

LA ESTADÍSTICA DE PASEO CON LA ESTILÍSTICA, O EL ESTADO DE LA CUESTIÓN EN EL ANÁLISIS TEXTUAL MULTIDIMENSIONAL*

Dolores González Álvarez & Javier Pérez Guerra
Universidade de Vigo

ABSTRACT

Biber's multidimensional multifactorial model of the functional characterisation of text types is one of the most significant approaches to the description of genres by way of the careful examination of the linguistic features occurring in a corpus. In this paper, on the one hand, we explore both the model and a variety of relevant studies which have made use of Biber's analysis. On the other hand, we comment on some wrinkles to be ironed out in Biber's approach which we have observed in our investigation. The article thus provides an insight into the intricacies of the model and the possibilities which multifactorial analysis has as far as the characterisation of texts is concerned.

KEY WORDS: text-type, genre, multidimensional analysis, multifactorial analysis, factor analysis, linguistic feature.

RESUMEN

El marco multidimensional y multifactorial de caracterización de tipos de textos desarrollado por Biber es uno de los modelos más significativos de descripción de géneros mediante el examen riguroso de los rasgos lingüísticos que aparecen en un corpus. En este artículo, por un lado, analizamos tanto el modelo como un buen número de estudios relevantes que han hecho uso de la propuesta de Biber. Por otro lado, realizamos una serie de comentarios sobre algunos detalles que, a la luz de nuestra propia investigación, deben ser subsanados en el sistema de Biber. En definitiva, este estudio supone una revisión de las bases del modelo y de las posibilidades que el análisis multifactorial de textos tiene en lo que respecta a la caracterización de los textos.

PALABRAS CLAVE: tipo de texto, género, análisis multidimensional, análisis multifactorial, análisis factorial, rasgo lingüístico.

1. INTRODUCCIÓN

El modelo multidimensional, desarrollado por Biber (1985, 1986, 1988, 1990, 1993a, b, c y 1994), Biber y Finegan (1991) y sus inmediatos seguidores, intenta elevar a la categoría de materias cuantificables, y por tanto contrastables,



una serie de características textuales de un pasaje, de un grupo de pasajes o incluso de un género estilístico determinado que hasta el momento habían quedado reservadas a la mirada de estudiosos de la retórica, el estilo o el análisis del discurso.

En este artículo¹ nos proponemos una triple 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 al plano más simple (véase, en este sentido, Pérez-Guerra y González-Álvarez 1999). En segundo lugar, recorreremos a modo de ilustración los diversos tipos de trabajos lingüísticos tanto sincrónicos como diacrónicos enmarcados dentro de esta metodología. Por último, 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. A pesar de los desaciertos que pondremos de manifiesto a lo largo de nuestra exposición, queremos dejar clara desde el principio 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 útil y operativa.

2. LA MECÁNICA DEL MODELO

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

Por “dimensión” Biber entiende 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 determinar en qué medida un pasaje, texto o género es más o menos elaborado con respecto a otro, o desde el ángulo opuesto, en qué proporción los referentes dispersos en el material textual a analizar están ligados a circunstancias externas a la propia realidad textual. Cada dimensión es un continuum con dos polos opuestos, que en la terminología de Biber están enlazados con *versus* (por ejemplo, “estilo impersonal versus estilo personal,” “producción narrativa versus producción no narrativa,” etc).

* La investigación que aquí ofrecemos se enmarca en los proyectos de investigación PB97-0507 y PB96-0955, financiados por el Ministerio de Ciencia y Tecnología. Deseamos dejar constancia de nuestro agradecimiento por esta colaboración.

¹ Agradecemos las intervenciones de los asistentes al taller “Análisis multifactorial y multidimensional y su aplicación al estudio de la lengua inglesa” (XXIV Congreso de AEDEAN, Ciudad Real, 2000), en el que presentamos una versión preliminar de este trabajo.

Pasemos a continuación a describir el proceso técnico que permite identificar las dimensiones y situar un texto o género en cada una de ellas. 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.

Para que esos rasgos sean representativos de un texto y a la vez puedan ser computados y reducidos a una mera expresión numérica han de poseer naturaleza lingüística no gramaticalizada. Por no gramaticalizada nos referimos al hecho de que la aparición de un rasgo lingüístico no pueda estar condicionada por las estructuras gramaticales invariables, esto es, fosilizadas, de la lengua en cuestión. A modo de ejemplo, si la inversión del sujeto y el verbo (*The postman_{SUJ} comes_V here* versus [*Here*] *comes_V the postman_{SUJ}*) 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 la inversión en las oraciones interrogativas directas o de aquellas estructuras con elementos negativos topicalizados (del tipo *Never did I smoke a cigar*) no pueden ser tenidas en cuenta a la hora de incluir el fenómeno de inversión en dicho estudio pues está se encuentra absolutamente determinada en estos casos por las circunstancias gramaticales de la construcción. 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.

Partiendo de estos preceptos, Biber analizó en su corpus de textos la aparición de una serie de rasgos lingüísticos que habían sido catalogados como representativos de un determinado tipo de discurso (oral, narrativo, técnico, formal, etc.) en estudios previos sobre análisis textual no necesariamente multidimensionales. En la Tabla 1 ofrecemos los 67 rasgos que, para Biber, son potencialmente importantes para el inglés por estar asociados con funciones comunicativas específicas en la lengua inglesa de nuestros días.

Una vez seleccionados los rasgos nos adentraremos en el proceso de recuento. Incluso en corpus pequeños de textos, la totalización de rasgos puede convertirse en una cuestión laboriosa y tediosa, a no ser que esté apoyada en 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 entre palabras/tipos de palabras puede obtenerse fácilmente mediante programas como, entre otros, TACT, WordCruncher o WordSmith; 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 problema para los estudiosos de la lingüística discursiva acostumbrados al uso de



TABLA 1 (DE BIBER 1995A: 95S). RASGOS LINGÜÍSTICOS
POTENCIALMENTE IMPORTANTES PARA EL ANÁLISIS DEL INGLÉS

MARCADORES DE TIEMPO Y ASPECTO

1. Verbos en pasado
2. Aspecto perfectivo
3. Verbos en presente

ADVERBIOS DE LUGAR Y TIEMPO

4. Adverbios de lugar (*above, beside, outdoors*)
5. Adverbios de tiempo (*early, instantly, soon*)

PROFORMAS

6. Pronombres de primera persona
7. Pronombres de segunda persona
8. Pronombres personales de tercera persona (excepto *it*)
9. Pronombre *it*
10. Pronombres demostrativos
11. Pronombres indefinidos
12. Proforma *do*

PREGUNTAS

13. Preguntas directas de tipo *wh*

FORMAS NOMINALES

14. Nominalizaciones (terminadas en *-tion, -ment, -ness, e -ity*)
15. Gerundios (formas de participio funcionando como nombres)
16. Total de nombres

PASIVAS

17. Pasivas sin agente
18. Pasivas con *by*

FORMAS ESTATIVAS

19. *Be* como verbo principal
20. *There* existencial

SUBORDINACIÓN

21. Complementos verbales introducidos por *that* (*I say that he went*)
22. Complementos adjetivales introducidos por *that* (*I am glad that you like it*)
23. Cláusulas del tipo *wh* (*I believed what he told me*)
24. Infinitivos
25. Cláusulas adverbiales con participio de presente (*Stuffing his mouth with cookies, Joe ran out*)
26. Cláusulas adverbiales con participio de pasado (*Built in a single week, the house would stand for 50 years*)
27. Cláusulas postnominales con participio de pasado (*the solution produced by this process*)
28. Cláusulas postnominales con participio de presente (*the event causing this decline was...*)
29. Relativo *that* en función de sujeto (*the dog that bit me*)
30. Relativo *that* en función de objeto (*the dog that I saw*)
31. Relativo *wh-* en función de sujeto (*The man who likes popcorn*)
32. Relativo *wh-* en función de objeto (*the man who Sally likes*)
33. Relativos precedidos por preposiciones (*the manner in which he was told*)
34. Relativos con antecedentes oracionales (*Bob like fried mangoes, which is the most disgusting thing I've ever heard of*)
35. Subordinación causal (*because*)
36. Subordinación concesiva (*although, though*)
37. Subordinación condicional (*if, unless*)
38. Otras conjunciones subordinantes (por ejemplo *since, while, whereas*)

FRASES PREPOSICIONALES, ADJETIVOS Y ADVERBIOS

39. Frases preposicionales
40. Adjetivos atributivos
41. Adjetivos predicativos



42. Total de adverbios

LÉXICO

43. Proporción palabras/tipo de palabras

44. Longitud media de las palabras

Clases de palabras (o categorías morfosintácticas)

45. Conectores “conjuntos” (*consequently, furthermore, however...*)

46. Downtoners (*barely, nearly, slightly...*)

47. Hedges (*something like, sort of...*)

48. Amplificadores (*absolutely, extremely, perfectly...*)

49. Formas enfáticas (*a lot, for sure, really...*)

50. Partículas discursivas (*well, now, anyway...*)

51. Demostrativos

MODALES

52. Modales de posibilidad (*can, may, might, could*)

53. Modales de necesidad (*ought, should, must*)

54. Modales de predicción (*will, would, shall*)

CLASES DE VERBOS

55. Verbos públicos (*assert, declare, mention...*)

56. Verbos de privacidad (*assume, believe, doubt, know...*)

57. Verbos de persuasión (*command, insist, propose...*)

58. *Seem* y *appear*

FORMAS REDUCIDAS Y CONSTRUCCIONES DE USO COLOQUIAL

59. Contracciones

60. Eliminación de *that* (*I think [that] he went*)

61. Preposiciones finales (*the candidate that I was thinking of*)

62. Infinitivos divididos (*He wants to convincingly prove that...*)

63. Separación de auxiliares (*They were apparently shown to...*)

COORDINACIÓN

64. Coordinación de frases (nombre+nombre, adjetivo+adjetivo, verbo+verbo, adverbio+adverbio)

65. Coordinación oracional

NEGACIÓN

66. Negación sintética (*No answer is good enough for Jones*)

67. Negación analítica (*That's not likely*)

herramientas informáticas de análisis textual, 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 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 extratextual 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.

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 efectua-



do 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 primeros recibe el nombre de “frecuencia” del rasgo.

Mediante el proceso estadístico de análisis factorial, las frecuencias son agrupadas en los llamados “factores,” de acuerdo con su coaparición en el texto. Más exactamente, esta técnica permite repartir todos los rasgos lingüísticos examinados en un número determinado de diferentes factores según el resultado de la correlación de todos ellos. Basándose en la asunción de que determinados rasgos coaparecen con mucha frecuencia porque poseen las mismas funciones comunicativas, los factores se interpretan en términos de las funciones situacionales, sociales y cognitivas comúnmente asociadas a los distintos rasgos que pertenecen al mismo grupo.

Empleemos como ejemplo la taxonomía de rasgos contemplada en la Tabla 2. 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 que, entre otras, existe la tendencia de que a mayor cantidad de nombres, mayor cantidad de preposiciones y 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 de estos rasgos guarda una proporción inversa con, por ejemplo, el número de verbos de privacidad, pronombrs 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 otras dimensiones.

Tras el análisis factorial, cada rasgo aparece asociado con una cifra de coaparición comprendida entre 0 y 1, 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 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: 113s), “when the features with positive loadings occur together frequently in a text, the features with negative loadings are markedly less frequent in that text and vice versa.” En resumen, 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. Partiendo de 67 rasgos lingüísticos y 481 textos orales y escritos (aproximadamente 960.000

TABLA 2. LAS CINCO DIMENSIONES BÁSICAS DEL INGLÉS
Y LOS RASGOS LINGÜÍSTICOS ASOCIADOS A ELLAS (BIBER 1988)

DIMENSIÓN 1: PRODUCCIÓN INFORMATIVA VS PRODUCCIÓN PARTICIPATIVA		DIMENSIÓN 2: PRODUCCIÓN NARRATIVA VS PRODUCCIÓN NO NARRATIVA	
nombres	0,80	verbos en pasado	0,90
longitud de las palabras	0,58	pronombres 3ª persona	0,73
frases preposicionales	0,54	verbos con aspecto perfecto	0,48
proporción palabras/tipos de palabras	0,54	verbos públicos	0,43
adjetivos atributivos	0,47	negación sintética	0,40
<hr/>		cláusulas participiales de presente	0,39
verbos de privacidad	-0,96	<hr/>	
eliminación de <i>that</i>	-0,91	verbos en presente	-0,47
contracciones	-0,90	adjetivos atributivos	-0,41
verbos en presente	-0,86	<hr/>	
pronombres segunda persona	-0,86	DIMENSIÓN 3: PRODUCCIÓN ELABORADA VS PRODUCCIÓN DEPENDIENTE DEL CONTEXTO	
proforma <i>do</i>	-0,82	relativos <i>wh-</i> como objeto	0,63
negación analítica	-0,78	relat. precedidos por preposición	0,61
pronombres demostrativos	-0,76	relativos <i>wh-</i> como sujeto	0,45
partículas enfatizadoras	-0,74	coordinación de frase	0,36
pronombres primera persona	-0,71	nominalizaciones	0,36
pronombre <i>it</i>	-0,71	<hr/>	
<i>be</i> como verbo principal	-0,71	adverbios de tiempo	-0,60
subordinación causal	-0,66	adverbios de lugar	-0,49
partículas discursivas	-0,66	otros adverbios	-0,46
pronombres indefinidos	-0,62	<hr/>	
<i>hedges</i>	-0,58	DIMENSIÓN 4: PRODUCCIÓN ARGUMENTATIVA VS PRODUCCIÓN NO CLARAMENTE ARGUMENTATIVA	
amplificadores	-0,56	infinitivos	0,76
relativos con antecedentes oracionales	-0,55	modales de predicción	0,54
preguntas <i>wh-</i>	-0,52	verbos de persuasión	0,49
modales de posibilidad	-0,50	subordinación condicional	0,47
coordinación oracional	-0,48	modales de necesidad	0,46
cláusulas <i>wh-</i>	-0,47	separación de auxiliares	0,44
preposiciones finales	-0,43	modales de posibilidad	0,47
<hr/>		<hr/>	
DIMENSIÓN 5: ESTILO IMPERSONAL VS ESTILO PERSONAL		[No hay rasgos complementarios]	
conectores	0,48		
pasivas sin agente	0,43		
cláusulas adverbiales de participio de pasado	0,42		
pasivas con <i>by</i>	0,41		
participios de pasado postnominales	0,40		
otros subordinantes adverbiales	0,39		

[No hay rasgos complementarios]

palabras), Biber (1988) identifica cinco dimensiones de variación para el inglés contemporáneo (en realidad, siete dimensiones de las que dos no resultan ser significativas). La Tabla 2 presentaba los rasgos asociados con cada una de las dimensiones y la interpretación funcional que Biber propone para ellas (véase Biber 1988: capítulos 6 y 7).

Aunque la identificación de las dimensiones es importante en sí misma en tanto que identifica los parámetros básicos de variación en inglés, el uso fundamental de las dimensiones es el de analizar las características lingüísticas de un texto, un grupo de textos o un género. Para ello ha de calcularse el “resultado factorial” (*dimension* o *factorial score*) de cada texto, lo cual lograremos si seguimos los pasos que expondremos a continuación. En primer lugar, y una vez halladas las frecuencias, o porcentajes normalizados, de los rasgos de cada uno de los textos a analizar, éstas se estandarizan según el mismo criterio. Esto permitirá, por una parte, poder manejar homogéneamente los resultados globales obtenidos de 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 podría 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 período 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 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 a analizar. Con esta cifra final estamos ya en disposición de establecer todo tipo de comparaciones tanto sincrónicas como diacrónicas entre géneros.

Consideremos a modo de ejemplo la Tabla 3, que presenta la media de los resultados factoriales de 11 géneros tanto orales como escritos del inglés para la dimensión 1, “producción informativa *versus* producción participativa (*involved*).” Los géneros con cifras negativas elevadas (como la conversación) tienen abundantes verbos en presente, verbos de privacidad, pronombres de primera y segunda persona, contracciones, etc. (rasgos negativos de la dimensión 1), pero relativamente pocos nombres y frases preposicionales (rasgos positivos de la dimensión 1). Por el contrario, aquellos géneros que obtienen resultados factoriales positivos muestran una gran proliferación de nombres, frases preposicionales y adjetivos atributivos, pero escasez de verbos en presente, contracciones, etc. Las exigencias de la producción informativa determinan el orden relativo de los géneros que muestra la Tabla 3. Los géneros como la conversación son fundamentalmente interactivos. En ellos los participantes no tienen normalmente tiempo para producciones de gran concentración informativa, ni tampoco suele ser éste el objetivo fundamental, de ahí que ocupen el extremo negativo de la escala. Los discursos espontáneos o las entrevistas son intermedios en este sentido porque tienen un propósito relativamente

informativo, aunque los participantes están limitados por la producción inmediata. Finalmente, los géneros como la prosa académica y los documentos oficiales tienen como intención principal la de informar y están además producidos en circunstancias muy controladas en que cabe la edición.

TABLA 3. RESULTADOS FACTORIALES DE 11 GÉNEROS CON RESPECTO A LA DIMENSIÓN 1 (ADAPTADA DE BIBER 1993C: 336)

Conversaciones	-35.3
Correspondencia privada	-19.5
Discursos espontáneos	-18.2
Entrevistas	-17.1
Discursos preparados	-2.2
Ficción	0.8
Correspondencia profesional	3.9
Editoriales de prensa	10.0
Prosa académica	14.9
Reportaje de prensa	15.1
Documentos oficiales	18.1

El análisis de una única dimensión como el que acabamos de acometer no es sin embargo suficiente para comparar géneros, registros o estilos. Así, por ejemplo, si sólo tenemos en cuenta la dimensión 1, la ficción y la correspondencia profesional parecen ser bastante similares; sin embargo los resultados factoriales que estos géneros obtienen en las cuatro dimensiones restantes son diametralmente opuestos, como pone de manifiesto la Tabla 4. En otras palabras, para que el análisis multidimensional sea completo debemos estudiar el comportamiento de los textos o géneros que queremos comparar con respecto a todas las dimensiones (en el caso del inglés, las 5 —o 7— identificadas por Biber).

TABLA 4. RESULTADOS FACTORIALES DE 6 GÉNEROS CON RESPECTO A LAS 5 DIMENSIONES DE BIBER (1988) (ADAPTADA DE BIBER 1995A: 316S)

GÉNERO	DIM1	DIM2	DIM3	DIM4	DIM5
Conversaciones	-35.3	-0.6	-3.9	-0.3	-3.2
Correspondencia privada	-19.5	0.3	6.5	1.5	-2.8
Discursos espontáneos	-18.2	1.3	1.2	0.3	-2.6
Ficción	0.8	5.9	-1.4	0.9	-2.5
Correspondencia profesional	3.9	-2.2	6.5	3.5	0.4
Documentos oficiales	18.1	-2.9	7.3	-0.2	4.7

En una rápida ojeada a los resultados factoriales de los seis géneros que figuran en la Tabla 4 (conversación, correspondencia privada, discursos espontáneos, correspondencia profesional, ficción y documentos oficiales), observamos, por ejemplo, que la conversación obtiene los resultados factoriales más bajos en



cuatro de las cinco dimensiones: es notablemente “participativa” (dimensión 1), “dependiente del contexto” (dimensión 3), “no claramente argumentativa” (dimensión 4) y “personal” (dimensión 5). En cuanto a la dimensión 2, sin embargo, su posición es intermedia, lo cual refleja el uso de estrategias narrativas y no narrativas. Los documentos oficiales tienen las características contrarias en tanto que son “informativos” (dimensión 1), claramente “no narrativos,” (dimensión 2), “elaborados” (dimensión 3) y “abstractos” (dimensión 5), aunque tienen el mismo comportamiento que la conversación en lo que se refiere a la dimensión 4, al ser ambos géneros no argumentativos. La correspondencia profesional es relativamente similar a los documentos oficiales, aunque la primera difiere de éstos por su carácter argumentativo (dimensión 4) y su estilo relativamente no abstracto (dimensión 5). En resumen, las relaciones entre los géneros son muy complejas. Géneros radicalmente diferentes en unos aspectos pueden ser muy similares en otros.

3. APLICACIONES DEL MODELO MULTIDIMENSIONAL

3.1. ESTUDIOS SINCRÓNICOS

En esta sección presentaremos la metodología y los resultados de los trabajos inspirados en el modelo multidimensional de Biber más significativos en lo que respecta a la investigación sincrónica de la lengua. Así, prestaremos atención a experiencias relativas a variación textual diatópica (inglés británico *versus* inglés americano) y a la caracterización de materiales textuales específicos (manuales escolares, artículos científicos), discurso oral (escolares), variación multilingüística, etc.

El análisis multidimensional ha sido utilizado para caracterizar y comparar dominios discursivos de tipo diverso. Biber (1987), por ejemplo, compara nueve géneros escritos británicos y americanos. El estudio demuestra que los géneros americanos hacen uso de registros y estilos más coloquiales e interactivos que sus equivalentes británicos, aunque estos últimos son menos impersonales que los primeros.

Biber (1991) analiza las características lingüísticas de los libros de texto de enseñanza primaria. Este trabajo muestra que hay notables diferencias de estilo entre los textos de distintos niveles y disciplinas, y que en general estos textos difieren enormemente de los dirigidos a adultos.

Biber, Conrad y Reppen (1998: 157-170), basándose en Conrad (1996), investigan las diferencias de estilo entre los artículos de investigación de historia y de biología. Este trabajo se enmarca en un proyecto que intenta descubrir las características específicas de los distintos registros especializados del inglés con el fin de diseñar materiales y actividades adecuados para ayudar a los estudiantes a entender y producir estos registros de manera efectiva. En una segunda etapa, analizan además los rasgos lingüísticos característicos de las cuatro partes de las que consta un artículo científico, esto es, la introducción, la metodología, los resultados y la discusión y conclusiones (véase también Biber y Finegan 1994). Algunos de los resultados son sorprendentes, como el hecho de que las secciones de discusión hagan uso de los mismos rasgos característicos del estilo impersonal que las secciones de metodología.



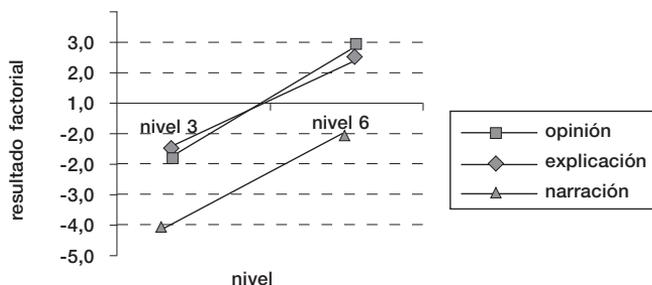


Gráfico 1. Resultados relativos a la dimensión 1 de las tres actividades (opinión, explicación y descripción), en los que se compara trabajos de tercer y sexto nivel educativo.

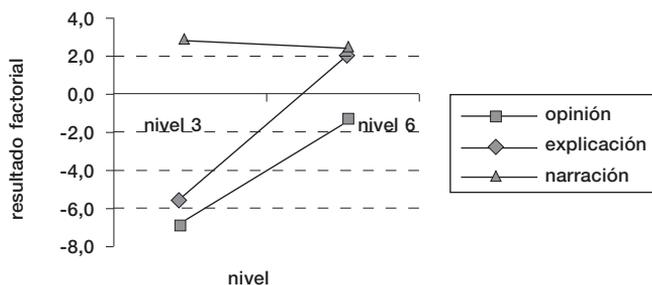


Gráfico 2. Resultados relativos a la dimensión 2 de las tres actividades (opinión, explicación y descripción), en los que se compara trabajos de tercer y sexto nivel educativo.

Los mismos autores (Biber, Conrad y Reppen 1998: 180-197), partiendo de Reppen (1995), analizan el discurso oral y escrito de los estudiantes de primaria, su evolución del tercer al sexto curso y las diferencias entre su lenguaje y el de los adultos. En este caso, Biber y sus colegas decidieron que no podían asumir como válidas para el lenguaje infantil las cinco dimensiones de variación que se venían utilizando desde Biber (1988) para el inglés producido por adultos, de manera que el primer paso en su estudio fue la selección de los rasgos lingüísticos potencialmente relevantes y la identificación de las dimensiones de los estudiantes. De nuevo son cinco los parámetros de variación en el inglés producido por y para niños.

Veamos, a modo de ejemplo, uno de los estudios incluidos en este trabajo, que abre las puertas del modelo multidimensional a los campos de la adquisición del lenguaje y del aprendizaje de segundas lenguas. Los Gráficos 1 y 2 plasman los resultados factoriales que tres tipos de redacciones de estudiantes de tercer y sexto curso obtienen con respecto a las dimensiones 1 (producción informativa editada *versus* producción informativa inmediata *on-line*) y 2 (producción narrativa léxicamente elaborada *versus* producción no narrativa).

Por lo que respecta a la dimensión 1, los estudiantes muestran en los tres casos un giro hacia estilos más informativos y cuidadosamente editados, lo cual se



traduce en resultados factoriales más elevados. Los cambios que se producen en la dimensión 2 son más complejos. En las redacciones en que opinan sobre la televisión y en las que nos explican su juego favorito aumenta claramente el uso de los recursos característicos de la narración elaborada, fundamentalmente porque en el sexto curso el léxico del alumnado es mucho más variado y especializado. Sin embargo, el estilo que los escolares utilizan en la redacción en que cuentan una historia a partir de viñetas apenas experimenta cambios de tercer a sexto grado. El discurso es relativamente elaborado ya en los alumnos más jóvenes.

Veamos ahora cómo se manifiestan los cambios que reflejan los Gráficos 1 y 2 en textos reales.

Texto 1. Explicación de un juego por un niño de tercer curso (Biber, Conrad y Reppen 1998: 190)

My favourite game is Namu. You write at least 20 names on a board. And then you write 5 names on paper. Then someone puts a checkmark on a name on the board. And if you have that name on your paper mark it with a checkmark. Then keep on doing it until someone gets all the names are marked off. And after someone has done it then they say Namu.

Texto 2. Explicación de un juego por un niño de sexto curso (Biber, Conrad y Reppen 1998: 190s)

I am going to explain how to play a game called Capture the Flag. You can have as many players as you want to. You put a flag *at either side of the court*. There is an *imaginary line* dividing the court. The people *on the 2 teams* are trying to get each other out. They do this by tagging each other. You can only tag someone if they are *on your side of the court*. Once you tag people, they are *in your jail*. The only way that they can get out is to have one *of their teammates* come and tag them. The object *of the game* is to get the *opposite team's* flag back *to your side* safely

La redacción del niño de tercero está mínimamente elaborada. La mayor parte de las oraciones comienzan con *and* o *then*. Algunos de los referentes son vagos, como en *doing it* y *has done it* en las últimas líneas, y el vocabulario se repite una y otra vez a lo largo del texto. La redacción del estudiante de sexto, sin embargo, muestra un número considerable de rasgos característicos de producciones editadas, como los adjetivos atributivos (*imaginary line*, *opposite team*) y las frases preposicionales (*at either side of the court*, *on the 2 teams*, *in your jail*, *to your side*). El léxico es además relativamente elaborado, con abundancia de palabras largas, como *imaginary*, *teammates*, *opposite*, y con variedad de vocabulario.

Los Gráficos 1 y 2 revelan además que los niños de tercer grado ya son conscientes de las diferencias de estilo entre las diversas producciones. Así, la narración de la historia es ya lingüísticamente diferente de los ensayos de opinión y de la explicación del juego con respecto a las dos dimensiones.

En todos los trabajos que hemos mencionado hasta ahora se describen y comparan géneros o registros predeterminados. Biber (1989, 1995a: capítulo 9), sin embargo, utiliza el modelo multidimensional para elaborar una tipología de tipos de textos basada exclusivamente en las características lingüísticas compartidas

por los miembros de cada uno de los tipos textuales. Las categorías a las que llega tras el análisis factorial no corresponden a las expectativas que a priori tenemos sobre la clasificación de textos en géneros. Por ejemplo, no hay una única categoría “conversación,” sino dos: la interacción interpersonal íntima y la interacción informativa. Igualmente, distingue tres tipos de textos expositivos: la exposición científica, que es extremadamente informativa, elaborada, técnica y abstracta; la exposición académica, que es similar a la científica excepto en su estilo, que es menos abstracto y técnico; y la exposición narrativa general, que combina formas narrativas con características expositivas e informativas.

En los últimos años el marco multidimensional se ha aplicado también al estudio comparado de los parámetros básicos de variación en distintas lenguas (Biber 1995a, Saiz 1998). En el volumen de Biber (1995a), en que se identifican las dimensiones del inglés, el somalí, el coreano y el tuvaluano, Biber concluye que, aunque algunas dimensiones, como la dimensión “estilo abstracto *versus* estilo no abstracto” del inglés o la dimensión “expresión de honra” del coreano, son específicas de una lengua, la mayoría tienen equivalentes en dos o más lenguas. Por ejemplo, las cuatro lenguas tienen en común dimensiones que diferencian los textos orales de los escritos y los narrativos de los no narrativos. El inglés y el somalí además comparten dimensiones asociadas con la argumentación y la persuasión. La comparación entre las dimensiones del inglés y el español arroja resultados similares (véase Saiz 1998). El problema que plantean, en nuestra opinión, muchos de estos estudios interlingüísticos es que comparan situaciones lingüísticas muy diferentes sin prestar demasiada atención a esta diversidad. Así, en Biber (1995a) no se tiene en cuenta, por ejemplo, que el corpus de tuvaluano representa el discurso de una pequeña comunidad agrícola de un atolón indonesio (Nukulaelae) de 310 habitantes y no la lengua estándar de una nación, como el corpus de inglés, que incluye textos de intelectuales y otras élites. Con toda seguridad, el atolón Nukulaelae es una comunidad lingüística radicalmente diferente de la formada por los individuos cuyo discurso reflejan los corpus *LOB* y *London-Lund*, lo cual pone en tela de juicio cualquier comparación entre ambas.

3.2. ESTUDIOS DIACRÓNICOS

El modelo multidimensional también ha sido utilizado para estudiar la evolución histórica de diversas producciones lingüísticas (e.g., Biber y Finegan 1989, 1992 y 1997; Biber 1995a: capítulo 8; Atkinson 1992, 1996; Meurman-Solin 1993: 276-306; Biber, Conrad y Reppen 1998: capítulo 8; González-Álvarez y Pérez-Guerra 1998, 2000²).

² Estos estudios relativos al periodo moderno han sido realizados gracias a la financiación facilitada por el Vicerrectorado de Investigación de la Universidade de Vigo a través del proyecto en áreas emergentes “Caracterización formal y discursiva de textos en el inglés moderno tardío.” Deseamos dejar constancia de nuestro agradecimiento por esta colaboración.



En el primer estudio de este tipo, Biber y Finegan (1989), se analiza la evolución histórica de tres géneros (el ensayo, la ficción y la correspondencia personal) desde 1650 hasta nuestros días con respecto a las dimensiones 1, 3 y 5. El estudio demuestra que en los tres géneros ha habido un cambio gradual hacia estilos más orales, pues los tres son en el siglo XX más interactivos (dimensión 1), más dependientes del contexto (dimensión 3) y menos abstractos (dimensión 5) que en el siglo XVII. En (1992), Biber y Finegan extienden este análisis a dos géneros más, en este caso orales, el diálogo de la novela y de la comedia, con idéntico resultado: el cambio hacia estilos más orales. Este cambio, argumentan, es el resultado de la creciente alfabetización de la población y por tanto de la necesidad que tiene el público de textos escritos más accesibles. La compilación del corpus *ARCHER* permite en (1997) ampliar el análisis a diez géneros. La inclusión de registros especializados lleva a Biber y Finegan a reconsiderar las conclusiones a las que habían llegado en los dos estudios previos. Efectivamente, las producciones más populares (cartas, diarios, comedias, sermones y ficción) evolucionan hacia estilos más interactivos y orales. Sin embargo, la prosa especializada muestra la tendencia opuesta. Los textos médicos, científicos y legales hacen cada vez más uso de rasgos lingüísticos característicos de producciones elaboradas e informativas. El motivo, según Biber y Finegan, es, de nuevo, el cambio que se ha producido en el público al que van dirigidos estos textos. Los lectores constituyen ahora un público especializado, que se ha preparado y entrenado para comprender estos textos de manera efectiva. Los textos explotan pues los recursos de la prosa elaborada e informativa hasta límites desconocidos hasta el momento.

A modo de ejemplo, procederemos a analizar con textos tomados del corpus *ARCHER R* la evolución de la comedia y la prosa médica desde el siglo XVII hasta nuestros días. Los Gráficos 3, 4 y 5 representan la evolución de estos dos géneros con respecto a las dimensiones informativa (1), elaborada (3) y abstracta (5), respectivamente. Los tres muestran un esquema evolutivo similar: en el siglo XVII, la comedia y la prosa médica son relativamente similares, al menos en lo que se refiere a los rasgos lingüísticos característicos de las dimensiones que nos ocupan, pero, a lo largo de los siglos siguientes, comienzan a diferenciarse cada vez más hasta alcanzar caracterizaciones totalmente opuestas en el siglo XX.

Veamos en textos reales el cambio que se produce en estas dos producciones. Los textos 3 y 4 ilustran las características lingüísticas de la comedia de los siglos XVII y XX, respectivamente:

Texto 3. Drama (1701TROT.D2)

Catharine Trotter. 1701. *Love at a Loss, or, Most Votes Carry It* (Biber y Finegan 1997: 262s)

GRANDFOY: You promise then to be dispos'd to me; I wear the Pledge of your Fidelity.

LESBIA: Won't it be a *bold Venture* to put my self in the power of a Man, I have injur'd?

GRANDFOY: Unless in your *unjust Suspicions* of me, you only have been injur'd, your Misfortune with Beaumine has a thousand Excuses on your part, as unhappy

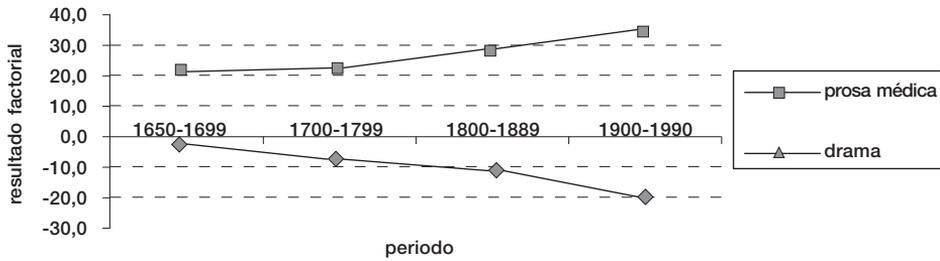


Gráfico 3. Producción interactiva versus producción informativa.

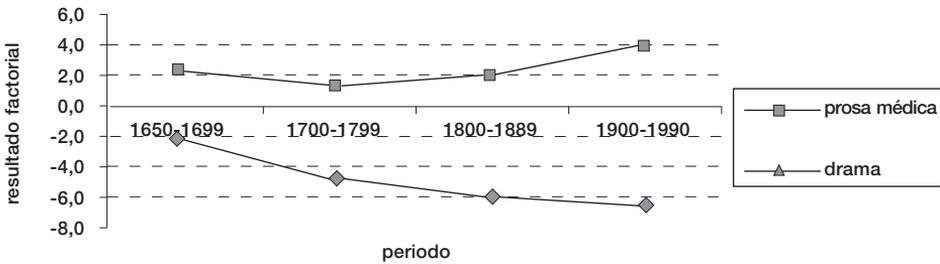


Gráfico 4. Producción dependiente del contexto versus producción elaborada.

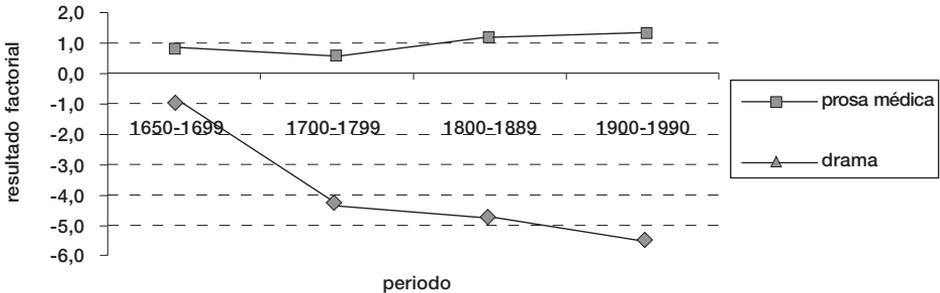


Gráfico 5. Producción personal versus producción impersonal.

as it makes me; O Lesbia, that I should ever think it reasonable to wish you anothers! to force a Man to deprive me of all I value!

LESBIA: 'Tis scarce reasonable indeed to value me now so much, as to care whose I am.

GRANDFOY: I have such an opinion of your Sincerity and Vertue, that even now, would you consent to be mine, I should receive you as the greatest *earthly Blessing*, but that you have refus'd me, unless Beaumine, by a *declar'd infidelity*, entirely release you.

Texto 4. Drama (1958)John.D9)

Denis Johnston. 1958. *The scythe and the sunset*. (Biber y Finegan 1997: 263)

MACCARTHY: Why pick on me, my dark Rosaleen?

ROISIN: I dunno. Mebbe it's because what you say is sort of ... original. D'ye know?

MACCARTHY: Original? My dear, I pride myself on never having made an *original remark* in my life.

ROISIN: That's a *queer thing* to be proud of.

MACCARTHY: The only people who make original remarks are my patients.

They're all completely original.

ROISIN: All the same I'd like to ask you something I've offn wunnered. Mebbe you'd tell me, an' you married all these years, an' ...

MACCARTHY: And had the nerve taken out. Go on.

ROISIN: Is there ... is there such a ting as true love at all, Doctor Mac? D'you know what I mean?

MACCARTHY: I know just what you mean —... unfortunately.

Estos dos pasajes difieren notablemente con respecto a la dimensión 1. Ambos hacen uso de pronombres de primera y segunda persona, contracciones y preguntas. Sin embargo, el segundo extracto usa un abanico más amplio de rasgos negativos o interactivos de la dimensión 1 que el primero, como, por ejemplo, verbos de privacidad (especialmente *know*), *hedges* (*maybe, sort of*), demostrativos (*that*) y cláusulas del tipo *wh* (*what you say, what I mean*). Por el contrario, en el primer texto es mayor el repertorio de recursos positivos o informativos de esta dimensión, como la frecuencia relativamente alta de nombres y adjetivos atributivos y la mayor ratio entre tipos de palabras y número de palabras.

Contrastemos ahora estos pasajes con textos médicos de los mismos períodos.

Texto 5. Prosa médica (1735)JAMI.M2)

James Jamieson. 1735. A large steatom passing with the Oesophagus. *Edinburgh Medical Journal*. (Biber y Finegan 1997: 263s)

A carpenter in this place, thirty six years of age, of a thin make of body, asked my advice about the middle of February 1732; his complaints were, pains at the heart, in the pit of the stomach, both sides of the thorax, and betwixt his shoulders; a great difficulty in swallowing, and frequent inclination to vomit. He told me it was about six years, since, upon a sudden grief, he was first seized with the pain, and frequent palpitations at his heart, and that the other symptoms came on gradually afterwards, but never kept him from his ordinary work till now. [...] He continued to turn gradually worse, till about the middle of March, when he was attacked by vomiting, which none of the different stomachicks nor opiates given him could restrain, but were thrown back again as soon as swallowed. His thirst became excessive and perpetual; but all his drink, though taken in very small quantities at a time, was vomited as quickly as his medicines, without increasing his pain; and a singultus soon came on, which continued till his death.

Texto 6. Prosa médica (1985POWE.M9)

D.A. Power *et al.* 1985. Acute renal failure: The tip of the iceberg? *Scottish Medical Journal*. (Biber y Finegan 1997: 264)

Approximately 30 patients per million population were dialysed annually for acute renal failure; 69 per cent of these patients (20.5 per million population per year) were dialysed for acute reversible intrinsic renal failure (ARIRF) and mortality in this group was 44 per cent. Patients with more severe disease at the time of presentation to the renal unit as defined by a clinical severity score, had significantly reduced survival rates. However, it was not possible to predict the outcome in individual cases; ten of 24 patients with clinical severity scores which indicated a poor prognosis survived the period of oliguria and were discharged from hospital. The fact that other renal units dialyse fewer patients per million population per year for ARIRF probably reflects a reluctance to refer patients whose general condition appears poor. As the overall mortality rate reported in this study does not differ significantly from rates reported previously from centres treating a smaller proportion of patients, such decisions may not be correct. It is well known that facilities in Britain for treating patients with end-stage renal disease are inadequate; it now appears likely that some patients who might benefit from acute dialysis are being denied treatment for a potentially reversible disease process.

El texto 5 hace uso de rasgos lingüísticos orales y no-orales de las tres dimensiones. Con respecto a la dimensión 1, por ejemplo, combina rasgos informativos, como los nombres, las frases preposicionales y los adjetivos atributivos con rasgos interactivos como los pronombres (primera y segunda personas, e *it*), la elisión de *that* o el uso de *hedges* y de coordinación clausal. El texto del siglo XX, sin embargo, muestra caracterizaciones extremas en las tres dimensiones. En cuanto a la dimensión 1, hace uso casi exclusivamente de rasgos lingüísticos del polo informativo de esta dimensión, como los nombres, los adjetivos atributivos, las frases preposicionales, el vocabulario especializado y la mayor longitud de las palabras. De las dimensiones 3 y 5 destaca el uso abundante de las cláusulas de relativo (rasgo del polo elaborado de la dimensión 3) y de construcciones pasivas (rasgo indicativo de estilo abstracto; dimensión 5).

En González-Álvarez y Pérez-Guerra (1998, 2000) extendemos el análisis de los 5 géneros incluidos en Biber y Finegan (1992) al inglés medio tardío y al período moderno temprano con el objetivo de determinar si estos géneros mostraban ya características distintivas en sus orígenes o si se diferenciaron con el tiempo. Estos estudios demuestran que hasta el siglo XVI todos los géneros excepto la prosa científica son muy similares con respecto a las dimensiones 3 (producción elaborada *versus* producción dependiente del contexto) y 5 (estilo impersonal *versus* estilo no-impersonal). Con respecto a la dimensión 1, sin embargo, ya en el siglo XV hay una marcada diferencia entre los géneros interactivos (comedia y cartas personales) y los no interactivos (ensayo, ciencia y ficción).

A diferencia de Biber y Finegan y de González-Álvarez y Pérez-Guerra, que comparan la evolución de diferentes géneros en un periodo determinado de tiempo, otros investigadores usan el modelo multidimensional para analizar el cambio lingüístico que experimenta un registro en particular. Así, Atkinson examina la



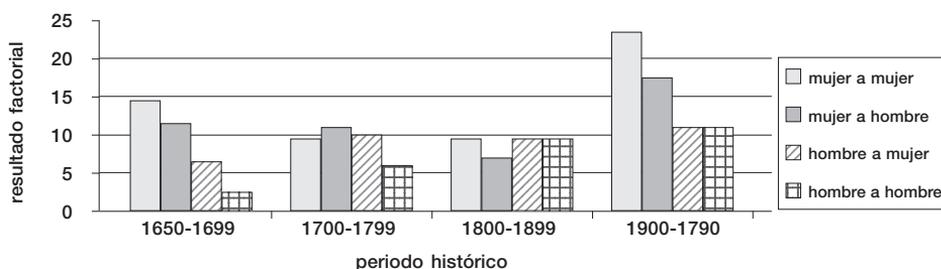


Gráfico 6. Evolución histórica en las cartas, por género, con respecto a la dimensión 1.

prosa médica del *Edinburgh Medical Journal* de 1735 a 1985 en (1992) y el lenguaje de la revista científica *Philosophical Transactions of the Royal Society of London* de 1675 a 1975 en (1996). Destaquemos el hecho de que Atkinson estudia el estilo y rasgos lingüísticos de cada una de las partes de las que consta un artículo científico como si se tratara de subgéneros independientes. En ambos casos concluye que la evolución estilística de la prosa médica ha sido un proceso gradual y continuo paralelo a los cambios que se han producido en la metodología investigadora a lo largo de estos siglos.

El modelo multidimensional puede utilizarse también para comparar la evolución histórica de diferentes variedades dialectales, tanto geográficas como demográficas (el lenguaje de las distintas clases sociales, sexos, edades, etc.). El único estudio de este tipo que conocemos es el de Biber, Conrad y Reppen (1998: 215-222), en el que se investiga la evolución que ha experimentado el discurso de hombres y mujeres en la correspondencia privada de los cuatro últimos siglos con respecto a la dimensión 1 (producción informativa *versus* producción interactiva). El Gráfico 6 muestra los resultados factoriales de cada una de las categorías que analizan en los cuatro siglos.

Todas las cartas obtienen resultados factoriales negativos, lo cual indica que el conjunto muestra características de producciones interactivas, aunque de manera más acentuada en el siglo XX. Se observan, sin embargo, diferencias importantes entre las cuatro categorías. Por ejemplo, en los siglos XVII y XX las cartas escritas por mujeres son más interactivas que aquellas escritas por hombres, y las cartas dirigidas a mujeres son también más interactivas que las dirigidas a hombres, cualquiera que sea el sexo del autor. En los siglos XVIII y XIX, sin embargo, las diferencias entre los sexos no son en general significativas. Textos como 7 y 8 ilustran las diferencias de estilo entre las distintas categorías en el siglo XVII.

Texto 7. Carta del siglo XVII de mujer a mujer (Biber et al 1998: 220)

Dear Sister Pue

Distance or Place, nor *Length of Time*, cannot lessen my *Natural*, or rather *Supernatural Affection* to you, for certainly my Love for you is more than a Sisters Love, nay, such a Love, as when I Lived with you, it could not choose but be somewhat Troublesome, by reason my Love was Accompanied with such Fears, as it would neither let you Rest, Pray, nor Eat in Quiet, for though it was a *Watchful Love*, yet

it was a *Fearful* Love, for I remember I have oftentimes Waked you out of your Sleep, when you did Sleep Quietly, with *Soft Breathing*, fearing you had been Dead, and oftener have I laid my Face over your Mouth, to feel if you Breath'd ...

Texto 8. Carta del siglo XVII de hombre a hombre (Biber et al 1998: 221)

Most worthy Sr,

I returne you *humble thancks for your courteous letter and the good newes of the hopefull recoverie of Mr Dugdale ...* It can very hardly fall into my apprehension how I can afford any addition unto your *worthy endeavors*, notwithstanding I have enclosed a list of such *tracts of that subject* wch I have by mee, most whereof I received from Dr Arthur Dee, *my familiar freind, sonne unto old Dr Dee* the mathematician; hee lived many yeares and dyed in Norwich, from whom I have heard *many accounts agreeable unto those* wch you have sett downe in your annotations, concerning his father and Kelley; hee was a *persevering student* in *Hermeticall philosophy* and had noe *small encouragement*, having seen projection made...

En el primero, escrito por Margaret Cavendish a su hermana, destaca la abundancia de pronombres de primera y segunda persona, del pronombre *it* y de cláusulas del tipo *wh*, y la preferencia por construcciones verbales más que por frases nominales elaboradas. El segundo, tomado de una carta de Sir Thomas Brown a Mr Elias Ashmole, hace un uso mayor del polo informativo de la dimensión 1, con rasgos tales como la mayor longitud de las palabras, la variedad de vocabulario y el uso de nombres modificados por adjetivos atributivos o frases preposicionales (por ejemplo, *Humble thancks, courteous letter, hopefull recoverie, worthy endeavors, Hermeticall philosophy, the good newes of the hopefull recoverie of Mr Dugdale, tracts of that subject*). Un análisis completo de las diferencias de género requeriría un estudio del lenguaje de los dos sexos en un repertorio amplio de registros con respecto a todas las dimensiones.

Otra de las aplicaciones que ha tenido el aparato multidimensional ha sido el análisis de los estilos de los autores. Los trabajos de este tipo parten de la asunción de que la manera más efectiva de analizar el estilo de un autor determinado es la comparación con los textos de otros autores, géneros y periodos. Así, por ejemplo, el Gráfico 7 muestra el rango de resultados factoriales para la dimensión 5 (estilo abstracto *versus* estilo no abstracto) de 51 ensayos y 33 textos de ficción de los siglos XVII a XX analizados en Biber y Finegan (1989).

El Gráfico muestra el giro gradual hacia estilos menos impersonales en los periodos más recientes de la lengua, y la gran variedad de estilos que caracteriza al siglo XVIII. En este contexto se reflejan los resultados factoriales de 8 textos de dos autores del siglo XVIII: dos ensayos y dos narraciones de ficción de Defoe y Johnson. Por lo que respecta a esta dimensión, la diferencia de estilo entre Defoe y Johnson es clara, sobre todo en el caso de los ensayos. Los textos de Johnson tienden a ser marcadamente impersonales, mientras que los de Defoe son relativamente no impersonales. Las diferencias son también notables en las demás dimensiones (véase Biber y Finegan 1994, en que se comparan textos de Defoe, Johnson, Swift y Addison).

A excepción de Biber (1995a: 311-313), en que se compara la evolución de varios géneros en inglés y somalí, no existen estudios diacrónicos interlingüísticos y



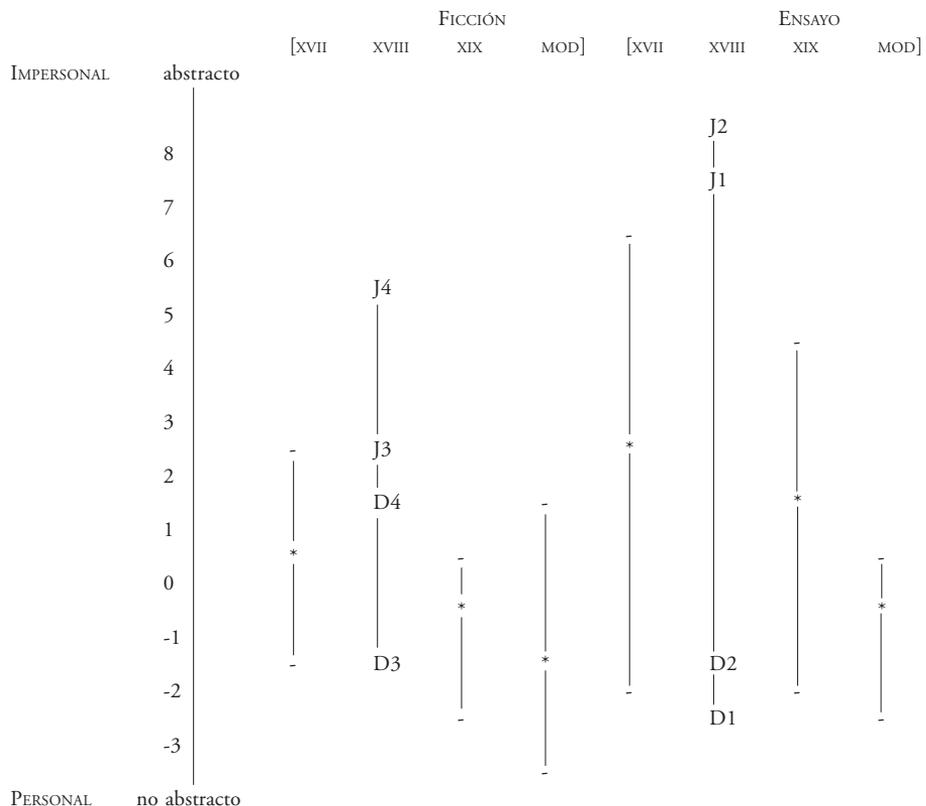


Gráfico 7. Caracterización de textos de Defoe y de Johnson relativa a la dimensión 5 (“estilo abstracto versus no abstracto”).

[Textos - D1: Defoe (1713) “Some thoughts on... commerce”; D2: Defoe (1698) “Essays on projects: Of academies”; D3: Defoe (1719) “Life of Robinson Crusoe” (capítulo 1); D4: Defoe (1722) *Moll Flanders* (capítulo 1); J1: Johnson (1750) “On Fiction”; J2: Johnson (1755) “Preface to *Dictionary of the English Language*”; J3: Johnson (1759) “The History of Rasselas: Prince of Abyssinia” (pasaje narrativo); J4: Johnson (1759) “The History of Rasselas: Prince of Abyssinia” (pasaje descriptivo)]

por tanto no es posible determinar si los cambios observados en los distintos géneros, registros y dialectos del inglés son compartidos por otras lenguas o si, por el contrario, cada lengua sigue sus propias pautas de evolución. El único trabajo de este tipo que conocemos, el ya mencionado Biber (1995a), no resulta muy útil en este sentido, pues aunque Biber defiende un modelo de evolución muy similar para el inglés y el somalí, que él relaciona con un cambio paralelo en la relación escritor-lector, los corpus que utiliza no nos parecen comparables. El del inglés abarca 350 años, mientras que el de somalí solamente comprende los 26 años que transcurren entre 1973, fecha a la que pertenece el primer documento escrito en esta lengua, y 1989.

Para concluir, debemos señalar que, a pesar de que todos los trabajos reseñados en esta sección han demostrado la utilidad de la propuesta multidimensional para el análisis histórico de géneros, registros, estilos y dialectos geográficos y sociales, creemos, como explicamos en Pérez-Guerra y González Álvarez (1999) y González-Álvarez y Pérez-Guerra (2000), que es necesario revisar el modelo al menos en dos aspectos fundamentales, (i) la validez para períodos más tempranos de la lengua de los rasgos lingüísticos que representan a cada dimensión en inglés contemporáneo, y (ii), como consecuencia de lo anterior, la posibilidad de que sean otras, y no las 5 establecidas desde Biber (1988), las dimensiones relevantes para el análisis de registros, estilos o dialectos de períodos distintos al inglés de nuestros días.

4. OBJECIONES AL MODELO

Mientras que en las secciones anteriores exploramos los principios que rigen el modelo multidimensional y sus múltiples aplicaciones para el análisis tanto sincrónico como diacrónico de variación textual y dialectal, en esta sección reflexionaremos sobre cuestiones específicas que, a nuestro entender, requieren reformulaciones y correcciones que, lejos de invalidar la teoría, pueden conducir a mejoras sustanciales en determinados aspectos del proceso.

Como hemos visto en las secciones anteriores, el análisis multidimensional consta de los siguientes estadios de investigación: (i) la creación del corpus de textos, (ii) la determinación y el recuento de los rasgos lingüísticos en nuestro corpus, y (iii) la identificación de factores, la interpretación funcional de dichos factores y la consecución de los resultados factoriales de cada texto y género. En los párrafos que siguen nos detendremos en cada uno de ellos.

4.1. CREACIÓN DEL CORPUS DE TEXTOS

Uno de los pecados del modelo multidimensional es la circularidad del proceso. Biber parte de conceptos prefijados que marcan todo el proceso multidimensional, como los conceptos de *register* y *genre*, 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, obtenemos un valor numérico general de toda la categoría discursiva, y comparamos ese valor con el obtenido por análisis similares en otro grupo de textos que ya hemos inicialmente caracterizado dentro de otra categoría textual (ficción narrativa, por ejemplo), los resultados posibles 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 las sospechas que a lo largo de la historia de la estilística y el análisis del discurso ya habían sido reflejadas en estudios previos. Sin embargo, las conclusiones finales pueden pecar de determinismo analítico, pues están condicio-



nadas por nuestro acierto o error en la determinación previa de las categorías textuales. En la clasificación de los textos empleados 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 texto académico entendemos estudios lingüísticos, 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 lingüístico. De hecho, Taavitsainen (1993: 190-91, 194) demuestra mediante el análisis factorial de varios géneros del inglés medio que las fronteras entre los géneros son imprecisas y que muchos géneros 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 es que la fiabilidad del modelo multidimensional a la hora de comparar géneros o registros está claramente condicionada por la homogeneidad de los textos agrupados en cada categoría. Dicho de otro modo, cuanto mayor sea la variedad de los textos incluidos en cada categoría, más relativos serán los resultados (sobre la clasificación en *text-types* que propone Biber, véase también Besnier 1998: 127).

Una segunda cuestión relativa a la composición del corpus es la extensión que ha de tener. Aunque Biber y Finegan (1997: 254) han dejado patente la necesidad de ampliar el número de textos manejados, los trabajos publicados por Biber, Biber y Finegan y Atkinson (1994), entre otros, han sido realizados sobre muestras, en nuestra opinión, reducidas, si las comparamos con estudios contemporáneos de lingüística de corpus en que se emplean fuentes textuales de millones de palabras (por ejemplo, Gómez-Guinovart y Pérez-Guerra 2000 analizan casi 3 millones de palabras). En la Tabla 5 mostramos a modo de ejemplo el número de palabras manejadas en algunos de los trabajos más representativos del marco multidimensional:

TABLA 5. NÚMERO DE PALABRAS EN TRABAJOS RELEVANTES
DEL MODELO MULTIDIMENSIONAL

ESTUDIO	EXTENSIÓN DEL CORPUS
Biber (1988)	aprox. 960.000 palabras
Biber y Finegan (1989)	aprox. 120.000 palabras
Atkinson (1992)	186.553 palabras
Kim (1990), Kim y Biber (1995)	aprox. 135.000 palabras
Besnier (1988), Biber (1995a)	aprox. 152.000 palabras
Conrad (1996), Biber, Conrad y Reppen (1998: 159)	aprox. 96.000 palabras

Es más, Biber (1995b: 364) ha defendido la irrelevancia de la cuestión del número de palabras y ha llegado a aseverar que el modelo multidimensional no requiere grandes cantidades de textos sino textos representativos (ver Biber 1990: 263-269). Dado que la representatividad de un texto en un corpus determinado es una asignatura todavía pendiente en lingüística de corpus, 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. Sobre la fiabilidad de los corpus utilizados por Biber, véase Altenberg (1989: 170), Baayen (1997: 66) o Besnier (1998: 127).

4.2. DETERMINACIÓN Y 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 fe ciega en estudios no estadísticos 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 de su posterior agrupamiento en factores tras el análisis factorial.

Por lo que respecta a la primera de las deficiencias, ya hemos reparado antes en el hecho de que todo el proceso parte de unos rasgos lingüísticos que más o menos intuitivamente han sido reconocidos como reveladores de hechos estilístico-funcionales en fuentes académicas relevantes (véase Watson 1994: 243ss). De esos rasgos, sólo aquellos susceptibles de recuento automático son incluidos en el análisis. El riesgo en el que el investigador puede caer si su listado de rasgos es incompleto es enorme, bien porque algún rasgo no haya sido destacado en la literatura o bien porque un determinado rasgo no pueda ser rastreado automáticamente. Pensemos en la incidencia numérica y, posteriormente, estadística de la ausencia del rasgo más determinante de, por ejemplo, la dimensión 1. Dicha carencia desencadenaría resultados estandarizados más bajos, y, de ahí, resultados factoriales reducidos. Lo que queremos decir con esto es que un análisis multidimensional no puede estar exclusivamente sostenido en una selección no estadística y en muchos casos informal de rasgos lingüísticos supeditada a mecanismos computacionales de recuento. Así, Biber, Conrad y Reppen (1998) y Gómez-Guinovart y Pérez-Guerra (2000) parten de su propio inventario de rasgos.

El segundo gran obstáculo es el proceso mismo de recuento de rasgos. Como ya hemos mencionado anteriormente, aunque el cómputo de algunos rasgos es fácilmente realizable mediante un simple editor de textos o alguna aplicación específica de análisis textual, el de otros ha de ser determinado mediante técnicas informáticas a la medida, especialmente en aquellos casos en que los corpus de textos no contienen ningún tipo de anotación gramatical añadida. Con el fin de automatizar el recuento en el mayor grado posible, Biber (1988) desarrolló su propio anotador y sistema de anotación que él mismo describe como bastante eficiente, aunque precisó en algunas ocasiones de desambiguación y corroboración manuales. Ni qué decir tiene que todo el proceso estadístico posterior confía en los resultados obtenidos tras la aplicación de los algoritmos desarrollados por Biber (la descripción de dichas fórmulas pueden encontrarse en Biber 1988: apéndice II). Sin embargo, el análisis de los datos que hemos realizado para nuestros estudios (González-Álvarez y Pérez-Guerra 1998, 2000) nos lleva a reconocer que algunos de los algoritmos no están libres de problemas. Por ejemplo, el analizador de Biber añade



erróneamente a la clase de la proforma *do* todas las apariciones de *do* como verbo principal (véase Altenberg 1989: 172). Del mismo modo, la fórmula propuesta por Biber (1988) para el recuento de oraciones *wh*- no identifica todas las apariciones de este 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, de privacidad y de persuasión. En consecuencia, aquellas apariciones de oraciones *wh* del tipo *What he said was a lie* son obviadas simplemente porque, en palabras de Biber (1988: 231), “they could not be identified reliably by automatic analyses.” Idéntico motivo se aduce para excluir a las oraciones de relativo con relativo \emptyset del recuento total de oraciones de relativo.

Una tercera deficiencia, que nos hace, como poco, dudar de la validez de ciertos resultados de Biber y Finegan, es la falta de actualización de los algoritmos originales de Biber (1988) en estudios posteriores de naturaleza diacrónica. A pesar de que la idoneidad para estudios diacrónicos de las fórmulas del trabajo de 1988 no es cuestionada en absoluto por los autores (véase Biber, Conrad y Reppen 1998: 181), hemos encontrado serios defectos en, por ejemplo, la confección de las 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, utterly* y *very*, en inglés medio tardío echamos en falta amplificadores como *all(e), heartely, sore* o *fervently*, lo cual puede desembocar en resultados incorrectos. A la falta de adaptación de los algoritmos ha de sumarse el descuido, que ya apuntamos anteriormente, de utilizar para el análisis de textos de periodos más tempranos de la lengua aquellos rasgos de Biber (1988) supuestamente válidos para el inglés contemporáneo. En otras palabras, los estudios diacrónicos llevados a cabo hasta el momento parten de la asunción de que aquellos rasgos que efectivamente conducían a determinadas generalizaciones funcionales en el inglés de nuestros días son los mismos que resultaban operativos en el inglés medieval o en el inglés moderno temprano.

A modo de conclusión, con los desaciertos que hemos puesto de relieve en los párrafos anteriores, hemos querido poner en duda la exactitud de los resultados numéricos crudos de Biber y Finegan. Aunque somos conscientes de que esos datos son sometidos a estrictos procesos de estandarización, lo cual aminora la relevancia de los posibles errores, queremos dejar constancia de la gravedad metodológica que, en nuestra opinión, supone supeditar la búsqueda de fiabilidad más estricta a la inmediatez instrumental.

4.3. SOBRE LA IDENTIFICACIÓN DE FACTORES, LA INTERPRETACIÓN FUNCIONAL DE DICHO FACTORES Y LA CONSECUENCIA DE LOS RESULTADOS FACTORIALES DE CADA TEXTO Y GÉNERO

En la sección 2 describimos el proceso propuesto por Biber para la identificación de los factores. En pocas palabras, se recontaban los rasgos lingüísticos relevantes para la caracterización funcional de un texto, se normalizaban sobre una

base común de 1.000 palabras y mediante el proceso estadístico de análisis factorial se agrupaban en los llamados factores de acuerdo con su co-aparición en el texto. Aquí surge el primer problema, pues el número de factores depende de la decisión del investigador.

Supongamos que aceptamos como válidos para el inglés los 5 factores (o 7) identificados por Biber (1988). El paso siguiente es la interpretación funcional de cada dimensión. 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 la Tabla 2 (producción informativa *versus* producción participativa, 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 añade “subjective interpretation and luck.”

Si aceptamos las interpretaciones funcionales que Biber asigna a cada uno de los factores, nos encontramos con un segundo elemento problemático: las dimensiones no son funcionalmente tan uniformes y transparentes como Biber quiere hacernos creer. La mayoría parecen reflejar una mezcla de parámetros funcionales y algunas incluyen rasgos que son difíciles de agrupar con los demás bajo una etiqueta funcional única. Por ejemplo, la dimensión 3 (producción elaborada *versus* producción dependiente del contexto) incluye entre los rasgos característicos del estilo elaborado dos construcciones —la coordinación de sintagma y las nominalizaciones— que también parecen indicar integración de información, un fenómeno que normalmente se asocia con el polo informativo de la dimensión 1. En el polo “dependiente del contexto” nos encontramos con rasgos como los adverbios de tiempo, lugar y “otros,” cuyas funciones son demasiado variadas como para definirlos con la etiqueta “dependientes del contexto.” De hecho, si fueran recursos puramente dependientes del contexto deberíamos encontrarlos entre los rasgos orales de la dimensión 1, en la que no aparecen. La dimensión 4, que comprende los infinitivos con *to*, los auxiliares modales, los verbos de persuasión y las cláusulas condicionales, es interpretada por Biber como “producción abiertamente argumentativa.” Esta interpretación es válida si estamos dispuestos a aceptar que todos los modales tienen esta función y que los infinitivos con *to* funcionan principalmente como complementos de adjetivos y verbos de actitud (lo mismo podría decirse de muchas completivas con *that*, y sin embargo están incluidas en la dimensión 6).

El tramo final del modelo multidimensional consiste en el cálculo, primero, del resultado factorial de cada texto por factor o dimensión, y posteriormente, de la media del género con respecto a cada una de las dimensiones. Aunque hemos detectado algunas incorrecciones significativas en el entramado estadístico, como, por ejemplo, que el criterio para la discriminación de rasgos sea únicamente la carga factorial de dichos rasgos y no algún concepto estadístico que integre además su frecuencia, o el hecho de que todos los rasgos que no hayan sido “expulsados” del análisis tengan la misma consideración dentro de los cálculos estadísticos posterior-



res independientemente de su carga factorial, no vamos a detenernos en los detalles más técnicos del proceso (véase, a este respecto, Pérez-Guerra y González-Álvarez 1998).

5. COMENTARIOS FINALES

En este trabajo hemos descrito las fases fundamentales del análisis multidimensional de Biber y sus múltiples aplicaciones tanto para los estudios de variación sincrónicos como diacrónicos, con el objetivo de acercarlo a una audiencia desconocedora de la compleja maquinaria estadística. En este sentido, creemos, con Kilgarriff (1996), que el hecho de que hasta ahora el modelo multidimensional haya sido utilizado casi exclusivamente por Biber y sus colaboradores se debe a que la metodología es técnicamente compleja y laboriosa. Al mismo tiempo que hemos manifestado 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, hemos señalado algunas incorrecciones de índole tanto teórica como metodológica en el modelo de Biber.



OBRAS CITADAS

- ALTENBERG, Bengt. Reseña de *Variation across Speech and Writing*. *Studia Lingüística* 43/2 (1989): 167-174.
- ATKINSON, Dwight. "The Evolution of Medical Research Writing from 1735 to 1985: The Case of the Edinburgh Medical Journal." *Applied Linguistics* 13.4 (1992): 337-374.
- . "The Philosophical Transactions of the Royal Society of London: A Sociohistorical Discourse Analysis." *Language in Society* 25 (1996): 333-371.
- BAAAYEN, R.H. Reseña de *Dimensions of Register Variation: A Cross-linguistic Comparison*. *Literary and Linguistic Computing* 12.1 (1997): 65-67.
- BESNIER, Niko. Reseña de *Dimensions of Register Variation: A Cross-linguistic Comparison*. *Language in Society* 27.1 (1998): 126-129.
- BIBER, Douglas. "Investigating Macroscopic Textual Variation Through Multi-feature/multi-dimensional Analyses." *Linguistics* 23 (1985): 337-360.
- . "Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings." *Language* 62 (1986): 384-414.
- . "A Textual Comparison of British and American Writing." *American Speech* 62 (1987): 99-119.
- . *Variation across Speech and Writing*. Cambridge: Cambridge UP, 1988.
- . "A Typology of English Texts." *Linguistics* 27 (1989): 3-43.
- . "Methodological Issues Regarding Corpus-based Analyses of Linguistic Variation." *Literary and Linguistic Computing* 5 (1990): 257-69.
- . "Oral and Literate Characteristics of Selected Primary School Reading Material." *Text* 11 (1991): 73-96.
- . "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8 (1993a): 1-15.
- . "Using Register-diversified Corpora for General Language Studies." *Computational Linguistics* 19 (1993b): 219-241.
- . "The Multi-dimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings." *Computers and the Humanities* 26 (1993c): 331-45.
- . "An Analytic Framework for Register Studies." *Sociolinguistic Perspectives on Register*. Ed. Douglas Biber & Edward Finegan. New York: Oxford UP, 1994. 31-56.
- . *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge UP, 1995a.



- “On the Role of Computational, Statistical, and Interpretive Techniques in a Multi-dimensional Analysis of Register Variation: A Reply to Watson.” *Text* 15.3 (1995b): 341-370.
- BIBER, Douglas & Edward FINEGAN. “Drift and the Evolution of English Style: A History of Three Genres.” *Language* 65.3 (1989): 487-517.
- “On the Exploitation of Computerized Corpora in Variation Studies.” *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. Ed. Karin Aijmer & Bengt Altenberg. London: Longman, 1991. 204-220.
- “The Linguistic Evolution of Five Written and Speech-based English Genres from the Seventeenth to the Twentieth Centuries.” *History of Englishes: New Methods and Interpretations in Historical Linguistics*. Ed. Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen. Berlin: Mouton, 1992. 688-704.
- “Multidimensional Analyses of Author’s Style: Some Case Studies from the Eighteenth Century.” *Research in Humanities Computing*. Volume III. Ed. Susan Hockey & Nancy Ide. Oxford: Oxford UP, 1994.
- “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*. Ed. Terttu Nevalainen & Leena Kahlas-Tarkka. Helsinki: Société Néophilologique, 1997. 253-275.
- BIBER, Douglas, Edward FINEGAN & Dwight ATKINSON. “ARCHER and Its Challenges: Compiling and Exploring a Representative Corpus of Historical English Registers.” *Creating and Using English Language Corpora*. Ed. Udo Fries, Gunnel Tottie & Peter Schneider. Amsterdam: Rodopi, 1994. 1-14.
- BIBER, Douglas, Susan CONRAD & Randi REPPEN. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge UP, 1998.
- CONRAD, Susan. *Academic Discourse in Two Disciplines: Professional Writing and Student Development in Biology and History*. Diss. Northern Arizona U, 1996.
- GÓMEZ-GUINOVART, Javier & Javier PÉREZ-GUERRA. “A Multidimensional Corpus-based Analysis of English Spoken and Written-to-be-spoken Discourse.” *Cuadernos de Filología Inglesa* 9.1 (2000): 39-70.
- GONZÁLEZ-ÁLVAREZ, Dolores & Javier PÉREZ-GUERRA. “Texting the Written Evidence: On Register Analysis in Late Middle English and Early Modern English.” *Text* 18.3 (1998): 321-348.
- “A Corpus-based Approach to the Multidimensional Variation of Four Late Middle English Genres.” *Proceedings of the 10th International Conference of SELIM*. Ed. Ana Hornero & Pilar Navarro. Zaragoza: Institución Fernando el Católico, 2000. 69-182.
- KILGARRIFF, Adam. Reseña de *Dimensions of Register Variation: A Cross-linguistic Comparison*. <http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/biber.asc> (1996).
- KIM, Yong-Jin. *Register Variation in Korean: A Corpus-based Study*. Diss. U of Southern California, 1990.
- KIM, Yong-Jin & Douglas BIBER. “A Corpus-based Analysis of Register Variation in Korean.” *Socio-linguistic Perspectives on Register*. Ed. Douglas Biber & Edward Finegan. Oxford: Oxford UP, 1994. 157-181.
- MEURMAN-SOLIN, Anneli. *Variation and Change in Early Scottish Prose*. Helsinki: Suomalainen Tiedekatemia, 1993.
- PÉREZ-GUERRA, Javier & Dolores GONZÁLEZ ÁLVAREZ. “Hacia una nueva aritmética del texto: unas reflexiones sobre el modelo multidimensional de análisis textual.” *RESLA* 13 (1999): 67-85.



- REPPEN, Randi. "A Multi-dimensional Comparison of Spoken and Written Registers Produced by and for Students." *Organisation in Discourse: Proceedings from the Turku Conference*. Ed. Brita Wärvik, Sanna-Kaisa Tanskanen & Risto Hiltunen. Turku: U of Turku, 1995. 477-486.
- SAIZ, Marta. "A Multivariate Analysis Using English and Spanish Corpora." *SEPLN 23* (1998): 69-75.
- TAAVITSAINEN, Irma. "Genre/subgenre Styles in Late Middle English?" *Early English in the Computer Age: Explorations through the Helsinki Corpus*. Ed. Matti Rissanen, Merja Kytö and Minna Palander-Collin. Berlin: Mouton de Gruyter, 1993. 171-200.
- WATSON, G.J. "A Multidimensional Analysis of Style in Mudrooroo Nyoongah's Prose Works." *Text* 14.2 (1994): 239-285.

