GIOVANNI PARODI
Pontificia Universidad Católica de Valparaíso
Chile
gparodi@ucv.cl

Variación en el corpus electrónico pucv-2003: análisis multidimensional

Introducción

Son relativamente escasos o nulos los estudios acerca del español que describen las características lingüísticas del discurso especializado didáctico escrito en lengua española. La mayoría de estos trabajos focaliza el llamado discurso divulgativo de especialidad (Ciapuscio 2000, 2003, Ciapuscio y Kuguel 2002, Calsamiglia 2000, Cassany López y Martí 2000, Cademártori, 2003) o aborda los marcadores discursivos (Portolés 1998, Martín Zorraquino y Portolés 1999, Montolío 2001); otros, se concentran en algunos pocos rasgos lingüísticos (Ciapusccio 1992, Harvey 2002); los menos, estudian la terminología especializada (Cabré 2000, 2002, Lorente 2002). Resulta aún mucho más difícil encontrar investigaciones del español en que se describa rasgos lingüísticos de registros especializados a partir de textos que se entregan como lectura obligatoria a alumnos de liceos técnico-profesionales. De hecho, los estudios que comparan registros escritos en diversas áreas del conocimiento técnico-científico especializado son casi inexistentes. Como resultado, los textos de registros escolares de nivel especializado han recibido sorprendentemente poca atención (Menéndez 1999, Parodi y Gramajo 2003, Parodi, 2004) o han sido virtualmente ignorados. Investigaciones en esta línea y que sigan principios de lingüística de corpus y utilicen apoyo computacional y estadístico también resultan singularmente poco frecuentes.

Ahora bien, en el presente artículo se indaga la variabilidad lingüística y funcional al interior del corpus PUCV-2003 (90 textos con un total de 1.466.744 palabras). Con el objetivo final de describir acuciosamente un subcorpus técnico-científico (PUCV-CTC), se realiza un análisis multirasgos y multidimensiones con apoyo de técnicas estadísticas multivariadas (Análisis Factorial de Componentes Principales). A partir de un total de 65 rasgos lingüísticos sobresalientes, explorados en el total del corpus, se busca determinar –en base a la alta co-ocurrencia de patrones de rasgos significativos– las «dimensiones» subyacentes a dichos rasgos.

De modo sintético, los objetivos de esta investigación son: (a) identificar los patrones lingüísticos relevantes y en co-ocurrencia en el total del corpus PUCV-2003, desde una perspectiva empírica cuantitativa, (b) comparar sistemáticamente tres corpora: uno especializado técnico-científico escrito, otro no especializado literario escrito y un tercero oral no especializado de entrevistas; todo ello a partir de dimensiones funcionales, determinadas a base de aquellos patrones lingüísticos de co-ocurrencia, y (c) identificar similitudes y diferencias –en términos de dimensionales– entre tres áreas de especialización del subcorpus técnico-científico.

1. Marco de referencia

Biber (1994) defiende dos ideas significativas respecto a la representatividad de los estudios lingüísticos basados en corpus que me parecen relevantes: 1) todo corpus debe tener una amplitud importante, y 2) un corpus debe contener registros o categorías textuales diversificadas. Ello conlleva mayor validez de las conclusiones y permite la comparación y la generalización. La primera de estas ideas ha sido destacada en la bibliografía acerca de lingüística de corpus en los últimos veinte años (Francis 1979, Leech 1991, 1992, Sinclair 1982, 1991, Johansson 1991, Stubbs 1996). La segunda resulta ser más original y constituye uno de sus importantes aportes en torno a la cual Biber ha desarrollado sus investigaciones en diversas lenguas: la variación lingüística a través de diferentes registros orales y escritos.

Biber (1988, 2003) y Biber, Conrad y Reppen (1998) han dado cuenta de interesantes variaciones sistemáticas de orden gramatical y léxico en diversos registros del inglés oral y escrito. Dos hallazgos entre muchos de los reportados parecen relevantes: por un lado, los rasgos lingüísticos individuales presentan una ocurrencia diversa en variados registros; por otro, los mismos o similares rasgos lingüísticos pueden tener funciones diferentes en textos pertenecientes a diversos registros. En este sentido, una de las fortalezas de este enfoque metodológico se funda en un principio lingüístico comunicativo que resulta extremadamente sensato: la variación entre registros no se explica únicamente por un solo parámetro o dimensión, lo que equivale a sostener que existen múltiples distinciones situacionales entre registros. Dicho de otro modo, no es posible pensar que un rasgo lingüístico o, incluso, unos pocos de ellos puedan explicar exclusivamente una determinada variación entre registros (por ejemplo: oral/escrito, formal/informal).

2. Metodología

2.1 Los objetivos de la investigación y los corpora

En esta investigación se pretende determinar estadísticamente –mediante análisis factorial– los patrones lingüísticos sobresalientes y co-ocurrentes en el corpus PUCV-2003 y realizar un estudio comparativo a partir de los tres diferentes grupos de textos. Por último, a través del enfoque multidimensional, se efectúa una comparación de las tres áreas técnicas del CTC. Para ello se hace uso de programas computacionales que apoyan la labor de etiquetaje de los textos y de interrogación de los mismos. La conformación general de Corpus PUCV-2003 se desglosa en 90 textos que equivalen a 1.466.744 palabras. Este corpus general está dividido en tres registros o subcorpora (Corpus Técnico-Científico –CTC, Corpus de Literatura Latinoamericana –CLL, y Corpus de Entrevistas Orales –CEO). La siguiente tabla muestra su distribución por número de textos y palabras.

Tipo de Corpus	Número de archivos o textos	Total de Palabras
Corpus PUCV-CTC	74 (82%)	626.790 (42%)
Corpus PUCV-CLL	12 (13%)	459.860 (32%)
Corpus PUCV-CEO	04 (5%)	380.094 (26%)
Totales	90 (100%)	1.466.744 (100%)

Tabla 1. Constitución Global del Corpus PUCV-2003

2.2 Rasgos lingüísticos

Respecto a los rasgos lingüísticos a indagar, se llevó a cabo un rastreo bibliográfico con el fin de identificar categorías gramaticales que mostraran relevancia funcional en español. Con esta información disponible, se construyó una matriz con un total de 65 rasgos que se entregan en la Tabla 2:

Rasgos Lingüísticos Proyecto Corpus PUCV-2003	
A. Marcadores de tiempo verbal	**H. Formas estativas activas**
1. Pretérito indefinido (indicativo)	33. Ser
2. Pretérito imperfecto (indicativo)	34. Estar
3. Pretérito perfecto (indicativo y subjuntivo)	**I. Tipos verbales**
4. Presente (indicativo y subjuntivo)	35. Públicos
5. Futuro (indicativo y subjuntivo)	36. Privados
6. Futuro perifrástico	37. Persuasivos
B. Marcadores de modo verbal	38. Perceptivos
7. Indicativo/imperativo	**J. Verbos modales**
8. Subjuntivo/imperativo	39. Posibilidad
9. Modo indicativo	40. Necesidad
10. Modo subjuntivo	41. Obligación
11. Modo imperativo	42. Volición
C. Desinencias verbales de persona	**K. Marcadores de modalidad**
12. Primera singular	43. Atenuadores
13. Segunda singular	44. Enfáticos
14. Tercera singular	**L. Adverbios**
15. Primera plural	45. De lugar
16. Segunda plural	46. De tiempo
17. Tercera plural	47. De modo
D. Pronombres personales	48. De cantidad
18. Primera persona singular	**M. Marcadores de subordinación**
19. Primera persona plural	49. Subordinadas sustantivas con «que»
20. Segunda persona singular	50. Subordinadas adjetivas pron. relativo
21. Segunda persona plural	51. Subordinadas adverbiales de razón o c/e
22. Tercera persona singular	52. Subordinadas adverbiales de concesión
23. Tercera persona plural	53. Subordinadas adverbiales condicionales
24. Demostrativos	54. Subordinadas adverbiales de tiempo
E. Formas nominales	55. Frases infinitivo en función nominal
25. Nominalizaciones	**N. Frases preposicionales y adjetivos**
26. Sustantivos (comunes y propios)	56. Frases prep. (compl. del nombre)
F. Formas Pasivas	57. Adjetivos atributivo s(calificativo)
27. Pasivas con «se»	58. Adjetivos predicativos
28. Pasivas con ser sin agente	59. Adjetivos demostrativos
29. Pasivas con ser con agente	60. Participios función adjetiva
30. Pasivas con estar	**Ñ. Marcadores de Coordinación**
G. Especificidad léxica	61. Conjunciones adversa., adit. y disyun.
31. Relación *type/token por forma*	**O. Marcadores de negación**
32. Relación *type/token por lema*	62. Adverbio de negación
	63. Adverbios de negación temporal
	64. Conjunción de negación
	65. Pronombres de negación

Tabla 2. Rasgos lingüísticos Proyecto Corpus PUCV-2003

Una descripción funcional de cada uno de estos rasgos conjuntamente a su relación con ciertos registros y un detallado apoyo bibliográfico referencial, se encuentra en Parodi (2004).

3. Resultados

A continuación: (1) se identifican los factores y se determinan las funciones comunicativas compartidas por los conjuntos de rasgos en co-ocurrencia (interpretación funcional), y (2) se analiza la distribución de estos grupos de rasgos a través de los tres registros (CTC, CLL y CEO) y de los tres ámbitos técnico-científico del CTC (marítimo, industrial y comercial).

3.1 Cinco Dimensiones

Los rasgos que constituyen la Dimensión 1 Foco Contextual e Interactivo son los más numerosos y ostentan los pesos estadísticos más altos. De los 23 rasgos iniciales con peso estadístico sobre 0,40, se desatacan quince rasgos con pesos superiores a 0,60. Los rasgos con pesos negativos son relativamente menores en número y sólo cuatro presentan pesos sobre 0,40. Comparativamente, los rasgos con peso negativo son menos e indican una clara interpretación (nominalizaciones, frases preposiciones con función de complemento del nombre, sustantivos y participios pasivos en función adjetiva). Ellos son clásicamente considerados portadores de la carga referencial del texto, permiten la integración y precisión de grandes cantidades de información y una alta frecuencia de ellos apunta a una fuerte densidad informacional. Por su parte, los rasgos con altos puntajes de ocurrencias positivas revelan una referencia directa al contexto físico y temporal, determinan marcos de orden de la sucesión de hechos,

establecen una vinculación con la acción y expresan motivos y consecuencias. Esta dimensión se concreta en la acción, en la sucesión de acontecimientos y en las relaciones interpersonales de tipo dialógico. En oposición, la Dimensión 2 Foco Narrativo denota una evidente orientación hacia un determinado tipo de trama textual de orden narrativo. Ello se refleja en un amplio espectro de rasgos con peso positivo (de igual número al de la Dimensión 1), tales como, algunos pronombres personales y las respectivas desinencias verbales; ellos denotan un marcado acento en la identificación de las personas del discurso. En directa relación con lo anterior, se aprecia la co-ocurrencia de los tiempos verbales del pasado: el pretérito imperfecto y el pretérito indefinido. En suma, esta dimensión se asocia con una sucesión de acontecimientos, que implica la precisión de circunstancias de tiempo y lugar, como también la participación de las personas del discurso. Ella permite identificar textos literarios orales o escritos, a diferencia de textos altamente especializados. La Dimensión 3 es interpretada como Foco Compromiso pues la alta ocurrencia de verbos tales como privados (decidir, adivinar, sentir, determinar, demostrar, estimar, reconocer) y de volición (querer + infinitivo) y de los pronombres personales y las desinencias verbales de primera persona constituyen marcas relevantes de la expresión del «yo». Este compromiso con el discurso y su contenido revela los afectos y los propósitos del escritor/hablante. Esta dimensión está asociada a textos en los que sobresale la intención y la actitud del emisor que revela su voluntad de involucrarse en el discurso de manera explícita y de asumir un rol preponderante. Para construir la interpretación de la Dimensión 4, Foco Modalizador, se atendió de manera especial a la alta co-ocurrencia significativa de atenuadores (parecer que, creer, tal vez, a lo mejor, quizás, quizá), verbos modales de posibilidad (poder) y adverbios modales (probablemente, posiblemente). Ellos revelan un parámetro funcional muy preciso: la regulación y atenuación de la información entregada, es decir, la expresión de la probabilidad y la incertidumbre de los hechos o acontecimientos descritos o narrados. La conjunción sistemáticas de estas marcas lingüísticas tienden a darse en textos con énfasis en cómo (*modus*) se dicen las cosas, más que en lo dicho *(dictum).* Esta distribución de rasgos presenta el contenido de un discurso como incierto y abierto a la verificación.

Finalmente, la quinta dimensión, Foco Informacional, se constituye a partir de once rasgos: cinco positivos y seis negativos. Los rasgos positivos en co-ocurrencia como verbos modales de obligación revelan la necesidad y certeza de los juicios expresados; el modo subjuntivo remite a organizaciones de mayor complejidad sintáctica. La presencia de nominalizaciones junto a participios en función de adjetivo, sustantivos y frases preposicionales son todos rasgos indicadores de integración y compactación de información altamente abstracta, típica del discurso especializado escrito. En resumen, los rasgos positivos agrupados en torno a esta última dimensión se encuentran básicamente orientados hacia la informatividad, entendida ésta como la concentración de información en unidades y estructuras lingüísticas compactas, que presentan los datos lo más concisa y precisamente posible.

3.2 *Variabilidad en el corpus PUCV-2003*

En el Gráfico 1 se presentan los puntajes promedio por dimensión para cada uno de los 3 registros del corpus PUCV-2003 (técnicos, literarios y orales) con respecto a las cinco dimensiones en cuestión.

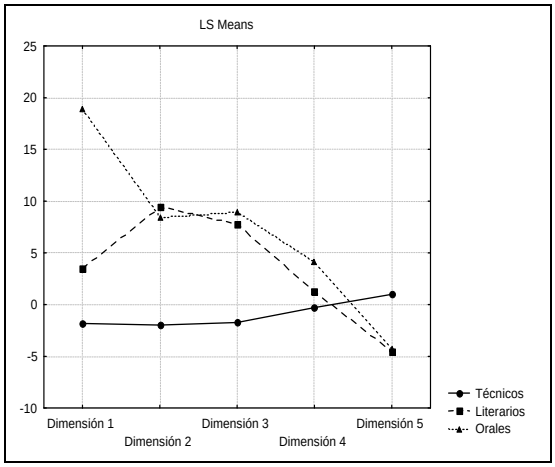

Gráfico 1. Dimensiones y registros.

Resulta interesante comprobar que los textos del CTC obtienen un puntaje factorial negativo muy similar a lo largo de las cuatro primeras dimensiones. Esta homogeneidad revela que, por un lado, los rasgos que identifican estas cuatro dimensiones deben presentar una ocurrencia similar en los textos de este corpus; según estos datos, la interactividad, contextualización, modalización y compromiso no serían rasgos característicos ni de alta ocurrencia en los textos del ámbito técnico y científico, comparado con los otros dos registros. Por otra parte, es esclarecedor que la Dimensión 5 Foco Informacional se presente con un puntaje promedio positivo (0,9), lo que permite identificar y distinguir significativamente el CTC del CLL y CEO. Las nominalizaciones, frases preposicionales como complemento del nombre y participios pasivos adjetivos, entre otros, son rasgos de alta relevancia en estos textos escritos. Estos rasgos contribuyen al cumplimiento del dispositivo referencial, el cual se aleja claramente de contenidos interpersonales y afectivos. Al mismo tiempo, como se observa, los textos de literatura latinoamericana escrita y los textos basados en entrevistas orales poco planificadas, con un grado importante de espontaneidad, muestran puntajes negativos en la Dimensión 5 y, en cambio, obtienen los más altos puntajes promedio positivos

en el resto de las otras cuatro dimensiones. Entre ellos, cabe destacar el elevado puntaje factorial que los textos orales alcanzan en la Dimensión 1 Foco Contextual e Interactivo (19); se hace evidente, entonces, que en las entrevistas orales cara a cara los rasgos como pronombres personales y desinencias verbales, adverbios de tiempo, lugar y modo, tiempo presente y pronombres y adjetivos demostrativos se constituyen en prototípicos.

Este análisis comparativo permite identificar las áreas de intersección entre la oralidad de las entrevistas y la narratividad de la literatura latinoamricana. También ayuda a separar distintivamente los textos de alta especialización en los que se detectan construcciones gramaticales más complejas y de mayor empaquetamiento y reducción de información de aquellos que involucran detalladamente a los participantes y sus relaciones interpersonales. En estos últimos textos (orales y escritos) se detecta y se expresa -de modo más explícito- el involucramiento del autor a través de marcas lingüísticas específicas (ciertos tipos de verbos, adverbios, pronombres, etc.). Hecho que también puede caracterizar el discurso especializado escrito, pero que muchas veces se implica a través de otros recursos.

Finalmente, en este último gráfico, se entregan los puntajes factoriales por dimensión para cada uno de las tres áreas de especialización del corpus CTC-PUCV-2003 (comercial, industrial y marítimo).

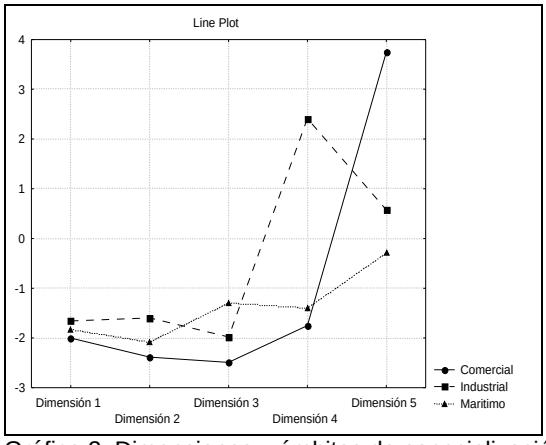

Gráfico 2. Dimensiones y ámbitos de especialización

Es interesante observar que también en el Gráfico 1 el CTC muestra una clara diferenciación en cuanto a la Dimensión 5 Foco Informacional del resto de los otros dos registros (CLL y CEO); en este nuevo análisis, más pormenorizado, es justamente esta quinta dimensión la que parece revelar la mayor distinción entre las áreas técnico-científicas. Según se aprecia, el ámbito comercial presenta el mayor puntaje promedio positivo en esta dimensión (3,8), hecho que muestra su mayor carga informacional a través de alta densidad léxica y complejidad sintáctica.

La dimensión cinco parece distinguir entre textos de un ámbito y otro. Seguidamente en la jerarquía, la Dimensión 4 también aporta a la distinción entre especialidades. Es ahora el área industrial la que revela mayor puntaje promedio positivo en la Dimensión Foco Modalizador (2,5), mientras que el marítimo y el comercial obtienen puntajes negativos muy cercanos y sin significatividad estadística entre ellos. Estos datos permiten inferir que los textos del ámbito industrial contienen mayor regularidad en los patrones sistemáticos de ocurrencia en torno a los rasgos distintivos de la atenuación e incertidumbre. En el resto de las dimensiones (1, 2 y 3), las tres áreas del CTC presentan cifras negativas y relativamente similares. Hechos que indican que estas dimensiones no aportarían gran explicación diferenciadora en la descripción del discurso técnico-científico de divulgación didáctico escrito. También se puede sugerir que los textos de estos ámbitos especializados no destacan por la ocurrencia de estos rasgos que denotan interacción, relaciones interpersonales e involucramiento de los participantes en el discurso.

4. Conclusiones

En este artículo se buscaba llevar a cabo una descripción inicial del corpus técnico-científico desde la perspectiva del análisis multidimensional. El análisis multidimensiones probó ser una metodología poderosa en el marco de la lingüística de corpus. Los beneficios y fortalezas del AMD implican más que la mera descripción de los rasgos superficiales de los textos, la determinación de regularidades sistemáticas que develan funciones comunicativas y aproximaciones hacia una descripción profunda de la función que tal o cual texto aporta en un contexto determinado de uso. La fortaleza que radica en el análisis multirasgos y la muestra relativamente grande de textos permite generar conclusiones robustas y con mayores implicancias. El hallazgo de las cinco dimensiones resultó en un recurso de extraordinario potencial descriptivo. Los aportes entregan una distinción fundamental entre los textos estudiados de la modalidad oral y escrita: los primeros aparecen anclados en la interactividad, contextualización y las relaciones interpersonales, esto es, fuertemente descritos por la Dimensión 1 Foco Contextual e Interactivo; al mismo tiempo, la oralidad aquí descrita (entrevistas) no se relaciona directamente con una prosa informacionalmente densa, tanto desde el punto de vista léxico o sintáctico. Si bien es cierto que el registro literario escrito coincide con el oral en las Dimensiones Foco Narrativo, Foco Compromiso y Foco Modalizador, la distinción manifiesta en la Dimensión 1 es suficiente como para establecer una diferencia importante. Evidentemente, esto no permite hablar de una fuerte polarización entre oralidad y escritura, lo que

reafirma la adscripción a la idea de un desplazamiento continuo entre los registros, a modo de categorías prototípicas y también difusas. Por otra parte, el AM también ha permitido identificar registros más expositivos y técnicos escritos de otros claramente narrativos literarios escritos.

Otro posible hallazgo relevante que se desprenden de los resultados de este estudio es la similitud detectada en varias lenguas en cuanto a la determinación de dimensiones lingüísticas que ayuden a la descripción y distinción entre registros (Kittredge 1982, Biber 1994, Aijmer, 2002, Reppen, Fitzmaurice y Biber 2002). El español, en términos generales, al igual que el inglés, el somalí y el coreano presenta múltiples dimensiones que reflejan -al menos- la interactividad, el foco informacional y el compromiso. En este sentido, la variabilidad registrada en usos lingüísticos al interior de las lenguas particulares indagadas, aunque existentes en este sentido de patrones, parecen tender a una cierta universalidad.

Los resultados aportados revelan la significatividad del estudio del uso lingüístico en entornos técnico-científicos escolares del ámbito profesional y confirman la necesidad de crear materiales especializados de instrucción sobre la base de descripciones empíricas de los registros en estudio.

5. Referencias bbibliográficas

Aijmer, K. 2002: «Modal adverbs of certainty and uncertainty in an English-Swedish perspective», en Hasselgard, H., Johansson, S., Behrens B. y Fabricius-Hansen, C. (eds.), *Information structure in a cross-linguistics perspective*, Amsterdam, Netherlands, págs. 97-113.

Biber, D. 1988: *Variation across speech and writing*, Cambridge, Cambridge University Press.

Biber, D. 1994: «Using register-diversified corpora for general language studies», en Armstrong, S. (ed.), *Using large corpora*, Cambridge, The MIT Press, págs. 180-201.

Biber, D. 2003: «Variation among university spoken and written registers: A new multi-dimensional analysis», en Leistyna, P. y Meyer, Ch. (eds.), *Corpus analysis. Language structure and language use*, Amsterdam, Rodopi, págs. 47-70.

Biber, D., Conrad, S. y Reppen, R. 1998: *Corpus linguistics. Investigating language structure and use*, Cambridge, Cambridge University Press.

Cabré, M. 2000: *La terminología: Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*, Barcelona, Instituto Universitario de Lingüística Aplicada, Universidad Pompeu Fabra.

Cabré, M. 2002: «Textos especializados y unidades de conocimiento: metodología y tipologización», en García, J. y Fuentes, M. (eds.), *Texto, terminología y traducción*, Barcelona, Almar, págs. 122-187.

Cademártori, Y. 2003: «La inscripción de las personas en textos de divulgación científica», *Revista Latinoamericana de Estudios del Discurso*. En prensa.

Calsamiglia, E. 2000: «Decir la ciencia: Las prácticas divulgativas en el punto de mira», *Revista Iberoamericana del Discurso y Sociedad*, 2(2) , págs. 3-8.

Cassany, D., López, C. y Martí, J. 2000: «La transformación divulgativa de redes conceptuales científicas. Hipótesis, modelo y estrategias», *Revista iberoamericana de Discurso y Sociedad, Lenguaje en contexto desde una perspectiva crítica y multidisciplinaria* 2(2), págs. 73-103.

Ciapuscio, G. 1992: «Impersonalidad y desagentivación en la divulgación científica», *Lingüística española actual* 2, págs. 183-205.

Ciapuscio, G. 2000: «Hacia una tipología del discurso especializado», *Revista Iberoamericana del Discurso y Sociedad* 2(2), págs. 39-71.

Ciapuscio, G. 2003: *Textos especializados y terminología*, Barcelona, Instituto Universitario de Lingüística Aplicada, Universidad Pompeu Fabra.

Ciapuscio, G. y Kuguel, I. 2002: «Hacia una tipología del discurso especializado», en García, J. y Fuentes, M. (eds.), *Texto, terminología y traducción* Salamanca, Almar, págs. 37-73.

Conrad, S. y Biber, D. 2001: «Multi-dimensional methodology and the dimensions of register variation in English», en Conrad, S. y Biber, D. (eds.), *Variation in English. Multidimensional Studies*, págs. 13-42.

Francis, N. 1979: «A tagged corpus: problems and prospects», en Greenbaum, S., Leech, G. y Svartvik, J. (eds.), *Studies in English linguistics for Randolph Quirk*, London, Longman, págs. 192-209.

Harvey, A. 2002: «Representación e imagen del quehacer científico en los Medios de Comunicación», en Parodi, G. (ed.), *Lingüística e Interdisciplinariedad,* Valparaíso, Ediciones Universitarias de Valparaíso, págs. 335-353.

Johansson, S. 1991: «Times change, and so do corpora», en Aijmer, K. y Altenberg, B. (eds.), *English Corpus Linguistics. Studies in honor of Jan Svartvik*, London, Longman, págs. 305-314.

Kittredge, R. 1982: «Variation and homogeneity of sublanguages», en Kittredge, R. y Lehrberger, J. (eds.),

Sublanguages. Studies of language in restricted semantic domains, Berlin, Walter de Gruyter, págs. 145-189.

Leech, G. 1991: «The state of the art in corpus linguistics», en Aijmer, K. y Altenberg, B. (eds.), *English Corpus Linguistics. Studies in honor of Jan Svartvik*, London, Longman, págs. 8-29.

Leech, G. 1992: «Corpora and theories of linguistic performance», en Svartvik, J. (ed.), *Directions in corpus linguistics: proceeding of Nobel symposium*, Berlin, Mouton de Gruyter, págs. 105-122.

Lorente, M. 2002: «Verbos y discurso especializado». [En línea]. Disponible en: http://elies.rediris.es/elies16/Lorente.html

Martín Zorraquino, M. y Portolés, J. 1999: «Los marcadores del discurso», en Bosque, I. y Demonte, V. (coords.), *Gramática descriptiva de la lengua española*, Vol. III, Madrid, Espasa Calpe, págs. 4051-4213.

Menéndez, S. 1999: «El discurso del libro de texto: Una propuesta estratégico-pragmática», *Revista Iberoamericana de Discurso y Sociedad* 1(2), págs. 85-104.

Montolío, E. 2001: *Conectores de la lengua escrita*, Barcelona, Ariel Practicum.

Parodi, G. 2004: «Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada en corpus computarizado», *Estudios Filológicos.* En revisión.

Parodi, G. y Gramajo, A. 2003: «Los tipos textuales del corpus PUCV-2003: una aproximación multiniveles», *Signos* 36(54), págs. 207-223.

Portolés, J. 1998: *Marcadores del discurso*, Barcelona, Ariel Practicum.

Reppen, R., Fitzmaurice, S. y Biber, D. 2002: *Using corpora to explore linguistic variation*, Amsterdam, John Benjamins.

Sinclair, J. 1982: «Reflections on computer corpora in English language research», Johansson, S. (ed.), *Computer corpora in English language research*, Bergen, Norwegian Computing Centre for the Humanities, págs. 1-6.

Sinclair, J. 1991: *Corpus, concordance, collocation*, Oxford, Oxford University Press.

Stubbs. M. 1996: *Text and corpus analysis. Computer-assisted studies of language and culture*, Malden, Massachusetts, Blackwell.