es, fosilizadas, de la lengua en cuestión. A modo de ejemplo, si la inversión de sujeto y verbo fuese un rasgo lingüístico relevante en inglés a la hora de profundizar en la caracterización funcional de cierto género textual, la incidencia, por ejemplo, de inversión en las oraciones interrogativas directas o de aquellas estructuras con elementos negativos topicalizados (del tipo *Never did I smoke a cigar*) no puede ser tenida en cuenta a la hora de incluir el fenómeno inversión en dicho estudio, pues este está absolutamente determinado en estos casos por las circunstancias gramaticales de esos tipos de construcciones. Dicho de otro modo, la inversión sujeto-verbo en preguntas o en oraciones iniciadas por *never* en inglés no es significativa desde una perspectiva estilística, pues no obedece ni a preferencias subjetivas del autor ni a condicionantes textuales de un cierto género. Son las propias estructuras gramaticales fijadas ya en la lengua las que no permiten la ausencia de inversión en esos casos<sup>6</sup>.

Partiendo de esos preceptos, Biber & Finegan analizaron en su corpus de textos la aparición de una serie de rasgos lingüísticos que habían sido catalogados como representativos en la caracterización de géneros en estudios previos sobre análisis textual no necesariamente multidimensionales (volveremos a esta cuestión en §3.3). Así, en la Tabla 1 ofrecemos una selección de los 67 rasgos iniciales que, para Biber & Finegan, están asociados con funciones comunicativas específicas en la lengua inglesa de nuestros días (el listado completo de rasgos puede verse en Biber 1988: 73-75, Tabla 4.4):

- DIMENSIÓN 1 (producción informativa *versus* producción participativa):
  - RASGOS POSITIVOS: número de nombres, preposiciones, adjetivos atributivos, longitud de las palabras, proporción palabras/tipo de palabras
  - RASGOS NEGATIVOS: verbos de privacidad, verbos en presente, *be* como verbo principal, modales de posibilidad, contracciones, eliminación de *that*, pronombres de primera y segunda persona, demostrativos, indefinidos, proformas *do* e *it*, negación analítica, formas enfáticas, amplificadores, partículas discursivas, formas de rodeo, subordinación causal, coordinación oracional, relativos con antecedentes oracionales, oraciones del tipo *wh*, preguntas de tipo *wh*, preposiciones finales
- DIMENSIÓN 2 (producción narrativa *versus* producción no narrativa):
  - RASGOS POSITIVOS: pronombres relativos del tipo *wh* en función de sujeto y objeto, relativos precedidos por preposiciones, coordinación no oracional, nominalizaciones
  - RASGOS NEGATIVOS: adverbiales de tiempo, lugar y otros
- DIMENSIÓN 3 (producción elaborada *versus* producción dependiente del contexto):
  - RASGOS POSITIVOS: conectores (conjuntos), pasivas, oraciones adverbiales con participios de pasado, eliminación de tipo *WHIZ* con participios de pasado, otros tipos de subordinación adverbial
  - RASGOS NEGATIVOS: (no hay)

Tabla 1. *Rasgos lingüísticos determinadores de las dimensiones de Biber & Finegan*

Dejemos para la siguiente sección la explicación del agrupamiento de los rasgos en dimensiones y en polaridades positiva-negativa, y adentrémonos a continuación en el proceso de recuento de dichos rasgos o, en términos más técnicos, de realización de “índices” (véase Sigley 1997a: 214s y 1997b: 205ss).

Incluso en corpus pequeños de textos, el recuento de rasgos puede convertirse en una cuestión laboriosa y tediosa, a no ser que esté apoyado por técnicas computacionales. Podríamos aquí distinguir tres grandes tipos de procedimientos de recuento de rasgos: (i) el de aquellos que cuentan con una lista cerrada invariable de elementos (por ejemplo, buscar *there* o *not* en un corpus de inglés contemporáneo), que puede realizarse mediante un editor o procesador de textos convencional; (ii) la determinación de los valores de aparición de algunos rasgos, que requiere el empleo de utilidades informáticas más específicas (por ejemplo, la proporción de palabras/tipos de palabras puede obtenerse fácilmente mediante programas como, entre otros, TACT o WordSmith; sobre herramientas de análisis textual, véase, por ejemplo, Pérez Guerra 1998); y (iii) el recuento de categorías gramaticales que no pueden determinarse mediante criterios exclusivamente formales. Mientras que la cuantificación de aquellos rasgos aludidos en (i) y (ii) no supone mayor problema que el uso de herramientas informáticas bastante habituales para los estudiosos de la lingüística discursiva, el recuento de los rasgos del tipo (iii) sí es particularmente problemático. Las únicas ayudas que el analista puede tener a su disposición a la hora de contar las apariciones de rasgos no identificables mediante un análisis superficial del material textual son la elaboración de programas *ad-hoc* que le permitan, con mayor o menor éxito, encontrar determinados rasgos, o la existencia de anotación o etiquetación gramatical en su corpus de textos, esto es, información extra-textual añadida directamente al texto mediante delimitadores específicos que no se confunden en modo alguno con el corpus en sí. Ya que, desafortunadamente, la gran mayoría de los corpus (especialmente diacrónicos) bien no contienen anotación alguna o bien la anotación de la que disponen es insuficiente para la cuantificación automática de los rasgos relevantes, el recuento manual o mediante programas a la medida es necesario en muchos casos (en §3.2 nos adentraremos en ciertas estrategias computacionales empleadas por Biber & Finegan).

### 2.3. *Hacia el análisis factorial de nuestros datos*

Partamos del supuesto de que ya hemos determinado qué rasgos lingüísticos no gramaticalizados desempeñan un papel importante en la aplicación de un tratamiento dimensional sobre nuestro corpus de textos y de que ya hemos efectuado el recuento de todos esos rasgos relevantes. El siguiente paso hacia la determinación del resultado factorial de cada texto y la media del género es la normalización de la aparición de los rasgos, de tal forma que la diferente extensión de los textos no pueda incidir erróneamente en la relevancia estadística de un rasgo sobre otro. De esta manera, el total real de la distribución de cada rasgo en cada uno de los textos es normalizado sobre una

base común de 1.000 palabras. El resultado de esta homogeneización de los datos resultantes del recuento de apariciones de los rasgos recibe el nombre de “frecuencia” del rasgo.

Mediante el proceso estadístico de análisis factorial<sup>7</sup>, las frecuencias son agrupadas en los llamados “factores”, de acuerdo con su co-aparición en el texto. Más exactamente, esta técnica permite repartir todos los rasgos lingüísticos examinados en un número limitado de diferentes factores según los resultados de la correlación de todos ellos. La premisa resultante de esta taxonomía de rasgos es la consideración de que estos factores representan dimensiones de variabilidad, a las cuales se da posteriormente una interpretación “situacional” a la vista de las funciones apuntadas por los distintos rasgos que pertenecen al mismo grupo. Volveremos a este aspecto problemático del modelo en §3.4.

Empleemos como ejemplo la taxonomía de rasgos contemplada en la Tabla 1. La aplicación del análisis factorial a las frecuencias de aparición de cada rasgo en cada texto, esto es, a los totales normalizados por 1.000 palabras, permite descubrir, entre otras, que existe la tendencia de que a mayor cantidad de nombres, mayor de preposiciones, adjetivos atributivos, mayor longitud media de las palabras y mayor ratio entre tipos de palabras y número de palabras (rasgos positivos de la dimensión 1). Por el contrario, la proliferación correlativa de estos rasgos guarda una proporción inversa con, por ejemplo, el número de verbos de privacidad, pronombres de primera y segunda persona, demostrativos, indefinidos, etc. (rasgos negativos de la dimensión 1). Asimismo, estas tendencias, que representan en realidad polos opuestos de la misma correlación, no guardan relación alguna con los rasgos de las dimensiones 2 y 3.

Tras el análisis factorial, cada rasgo aparece asociado con una cifra de co-aparición comprendida entre ‘0’ y ‘1’<sup>8</sup>, la cual, técnicamente, recibe el nombre de “carga factorial” del rasgo. Aquellos rasgos con baja carga factorial, esto es, con mínima incidencia a la hora de contribuir a la definición funcional de un texto o género, serán despreciados. Así, los rasgos de cargas factoriales inferiores a 0,3, 0,35 o 0,4 son excluidos de los estudios de Biber (1988, 1990) y Biber y Finegan (1997), respectivamente. Por otro lado, a las cargas factoriales de aquellos rasgos que, como observamos en el párrafo anterior, contribuyen positivamente en la correlación se les asigna polaridad positiva, mientras que aquellos que son marcadamente escasos en el texto reciben un signo negativo. En palabras de Biber (1995a: 113-114), “when the features with positive loadings [N.A.: cargas positivas] occur together frequently in a text, the features with negative loadings are markedly less frequent in that text, and vice versa”.

De todo el proceso descrito en esta sección, únicamente nos interesa el hecho de que la técnica del análisis factorial nos permite, por un lado, establecer grupos de rasgos según sus tendencias de co-aparición positivas o negativas, y, por otro, descartar aquellos rasgos que, aunque destacados como tales en estudios previos, no son especialmente pertinentes en la caracterización funcional de un texto. De las cargas factoriales de cada rasgo necesitaremos en las secciones que siguen exclusivamente la polaridad.

#### 2.4. *Cuando 2+2 es ficción: el nacimiento de la “media” de un género*

Hasta el momento, de cada texto del corpus disponemos de un listado de grupos correlativos de rasgos relevantes, cada uno de ellos con una polaridad determinada. Carecemos de cifras globales de textos o incluso de géneros, lo cual lograremos precisamente en los pasos restantes que exploraremos a continuación.

Todas las frecuencias, esto es, porcentajes normalizados, de los rasgos de cada uno de los textos han de ser estandarizadas según el mismo criterio. Esto permitirá, por una parte, poder manejar homogéneamente los resultados globales resultantes en cada texto, y, por otra, descartar aquellas distribuciones idiosincrásicas de un rasgo determinado en un texto, que perfectamente pudieran distorsionar los resultados finales. A este respecto, imaginemos las incorrecciones a las que pudiera llevar el tomar como dato absoluto el hecho de que un autor concreto haga un uso excesivo de las contracciones, mientras que el resto de los autores estudiados del mismo periodo muestran proporciones muy reducidas de contracciones en sus escritos. La técnica que se emplea desde Biber (1988) ha sido la estandarización a una media de 0,0 y una desviación típica de 1,0, disponible en la mayoría de los paquetes de estadística, e incluso realizable manualmente con facilidad. Los resultados de esta armonización estadística reciben la denominación de “resultados estandarizados” de cada rasgo.

A continuación, la simple adición de los resultados estandarizados de aquellos rasgos con carga factorial positiva y la sustracción de los resultados estandarizados de los rasgos de carga factorial con polaridad negativa nos devuelve un valor que representa la media del texto completo con respecto a un determinado agrupamiento de correlación de rasgos, o, en otras palabras, el “resultado factorial” del texto.

Obtener a partir de los resultados factoriales individuales de los textos la “media” del género no es una tarea complicada. Para ello, simplemente calculamos la media aritmética de los textos que conforman el género concreto. Gracias a esta cifra final, estamos ya en disposición de establecer todo tipo de comparaciones tanto sincrónicas como diacrónicas entre géneros.

### 3. RIZANDO EL RIZO: OBSERVACIONES “A COSTA” DEL MODELO

Mientras que en §2 exploramos los principios que rigen el modelo multidimensional de Biber & Finegan, en esta sección reflexionaremos sobre posibles desajustes del mecanismo de cuantificación propuesto, especialmente, en Biber (1988), que, lejos de invalidar la teoría, pueden conducir a mejoras sustanciales en determinados aspectos del entramado estadístico.

La propuesta multidimensional puede ser reducida a los siguientes estadios de investigación: la composición del corpus de textos y la determinación de los rasgos lingüísticos (§3.1), el reconocimiento de los rasgos lingüísticos en nuestro corpus (§3.2), la consecución de los resultados factoriales de cada texto y la media de cada género (§3.3), y

la interpretación funcional de cada dimensión (§3.4). En cada uno de ellos nos detendremos en los párrafos que siguen.

### 3.1. *Creación del corpus de textos y determinación preliminar de rasgos lingüísticos*

Uno de los pecados del modelo multidimensional es la circularidad del proceso, lo cual puede derivar en obviedades no necesariamente pertinentes incluso en grandes proyectos que hagan uso del método. Más concretamente, especialmente cuando abordan estudios generalistas de organización textual (no siempre es este el caso), Biber & Finegan parten de conceptos prefijados que marcan todo el proceso multidimensional. Entre esos conceptos figura su concepción de *genre* o *text type* (utilizaremos ambas etiquetas indistintamente), esto es, “género” estilístico, textual, discursivo, etc. Si escogemos una serie de textos que, a nuestro juicio, son representativos de una categoría genérica concreta (“estilo académico”, por ejemplo), los filtramos mediante una serie de tamices lingüísticos cuantificables (estadio de la determinación de rasgos lingüísticos), obtenemos un valor numérico general de toda la categoría discursiva, y comparamos ese valor con el obtenido de análisis similares en otro grupo de textos que ya hemos incipientemente caracterizado dentro de otra categoría textual (“ficción narrativa” en esta ocasión), los resultados posibles del contraste se limitarán a dos opciones: lingüísticamente ambas categorías son similares o, por el contrario, están alejadas. Independientemente del juicio obtenido, el proceso ya gozaría de repercusión científica, pues demostraría empíricamente posibles sospechas que a lo largo de la historia de la estilística y el análisis del discurso ya habrían sido materializadas en estudios previos. Sin embargo, las conclusiones finales pueden pecar de determinismo analítico, pues están limitadas bien por nuestro tino o bien por nuestro error en la determinación previa de las categorías textuales.

Biber y Finegan (1997: 254) han dejado patente la necesidad de ampliar el número de textos manejados, lo cual es fácilmente justificable. Podemos imaginar los resultados de un estudio sobre la ficción narrativa que únicamente maneje un corpus formado por una novela de un autor. Obviamente, el resultado no será representativo de ninguna categoría textual, sino de “el” estilo de “ese” autor en “esa” novela. Este aspecto no merece mayor detención, pues a todas luces la búsqueda de representatividad debe impregnar cualquier estudio (lingüístico o no) que se enmarque dentro de la tradición de estudios de corpus textuales. Aceptemos, pues, sin reservas el uso de numerosos textos por cada una de las categorías bajo observación. Dos reflexiones surgen en este momento. En primer lugar, el lector podrá pensar que el método multidimensional requerirá una gran cantidad de textos. Lejos de esa intuición, los experimentos publicados de Biber, Biber y Finegan y Atkinson, entre otros, han sido realizados sobre muestras tremendamente reducidas de material textual, donde “tremendamente reducidas” posee un significado claramente mediatizado por otros estudios contemporáneos de lingüística de corpus que emplean fuentes textuales de millones de palabras. En la siguiente tabla, mos-

tramos gráficamente el número de palabras manejadas en algunos de los trabajos representativos dentro del marco multidimensional:

Estudio	extensión del corpus
Biber (1988)	aprox. 960.000 palabras
Biber y Finegan (1989)	aprox. 120.000 palabras
Atkinson (1992)	186.553 palabras
Atkinson (1996)	aprox. 250.000 palabras
Biber y Finegan (1997)	aprox. 1.700.000 palabras

Tabla 2. *Extensión de los corpus sobre los que se ha aplicado el análisis multidimensional*

Es más, sobre la cuestión del número de palabras, Biber (1995b: 364) ha defendido la irrelevancia del tema, y ha llegado a aseverar que el modelo multidimensional no requiere grandes cantidades de textos sino textos representativos. En sus propias palabras, “studies have shown that the dimensions of variation identified in the 1988 multidimensional analysis can be replicated in much smaller corpora, if the corpus represents the same range of register variation” (véase, igualmente, Biber 1990: 263-269)<sup>9</sup>. La cuantificación de la representatividad de un texto en un corpus determinado es una asignatura todavía pendiente en lingüística de corpus y provoca vivas discusiones allí donde surge. Aparquemos esta cuestión y limitémonos a precisar que, ya que la normalización por 1.000 palabras es uno de los procedimientos homogeneizadores de la propuesta multidimensional, debemos rechazar cualquier pasaje que no llegue a la cifra de 1.000 palabras, cuya normalización sería técnicamente imprecisa. Esta observación no debe jamás interpretarse como una recomendación sobre el número de palabras de cada texto; al contrario, nuestro techo intuitivo es bastante más restrictivo que el de Biber en lo que se refiere a número mínimo de palabras.

En segundo lugar, como ya hemos avanzado más arriba, los textos empleados (sean muchos, pocos o los estrictamente necesarios) ya están de antemano clasificados en categorías textuales prefijadas. En esto, la tradición juega un papel fundamental. Lejos de cuestionarnos la pertenencia de un texto determinado a una u otra categoría, nuestro instinto lingüístico lo etiqueta inmediatamente. Así, por ejemplo, por “textos académicos” entendemos estudios lingüísticos, literarios, médicos, físicos, etc.; en “textos religiosos” incluimos pasajes bíblicos, sermones, etc. Sin embargo, un texto sobre geometría puede estar más cercano a la ficción narrativa que a un ensayo filosófico; un sermón puede parecerse más a una obra de teatro que a una disertación sobre el pecado original, etc. De hecho, Taavitsainen (1993: 190-191, 194) demuestra mediante el análisis factorial de varios géneros del inglés medio que las fronteras entre los géneros son imprecisas y que muchos textos podrían encuadrarse en más de una categoría. Así, *The Parson's Tale* se comporta como los sermones en unos aspectos y como ficción en otros. La conclusión

inmediata es que la fiabilidad del modelo multidimensional está claramente condicionada por la homogeneidad de los textos a la hora de su agrupamiento en categorías. Dicho de otro modo, cuanto mayor sea la variedad (aunque controlada por los prejuicios de la tradición) de las categorías textuales a analizar más relativos serán los resultados. Únicamente mediante el análisis y posterior contraste de categorías textuales muy concretas podremos extraer conclusiones válidas. Sobre la fiabilidad de los corpus utilizados por Biber & Finegan, véase Altenberg (1989: 170), Baayen (1997: 66) o Besnier (1998: 127).

### 3.2. *El recuento de los rasgos lingüísticos*

El proceso de determinación y recuento de rasgos está, en nuestra opinión, viciado no sólo por la preselección interesada del material y la (quizás errónea) fe ciega en estudios no estadísticos (obviamente, no multidimensionales) anteriores sino también por el procedimiento seguido a la hora de la cuantificación de las ratios de aparición de los propios rasgos en el corpus y su agrupamiento mediante el análisis factorial de los resultados obtenidos en inglés contemporáneo.

Por lo que respecta a la primera de las deficiencias apuntadas, ya hemos reparado en §2.2 en el hecho de que, mientras que el recuento de algunos rasgos es fácilmente realizable mediante un simple editor de textos o alguna aplicación específica de análisis textual, otros han de ser determinados a mano o mediante técnicas informáticas a la medida, especialmente en aquellos casos en los que los corpus de textos contienen exclusivamente, justo eso, textos, sin ningún tipo de anotación gramatical añadida. Así, Biber (1988), con el fin de automatizar el recuento en el mayor grado posible, desarrolló su propio anotador y sistema de anotación que, aunque bastante eficiente, precisó en algunas ocasiones de desambiguación y corroboración manuales<sup>10</sup>. Ni que decir tiene que todo el proceso estadístico posterior confía ciegamente en los resultados obtenidos tras la aplicación de los algoritmos desarrollados por Biber (la descripción de dichas fórmulas puede encontrarse en Biber 1988: Apéndice II).

La elegancia del procedimiento, sin embargo, está reñida en algunos casos con su eficacia. Así, el análisis manual de los datos que hemos realizado en nuestro estudio (en prensa) nos lleva a reconocer que algunos algoritmos no están libres de problemas. Por ejemplo, la fórmula propuesta por Biber (1988) para el recuento de oraciones *wh* no identifica todas las apariciones de ese rasgo. El algoritmo PUB/PRV/SUA + WHP/WHO + xxx (donde xxx es NOT = AUXILIARY) sólo se detiene ante ejemplos de oraciones *wh* que funcionan como objetos de verbos públicos (PUB), de privacidad (PRV) o de persuasión (SUA). En consecuencia, aquellas apariciones de oraciones *wh* del tipo *What he said was the truth* son obviadas simplemente porque, en palabras de Biber (1988: 231), “they could not be identified reliably by automatic analyses”<sup>11</sup>. Del mismo modo, el analizador de Biber & Finegan añade erróneamente a la clase de la proforma *do* todas las apariciones de *do* como verbo principal (véase Altenberg 1989: 172).

Otro grave obstáculo que nos hace, como poco, dudar de la validez de ciertos resultados de Biber & Finegan es la falta de actualización de los algoritmos originales de Biber (1988) en estudios posteriores, normalmente de naturaleza diacrónica. A pesar de que la idoneidad de las fórmulas del trabajo de 1988 no es cuestionada en absoluto en publicaciones posteriores, hemos encontrado serios defectos en la confección de, por ejemplo, listas de elementos que configuran un rasgo lingüístico determinado, sobre las que se basa el recuento computacional. Así, mientras la lista propuesta de amplificadores incluye *absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly* y *very*, en inglés medio tardío echamos en falta amplificadores como *all(e), he(a)rtely, sore* o *fervently*, lo cual puede desembocar en resultados incorrectos.

La segunda gran deficiencia, esto es, la aplicación de aquellos rasgos de Biber (1988), supuestamente válidos en inglés contemporáneo, a textos de periodos anteriores en estudios como Biber y Finegan (1989, 1992, 1997) es, en nuestra opinión, un nuevo descuido de planteamiento teórico inicial. La técnica de análisis factorial, a la que nos hemos referido en §2.3, nos permite agrupar rasgos lingüísticos en virtud de tendencias de co-aparición (o co-exclusión) que quedan determinadas por las cifras reales obtenidas tras la normalización del recuento de cada uno de los rasgos. Los listados de rasgos y su asignación a polaridades dentro de cada uno de los factores (véase Tabla 1) han sido realizados tomando como premisa los datos reales de Biber (1988). Independientemente de la acusada circularidad en la toma de datos, ya comentada en §3.1, a la que puede conducir el seleccionar como punto de partida tanto una tipología textual determinada como asunciones previas sobre qué rasgos inducen qué dimensión en un texto, los resultados arrojados tras la aplicación del análisis factorial a los datos del inglés contemporáneo serán sólo válidos en estudios centrados sobre corpus de inglés contemporáneo. Las conclusiones que uno pueda extraer del examen de material textual de periodos anteriores de la lengua inglesa desde la óptica de las dimensiones que (aceptémoslo así) son operativas en inglés contemporáneo serán inválidas en tanto que están sustentadas en preceptos teóricos de difícil justificación. Dicho de otro modo, el contraste entre los resultados factoriales obtenidos tras la aplicación del aparato teórico de Biber (1988) en textos actuales y en corpus históricos de la lengua inglesa adolece no sólo de la inclusión prematura de unos textos en un género específico sino también de la aceptación de que aquellos rasgos que efectivamente conducían a determinadas generalizaciones funcionales en el inglés de nuestros días son los mismos que resultaban operativos en inglés medieval o inglés moderno.

A modo de conclusión, mediante estos detalles específicos que hemos puesto de relieve en los párrafos anteriores, hemos querido poner en duda la exactitud de los resultados numéricos “crudos” de Biber & Finegan. Aunque somos conscientes de que esos datos son sometidos a estrictos procesos de estandarización, lo cual aminorará la relevancia de los posibles errores, queremos dejar constancia de la gravedad metodológica que, en nuestra opinión, supone supeditar la búsqueda de la fiabilidad más estricta a la inmediatez instrumental. En otras palabras, hemos demostrado que, por lo que respecta

a la computación de la aparición de rasgos en el corpus, el resultado final de las investigaciones de Biber & Finegan se ha puesto al servicio de la automatización del proceso y no al contrario, como cabría desear, esto es, que los mecanismos computacionales de recuento estuvieran condicionados por los resultados buscados.

### 3.3. *Sobre la reducción del texto a un número*

En §2.4 describimos el proceso propuesto por Biber & Finegan para la reducción de todos los rasgos lingüísticos de un texto a un único dígito. En pocas palabras, se sumaban las frecuencias normalizadas y estandarizadas con proporciones positivas de co-aparición y se restaban las negativas. La polaridad de estas se obtenía tras el análisis factorial, el cual caracterizaba las cargas factoriales de los rasgos como positivas o negativas.

Los procesos estadísticos comprendidos en esta fase del análisis multidimensional son extremadamente importantes pues determinarán la media de cada género, esto es, la base empírica fundamental para posteriores digresiones científicas. Precisamente por ello es por lo que haremos especial hincapié en lo que, creemos, pueden ser incorrecciones significativas del modelo.

Un primer hecho que nos sorprende es la aleatoriedad de la discriminación de factores tras el análisis factorial. Como hemos indicado más arriba, no todos los factores son procesados en esta fase, pues aquellos con cargas factoriales bajas son descartados y “expulsados” del proceso estadístico. Así, en Biber (1990), aquellos rasgos con carga factorial absoluta (esto es, independiente de la polaridad) inferior a 0,3 son rechazados, en Biber (1988) aquellos que no alcanzan el 0,35, mientras que en Biber y Finegan (1997) el listón está colocado en la carga factorial de 0,40. Lo que pudiera parecer un hecho sin demasiadas consecuencias, sí puede tenerlas, pues el número de rasgos en juego desempeña un papel significativo. Más concretamente, la carga factorial no tiene por qué ser paralela a la frecuencia de los rasgos. Mientras que la primera determina la incidencia de un factor en el esquema factorial o dimensional de sus compañeros de grupo, la frecuencia se limita a expresar el número de veces que un rasgo determinado aparece por cada 1.000 palabras de texto. No es pues extraño que un rasgo con una frecuencia no despreciable pueda tener una carga factorial pequeña, o, al contrario, que un rasgo cuyo grado de aparición en el texto sea mínimo sea altamente relevante en cuanto a la caracterización dimensional del pasaje, y, de ahí, su carga factorial sea alta. El hecho de no considerar el primero implicará que su frecuencia (insistimos, no su carga factorial) no será ni añadida ni sustraída en el resultado factorial del texto, lo cual puede derivar en una oscilación cuantitativa que podría ser importante en su contribución a la media del género.

Por lo expuesto anteriormente sugerimos que la discriminación de rasgos no ha de hacerse teniendo en cuenta las cargas factoriales, sino, en todo caso, las frecuencias de los rasgos. Así, en nuestro trabajo (en prensa), hemos descartado aquellos rasgos cuya distribución en cada texto es inferior a una aparición por cada 1.000 palabras. Es más,

somos conscientes de que, desde una perspectiva teórica, resulta difícil justificar la eliminación de aquellos rasgos que el análisis factorial ha caracterizado como relevantes, en mayor o menor cuota, a la hora de definir las circunstancias lingüísticas que conducen a una tipología textual concreta.

Un segundo elemento problemático que merece ser mencionado aquí es el modo en el que se calcula el resultado factorial de cada texto. Como hemos descrito en secciones anteriores, el dígito unificador de la relevancia estadística de los rasgos lingüísticos definidores de la dimensión bajo análisis se obtiene mediante sumas y restas de frecuencias. Las cargas factoriales son prácticamente obviadas, pues de ellas únicamente se hace uso de la polaridad (signo positivo implica adición; signo negativo, sustracción). A nuestro entender, es sinónimo de pobreza estadística relegar el papel del índice de correlación de los rasgos, esto es, el grado de importancia de un rasgo a la hora de determinar si el estilo de un texto concreto es más informativo o más participativo, a una mera dicotomía positivo *versus* negativo. No sólo pensamos que la carga factorial de un rasgo es extremadamente importante en lo que respecta a la averiguación del resultado factorial del texto, sino que la frecuencia ha de estar supeditada a esta. Nuestra propuesta conciliadora pasa por multiplicar la carga factorial (entre '-1' y '+1') por la frecuencia como paso previo incluso a la estandarización factorial de cada rasgo. Dicha multiplicación no sólo reducirá sensiblemente aquellas frecuencias con grados bajos de significación sino que implementará automáticamente la polaridad de cada rasgo. A efectos prácticos, aquellos rasgos con mayor significación prácticamente mantendrán el valor numérico absoluto de sus frecuencias normalizadas (cuanto más cercanas a '+1' o '-1' sean sus cargas factoriales, menor reducción sufrirán sus frecuencias), mientras que las frecuencias de los rasgos menos pertinentes para el análisis multidimensional decrecerán sustancialmente (cuanto más cercanas a '0' sean sus cargas factoriales, más se reducirá su frecuencia). Con ello conseguimos que ningún rasgo desaparezca, a no ser que bien su frecuencia o bien su carga factorial sean nulas, esto es, que su aparición o su significación no sean relevantes.

El mecanismo orquestado por Biber & Finegan no permite hacer distinciones entre rasgos, pues todos aquellos no rechazados por la varita mágica del 0,3, 0,35 o 0,40 tendrán la misma consideración en cálculos estadísticos posteriores. Nuestra propuesta permite, con éxito, repercutir la escala de afección (o, si se prefiere, prototipicidad) de los rasgos en cada dimensión en la cifra caracterizadora del rasgo. Es más, proponemos que el paso siguiente a la obtención de estas, llamémosles, "cifras reguladoras de la significación de cada rasgo" sea la estandarización a una media de 0,0 y una desviación típica de 1,0, lo cual evitaría los llamados "casos perdidos" mediante el uso exclusivo de las frecuencias (nuestra propuesta anterior) o cargas factoriales (en Biber & Finegan).

Observemos las diferencias entre nuestra propuesta y la de Biber & Finegan en la siguiente Tabla:

HACIA UNA NUEVA ARITMÉTICA DEL TEXTO

DIMENSIÓN 1	FRECUENCIAS (/1.000 palabras)	CARGA FACTORIAL	RESULTADOS ESTANDARIZADOS	CIFRAS REGULADORAS
<i>Rasgos positivos:</i>				
- rasgo 1:	35,60	0,96	-0,31	34,17
- rasgo 2:	8,21	0,78	0,33	6,40
- rasgo 3:	3,83	0,62	0,14	2,37
- rasgo 4:	2,19	0,47	-0,12	1,02
<i>Rasgos negativos:</i>				
- rasgo 1:	187,50	-0,80	0,46	-150
- rasgo 2:	105,00	-0,54	0,17	-56,7
- rasgo 3:	10,41	-0,39	0,27	-4,05

Tabla 3. *Resultados estandarizados (Biber & Finegan) versus "cifras reguladoras de la significación de cada rasgo" (simulación)*

Para finalizar, retomaremos un desacierto del modelo de Biber & Finegan ya insinuado anteriormente, a saber, la selección de rasgos lingüísticos. En secciones anteriores habíamos criticado el hecho de que todo el proceso parte de unos géneros predeterminados y de unos rasgos que, más o menos intuitivamente, han sido reconocidos como reveladores de hechos estilístico-funcionales en fuentes académicas relevantes.<sup>12</sup> Así las cosas, el riesgo en el que el investigador puede caer si su listado de rasgos es incompleto es enorme. En otras palabras, pensemos en la incidencia numérica y, posteriormente, estadística de la ausencia del rasgo más determinante de, por ejemplo, la dimensión 1. Tanto en la propuesta de Biber & Finegan como en la nuestra (reconocemos que el daño sería mayor en nuestra propuesta pues esta refleja más fielmente la relevancia de los rasgos) dicha carencia desencadenaría resultados estandarizados más bajos, y de ahí, resultados factoriales reducidos. "El" rasgo más influyente en la caracterización funcional de un tipo de textos con toda seguridad ya ha sido destacado por los estudiosos del tema. Sin embargo, la posibilidad de que algún rasgo menos importante (tanto positivo como negativo) se halle aún en el baúl de los recuerdos no parece ser una hipótesis descartable a todas luces. La influencia de este desacierto no sería tan drástica como en el ejemplo anterior, aunque bien pudiera distorsionar en buena medida todo el sistema dimensional. Una vez más, la conclusión apunta hacia el hecho de que un análisis multidimensional no puede estar (exclusivamente) sostenido en una selección *a priori* (no estadística y, en muchos casos, informal) de rasgos lingüísticos. Es más, cualquier estudio de coordenadas funcionales como "contenido participativo", "producción elaborada", "producción dependiente del contexto", etc. podrá echar en falta algunos rasgos que, intuitivamente, pudieran ser clarificadores a la hora de situar un texto (o género) en una escala dimensional determinada.

### 3.4. Interpretación funcional de las dimensiones

El tramo final del modelo multidimensional de Biber & Finegan consiste en el cálculo de la media aritmética correspondiente a cada género en la escala del factor o dimensión en cuestión. Si bien el mecanismo estadístico escogido es absolutamente congruente con el fin perseguido, el paso siguiente, esto es, la interpretación funcional de cada dimensión está bastante alejado de la justificación estadística y base empírica que impregnan todo el modelo. Así, resulta difícil pasar de listados de rasgos con incidencia positiva o negativa a explicaciones globales de carácter extra-textual como cualquiera de los enunciados de las dimensiones de §2.1 (producción informativa *versus* producción participativa, producción narrativa *versus* producción no narrativa, producción elaborada *versus* producción dependiente del contexto, etc.). Es aquí donde, a nuestro parecer, el modelo carece de base científica suficiente para fundamentar las conclusiones funcionales en torno a las que conforma cada una de las dimensiones. Besnier (1998: 126) es absolutamente contundente cuando dice que en la descripción funcional de los factores Biber & Finegan añaden “subjective interpretation, and luck”.

## 4. CONCLUSIONES Y OBSERVACIONES FINALES

En este artículo hemos descrito el análisis multidimensional aplicado por Biber & Finegan y hemos reparado en cuestiones específicas que, a nuestro entender, requieren reformulaciones y correcciones, a la vez que hemos evitado entrar en cuestiones más específicas de los trabajos de estos autores, algunas de ellas, como la inadecuada composición del corpus o fallos adicionales del anotador computacional de Biber (1988), referenciadas con mayor o menor acierto en Watson (1994).

Al mismo tiempo que hemos demostrado nuestra fe en la necesidad de un mecanismo de cuantificación textual mediante el análisis factorial de las proporciones de coaparición de ciertos rasgos lingüísticos que permita situar un texto, pasaje o género en una escala funcional, creemos que el modelo de Biber & Finegan adolece de incorrecciones de peso que pueden conducir a la inutilidad de sus resultados. En particular, hemos criticado la circularidad del proceso, que toma como punto de partida una serie de asunciones tanto en lo que respecta a la tipología textual como a los rasgos lingüísticos definidores. Se echa así de menos un tratamiento aún más profundo del concepto de género o prototipo textual y una reclasificación de materiales textuales concretos en base a principios lingüísticos rigurosos. Igualmente, hemos hecho propuestas concretas de mejora de ciertas fases estadísticas del modelo, a nuestro entender, con gran incidencia positiva en los resultados finales (véase, al respecto, Baayen 1997: 66). Finalmente, hemos mostrado nuestra reticencia a aceptar interpretaciones funcionales más o menos fundadas de las dimensiones resultantes del análisis factorial. Creemos firmemente que el salto de la estadística textual a la extracción de conclusiones de índole extralingüística es un campo mínimamente desarrollado por Biber & Finegan, que requiere futuras investigaciones complementarias.

## NOTAS

1. Esta investigación ha sido financiada por el Ministerio de Educación y Cultura (MEC), a través de su Dirección General de Enseñanza Superior (DGES), proyectos PB97-0507 y PB96-0955, cuya colaboración aquí agradecemos.
2. En lo que sigue, a no ser que indiquemos lo contrario, emplearemos la autoría “Biber & Finegan” para referirnos, en general, a las publicaciones referenciadas al final del artículo de ambos autores, tanto individuales como conjuntas.
3. Además de estas cinco dimensiones, Biber (1988) también ha tratado dos adicionales, que dan lugar a las interpretaciones “elaboración informativa ‘en línea’” y “mitigación académica”, respectivamente [las traducciones son nuestras]. Sin embargo, estas dimensiones no han gozado del rendimiento científico de las cinco enumeradas en el texto principal, de las cuales las 1, 3 y 5 han sido a las que mayor atención se ha prestado en trabajos posteriores.
4. Ghadessy (1997) critica el hecho de que los rasgos de Biber & Finegan tengan que ser de naturaleza exclusivamente formal. En su opinión, circunstancias puramente superficiales no pueden contribuir en absoluto a la caracterización estilística (y, menos, funcional) de un texto o género. Su propuesta, que hace uso de las propiedades gramaticales y léxico-semánticas de los “temas” de Halliday (1994), conduce a una tipología textual a todas luces diferente de la de Biber & Finegan.
5. En palabras de Biber (1996: 174), estas “[n]on-linguistic association patterns [...] describe how certain features are differentially associated with registers, dialects, or historical change”, donde *non-linguistic associations* se refiere a aquellos rasgos no gramaticalizados, esto es, no prefijados por las estructuras no marcadas de la lengua objeto de análisis.
6. Como un revisor anónimo de este trabajo ha sugerido con gran acierto, la irrelevancia estadística de la incidencia de, por ejemplo, inversión en las estructuras antes mencionadas no niega la importancia de la construcción en general. En otras palabras, si bien la aparición de sujetos postverbiales en oraciones introducidas por elementos léxicos negativos no es significativa en la lengua inglesa, la existencia de estructuras invertidas en una taxonomía discursiva determinada sí puede ser influyente en cuanto a su caracterización lingüístico-funcional.
7. Sobre la técnica de “análisis factorial”, véase, entre otros, Biber (1988: §§ 5.2, 5.3; 1995b: §5.6.1), Hatch y Farhady (1982: 255-260) o Woods et al. (1986: 290-295). Por razones obvias, en este trabajo no podemos detenernos en los detalles técnicos de este análisis, que, como ya advierte Kilgarriff (1996), suponen la mayor complejidad del proceso.
8. Una carga factorial de ‘0’ implica que no hay correlación alguna entre el rasgo asociado con dicha carga y los restantes rasgos del grupo. Por el contrario, cargas factoriales equivalentes a ‘1’ son típicas de aquellos rasgos que mantienen una correlación de co-aparición perfecta con los demás de su clase dimensional.
9. Recordemos que la técnica estadística empleada por Biber & Finegan es válida, como indica Sigley (1997b: 231), incluso con textos de sólo 100 palabras.
10. La necesidad de disponer de un procedimiento automatizado ha sido criticada por, por ejemplo, Altenberg (1989: 171).
11. Igualmente, Altenberg (1989: 172) duda de la validez estadística de un recuento de cláusulas de relativo en el que se excluye aquellas con relativo cero porque estas no pueden ser rastreadas automáticamente.
12. Crystal (1991) retoma la cuestión de la procedencia y selección de rasgos lingüísticos en un estudio estadístico de análisis del discurso. El mismo tema es discutido en Watson (1994: 243ff), quien afirma: “one criticism of this approach of Biber’s is that the linguistic features chosen for investigation are based upon previous research and interpretative assumptions of the function of those features”.

## BIBLIOGRAFÍA

- Altenberg, B. 1989. "Review of Douglas Biber, *Variation Across Speech and Writing*". *Studia Linguistica* 43, 2: 167-174.
- Atkinson, D. 1992. "The Evolution of Medical Research Writings from 1735 to 1985: The Case of the *Edinburgh Medical Journal*". *Applied Linguistics* 13, 4: 337-374.
- Atkinson, D. 1996. "The Philosophical Transactions of the Royal Society of London, 1675-1975: A Sociohistorical Discourse Analysis". *Language in Society* 25: 333-371.
- Baayen, R.H. 1997. "Review of D. Biber, *Dimensions of Register Variation: A Cross-Linguistic Comparison*". *Literary and Linguistic Computing* 12, 1: 65-67.
- Besnier, N. 1998. "Review of Douglas Biber, *Dimensions of Register Variation: A Cross-Linguistic Comparison*". *Language in Society* 27, 1: 126-129.
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. 1990. "Methodological Issues Regarding Corpus-based Analyses of Linguistic Variation". *Literary and Linguistic Computing* 5, 4: 257-269.
- Biber, D. 1995a. *Dimensions of Register Variation. A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.
- Biber, D. 1995b. "On the Role of Computational, Statistical, and Interpretive Techniques in a Multi-dimensional Analysis of Register Variation. A Reply to Watson". *Text* 15, 3: 341-370.
- Biber, D. 1996. "Investigating Language Use through Corpus-based Analyses of Association Patterns". *International Journal of Corpus Linguistics* 1, 2: 171-198.
- Biber, D., S. Conrad y R. Reppen. 1998. *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D. y E. Finegan. 1989. "Drift and the Evolution of English Style: A History of Three Genres". *Language* 65, 3: 487-517.
- Biber, D. y E. Finegan. 1992. "The Linguistic Evolution of Five Written and Speech-based English Genres from the 17th to the 20th Centuries". *History of Englishes: New Methods and Interpretations in Historical Linguistics*. Eds. M. Rissanen, O. Ihalainen, T. Nevalainen y I. Taavitsainen. Berlín: Mouton de Gruyter. 688-704.
- Biber, D. y E. Finegan. 1997. "Diachronic Relations among Speech-based and Written Registers in English". *To Explain the Present. Studies in the Changing English Language in Honour of Matti Rissanen*. Eds. T. Nevalainen y L. Kahlas-Tarkka. Helsinki: Société Néophilologique. 253-275.
- Crystal, D. 1991. "Stylistic Profiling". *English Corpus Linguistics*. Eds. K. Aijmer y B. Altenberg. Londres: Longman. 221-238.
- Ghadessy, M. 1997. "Thematic Organisation as a Criterion for Establishing Text-types". *Estudios Ingleses de la Universidad Complutense* 5: 129-145.
- González Álvarez, D. y J. Pérez Guerra. 1998. "Texting the Written Evidence. On Register Analysis in Late Middle English". *Text* 18, 2: 321-348.

- González Álvarez, D. y J. Pérez Guerra. en prensa. "A Corpus-based Approach to the Multidimensional Variation of Four Late Middle English Genres". *Proceedings of the 10th International SELIM Conference*. Zaragoza: Universidad de Zaragoza.
- Halliday, M.A.K. 1994 (1985). *An Introduction to Functional Grammar*. Londres: Edward Arnold.
- Hatch, E. y H. Farhady. 1982. *Research Design and Statistics for Applied Linguistics*. Rowby, Mass.: Newbury House.
- Kilgarriff, A. 1996. "Review of Douglas Biber. *Dimensions of Register Variation: A Cross-Linguistic Comparison*". [Documento de Internet disponible en <http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/biber.asc>, también en *Journal of Natural Language Engineering*]
- Meurman-Solin, A. 1993. "Involvement Markers in the Characterization of Early Prose Genres". *Variation and Change in Early Scottish Prose: Studies Based on the Helsinki Corpus of Older Scots*. Ed. A. Meurman-Solin. Helsinki: Suomalainen Tiedeakatemia. 276-312.
- Pérez Guerra, J. 1998. *Análisis computarizado de textos. Una introducción a TACT*. Vigo: Universidade de Vigo (Servicio de Publicacións).
- Sigley, R. 1997a. "The Influence of Formality and Channel on Relative Pronoun Choice in New Zealand English". *English Language and Linguistics* 1, 2: 207-232.
- Sigley, R. 1997b. "Text Categories and Where You Can Stick Them: A Crude Formality Index". *International Journal of Corpus Linguistics* 2/2: 199-237.
- Taavitsainen, I. 1993. "Genre/Subgenre Styles in Late Middle English". *Early English in the Computer Age: Explorations through the Helsinki Corpus*. Eds. M. Rissanen, M. Kytö y M. Pallander-Collin. Berlín: Mouton de Gruyter. 171-200.
- Taavitsainen, I. 1997. "Genre Conventions: Personal Affect in Fiction and Non-fiction in Early Modern English". *English in Transition. Corpus-based Studies in Linguistic Variation and Genre Styles*. Eds. M. Rissanen, M. Kytö y K. Heikkonen. Berlín: Mouton de Gruyter. 185-266.
- Watson, G.J. 1994. "A Multidimensional Analysis of Style in Mudrooroo Nyoongah's Prose Works". *Text* 14, 2: 239-285.
- Woods, A., P. Fletcher y A. Hughes. 1986. *Statistics in Language Studies*. Cambridge: Cambridge University Press.