

GIOVANNI PARODI

Editor

*Discurso
Especializado
e*

Instituciones
Formadoras



EDICIONES UNIVERSITARIAS DE VALPARAÍSO
Pontifical Universidad Católica de Valparaíso

- Capítulo 1.** Trabajo elaborado a partir de la Conferencia Magistral Inaugural, dictada con ocasión del Segundo Congreso Internacional sobre Lengua Escrita y Textos Académicos, Pachuca, México, abril, 2004.
- Capítulo 2.** Versión ampliada y revisada del artículo publicado en *Revista Signos*, (36)54, 2003.
- Capítulo 3.** Versión ampliada y revisada del artículo publicado en *Revista Estudios Filológicos*, 29, 2004.
- Capítulo 4.** Versión ampliada y revisada del artículo publicado en *Revista Española de Lingüística*, 1, 2005.
- Capítulo 6.** Versión ampliada y revisada del artículo publicado en *Revista Signos*, (37)55, 2004.
- Capítulo 9.** Versión ampliada y revisada del artículo publicado en *Revista Literatura y Lingüística*, 15, 2004.

"You cannot find peace, by avoiding life"
Virginia Wolf

© Giovanni Parodi, 2005

ISBN 956-170367-0

Inscripción N° 144.794.

Derechos Reservados

Ediciones Universitarias de Valparaíso

Pontificia Universidad Católica de Valparaíso

Calle 12 de Febrero 187, Valparaíso

Teléfono (32) 27 30 67 - Fax (32) 27 34 29

E-mail: euvc@uv.cl

www.euvc.cl

Jefe de Diseño: Guido Olivares

Asistente de Diseño: Mauricio Guerra

Corrección: Ovaldo Oliva

Impreso en Solsticios S.A.

Índice

- 9 **Prólogo: Daniel Cassany**
- 11 **Introducción: Giovanni Parodi**
- 15 **1 Discurso especializado e instituciones formadoras:
Encuentro entre lingüística e informática
Giovanni Parodi**
- 45 **2 Los tipos textuales del corpus técnico-profesional PUCV-2003: Una aproximación multíniveles
Giovanni Parodi y Aída Gramajo**
- 61 **3 Textos de especialidad y comunidades discursivas
técnico-profesionales: Una aproximación basada
en corpus computarizado
Giovanni Parodi**
- 83 **4 Lingüística de corpus y análisis multidimensional:
Exploración de la variación en el Corpus PUCV-2003
Giovanni Parodi**
- 127 **5 Hacia una identificación automatizada de rasgos
argumentativos en corpus
René Venegas**

Prólogo

- 159 **6 Especificidad, especialización y variabilidad verbal:
Una aproximación computacional en estadística léxica**
Omar Sabaj
- 181 **7 Foco narrativa y foco informativo: Dos dimensiones para
una descripción de los manuales en la formación
técnico- profesional**
Juanita Marinkovich y Yanina Cademártori
- 193 **8 Las estrategias de reformulación: El paso desde un
texto-fuente a un texto de divulgación didáctica**
Juanita Marinkovich
- 217 **9 BUCÓLICO: Aplicación computacional para el análisis de
textos. Hacia un análisis de rasgos de la informatividad**
Giovanni Parodi y René Venegas
- 239 **Postfacio: Elvira Narvaja de Arnoux**
- 243 **Referencias bibliográficas**

Aunque nunca se diga, todos sabemos que escribir y leer el prólogo de un libro es a menudo una tarea protocolaria, en la que las conocidas máximas de cortesía pragmática encuentran uno de sus mejores ejemplos. El prologuista, halagado por la invitación de los autores, se siente comprometido a elogiarlos más allá del valor intrínseco del libro. Los lectores, que no ignoran el caso, se acercan al texto esperando encontrar en él los tópicos al uso sobre la importancia, la idoneidad y el interés del volumen, etc., pero con la esperanza de saber que el género es, por suerte, breve.

Nada de esto ocurre aquí –lo espero que no ocurra! (excepto lo de la extensión breve)–, aunque no podamos suscribirnos a la tradición sociocultural del género. Para mí es un gran placer y un honor escribir estas palabras y mi intención es poder transmitirles a ustedes las razones de este sentimiento, las académicas y las personales, de modo que puedan acercarse al auténtico principio de este volumen con un poco más de apertura.

Entre las razones académicas, creo que *Discurso Especializado e Instituciones Formadoras* recoge algunas de las perspectivas y de las preocupaciones más actuales en el ámbito de la lingüística aplicada o, más en concreto, de la didáctica de la lectura y la escritura. En primer lugar, dicho de modo metafórico, asume que “leer y escribir son verbos transitivos”, o sea, que procesamos de modo diferente el discurso, según el contexto, la finalidad, el destinatario, el propósito, etc., y que estas diferencias son relevantes tanto en la investigación de estas habilidades comunicativas como en su enseñanza.

aprendizaje en el que. Hoy en día una didáctica eficaz de la lectura y la escritura debe considerar la gran diversidad de discursos que manejan en los diferentes esferas de la actividad humana, personal y profesional.

En este sentido, creo muy acertada la elección de analizar las prácticas de lectura que realizan los estudiantes chilenos de secundaria en los ámbitos técnico-profesionales, puesto que, si bien se trata de las ramas del saber que cada día atraen a más jóvenes, carecemos –no solo en Chile– de estudios serios y detallados que nos iluminen al respecto. Por esta razón, este conjunto de artículos puede –debe– resultar útil al colectivo de docentes implicado en la formación lingüística de estos ámbitos.

En segundo lugar, la investigación incluye varios niveles de análisis, de modo que los resultados ofrecidos son amplios y sugerentes. La aproximación a las prácticas lectoras técnico-profesionales adoptó los conceptos de género y comunidad discursiva, para pasar a identificar, clasificar y describir tipos de texto, a caracterizarlos gramatical y discursivamente, y a describir sus constantes y sus variaciones, de modo comparativo en cada uno de los tres ámbitos temáticos elegidos. Este espectro tan variado no solo muestra la riqueza del proyecto actual de investigación, sino también las posibilidades de futuras o posteriores exploraciones.

En tercer y último lugar, las investigaciones presentadas se inscriben en una propuesta metodológica decididamente empírica, moderna y sofisticada, en el ámbito de la lingüística del corpus. La descripción de las prácticas mencionadas se basa en el análisis computacional de regularidades en un corpus de textos que previamente han sido procesados etiquetados gramaticalmente. El volumen de textos (corpus) y su notable representatividad son una garantía de la fiabilidad de los resultados. Pero también hay que destacar el hecho de que tanto la metodología como los corpus procesados quedan también a disposición de los lectores y de los usuarios que quieran realizar nuevas investigaciones, a partir de las que se presentan ahora.

En el plano más personal, quiero recordar que conozco al equipo de docentes e investigadores de la PUCV desde 1998 y que he sido testigo silencioso de su paulatino crecimiento y de su incansante maduración. El mundo universitario, y especialmente los departamentos de letras, no siempre pueden presentar equipos cooperativos de investigadores maduros y jóvenes, docentes y estudiantes doctorandos, que trabajan todos a una en un mismo proyecto, desde diferentes intereses personales y formaciones académicas.

Por todo lo dicho, cabe felicitar a los autores y al editor de este volumen por el fruto de un trabajo tan actual y sugerente.

DANIEL CASSANY
Universidad Pompeu Fabra
España

Introducción

Ineludiblemente existen muy pocas actividades, valoradas como signos de cultura, que no se vinculen de manera directa a la modalidad escrita de la lengua. El manejo adecuado de la lectura y de la escritura ha representado durante un largo periodo de tiempo señal inequívoca de estatus y progreso intelectual en las culturas alfabetizadas, tal como lo muestra. No obstante ello, es bien sabido que los esfuerzos por lograr que los estudiantes de diversos niveles educativos alcancen una competencia comunicativa escrita suficientemente aceptable han enfrentado serios inconvenientes. Si bien es cierto que el panorama no es favorable en lo que respecta al dominio del lenguaje general, preocupación del sector científico-humanístico, se detecta un mayor desarrollo y apoyo institucional en esta área, pero se percibe un menor acento respecto de las habilidades lingüísticas que se relacionan con el discurso técnico-profesional de comunidades especializadas.

Este libro, uno de los primeros en Chile y en Latinoamérica, que se adentra en la problemática de este tipo de discurso especializado, intenta abrir un camino inicial que aporta de manera parcial a acortar la brecha en el conocimiento disponible acerca de los materiales escritos que circulan en tres áreas técnicas. Por cierto, no pretende ofrecer soluciones definitivas, sino más bien entregar una primera aproximación al ámbito.

En este contexto, la construcción de un conocimiento específico que dé cuenta del estado actual en Chile de los tipos de textos y sus rasgos lingüísticos y discursivos característicos a partir de los materia-

les escritos que se entrega como lectura obligatoria en diversos áreas del sistema de formación técnico profesional secundario se constituye en una tremenda oportunidad de indagación científica. Como se decía, la carencia de estos conocimientos y su disponibilidad es actualmente, sin lugar a dudas, una gran debilidad para cualquier tipo de decisión que se desee enfrentar. Sea ésta de corte instruccional, metodológico, curricular o puramente investigativa.

El foco de este libro está en lo que denominamos "discurso de especialidad o especializado". Nuestra opción por este término no es antojadiza. Creemos que con gran precisión capta y da cuenta de, por una parte, una amplia heterogeneidad de textos que circulan en ciertos ámbitos técnico-profesionales de alta especialización pero, por otra, incluye una serie de discursos que pueden ubicarse en lo que denominamos divulgación de la ciencia y la tecnología y acoge hasta aquellos de divulgación más didáctica propiamente tal. Al mismo tiempo, dentro de este *continuum*, también se incorporan discursos netamente profesionales que circulan en la formación técnico-profesional pero que no necesariamente se produce en dicho ámbito, sino que se crean y circulan en ambientes exclusivamente laborales. Una aproximación a esta cuestión terminológica se encontrará a la larga de diversos capítulos del libro, pero más acorazadamente en el Capítulo 4, en donde Giovanni Parodi y Aida Gramajo aplican una taxonomía multiniveles compleja en un intento por tipologizar los textos técnico-profesionales del Corpus PUCV-2003. También en los Capítulos 1 y 2, se busca afinar esta concepción de especialización.

Este libro reúne una colección de trabajos, todos artículos orgánicamente presentados en torno a un tema nuclear: el discurso escrito especializado que circula en el sistema de formación educacional técnico-profesional.

Estoy seguro que la fortaleza de contar -en un solo volumen- con un conjunto de nueve artículos científicos, algunos de ellos publicados originalmente en revistas de especialidad y otros inéditos y, por ende, dispersos en diversos espacios, constituye un aporte sustancial al ámbito. A partir de ello, se logra tener una visión de conjunto de los desarrollos en una misma línea de investigación en ejecución en la Pontificia Universidad Católica de Valparaíso (Chile), conducidos todos a partir de un mismo corpus de textos (Corpus PUCV-2003) y desde enfoques -en algunos casos algo diversos- pero con el mismo propósito general de describir y explicar los tipos de textos en el universo recolectado, sus rasgos lingüísticos co-ocurrentes más prototípicos y sus diversas tramas textuales; todo ello a partir de lo que leen y dicen oralmente alumnos de tres ámbitos técnico-científicos de establecimientos de la ciudad de Valparaíso: marítimo, comercial e industrial.

Si partimos del supuesto que cada comunidad discursiva, en este caso de índole técnico-profesional, genera un tipo de material escrito que representa su particular modalidad de comunicación especializada intragrupal y que es a través de este medio que se transmite, valida y aprende ese mismo conocimiento, debemos concordar que el estudio de esa forma particular de lengua reviste importancia vital. Es, como se comprende, mayoritariamente, por medio del discurso especializado escrito de una comunidad específica que los interesados en ingresar a formar parte de ese grupo humano tienen que acceder al conocimiento disciplinar e integrarse así paulatinamente a formar parte del modo en que opera dicha comunidad.

En Chile, la realidad del sistema de enseñanza/aprendizaje del discurso técnico-profe-

sional en comunidades especializadas de establecimientos de educación secundaria es compleja y diversa. Hasta hace muy poco tiempo se disponía de escasos o nulos antecedentes que permitieran esbozar un panorama lingüístico desde una óptica integradora y transdisciplinaria de modo más o menos completo. Por lo tanto, la alfabetización y la planificación lingüísticas de estas comunidades son urgencias fundamentales.

En el escenario actual, de inminente necesidad de formación de recursos humanos altamente capacitados que refuercen las líneas existentes en investigación en las universidades chilenas y que a la vez sean capaces de generar nuevas fronteras de desarrollo del conocimiento, dos alumnos del Programa de Doctorado en Lingüística de la PUCV y, al mismo tiempo, becarios del Proyecto MECESUP UCV/004 han sido invitados como co-autores y autores de capítulos aquí incluidos. Ellos son Omar Sobaj y René Venegas. De más está referirse a los beneficios de dar espacio a estas nuevas generaciones y acogerlos en grupos consolidados ya existentes de los cuales se hacen parte activa paulatinamente. La riqueza humana y académica de su incorporación no ha sido casual, sino que se inscribe en el programa de desarrollo estratégico del postgrado en lingüística de la PUCV. Su participación ha resultado muy valioso y ha venido a demostrar que hemos acertado en las decisiones; todo ello también demuestra fehacientemente el modo en que debemos continuar: interactuando en la academia científica: hacia la búsqueda de la verdad y el conocimiento pero con el corazón en la mano, ya que somos hombres y mujeres trascendentes.

Tan hondo ha calado el trabajo disciplinar y el esfuerzo por constituir una línea de investigación emergente que nos identifique que hemos decidido constituir un grupo de investigación, denominado Grupo ALADE (*Análisis Lingüístico del Discurso Especializado*). La producción científica de este Grupo de Investigación es relativamente nueva, dado que su creación data del año 2003. Ahora bien, a pesar de su reciente nacimiento como Grupo de Investigación, el equipo de investigadores fundantes ya lleva más de tres años abordando el discurso especializado y cuenta, a su vez, con una larga tradición internacional en temas de psicolingüística del discurso escrito. De modo más general, algunos de los investigadores del grupo son lingüistas de trayectoria en Latinoamérica, cuyas investigaciones han dedicado atención particular a los procesos de comprensión y producción del discurso escrito.

En el contexto del Grupo ALADE, atenderemos al discurso especializado desde una óptica -muchas veces- decididamente transdisciplinaria, incorporando de manera fundamental el empleo de las tecnologías de la computación como parte vital de nuestras investigaciones (tal es el caso de la lingüística computacional de corpus [al respecto, ver Parodi, Sobaj & Venegas, 2005]). No obstante ello, también nos interesa corroborar los hallazgos empíricos de corte cuantitativo y estadístico a través de exploraciones eminentemente cualitativas en las que se ponga en evaluación el producto de las certificaciones y se devela avulsalemente la complejidad de los textos en estudio. Este proceso recursivo es fundamental para, por una parte, comprobar los resultados descritos de modo preliminar y para, por otra, aportar una mirada más enriquecedora del fenómeno discursivo de naturaleza compleja, como el que nos preocupa. En esta última línea, se sitúan los Capítulos 7 y 8, escritos por Juanita Marinkovich y Yanina Cademartori.

El desarrollo vertiginoso del Grupo en cuestión y de la tecnologización de la investigación en el tipo de lingüística que aquí se cultiva, asociada a los proyectos con fondos gubernamentales que actualmente se ejecutan, se ha visto concretado en un卑ivisimo lapso de

tiempo. Para ello se ha debido desplegar opciones de diversa naturaleza tales como: constitución de equipos transdisciplinarios (lingüistas, profesores de ciula, especialistas técnico-profesionales, ingenieros computacionales, estadísticos, diseñadores gráficos, entre otros). Gracias a los recursos económicos obtenidos a través de concursos provenientes del Proyecto FONDECYT 1020786, del Proyecto MECESUP UCV/004, del Proyecto MECESUP UCU/0311, del Proyecto FONDEF 1051 y de Proyecto Andes C-13955[4], además de recursos aportados por la PUCV, se ha hecho posible fortalecer la constitución del equipo investigador, implementar un *laboratorio Computacional en Ciencias del lenguaje* con alta tecnología y adquirir paquetes de programas computacionales. También en el marco de estos proyectos en curso, se han desarrollado y se encuentran en estudio plataformas computacionales poderosas [tal como, *El Grial*] y herramientas específicas para el tratamiento de grandes muestras de textos (ver Capítulo 9). Al mismo tiempo, se han recolectado otros grandes corpus de textos auténticos, aparte del que se da cuenta en este libro.

Es significativo aclarar que los trabajos que emergen del Corpus PUCV-2003 no representan de manera alguna una variedad del español de Chile. Nunca fue ese el propósito en la recolección de estos textos. Por una parte, la heterogeneidad en naturaleza y autoría de los textos escritos que circularan en el sistema educacional secundario técnico-profesional no permite tal ambición. Por otra, más bien motivados por el deseo de recuperar un tipo auténtico de textos que efectivamente leyeron los estudiantes de estos ámbitos nos alejó de la idea de dar cuenta de una variedad determinada del español de Chile.

También es de justicia reconocer el importante aporte que alumnos de nuestra universidad, tales como tesistas del Magíster en Lingüística Aplicada así como en proyectos de titulación de pregrado en la carrera de Castellano y Comunicación. Sus nombres no aparecen como autores de capítulos del libro, pero no cabe duda que muchos de estos apartados no habrían visto la luz sin su entusiasta y decidida colaboración.

El libro, que estas palabras abren, es un primer acercamiento tanto a la descripción de los tipos de textos como hacia una caracterización de los rasgos lingüísticos, textuales y comunicativos del material escrito que se lee en tres ámbitos de la educación técnico-profesional chilena. Tal como se adelantaba, son escasos los antecedentes en estos temas en nuestro país; de modo que esperamos que esta contribución inicial pueda constituirse en un factor estimulante para que se genere mayor interés por el ámbito y se apoye decididamente a los alumnos en su paulatina incorporación en la construcción de un conocimiento disciplinar determinado.

Ciertamente que algunas conclusiones y propuestas esbozadas en este libro son de carácter tentativo, pero las metodologías empleadas y los marcos de referencia constituyen apartes certeros y modelos que emular. Todo ello, abre fronteras que el estudio científico de la palabra escrita, llave singular del conocimiento disciplinar, ofrece esperanzadoramente.

GIOVANNI PARODI S.
Editor

Capítulo 1

Discurso especializado e instituciones formadoras: Encuentro entre lingüística e informática*

GIOVANNI PARODI

INTRODUCCIÓN

En este primer capítulo del libro pretendo esbozar un marco referencial general que enmarque las investigaciones científicas de naturaleza empírica y las desarrollos computacionales que damos cuenta en los capítulos posteriores. Para ello, efectuaré un recorrido a través de ciertos hitos que juzgo relevantes para situar el resurgimiento de los estudios de corpus, ahora desde una óptica computacional y con gran apoyo tecnológico en estadística. La perspectiva adoptada es claramente desde una lingüística transdisciplinaria, núcleo convergente que a mi juicio está determinando de manera fundamental el devenir de nuestra disciplina y que permite perfilar algunos de los desafíos que debemos enfrentar para continuar de manera certeza los desarrollos en el área. De otro modo, estoy seguró que la lingüística de corte más clásico corre riesgos como nunca antes.

Posteriormente, abordaré el concepto de discurso especializado, en el cual se inscriben los estudios de este libro. Desde allí, realizaré un panorama diagnóstico de la situación que se vive actualmente en algunos liceos técnico-profesionales de las ciudades de Valparaíso y de Viña del Mar (Chile) en cuanto al proceso de alfabetización técnico-profesional con base en la lengua escrita y, en este contexto, me ocuparé de delinear algunas proyecciones hacia una didáctica del discurso especializado.

* Todas las traducciones son responsabilidad del autor.

SALIENDO DEL CASCARÓN: HACIA LA LINGÜÍSTICA DEL USO

Es muy probable que no en todo el mundo occidental el movimiento generativista chomskiano lograra una cobertura importante y, por ende, su impacto no constituyó una influencia decisiva y hegemónica en el devenir científico de las ciencias del lenguaje. A pesar de ello, en mi opinión y experiencia, si estimo que en parte importante de los polos de desarrollo científico las ideas esgrimidas inicialmente por Noam Chomsky constituyeron y aún proyectan una concepción del lenguaje humano y un andamiaje para su indagación que impidieron de manera certera el potenciamiento de posturas que abordaban el estudio del lenguaje desde otras ópticas más diversas. Razones se puede buscar muchas. Ya veremos algunas.

Está de más referirse a que, en lingüística, un paradigma científico emergente, en términos khuneanos, no reemplaza ni desplaza de manera categórica a otro precedente. Esto, muy probablemente, se aplica a otras áreas del conocimiento, mas no se produce necesariamente así en las ciencias sociales. Es sabido que la coexistencia paradigmática es una cuestión normal en lingüística y que varias opciones tanto teóricas como metodológicas pueden coexistir. De hecho, hoy en día está claro que se requiere de una necesaria integración de aproximaciones diversas para dar cuenta más acabadamente de un fenómeno determinado.

No obstante esta aclaración, es evidente que la impronta generada por Chomsky y la fuerza de sus ideas imperaron con bastante rigor en lingüística por casi tres décadas. Sin lugar a dudas, la figura de este lingüista y su decisiva aportación permitió, en un momento, abrir una brecha hacia una mirada más integral del ser humano y posibilitó la indagación de la mente humana, superando así posturas muy radicales que, debido a principios epistemológicos o a impedimentos metodológicos, restringieron las investigaciones lingüísticas exclusivamente al estudio de los conductos observables. Esto apertura, por supuesto, no solo generada por Chomsky sino por otros científicos y movimientos multidisciplinarios, posibilitó que algunas nacientes interdisciplinas, tal como la psicolingüística, se posicionaran y avanzaran decididamente hacia el estudio científico de los procesos de comprensión y producción de las lenguas naturales.

Si bien es cierto que el generativismo aportó de manera crucial en cuestiones diversas, no es menos cierto que -entre otras- la visión idealizada del lenguaje (a saber, el estudio de la competencia lingüística) bloqueó u opacó seriamente -tal vez sin proponérselo- a los verdientes proponentes del estudio del lenguaje en uso y de la investigación de la variabilidad lingüística. Ello produjo una discontinuidad de ciertas líneas de investigaciones en lingüística. Sinclair (1991:11) ilustra con claridad los efectos de lo limitado del enfoque generativista:

"Sedienta de información adecuada, la lingüística lingüidecía -de hecho- se volvió totalmente introversa. Se hizo una moda mirar hacia adentro de la mente más que hacia la sociedad. La intuición se volvió la clave y se enlatizó la similitud de la estructura del lenguaje y varios modelos formales. El rol comunicativo del lenguaje fue escasamente mencionado".

Buscando una explicación a la falta de preocupación por el uso lingüístico, Chafe

(1992) arguye que la naturaleza modular de la teoría impulsada por Chomsky, cuyo núcleo se fundamenta en que el sistema lingüístico opera de manera independiente del sistema cognitivo humano, se constituye en un impedimento al estudio del uso lingüístico. Chafe (1992: 81) aclara que:

"Una consecuencia de la visión modular del lenguaje humano es que sus adherentes no están interesados en la observación del uso del lenguaje cotidiano ya que consideran que lo más interesante acerca del lenguaje existe independientemente de su uso".

Del mismo modo que la hegemonía generativista desestimó el estudio del lenguaje a través de corpus de textos naturales, también evadió un enfoque de dimensiones probabilísticas. Enfatizando esta postura, Chomsky (1969:38) opinaba que:

"Se debe reconocer que la noción de 'probabilidad de una oración' es completamente inútil, sea cual sea la interpretación de este término".

Así, la autoridad detentada por Chomsky impuso restricciones metodológicas serias a los estudios lingüísticos y se excluyó -por un largo tiempo- en parte importante de las investigaciones acerca del lenguaje la incorporación de herramientas estadísticas, de diversa índole, como un componente central de una concepción particular de lenguaje.

El renacimiento de los estudios basados en corpus se puede fijar a comienzos de la década del sesenta, marcado -en parte- por la fuerte irrupción de los computadores en el ámbito lingüístico y el desarrollo de grandes proyectos de investigación en Inglaterra y en los países escandinavos, a partir de la construcción de grandes corpus lingüísticos digitales del inglés (por ejemplo, SEU Corpus, Cobuild Corpus, IL Corpus, Brown Corpus); ellos constituyeron el eje de avanzada de esta nueva reposición. Al respecto, Leech (1991) argumenta que esta renovada Lingüística de corpus se desarrolló rápidamente hasta volverse una rama independiente dentro de la lingüística.

En esta línea, los requerimientos de análisis semiautomáticos y exhaustivos de textos sobre la base de herramientas computacionales (tales como etiquetadores morfosintácticos) derivó en descripciones en términos probabilísticos y llevó al desarrollo de gramáticas independientes del contexto (*context-free grammars*). Desde el enfoque probabilístico, la variación es tomada como parte integral del funcionamiento lingüístico en la formulación de los mecanismos de selección, ya que ellos emergen de distribuciones observables, frecuencias relativas y correlaciones estadísticas. En estos términos, una gramática probabilística es muy similar a algunas gramáticas convencionales, excepto que además de asignar un conjunto de estructuras para cada secuencia de palabras de una lengua, también entrega una probabilidad para cada una de ellas (Halliday, 1992; Aarts, 1991; Stubbs, 1996). Una característica importante de las gramáticas y de los etiquetadores probabilísticos es que se van construyendo a partir de la interacción entre unos resultados preliminares y la revisión de expertos que retroalimentan los posibles problemas del sistema, de modo que el etiquetador o la gramática en cuestión se vuelve cada vez más preciso y robusto.

Parte del poder de los programas computacionales probabilísticos reside en la desambigüación de los términos en el análisis lingüístico, eligiendo la alternativa más plausible.

sible según sea el caso. Tal elección se basa en las frecuencias relativas de ocurrencia de todas las combinaciones posibles en un *corpus* representativo (Brusciano & Carroll, 1994; Minker, Wabel & Mariari, 1999; Juolasky & Morfin, 2000; Bunt & Nijholt, 2000).

Ahora bien, la sucesión de estos cambios provocó una nueva manera de enfrentar la investigación científica, revitalizando el interés por los usos del lenguaje natural y cotidiano y su inherente variabilidad. Esta renovada mirada alternativa nos enfrenta al renacimiento del empirismo, pero no necesariamente bajo la influencia de la lingüística estructural de corte behaviorista ni de la psicología conductista imperantes en los años cincuenta: la oposición entre métodos basados en el conocimiento (Church & Mercer, 1993) y métodos empíricos, así como la oposición entre una llamada "lingüística del sillón," versus una "lingüística de *corpus*" (Fillmore, 1992), son distinciones dicotómicas que ya no tienen cabida ante las visiones inter y transdisciplinarias, en donde se oponen hacia integraciones y colaboraciones más eficientes entre los distintos ámbitos de las ciencias. Por supuesto, ello no impide la existencia de posturas extremadamente radicales en uno y otro polo de una opción deductiva o inductiva.

EN BUSCA DE HERMANDAD: LINGÜÍSTICA E INFORMÁTICA

La lingüística, denominada clásica en algún sentido, en muchos ámbitos latinoamericanos y en otras partes del mundo, se ha mantenido algo ajena a los avances en las ciencias de la computación. La irrupción de los computadores y los emergentes y poderosos programas computacionales han corrido -muchas veces- paralelamente a las fuentes principales del desarrollo de las ciencias del lenguaje. Al mismo tiempo, es claro que los *corpus* de textos naturales digitalizados han mostrado un importante impacto en los análisis lingüísticos en las últimas décadas. Por ejemplo, un aspecto altamente destacable es que las indagaciones desde esta óptica han demostrado que los estudios lingüísticos basados en grandes muestras de textos digitales no siempre corroboran las intuiciones iniciales de los investigadores. Entre otras razones, ello ha impulsado el uso del *corpus* con sustento computacional así como la creación y utilización de programas que posibilitan el tratamiento y posterior análisis de los textos. Estos avances han llevado a las investigaciones lingüísticas hacia nuevas y desgarradas fronteras, superando así la era de los estructuralismos estrictos y la era de los idealismos exclusivos.

Como se ha dicho, el empleo conjunto de aproximaciones lingüísticas basadas en *corpus* y de enfoques computacionales no ha sido una combinación habitual en el ámbito de la lingüística; en parte, la ausencia de esta interacción, ha limitado a las investigaciones a centrarse en el estudio y análisis de reducidos números de textos, realizando dichos análisis de manera manual, con gran costo de tiempo y energía. Así, se ha desaprovechado los beneficios de *corpus* progresivamente mayores, adecuadamente digitalizados y etiquetados, junto a robustos programas computacionales de análisis automático. Queda claro, entonces, que los alcances de cualquier indagación a partir de muestras excesivamente reducidas de textos puede ofrecer serias limitantes, a no ser -obviamente- que solo constituyan aproximaciones exploratorias preliminares o que se amparen en paradigmas metodológicos *ad hoc* (tal como algunas metodologías en el marco del análisis crítico del discurso).

De este modo, se hace evidente la necesidad de buscar la complementación entre ambas posturas (lingüística e informática), tendiendo al equilibrio y evitando posibles asimilamientos perjudiciales. De no ser así, las investigaciones deben asumir limitaciones a su potencial. El estudio de pequeñas muestras de texto puede llevar a que los resultados obtenidos pierdan robustez y confiabilidad. Como es de esperar, esta tendencia no ha sido hegemónica en el panorama de la lingüística actual y algunos investigadores, cuando esfuerzos entre la lingüística y la informática, han capitalizado los recursos disponibles y con esfuerzos mancomunados están logrando avances sustantivos.

Desafortunadamente, por el contrario, se ha tendido a la creencia de que el caso o el estudio cualitativo ejemplar no presenta repercusiones importantes. No estamos sosteniendo que las investigaciones cualitativas se oponen a estudios de corte informático o estadístico. Es evidente que incluso los grandes *corpus* no logran dar cuenta de la lengua como en todo ni tampoco se pretende que así sea. La cuestión de la llamada representatividad de un *corpus* es un problema no menor. Como bien sabemos, la lengua en su dinamismo y heterogeneidad es mucho más rica de lo que se puede imaginar y no logró ser captada en un solo *corpus*, por gigantesco que sea su tamaño. Como bien apunta Leech (2002), un *corpus* puede ofrecer información detallada acerca de una lengua particular, pero es imposible recolectar un *corpus* que abarque *toda* una lengua. Si ese fuera el caso, sería necesario recolectar *todos* los usos de dicha lengua. De este modo, se debe siempre tener presente que un *corpus* es solo una colección finita de un universo *infinito*. Por ello, el desafío de contar con un *corpus* representativo de una variedad determinada de lenguas -incluso de un solo registro específico de tal o cual lenguaje- es una cuestión compleja debido a la enorme diversidad y variedad inherente a cada lengua particular. Por tanto, se debe hacer un esfuerzo e intentar alcanzar la mejor representatividad del objeto en estudio con el fin de lograr que los análisis de esos textos posibiliten cierto grado de generalización acerca de la variedad de lengua en estudio.

En el marco de lo dicho hasta aquí, cabe preguntarnos: ¿qué beneficios nos brinda a los investigadores del texto y del discurso la utilización de métodos asistidos por computadores? Opino que la respuesta a esta pregunta tiene múltiples alternativas. Algunas de ellas son:

- Accesibilidad a un análisis sistemático de grandes cantidades de textos de manera rápida y con alta confiabilidad, es decir, se fortalece la tecnologización de la investigación.
- Contrastación de hipótesis a través de evidencia empírica a gran escala, superando -en parte- los estudios de textos ejemplares.
- Validación de hallazgos preliminares (pequeñas muestras de textos) en *corpus* extensos.
- Acceso a tecnología que ahorra y acorta los tiempos de recolección, organización, marcación y análisis de las categorías a explorar.
- Aplicación de tecnología computacional a todos los niveles de la lengua: fonética, morfología, sintaxis, lexicología, pragmática, semántica y discurso.
- Posibilidades infinitas de exploración en textos etiquetados y no etiquetados.

Ahora bien, un principio empírico -extraordinariamente novedoso, pero si muy valioso- que guía esas investigaciones dice que en tanto mayor es la cantidad de información disponible para el análisis, mayor poder tendrán las conclusiones a las que se arriba. A

pesar de que en la actualidad gran parte de las investigaciones acerca del discurso se basan en textos auténticos, muchos de estos trabajos no se apoyan en bases de datos significativas en la línea de la lingüística de *corpus* (Sinclair, 1991; Stubbs, 1996; Biber, 1988; Church & Mercer, 1993; Reppen & Simpson, 2002; Rojo, 2002).

Biber, Reppen, Clark y Walter (2001) aportan cuatro ventajas importantes del enfoque de la lingüística de *corpus*. Ellas pueden ser expresadas de la siguiente manera:

1. La adecuada representación del discurso en su forma de ocurrencia natural a través de muestras amplias y representativas compiladas a partir de textos originales.
2. El procesamiento lingüístico (semiautomática) de los textos mediante el uso de computadores, lo que permite un análisis mucho más amplio y profundo de los textos mediante un vasto conjunto de rasgos lingüísticos caracterizadores.
3. Mayor confiabilidad y certeza en los análisis cuantitativos de los rasgos lingüísticos de grandes muestras de textos.
4. La posibilidad de contar con resultados acumulativos y altamente fiables. Posterioras investigaciones pueden utilizar los mismos *corpus* u otros *corpus* pueden ser analizados con las mismas herramientas computacionales.

Todo ello implica diseñar un *corpus* que sea medianamente aceptado como representativo, contar con programas computacionales automáticos para etiquetar morfosintácticamente los textos, disponer de una interfaz computacional poderosa de consulta en línea y utilizar programas estadísticos para determinar las relaciones entre las variables a estudiar. Es muy cierto que –particularmente para el caso del español- aún la tecnologización de la investigación es escasa y que tanto el desarrollo de programas computacionales como de tecnologías requeridas aún no se encuentran accesibles de manera expedita. A mi juicio, la superación de esta barrera metodológica y tecnológica no puede esperar si queremos, efectivamente, producir investigación científica competitiva y de primer orden. La superación de la brecha digital debe dejar de ser una utopía en nuestro medio, de modo que el acceso al conocimiento especializado esté disponible más democráticamente y nos permita posicionarnos a nivel internacional.

Desafortunadamente, son singularmente escasos los estudios acerca del español que indagan las características lingüísticas del discurso especializado escrito en nuestra lengua desde la perspectiva en cuestión. Resulta difícil encontrar investigaciones acerca del español en que se describe una amplia variedad de rasgos lingüísticos acerca de registros especializados a partir de grandes *corpus* de textos. De hecho, los estudios que comparan registros escritos en diversas áreas del conocimiento técnico-científico especializado son casi inexistentes; sobre todo, los que se enmarcan desde los principios de la lingüística de *corpus* (Parodi, 2005a y b; Parodi, Sabaj & Venegas, 2005).

LINGÜÍSTICA DE CORPUS COMPUTACIONAL: ¿POSIBLE FACULTAD PROBABILÍSTICA?

Como hemos venido comentando, lingüística, informática y estadística, entre otras disciplinas, constituyen hoy la denominada lingüística de *corpus* computacional o, para otros, lingüística computacional de *corpus*; ello dependiendo de donde se quiera poner el énfasis.

s. Ahora bien, existe un debate evidentemente no acabado acerca del *status* de esta posible subdisciplina de la lingüística o transdisciplina lingüística. La cuestión central está en si la lingüística de *corpus* se ha desarrollado como para alcanzar un grado de independencia tal que le permita constituirse en un nuevo paradigma o si ha de permanecer restringida como una potente metodología de investigación sin llegar a un rango mayor.

Por una parte, si nos posicionamos exclusivamente desde el nivel de la herramienta, innegablemente los aportes de esta lingüística de *corpus* son innovadores y brindan gran soporte para un número creciente de investigaciones cuyos resultados, entre otros, se capitalizan hacia la elaboración de gramáticas y materiales didácticos, la construcción de diccionarios, diversas aportaciones a la ingeniería lingüística, a las tecnologías del habla, a los sistemas de recuperación de información y también, por supuesto, para las investigaciones de interés lingüístico perse. Es oportuno hacer notar que la aceptación y adhesión a este cambio metodológico, de enorme importancia, acarrea dificultades o (pseudo)problemas que conviene tener presentes pues su consideración hará más potente sus desarrollos (Rojo, 2002).

También es pertinente la distinción entre un enfoque "basado en *corpus*" (*corpus-based*) y otro "guiado por el *corpus*" (*corpus-driven*). Según Tognini-Bonelli (2001), tal distinción compromete una aproximación en que las categorías de análisis están previamente determinadas y enmarcadas en una opción teórica y otra en que las categorías emergen del análisis y dan sustento a la construcción de una teoría decididamente guiada por los datos. En la primera opción, el *corpus* y las herramientas disponibles actúan como un método de investigación, indagación y corroboración de ideas preexistentes; en el segundo, el *corpus* es parte integral de la investigación y dará origen a un conocimiento totalmente nuevo no disponible previamente. Se hace evidente que lo que distingue a estos dos enfoques es una cuestión de objetivos, en el sentido del rol que cumple el *corpus* desde una y otra mirada. De este modo, el *corpus* puede cumplir el objetivo de validar, de exemplificar o de construir una teoría del lenguaje. En el primer caso, el objetivo es manejar una "metodología" ("basada en *corpus*") que permita poner a prueba o exemplificar teorías y descripciones ya formuladas, pero que buscan sustento en grandes muestras de textos digitales hoy disponibles. En el segundo caso, el lingüista busca ir más allá de los ejemplos para dar sustento a sus argumentos; así, desde el enfoque "guiado por el *corpus*" de la lingüística de *corpus*, la teoría no existe de manera independiente de la evidencia: la observación conduce a las hipótesis, las que conducen a las generalizaciones, las que conducen a la unificación de una afirmación teórica (Tognini-Bonelli, 2001:85).

Como siempre, toda investigación no está drásticamente obligada a discutir por los lineamientos de una y otra opción. Caminos intermedios, alternativos, también son factibles de explorar, tal es el caso de las investigaciones que damos cuenta en este libro (ver, entre otros, Capítulos 2 y 4).

Ahora bien, desde una mirada más ambiciosa, si se busca posicionar a la lingüística de *corpus* como una teoría explicativa de –al menos- parte del funcionamiento de la mente, las exigencias son mayores. De hecho, si se concibe el lenguaje humano como una facultad probabilística (Charniak, 1996; Manning & Schütze, 1999; Bod, 2003; Jurafsky, 2003) y se acepta el procesamiento estadístico del lenguaje natural como un modo de operar de la mente, nos encontramos frente a un paradigma emergente. Desde luego, se deberá decidir si su visión más radical, posiblemente anclada en concepciones conexionistas del

cerebro, con la consecuente negación de la mente con capacidad de representación simbólica del lenguaje es una alternativa plausible. Desde esta óptica extrema, la mente podría no existir y se restringiría el procesamiento lingüístico a una compleja red neuronal amparada en la metáfora de múltiples sistemas vectoriales interrelacionados.

Posturas intermedias, llamadas *híbridas* (Kintsch, 1998), parecen encontrar por ahora mayor acogida. Aunque el modo en que relacionan representaciones proposicionales simbólicas con modelos conexiónistas no están aún suficientemente explicitado (Parodi, 2003a).

Como es sabido, la noción de probabilidad proviene de la teoría matemática y se afincó en estudios lingüísticos a partir de la propuesta inicial de Shannon (1949), quien se basó en los modelos de cadenas ocultas de Markov (Manning & Schütze, 1999; Jurafsky & Martin, 2000). Desde este marco, los enfoques probabilísticos o estocásticos para el estudio del lenguaje humano operan básicamente a partir de técnicas estadísticas basadas en *corpus* de lenguaje natural (Oakes, 1998; Jurafsky & Martin, 2000).

Me ocuparé, entonces, en lo que sigue de este capítulo, de las características del discurso especializado, de las consecuencias más destacadas para una didáctica de este tipo de discurso particular y de sus proyecciones en investigación desde la lingüística de *corpus* computacional.

Una aclaración es válida: No es un error por parte nuestra ni una confusión ontologizada el que el uso en posición sintáctica específica del adjetivo *computacional*. La inclusión postmodificadora el nombre "*lingüística*" la hace muy diferente a la de su inscripción dentro de la frase preposicional complemento del nombre. Debe quedar claro que la adscripción a una "*lingüística de corpus computacional*" no reviste los mismos principios ni compromisos que una "*lingüística computacional de corpus*". Sin entrar en mayores profundidades, baste apuntar que la primera puede circunscribirse a un trabajo que preferentemente maneje textos digitales y se adhiera a un conjunto de principios empírico-metodológicos; en la segunda opción, el centro de la mirada proviene desde una lingüística computacional misma y parte de su material de trabajo son los *corpus* (obviamente digitales), pero su foco está en la creación de modelos computacionales del lenguaje y en el análisis automático de la comprensión y producción del discurso. Por ello, también es conocida como *ingeniería lingüística* o procesamiento del lenguaje natural (Parodi, Sabaj & Venegas, 2005).

DISCURSO ESPECIALIZADO: HACIA UNA CARACTERIZACIÓN

El término "discurso especializado" ha desplazado paulatinamente a otros que desde el terreno de las llamadas "lenguas para propósitos específicos" inicialmente dieron pautas para vislumbrar esta parcela de investigación de las lenguas naturales. Esta evolución terminológica no es una cuestión de moda. Resulta ser un cambio radical en la visión del lenguaje y de los dominios académico y profesional en el que se atiende a una óptica integradora y menos restringida exclusivamente a lo lingüístico. El posicionamiento y uso de este término, en mi opinión, se ha producido por su carácter más genérico y englobante en sustitución de otros más restringidos y menos ricos. Uno de los objetivos específicos que explican su aceptación y uso extendido es destacar su focalización en la lingüística del discurso y así dar cabida a este objeto multifacético de estudio como núcleo de las investigaciones en este terreno. Al mismo tiempo, se busca con esto desprenderse y deslindarse

de otras visiones más idealizadas del lenguaje humano y hacer sentir de este modo la preocupación por los textos en uso y la variabilidad inherente a ellos.

También me parece oportuno el uso del término "discurso especializado" ya que en sí mismo conlleva la idea de una gradiente que permite visualizar una variedad de textos que circulan en los ámbitos de especialidad. Este término apunta claramente a la heterogeneidad que los caracteriza, desde aquellos más altamente especializados (dirigidos exclusivamente a los especialistas), aquellos que podríamos catalogar como claramente de divulgación científica didáctica (para audiencias en formación) hasta otros claramente divulgativos en sí (enfocados al público general). Este aspecto tiene que ver definitivamente con la intención o propósito del hablante/escritor y el nivel de especialización de la audiencia a la que intenta dirigir su texto. Es, por tanto, el hablante/escritor quien elige y significa. Estos grados de especialización se materializan en diversos niveles de la comunicación especializada e imponen -entre otros- estructuras lingüísticas y terminología diferenciada. La mayor abstracción y hermetismo de un discurso especializado implica, como ya se insinuó, la decisión (consciente o no) por parte de un hablante/escritor de planificar un texto para un público específico y con un propósito particular.

La siguiente figura intenta captar esta idea de multiplicidad de textos de diverso grado de especialización a lo largo de una línea continua, en que se podrían disponer organizadamente según su cercanía o lejanía de los polos más prototípicos.

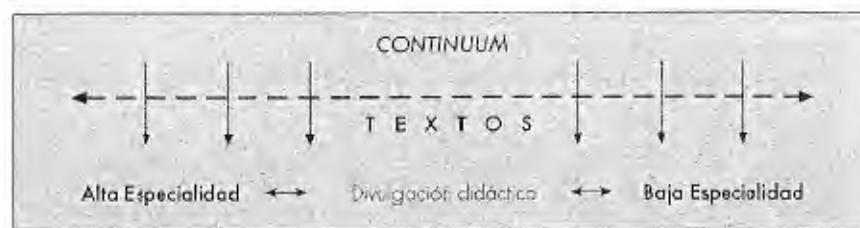


Figura 1. Gradiante diversificado de los textos de especialidad.

Como se aprecia en la Figura 1, la variedad de textos en el ámbito de especialización es grande y la determinación de sus límites -muchas veces- es un asunto definitivamente complejo. Las temáticas abordadas, las audiencias implicadas, los contextos de circulación y, entre otros, las configuraciones lingüísticas utilizadas para vehicular los significados textuales, son todas variables interviniéntes que se conjugan de manera certera en la construcción de un discurso especializado. Por ello, queda claro, que la clasificación apriorística y/o dicotómica de si un texto se debe considerar como de especialidad o no especialidad o de tipo más general es -por supuesto- un problema de envergadura teórica y metodológica. La precisión exigida a las categorías a partir de las cuales debería organizarse el sistema de oposiciones, rasgos discretos y valores se vuelve, por un lado, una gran dificultad pero, a la vez, un interesante desafío. Durante los últimos años, lingüistas de diversas ramas han buscado construir taxonomías complejas que logren dar cuenta de la multiplicidad de vo-

riables que se conjugan y explican la constitución de un texto especializado dentro del *continuum* de posibilidades. La tarea no ha sido sencilla, pero existen logros destacables [Charaudeau, 2004].

Tal como muestra la Figura 1, hoy en día acordamos, consensuadamente, en favor de un *continuum* de textos que se distribuyen de manera progresiva desde un dominio altamente especializado hacia otro extremo mucho más divulgativo y general [Gläser, 1982; Schröder, 1991; Halliday, 1993; Jeanneret, 1994; Perani, 1998; Ciapuccio, 1994, 2000, 2003; Cabré, 2002; Goldman & Bizon, 2002; Gott, 2003]. Esta idea de categorías de límites difusos (*Fuzzy Categories*) fue inicialmente propuesta por Lotoff (1972) y proviene de lo difícil que resulta aceptar en la realidad la idea de categorías estrictas y claramente dicotómicas entre un dominio y otro. Es un hecho que establecer límites precisos entre un tipo de texto y otro es una cuestión de envergadura. Sin importar el foco atencional en uno u otro criterio clasificatorio, siempre existirán casos mixtos o límiteles, pero -en la mayoría de los casos- el discurso especializado resulta ser una clase textual más o menos reconocible para cualquier integrante de la comunidad discursiva en la que éste se inscribe.

No obstante lo anterior, también es posible distinguir un eje vertical [no estrictamente en el sentido de Cabré, 1993, 1999, 2000] que se suma al horizontal, presentado en la Figura 1. En esta nueva mirada ampliada y más rica, ubicamos el grado de formalidad y la modalidad de lengua (oral o escrita). En la Figura 2, se grafica estas múltiples relaciones complejas que coexisten en el *continuum* de textos que caben clasificar como especializados.

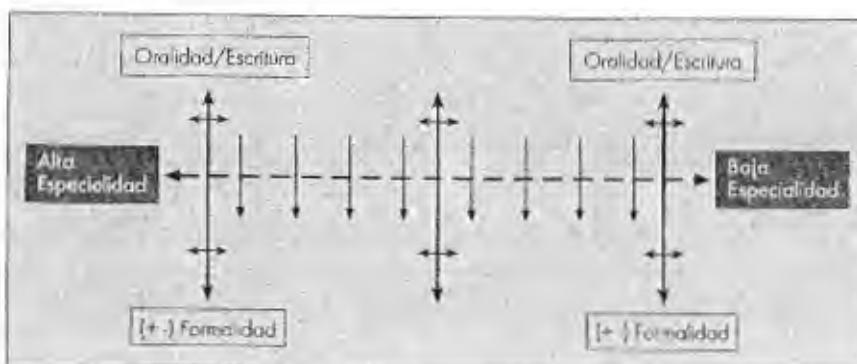


Figura 2. Relaciones horizontales y verticales en el discurso de especialidad

Como se visualiza en esta presentación diagramática, la variación discursiva al interior de los textos de especialidad es -como se ha dicho- compleja y multiforme y genera polémica de alto interés. La lengua, en su riqueza y variedad, está constituida por una amplia variedad de textos que pueden agruparse en subconjuntos de códigos más constreñidos, en los cuales los escritores/hablantes emplean de manera (muchas veces inconsciente) determinadas organizaciones discursivas y formas lingüísticas, haciendo selecciones adaptativas al tenor de sus necesidades y de los rasgos específicos de los contextos

comunicativos en que se encuentren. La distinción especializado/general (eje horizontal) debe conjugarse con la del registro (formal/informal) y con la modalidad escrita u oral de la lengua (eje vertical), los que en ningún caso tampoco presentan alternativas dicotómicas. Lo que trata de capturar el eje vertical es la idea de que, en un dominio especializado, puede inscribirse un texto ya sea oral o escrito, cuyo grado de formalidad puede ser variable de acuerdo a múltiples factores. Aunque existen antecedentes (ver Capítulo 4) que avalan una postura en favor de una fuerte relación prototípica entre alta especialización y modalidad escrita de la lengua, es cierto que también se puede encontrar textos especializados menos formales y orales; esto es, específicamente, lo que intenta mostrar la Figura 2.

Ahora bien, de manera más circunscrita, también he optado por el término discurso de especialidad o especializado, pues me interesa el discurso de la ciencia y de la tecnología en contextos institucionales de tipo educativo en donde, hipotéticamente, puede detectarse un conjunto de textos en una pendiente hacia lo más divulgativo (Gläser, 1982; Martín, Christie & Rothery, 1987; Halliday & Martín, 1993; Rose, 1997; Christie, 1998; Goldman & Bizon, 2002; Cameron, 2003; Cabré, 2002; Ciapuccio, 2003). Desde esta óptica, coincido con diversos investigadores para quienes el discurso de especialidad se conforma por un conjunto de textos que involucran tópicos que son prototípicos de un área determinada del conocimiento científico y tecnológico y se ajustan a convenciones particulares de índole lingüística, funcional y situacional, entre otras, provenientes de las tradiciones de la comunidad discursiva específica (Brinker, 1988; Schröder, 1991).

Delimitar -de modo certero- algunas de las características constitutivas del discurso especializado es complejo, pero es un asunto imposible de evadir. Una cuestión relevante es tener presente que para considerar un texto como especializado siempre deben concurrir un conjunto de variables; no basta con una definición operacional desde una sola esfera. Algunos de estos principios caracterizadores que determinan un discurso especializado prototípico, no necesariamente en orden jerárquico, son:

- Rasgos lingüísticos específicos de una variedad de lengua natural que co-ocurren sistemáticamente en determinados textos y no del mismo modo en otros.
- Rasgos constitutivos que van más allá de lo meramente léxico-sintáctico: se incorpora la semántica, la pragmática y todos los rasgos comunicativos del español como un sistema semántico complejo.
- Distinción de funciones comunicativas predominantes, específicamente distintivas e idiosincráticas, que se activan dentro de la misma lengua particular a través de organizaciones lingüísticas y no lingüísticas concurrentes sistemáticamente; por ejemplo, la referencialidad.
- Temáticas abordadas y léxico empleado son para construir significados exactos, restrictivos y cerrados en dominios determinados y acotados (monorreferencialidad, precisión); lejanos al más típicamente cotidiano.
- Coexistencia de una diversidad de tipos de textos que exhiben grados variados de especialidad, constituyendo un *continuum* en el que se incorporan textos de mayor especialización hasta algunos de franca divulgación.
- Contextos de uso de circulación de los textos muy particulares, es decir, se detectan rasgos situacionales identificables.

- Exigencia de conocimientos previos compartidos en áreas del conocimiento disciplinar, dado que se transmite un saber altamente específico
- Predominio de organizaciones discursivas o retóricas, tales como clasificación, definición, listado

Desde esta óptica, defino el discurso especializado, por una parte, como un conjunto de textos que se distinguen y se agrupan por una co-ocurrencia sistemática de rasgos lingüísticos particulares en torno a temáticas específicas no cotidianas en los cuales se exige experiencia previa disciplinar de sus participantes (formación especializada dentro de un dominio conceptual particular de la ciencia y de la tecnología); por otra, son textos que revelan predominantemente una función comunicativa referencial y circulan en contextos situacionales particulares: todo ello implica que sus múltiples rasgos se articulan en singulares sistemas semióticos complejos y no de manera aislada y simple.

En este sentido, no estoy de acuerdo con Charaudeau (2004) en cuanto a que los rasgos lingüísticos formales no podrían caracterizar un género de manera certeza ya que estarían presentes en otros tipos de discursos paralelamente. El argumento de Charaudeau puede ser cierto, pero solo parcialmente. Lo que hace que los rasgos formales sí puedan constituirse en identificadores importantes de un tipo de discurso es su co-ocurrencia sistemática a través de las diversas tramas de un texto. Esto quiere decir que no basta con que, por ejemplo, nominalizaciones y participios en función adjetiva (ver Parodi y Venegas, Capítulo 9) aparezcan en un determinado tipo de texto para caracterizarlo como tal o cual, sino que se debe explorar la sistemática concurrencia de estos y otros rasgos de manera no aleatoria en un texto para que su significativa aparición nos indique algo acerca del tipo de texto que pretenden identificar. Charaudeau (2004) está en el camino correcto al estimar que deben manejarse varios tipos de criterios para determinar un género discursivo, y esto es justamente lo que venimos argumentando nosotros en este capítulo. No obstante ello, sostenemos que los rasgos formales con significatividad funcional co-ocurrentes sistemáticamente en un texto son un aporte sustancial en el estudio de, por ejemplo, el discurso especializado.

Coincidientemente, según Sinclair (1996), el discurso especializado está determinado por principios de limitación, restricción, selección y simplificación. Ellos permiten especificar de manera decisiva la distinción con un discurso general que se aplica en contextos cotidianos y busca la expresión y la satisfacción de otros propósitos comunicativos predominantes, fuera de la esfera de lo más típicamente referencial. En este mismo sentido, estoy de acuerdo con Gotti (2003) en cuanto a que algunos rasgos relevantes del discurso especializado lo constituyen la monorreferencialidad, la carencia o escasez de emoción, la precisión y la concisión. Justamente ello se revela a través de la ocurrencia predominante y sistemática de ciertas características lingüísticas específicas, tales como las que señalaremos en los capítulos posteriores.

Naturalmente, las relaciones que se generan entre las áreas temáticas, los tipos de textos que circulan en contextos institucionales y el grado de especialización de los materiales escritos que circulan a lo largo del *continuum*, tanto en su eje horizontal como vertical (Figura 2), son un foco particular de interés para comprender las tensiones propias del ámbito. Si consideramos solo el criterio de accesibilidad propuesto por Lassen (2003),

debería existir en cada ámbito especializado una heterogeneidad en los tipos textuales y mayor diversidad en cuanto a la relación entre aquellos altamente especializados y otros divulgativos con el fin de considerar una cierta graduación jerárquica en los contenidos técnico-científicos, hecho que permitiría en definitiva un mejor acceso al conocimiento científico, propio de cada ámbito.

El esquema siguiente pretende captar las relaciones dialécticas que se producen entre estos tres ejes, con direcciones divergentes entre unas y otras.



La tensión que se revela entre el ámbito de especialidad y el *continuum* de discursos, entre los altamente técnico-científicos hasta los de mayor procesamiento reformulativo que tienen como objetivo regular la entrega de información y, por ende, aportar a la transparencia del conocimiento especializado se constituye en una relación bidireccional de extrema importancia. Por otra parte, la proyección unidireccional en atención a audiencias y contextos más restringidos y exclusivos entre expertos y aquellos de público más amplio y diverso es lo que permite la paulatina incorporación a las comunidades discursivas objetivas, por medio de materiales iniciales e idealmente más transparentes hasta aquellos progresivamente más especializados y de –supuestamente– mayor opacidad.

En este contexto, la respuesta a la pregunta si existe un español de especialidad es en mi opinión negativa. Sostengo que si existen usos específicos de español escrito especializado determinados en gran medida por los lectores y escritos involucrados y por los contextos de uso, a diferencia de usos orales o escritos no especializados. El discurso especializado constituye un modo particular de uso en que determinados recursos léxicos y gramaticales de la lengua española se seleccionan, organizan y distribuyen con cierta regularidad. Los rasgos lingüísticos rara vez co-ocurren aleatoriamente en los textos. Si ciertos rasgos co-ocurren consistentemente, entonces, es razonable indagar una influencia funcional subyacente que determina dicho uso particular.

Distinguir entre discurso especializado y discurso general puede –bajo una primera mirada– no resultar nada fácil; es más, se puede visualizar como una problemática altamente compleja y polémica. Todos quienes nos hemos adentrado en esta distinción hemos sufrido más de una crisis en búsqueda de criterios ideales y orientadores que permitan trazar una clara línea divisoria entre textos de especialidad y textos generales. Por una parte, como ya hemos indicado, hoy en día está suficientemente claro que la distinción dicotómica entre uno y otro discurso no se aplica; por otra, como se ha venido argumentando, la definición de uno u otro tipo de discurso no radica exclusivamente en cuestiones estrictamente lingüísticas, pues los aspectos pragmáticos o extralingüísticos constituyen un eslabón fundamental en la complementación de criterios consensuados, no materializados aisladamente.

Buscando precisar aún más la idea de discurso especializado, tomó la noción de "síndrome" acuñada por Halliday (1993) para mostrar que a través de grupos de rasgos co-ocurrentes de variación se puede identificar un tipo de registro determinado. En palabras de Halliday (1993:4):

"Cualquier variedad del lenguaje, sea funcional o dialectal, ocupa un espacio extendido, una región cuyos límites son difusos, dentro de los cuales puede existir una variación considerable. Pero puede ser reconocida y definida por ciertos síndromes, patrones de co-ocurrencia entre rasgos en uno u otro nivel lingüístico - típicamente rasgos de la expresión en el caso de un dialecto y rasgos de contenido en el caso de una variedad funcional del registro. Estos síndromes son los que hacen posible hablar de un 'lenguaje de la ciencia'."

Dentro de los discursos de especialidad, los síndromes de rasgos lingüísticos varían en cada caso para satisfacer las necesidades de los hablantes al construir nuevos significados. Estos rasgos pueden ser pesquisados a través de las ya clásicas variables del registro (campo, tenor y modo), cuyas dimensiones son relevantes ya que ellas, según Halliday (1985, 1993), tienen impacto en el uso lingüístico, en el modo en que usamos el lenguaje, esto es, producto de las variaciones en el contexto situacional.

Uno o solo un par de rasgos lingüísticos por sí solos no son capaces de dar cuenta de las características de un determinado registro. Por ello, la indagación de rasgos gramaticales y estructuras lingüístico-textuales que presentan claras implicancias de tipo funcional, comunicativo y cognitivo se vuelve una cuestión fundamental. Son estos rasgos y sus respectivos análisis en cuanto a su poder explicativo conjunto los que se constituyen en poderosas fuentes que permiten posteriormente la caracterización más acotada de ciertos registros altamente especializados o más divulgativos.

Desde una visión estadística y funcional, pero enfocado hacia el mismo principio, los estudios de Biber (1986, 1988) han enfatizado el estudio de la variación lingüística a través de diversos registros orales y escritos y lo han abordado desde el denominado Análisis Multirrasgos (AMR) y Multidimensiones (AMD). Para Biber la exploración de un amplio conjunto de rasgos lingüísticos a través de grandes corpora de textos diversificados, marcados y etiquetados morfosintácticamente, es una opción para describir y comparar registros. Lo que se persigue -inicialmente- es la construcción de una matriz estadística a partir de las frecuencias de ocurrencia de aquellos rasgos seleccionados que co-ocurren sistemáticamente en cada corpus. Para ello, se opera por medio de técnicas estadísticas multivariadas (Hair, Anderson, Tatham & Black, 2001). Los análisis estadísticos reducen el total de rasgos en estudio a determinados grupos específicos que han revelado principios de co-ocurrencia particular. Cada agrupamiento de rasgos lingüísticos (factor), emergido en base a estos procedimientos y luego interpretadas funcionalmente, es lo que Biber denomina "dimensión". Esto quiere decir que una dimensión se constituye mediante la interpretación funcional del conjunto de los rasgos co-ocurrentes sistemáticamente en textos de un corpus (factor), agrupados mediante análisis factorial (Parodi, 2005b).

El principio de la co-ocurrencia de rasgos lingüísticos y su relación con funciones subyacentes compartidas constituye un supuesto básico para el análisis multidimensional. En este tipo de análisis los patrones de co-ocurrencia de rasgos se interpretan en términos de fun-

ciones situacionales, sociales y cognitivas comunes, pues se asume que la ocurrencia de los mismos rasgos lingüísticos en textos determinados se debe a que cumplen funciones compartidas específicas. En este sentido, se entiende que un grupo de rasgos co-ocurren frecuentemente en ciertos textos porque ellos expresan un conjunto de funciones comunicativas determinadas.

Ahora bien, no cabe duda que la investigación en torno al discurso especializado, en particular, aquella que se centra en sus diferentes modalidades más divulgativas científico-tecnológica (Glaser, 1982; Calzamiglio, 2000; Clopuscio, 1994, 2000, 2003; Cassany, López & Martí, 2000; López, 2002; Gunnarsson, Linell & Nordberg, 1997) y en su uso al interior de la sala de clases en el nivel secundario y superior, constituye un elemento nuclear y estratégico en la planificación lingüística y en el proceso de alfabetización progresiva de estos grupos sociales en formación (Christie & Martin, 1997; Vee, 1998; Otero, León & Graesser, 2002; Biber, Reppen, Clark & Walter, 2001; Cameron, 2003). La indagación de los rasgos identificatorios de diversa índole de este tipo de discurso es un ámbito fructífero de indagación desde una opción semiótica compleja.

Desde estas ideas, movámonos hacia el modo en que los discursos especializados se relacionan con las comunidades discursivas y los mundos posibles del hombre.

■ COMUNIDADES DISCURSIVAS Y/O MUNDOS POSIBLES?

El término "comunidad discursiva" es ya bien conocido hoy en día en el ámbito de los discursos especializados. El cuño de John Swales avanzó un concepto que a la luz de lo expuesto hasta ahora se hace necesario e interpretar a muchos que abogábamos por superar visiones ancladas en otros paradigmas lingüísticos. De las seis características clásicas y definitorias de las que Swales (1990) se vale para conceptualizar su propuesta, ninguna resulta prescindible.

Así, una comunidad discursiva:

1. Tiene un amplio conjunto de objetivos públicos comunes en que está de acuerdo.
2. Tiene mecanismos de intercomunicación entre sus miembros.
3. Usa sus mecanismos participativos esenciales para proveer información y retroalimentación.
4. Utiliza y, por ende, posee uno o más géneros en el logro comunicativo de sus objetivos.
5. Aparte de los géneros que le son propios, ha adquirido un léxico específico.
6. Tiene un nivel de umbral de miembros con un grado adecuado de contenido relevante y experiencia discursiva.

A pesar de esta rica visión entregada por Swales, es indiscutible que su conceptualización carece de algunos componentes vitales en el marco de las ciencias transdisciplinarias que buscan una concepción del hombre, en cuanto persona, mucho menos homogeneizante. Desde allí, quisiera proponer -complementariamente- la noción de "mundo posible" como una extensión y enriquecimiento de la concepción propuesta por Swales.

Siguiendo a Gómez Macker (1998), entendemos por "mundos posibles" aquellas maneras globales y relativamente coherentes que poseemos las personas a modo de marcos de referencia para interpretar la realidad percibida o inventada y que se apoyan tanto

en la naturaleza de la realidad como en los procesos mentales creativos de cada ser humano particular.

A pesar de la necesaria incorporación a una comunidad discursiva por parte de cada sujeto, se debe reconocer que esta área involucra aspectos insospechados propios de la naturaleza humana. La tendencia moderna a parcelar la realidad con el propósito ingenuo de estudiarla o comprenderla mejor fuera de su contexto, puede estar atomizando sin discriminación diversas dimensiones de la realidad. La construcción de mundos posibles exige el consenso, pero también el respeto a la diversidad y a la interioridad propia de cada sujeto.

De las variadas consideraciones respecto a los mundos posibles del ser humano (Gómez Macker, 1998), destacamos las siguientes:

1. Las visiones de mundo son esquemas mentales globales de la percepción de la realidad que las personas construyen y mediante los cuales interpretan, ordenan y valoran lo específico.
2. Toda persona adulta posee por lo menos una visión de mundo.
3. Toda persona adulta tiene la necesidad y el derecho a poseer su(s) propia(s) visión(es) de mundo(s) de acuerdo a sus percepciones, prejuicios y valores.
4. Cada visión de mundo se configura de acuerdo a las capacidades psíquicas de las personas.
5. Es normal que entre las visiones de mundo de un individuo y otro existan diferencias e, incluso, discrepancias más o menos profundas.
6. Las visiones de mundo de cada persona no son fijas ni definitivas.

Rescatamos de esta propuesta los aspectos más diferenciadores entre los seres humanos y damos por supuesto aquellos en que Swales y Gómez Macker coinciden en cuanto a lo elaborado compartidamente por el contacto interpersonal en el contexto de lo social y de la cultura. Desde esta óptica, no se argumenta que Swales niegue los aspectos apoyados por Gómez Macker, sino que muy probablemente los da por supuestos. No obstante ello, me parece que la idea de comunidad discursiva se ve enriquecida al tener presente las restricciones que la propia naturaleza humana impone a las construcciones de visiones estrechamente compartidas. Estas dos cuestiones, aparentemente antagónicas, coexisten simultáneamente en el ser humano: lo social y lo individual. Conviene tener presente ambas ideas al momento de estudiar los discursos de especialidad pues se hace evidente que ellos no develan un conjunto de principios total y absolutamente homogéneos.

Las restricciones impuestas a la visión de Swales, también coinciden con lo que Ivanic (1998) argumenta en torno a la construcción de la identidad del sujeto a través de la escritura. La autora también define un punto de vista más relativo respecto a la construcción socialmente determinada y propone considerar los diferentes "yo" que conviven en todo ser humano.

Cuestionar la posibilidad de una construcción discursiva extrema del sujeto en contextos institucionales parece un asunto sensible pero absolutamente viable. Se debe resaltar que durante el progresivo proceso de construcción de todo sujeto discursivo y de su paulatina integración y aporte a algunas de sus comunidades y/o mundos posibles se debe tener presente que ello solo puede llegar a concretarse desde las múltiples identidades internas del "yo".

Todo esto apunta, en mi opinión, al concepto ya discutido de representatividad de un *corpus* y obliga a ser cuidadoso al momento de construirlo y generalizar extensamente a partir del mismo en cuenta a la homogeneidad de los rasgos y sus repercusiones sobre los sujetos discursivos involucrados. Se hace necesario relativizar adecuadamente los hallazgos emocionales surgidos de los enfoques lingüísticos computacionales que comentamos y distinguir lo más nuclear de lo accesorio. Evidentemente lo anterior no implica un debilitamiento de los resultados alcanzados ni de las posibles proyecciones, sino que más bien nos hace estimarlos en una medida ajustada.

Desde este escenario, estimo que la idea de los mundos posibles del hombre brinda una visión más integral y realista de la constitución e integración a las comunidades discursivas especializadas y nos hace ver un abanico de alternativas más productivas. Si son diversos los mundos posibles en que nos movemos los seres humanos (incluso, algunas antagonistas entre sí mismas), es obvio que podemos integrar múltiples comunidades -potencialmente virtuales de manera paralela a lo largo de nuestras vidas, abandonar algunas y paulatinamente profundizar en otras, pero también debemos relativizar la construcción homogeneizante de esos mundos o comunidades.

DESARROLLO DE COMPETENCIAS ESPECIALIZADAS

A modo de diagnóstico

A continuación me referiré a algunas cuestiones que dicen relación con el desarrollo de habilidades lingüísticas en el marco de los trabajos en discurso especializado a nivel técnico-científico secundario, a la luz de lo que denominamos *estadilingüística de corpus computacional*.

Un modo de acercarse al registro especializado escrito es partir del supuesto de que los textos escolares técnico-científicos a que se ven enfrentados los estudiantes en su práctica escolar diaria reflejan tanto los tipos de textos como los rasgos representativos de esa variedad de lengua especializada. En la actualidad los investigadores del discurso conocemos relativamente poco acerca de los patrones de uso lingüístico que son característicos de los registros especializados, sobre todo de los escritos en español. Como es fácil comprobar, la mayor parte de la investigación lingüística en español se ha concentrado en esferas que han estado ajena al rol más funcional y comunicativo de la lengua escrita y se han orientado clásicamente hacia una descripción de índole más bien gramatical pura.

En esta línea, me preocupa específicamente la escasa información disponible actualmente en nuestro país acerca de los textos que circulan en las comunidades especializadas en formación. Me parece urgente -en este contexto- poner de relieve que en Chile no conocemos los tipos de textos que circulan en el sistema educacional secundario técnico-profesional diferenciada. Tampoco contamos con información, aunque fuera preliminar, acerca de los rasgos lingüísticos y textuales que caracterizan dichos textos escritos. Como es obvio, es nula la información disponible acerca de las relaciones entre diversas estructuras o rasgos lingüísticos de estos textos especializados y las implicancias cognitivas que ellos llevan. Del mismo modo, es claramente inexistente una descripción del perfil de competencia en lengua escrita especializada (comprensión y producción lingüística) de los sujetos que egresan de estas instituciones técnico-profesionales.

Como se desprende, si la (psicollingüística de corpus computacional puede contribuir a ello, una tarea primordial ha de ser la recopilación y descripción de los textos que se emplean en estos ámbitos y de sus rasgos estructurales y funcionales. Para, desde allí, aproximarnos con instrumentos válidos a conocer las capacidades de lectura y escritura especializada.

En la Figura 3 se intenta dar cuenta del modo asimétrico en que ciertas relaciones operan actualmente en el sistema educativo técnico-científico. Por asimétrico entiendo que lo que aparece en el círculo mayor es lo que efectivamente sucede y, en el menor, lo que idealmente debiera ocurrir; ello se refuerza con las flechas que van desde lo que opera hoy en día hacia el ideal al que debiéramos propender.



Figura 3. Relaciones asimétricas en la planificación lingüística

Según se infiere de lo anterior, el desconocimiento y falta de accesibilidad a información pertinente que posibilite una planificación lingüística del discurso especializado es un fenómeno extendido en muchos establecimientos educacionales profesionalizantes. El énfasis detectado en este tipo de sistema educativo no parece radicar en una clara y organizada planificación progresiva desde textos más transparentes y que han sido objeto de intervención divulgativa hacia otros más altamente especializados; todo esto con el fin de potenciar conocimientos profesionales que apoye el desarrollo del estudiante. Elio se explica -en parte- porque no se tiene conciencia de los tipos de textos que efectivamente se emplean ni se cuenta con una jerarquización didáctica de ellos. Del mismo modo, tal como ya se ha señalado, desconocemos las posibles relaciones entre determinadas estructuras lingüísticas -típicas de estos textos- y sus exigencias de procesamiento cognitivo, hecho que impide contar con herramientas certeras que apoyen la progresión ideal que comentamos.

En mi opinión, casi está de más insistir en el tipo de relación asimétrica que está operando en el sistema educativo institucional. Según se muestra en la Figura 3, si la información requerida no es posible, por el momento, una planificación lingüística adecuada.

Ante esto, urge la necesidad de contar con investigaciones que aporten al conocimiento de los textos que circulan en estos ámbitos y su caracterización desde diversos ópticas. Es cierto que hoy en día existen numerosas investigaciones en curso, pero se debe hacer un

esfuerzo para que los resultados, por una parte, alcancen un nivel de aceptada representatividad y, por otra, que sus aportaciones se encuadren en líneas más o menos homogéneas de modo que sus datos sean comparables y permitan construir un panorama relativamente congruente. Estos datos serán soporte imprescindible para cualquier tipo de trabajo posterior de corte aplicado que se quiera implementar. Una alternativa metodológica como la aquí propuesta en el marco de la lingüística de corpus computacional es un medio para contar con resultados científicamente comparables. Otra cuestión será si, una vez construido el panorama, los datos se considerarán por las instancias pertinentes.

A continuación, en la Figura 4, se grafica el modo en que se implementan en la actualidad, en muchos establecimientos educativos especializados, los mecanismos de alfabetización de las comunidades técnico-científicas.

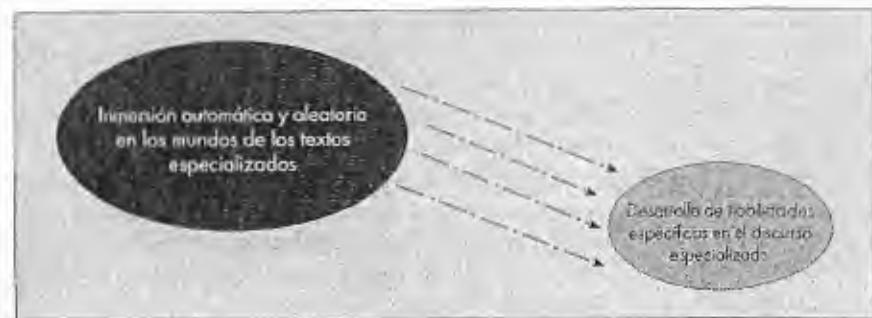


Figura 4. Inmersión automática en la comunidad discursiva y los mundos especializados del hombre

Lo anterior indica que el procedimiento de aproximación a los textos especiales se realiza de manera indirectamente proporcional a lo que podríamos esperar. Esto quiere decir que existe una postura en favor de una inmersión casi automática y a través de textos aleatorios. Todo ello revela una carencia de métodos que progresivamente devuelven el conocimiento disciplinar junto al pertinente desarrollo de habilidades lingüísticas que apoyen la transición desde una comunidad lega a una semi-especializada. Al respecto, contamos con información que indica que -en la mayoría de los casos- a los sujetos se les sumerge directamente en textos de alta especialización sin instrucción previa específica (muy posiblemente, ello acontece de modo no planificado conscientemente por parte del docente especialista, ya que no dispone del tipo de conocimiento que aquí discutimos).

Obviamente el modo de proceder que grafica la Figura 4 implica una dificultad enorme en el procesamiento lingüístico de los textos escritos y restringe su posible utilidad como materiales que apoyen su acceso al conocimiento especializado. De este modo, se bloquea el posible carácter introductorio del material iridíctico que posibilita la integración paulatina de los sujetos en las correspondientes comunidades discursivas y se les impulse, de este modo, hacia nuevos mundos posibles.

Por su parte, el profesor de lengua materna, en una labor paralela a la del especialista

técnico-científico, no cuenta ni con los conocimientos ni con medios disponibles para apoyar una paulatina aproximación a textos con características particulares, ya que el docente de lengua los desconoce porque en su formación inicial no recibió entrenamiento alguno en ello. Al mismo tiempo, la inexistencia de nexos entre el docente de lengua y el docente especialista tampoco posibilita que el primero comprenda los problemas del segundo y pueda buscar medios para desarrollar estrategias didácticas paliativas.

Ante la imposibilidad de comprensión de estos materiales por parte de los alumnos de especialidad, debido al tipo de relaciones asimétricas y no enfocadas en el correcto sentido, se niega al alumno la asimilación de los conocimientos técnicos y científicos y se le impide una construcción del mundo de la ciencia y la tecnología, herramienta vital para su desempeño profesional futuro.

La focalización en uno u otro polo ha sido inestable y, por ende, la direccionalidad de los esfuerzos ha estado desequilibrado. Desde esta óptica, se está propendiendo a la formación de técnicos y profesionales literalmente analfabetos en el manejo especializado de los materiales escritos y, por ende, se está creando un daño a su desempeño laboral eficiente. En palabras de algunos docentes especialistas, este proceder está produciendo futuros profesionales cesantes, que no califican para el mundo empresarial.

Así, el docente especialista -según confiesan muchos- se ve en la imperiosa necesidad de comentar oralmente los textos escritos especializados y hacer de "traductor" de éstos para sus alumnos. También se dice que estos docentes acaban confeccionando resúmenes extremadamente sintéticos de los contenidos de los libros especializados con el fin de que sus alumnos asimilen una parcialidad del total por abordar.

En la siguiente figura, se da cuenta de la focalización y tipo de habilidad psicolingüística que se detecta en el proceder de los docentes tanto de lengua materna como de especialidad.

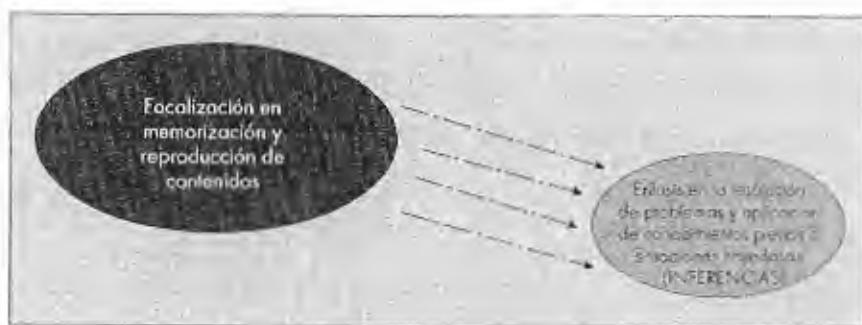


Figura 5. Énfasis en habilidades de menor rango

Como se aprecia en esta figura, los fundamentales procesos inferenciales -núcleo de un procesamiento que conduzca a una correcta comprensión lingüística- no son foco particular del trabajo con textos escritos en los avales de especialización. Se privilegian los ya clásicos procedimientos de menor jerarquía psicolingüística y no se potencia el razona-

miento y creatividad a partir de la lectura de materiales técnicos. Esta tradición, que ya ha sido evidenciada por numerosas investigaciones en diversos ámbitos del sistema educacional (Péronard, Gómez, Parodi & Núñez, 1998; Parodi, 2003a), debe ocupar un lugar menos enfatizado en el desarrollo de habilidades lingüísticas de alta exigencia cognitiva, como son la comprensión y producción de material escrito. Por cierto que los procesos de memoria están a la base de otros procesos de mayor envergadura como son las inferencias, por ende, no se critica su inclusión en el currículo, sino su exagerado énfasis y, a veces, exclusiva focalización.

Es muy cierto que los procesos inferenciales en lengua española han sido escasamente indagados a través de tipos textuales especializados. Por ende, desarrollar una capacidad inferencial efectiva con el fin de construir la coherencia de base de un texto con características especiales ha de ser una tarea inminente.

Los tipos de significados científicos que deben explotarse en los alumnos se pueden explorar a través de los modos de evaluar el trabajo académico que los docentes especialistas ejecutan. Cuando se pregunta a los docentes por las expectativas que tienen de sus alumnos, emergen muchos conceptos tales como "habilidad para generalizar", "relaciones de causa y efecto", "terminología técnica", "relaciones lógicas". La evidencia de los tipos de evaluaciones que ellos efectúan y de lo que también aparece presente en los manuales y guías didácticas indica una concentración en técnicas de escaso relación con estos conceptos; de modo que las actividades de explotación de los materiales no apunta en la línea de lo deseado.

No es de extrañar lo detectado. No existe una formación que aporte a la generación de materiales didácticos especializados que oriente a docentes que mayoritariamente provienen de profesiones ajena a la docencia. Y, aunque no fuera el caso, bien sabemos que no contamos con un detallado organigrama de principios metodológicos consensuados que conduzcan claramente hacia una didáctica lingüística de los textos escritos de mediano y alta especialización en diversos ámbitos del conocimiento técnico-científico.

Existe evidencia que muestra que los manuales utilizados en esta formación técnico-científica enfatizan un tipo de "conocimiento a través del hacer" y que escasamente se focalizan en un "conocimiento a través de la lectura". Tanto la información aportada por los docentes especialistas como por el análisis de las instrucciones de los manuales especiales se detecta que un grado importante de las actividades en ellos contenidas apunta a la memorización y reproducción de información más que a las explicaciones fundadas, razonamientos convergentes y divergentes, y discusiones con aplicación de conocimientos previos atípicos. Son escasas o nulas las actividades en que se impone a los alumnos a efectuar procesos inferenciales con fuerte énfasis en la resolución de problemas y aplicación de conocimientos previos a situaciones innovadoras.

Existe amplio consenso hoy en día entre los especialistas en cuanto a que las pruebas de competencia lingüística a que se somete a una determinada población deben atenerse a la variedad de uso lingüístico que esa comunidad discursiva emplea. Similarmente, si se busca evaluar el desempeño lingüístico de estudiantes secundarios de formación diferenciada técnico-profesional en su comprensión de la variedad de lengua escrita especializada, se debe contar con una descripción tanto de los tipos de textos que circular en ese sistema como de los rasgos lingüísticos característicos de los registros implicados.

Afirmar que el desarrollo de las habilidades especializadas en lectura y escritura de textos técnico-científicos constituye un eje fundamental en el acercamiento de estos profesionales a sus comunidades discursivas y, por ende, en su incorporación al mundo laboral es posible que sea extremadamente obvio. En particular, dado que en Chile aún no alcanzamos los niveles esperados en el manejo del lenguaje escrito en el ámbito de la educación general primaria y secundaria, podría suponerse postergable la preocupación por la alfabetización técnico-profesional. Desde mi punto de vista, tal suposición es evidentemente errónea. Cualquier intento por apoyar la participación como miembros activos de estos profesionales en sus comunidades discursivas está determinado por un manejo eficiente del discurso especializado a través del cual se genera, se transmite, se disemina y se adquiere el conocimiento de cada ámbito. El conocimiento acabado de los textos que se generan y que se emplean tanto en esos ambientes educativos profesionales como en las organizaciones e instituciones laborales es prioritario para un proyecto de educación técnico-científico secundaria.

Desafortunadamente, el panorama que describo no es alentador, pero estoy seguro que ante ello debemos adoptar una actitud positiva y proactiva. Estimo que -muy por el contrario de la desilusión- su adecuado conocimiento debe llevarnos a una suerte de encantamiento. Estos datos deben ser las llaves que se vuelvan estímulos alentadores que nos impulsen a buscar caminos creativos que permitan que el conocimiento de los textos especializados se haga accesible a los usuarios para, por un lado, asegurar su efectiva incorporación a las comunidades de especialistas y, por otro, lograr transformarlos en sujetos independientes que sean capaces de leer y comprender autónomamente los materiales que los insertarán de manera definitiva en su futuro laboral y les permitirán ser profesionales idóneos.

ALGUNAS PROPUESTAS

A la luz de este diagnóstico no sistemático, entregamos once claves, a modo de propuestas estratégicas, que apuntan hacia una didáctica del discurso especializado escrito, articuladas no necesariamente en orden jerárquico.

1. Constituir vínculos entre los departamentos de lenguaje y comunicación y los departamentos de especialidad con el fin de que se apoyen mutuamente.
2. Construir bases de datos internacionales acerca de los textos más usados en algunas especialidades técnico-profesionales.
3. Elaborar materiales de apoyo al conocimiento de los tipos textuales más prototípicos de ciertos ámbitos de especialidad.
4. Desarrollar principios metodológicos orientadores generales para una didáctica del discurso especializado que sirvan al docente especialista.
5. Producir materiales desde los principios de la lingüística del texto que otorguen pistas básicas acerca de lo que es un texto y sus principales mecanismos de cohesión.
6. Elaborar un conjunto de estrategias didácticas generales para el desarrollo de las habilidades interconectadas de la comprensión y la producción de textos escritos.
7. Proyectar una planificación lingüística que muestre el camino desde aquellos textos más divulgativos y genéricos hacia aquellos de más alta especialización.

8. Configurar conjuntos de textos particulares, articulados orgánica y temáticamente, que se constituyan en "pasajes hacia el conocimiento" para los alumnos.
9. Organizar el conocimiento disciplinar en torno a clases textuales que se conecten progresivamente con otras relacionadas y paulatinamente de mayor envergadura especializada.
10. Elaborar materiales en que se describa -con apoyo técnico- las formas textuales más prototípicas de la especialidad.
11. Propender a un cambio progresivo en el énfasis desde la observación y experimentación, hacia la incorporación de la explicación, la discusión y el debate en torno a los problemas de cada especialidad.

Un enfoque como el aquí esbozado resulta ineludible si esperamos que los estudiantes de comunidades técnico-profesionales obtengan una mínima comprensión del terreno de la ciencia y la tecnología contemporánea, apuntando hacia la necesaria accesibilidad y transparencia del conocimiento disciplinar que la formación académica exige.

Evidentemente, la mayoría de estas propuestas descansan -de manera más o menos definitiva- sobre la base de investigaciones que aporten datos en cierta medida hoy casi inexistentes, pero fundamentales para el desarrollo del área. Se vuelve obvio destacar que nuestro compromiso -inicialmente- ha de centrarse en las etapas previas. Pero no necesariamente, existen tareas que es posible implementar paralelamente al desarrollo del conocimiento descriptivo. Estoy seguro que nuestro entusiasmo debe ser grande, ya que la interesante tarea que tenemos por delante es tremendo y el futuro se abre promisoriosamente para los lingüistas interesados en esta parcela del conocimiento.

Finalmente, en la Figura 6, pretendo mostrar -resumidamente- el camino que estimo se debe seguir en la construcción de una didáctica del discurso especializado a la luz del paradigma lingüístico aquí esbozado.

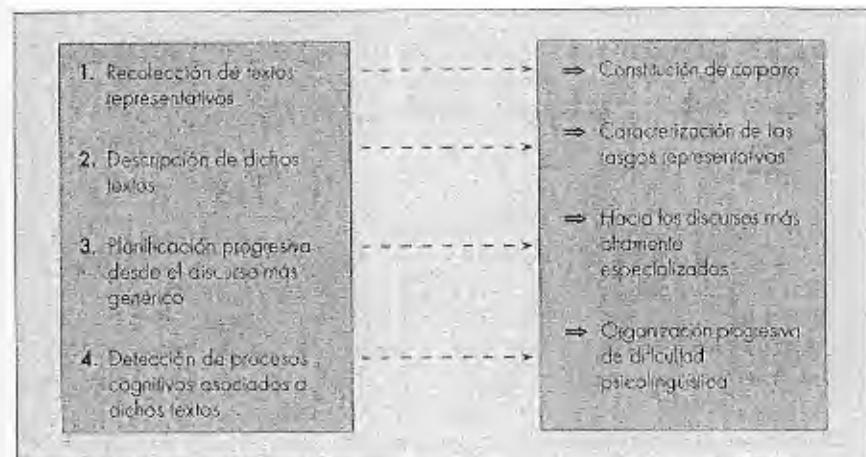


Figura 6. Énfasis hacia la planificación de inmersión progresiva

Estos cuatro ejes de acción sugieren un camino de investigación aplicado a la educación lingüística que exige contar con un conocimiento básico de lo que ocurre en las diversas especialidades. Es muy cierto que ello también requiere de una batería de procesos estratégicos adecuadamente secuenciados para el efectivo desarrollo de las habilidades de lectura y escritura. En este contexto, estoy seguro que el poder de un programa de intervención didáctica reside en el conocimiento psicolingüístico de las habilidades de comprensión y producción lingüísticas de un tipo de textos particular, tal como el discurso especializado. Así, este se constituye en el eje singular sobre el que descansa todo el sistema de formación educacional, el cual -aparentemente- hasta ahora no ha recibido la debida atención ni se ha sabido con exactitud cómo abordarlo eficientemente para potenciar las capacidades de los lectores y escritores en formación profesional.

Desde este ángulo, defiendo tanto teórica como metodológicamente la idea de un programa estratégico jerarquizado con potencial adaptativo. Esto quiere decir que creé fuertemente que para elaborar un programa orientado al desarrollo de estrategias de lectura comprensiva de textos especializados se debe proceder a través de una gradiente jerárquico que conduzca al sujeto lector-escritor a través de un "pasadizo de acceso a la competencia comunicativa especializada". Ello debe ocurrir de manera secuenciada y paulatinamente estructurada sobre la base de una taxonomía psicolingüística y una distribución de textos desde aquellos de corte más divulgativo didáctico en el área disciplinar hasta aquellos altamente especializados.

Se busca así impulsar en los alumnos en formación de todos los niveles tanto secundario terminal como universitario inicial en áreas de especialidad el movimiento desde un "lenguaje primario" a un "lenguaje experto". En este sentido, preocupa intervenir de manera decisiva en el proceso de alfabetización del discurso especial en estudiantes inmersos en sus propios contextos particulares y disciplinares. El desarrollo de la comprensión y producción lingüística con eje en el pensamiento metacognitivo y metarreflexivo ha de ejecutarse desde las propias prácticas discursivas cotidianas, no como un ente exógeno que lleve a una competencia pasajera, superflua y casi lúdica.

Esta dinámica combinatoria entre una jerarquía de procesos psicolingüísticos junto a una gradiente de textos de especialidad resulta en una herramienta poderosa desde el ámbito de las ciencias del lenguaje. Sucintamente, puede formularse así:

1. Jerarquía de los procesos de lectura comprensiva de textos especializados desde una óptica psicolingüística (para el fomento paulatino de estrategias lectoras)
2. Apoyo tecnológico a través de herramientas computacionales que permitan capturar los rasgos relevantes de los discursos de especialidad para así caracterizarlos y contar con datos fidedignos que apoyen los diseños metodológicos.
3. Exploración de las capacidades metacognitivas y metarreflexivas en el contexto de la lectura comprensiva.
4. Gradiente desde textos más divulgativo (e.g. manuales) hasta textos de alta especialización (e.g. artículos de investigación técnico-científico).
5. Lectura y escritura en la disciplina.

A MODO DE CIERRE

Espero no haber sido excesivamente disperso en el recorrido que he intentado delinear y que las diversas aristas que he buscado conectar pudieran alcanzar la anhelada cohesión textual y potencial coherencia que como escritor he pretendido. Los lingüistas bien sabemos que la coherencia no es una característica inherente y pre-existente de los textos, sino un contrato que se establece entre escritor y lector o hablante y oyente, en base a un sinúmero de variables interviniéntes.

Ahora bien, la utilidad de aproximaciones basadas en corpora lingüísticos, digitalizados y marcados, puede presentar propósitos puramente de investigación pero también con importante repercusión hacia fines didácticos múltiples. Al respecto, Kucer (1985) sostiene que la comprensión y producción escrita son actos de construcción de significado, en los cuales ambos procesos se ejecutan en paralelo y aportan a la elaboración de un "mundo textual", lo que para Gómez Macker (1991) equivaldría a las habilidades de base que ayudan a construir los mundos posibles del hombre. Si, como sabemos, el discurso escrito vehicula los significados -en este caso técnicos y especializados- es innegable su participación en la construcción de los mundos posibles de alta especialización y se hace vital su rol central en el proceso paulatino de incorporación de los sujetos en formación a las correspondientes comunidades discursivas.

Las reflexiones y propuestas avanzadas en este capítulo dan cuenta, por una parte, de la necesidad de elaborar e incorporar metodologías integradoras y compatibles en lingüística e informática para dar paso así a la lingüística de corpus computacional como una herramienta poderoso que apoya la labor de los lingüistas interesados en profundizar las investigaciones en la línea del discurso especializado, por ejemplo. Por otra, la realidad comentada en los establecimientos técnico-profesionales revela que existen urgentes necesidades por producir cambios que remuevan las bases del actual sistema educativo. Los datos disponibles hacen ver que existen síndromes de rasgos lingüísticos en los textos de especialidad y que estos síndromes tienden a producir un tipo de conocimiento que identifica los mundos posibles de las áreas de especialización.

Estoy seguro la realidad en otras comunidades discursivas y con otros mundos posibles es muy similar a la delineada aquí. Por ello, el desafío es de proporciones, hecho que sin lugar a dudas alienta la investigación científica desde la lingüística en múltiples ámbitos del conocimiento y dibuja un horizonte promisorio de múltiples investigaciones verdaderas. La tarea de diseño y elaboración de, por ejemplo, gramáticas, diccionarios y materiales didácticos especiales son los recursos vitales para hacer de puente hacia la construcción del conocimiento de la ciencia y la tecnología que los estudiantes de diversas disciplinas requieren. La organización de textos particulares en la jerarquía adecuada ha de posibilitar la adquisición y/o aprendizaje del conocimiento disciplinar que puede ocurrir fundamentalmente a través de la lectura y la escritura estratégicamente pavimentada. Sin lugar a dudas es por medio de estos instrumentos que se posibilita la construcción -en parte colaborativamente- de la identidad profesional discursiva, a la vez que se potencian las habilidades que permiten un desarrollo del pensamiento razonado y de alto nivel.

Estoy convencido de que si se fortalece, dinamiza y focaliza desde la disciplina misma el desarrollo de estrategias lectoras en textos de especialidad y se utiliza como eje fundante

del lenguaje escrito como un elemento clave en la formación de las generaciones futuras de técnicos-profesionales y universitarios los resultados serán muy diversos. Todo ello favorecerá certamente el acceso y la transparencia hacia el conocimiento de la ciencia y de la tecnología.

En los capítulos que siguen, diversas indagaciones empíricas esgrimen, exploran y ponen a prueba algunos de los conceptos aquí esbozados, básicamente nos hemos concentrado en textos escritos, por considerarlos más prototípicos del objeto que buscamos estudiar, aunque el discurso oral ocupa un lugar importante llegado el momento de profundizar en distinciones con la modalidad escrita de la lengua.

Capítulo 2

Los tipos textuales del corpus PUCV-2003: Una aproximación multiniveles

GIOVANNI PARODI
AÍDA GRAMAJO

INTRODUCCIÓN

Existe amplio consenso hoy en día entre los especialistas en cuanto a que las pruebas de competencia lingüística a que se somete a una determinada población deben atenerse a la variedad de uso lingüístico que esa comunidad discursiva emplea. Similmente, si se busca evaluar el desempeño lingüístico de estudiantes secundarios de formación diferenciada técnico-profesional en su comprensión de la variedad de lengua escrita especializada, se debe contar con una descripción tanto de los tipos de textos que circular en ese sistema educativo como de los rasgos lingüísticos característicos de los registros implicados.

Un modo de acercarse al registro especializado escrito es partir del supuesto de que los textos escolares técnico-científicos a que se ven enfrentados los estudiantes en su práctica escolar diariamente reflejan tanto los tipos de textos como los rasgos representativos de esa variedad de lengua especializada. En la actualidad los investigadores del discurso conocemos relativamente poco acerca de los patrones de uso lingüístico que son característicos de los registros técnico-profesionales, particularmente de los registros especializados en español escrito. Como es fácil comprobar, la mayor parte de la investigación lingüística en español se ha concentrado en esferas que han estado ajena al rol más funcional y comunicativo del lenguaje escrito y se han orientado clásicamente hacia una descripción de índole estrictamente gramatical. Es posible distinguir una línea de trabajo que se ha mantenido en un tipo de lingüística que –en cierto

sentido – mira más a la mente del ser humano que a la sociedad, mientras otro tipo que se ha interesado en un descriptivismo, a veces, algo exagerado, con riebles de immanentismo. Ambas posturas han privilegiado una lingüística del sistema y no una lingüística del uso.

Por su parte, la lingüística americana como la denominan muchos, o más bien la de los Estados Unidos de Norteamérica, es la que ha preconizado esta lingüística "introvertida" (Sinclair, 1991). Sin lugar a dudas, es el nombre de Noam Chomsky el que concilia las grandes disputas. Si bien es cierto que sus innovadores aportes estimularon avances lingüísticos radicales desde una determinada perspectiva y contribuyeron de manera decisiva –entre otros– a una mirada más proactiva de la mente humana y al desarrollo de nacientes interdisciplinas, también se debe reconocer que incipientes proyectos se vieron bloqueados momentáneamente (Coseriu, 1977; Leech, 1991; Caravedo, 1990; Stubbs, 1996; McEnery & Wilson, 1996; Kennedy, 1998). Por supuesto que toda generalización en este contexto es injusta: aquí nos referimos a los grandes proyectos que han concitado los desarrollos más populares y que han recibido gran apoyo de todo índole.

En Europa la situación ha sido algo diferente y se han mantenido enclaves que han defendido una lingüística del uso. En Latinoamérica también se detectan polos de desarrollo en este tipo de la lingüística, no obstante ello, no se aprecia un avance en la vertiente más moderna que acoge los planteamientos de la lingüística de corpus asistido por computadores. En sentido último, si bien es cierto que existe una importante tradición en investigaciones con apoyo de corpus lingüísticos en nuestro ámbito regional hispanohablante, con proyectos de envergadura (entre otros, Lape Blanch, 1977; Rabanal & Contreras, 1986; Bentivoglio, 1992; Barrientos, Cifuentes, Lagos, Muñoz & Tessera, 1996; Wagner, 1998, 1999; Cepeda, 1998; Valencia, 2002; Matus, 2002; Oyanedel & Samaniego, 1998; Lagos, Cifuentes, Tassara, Fuentes, Reyes, & Venegas, 1999; Bernales, 2002) no es menos cierto que sus alcances se han visto restringidos –en parte– justamente por no haber incorporado las tecnologías de vanguardia.

Ahora bien, volviendo al ámbito del discurso especializado desde el terreno de la formación profesional, no cabe duda de que éste ha sido poco a nada indagado en Chile. Es escaso el conocimiento que poseemos de los tipos de textos que el profesor entrega a sus alumnos para que sean leídos y logren así acceder al conocimiento disciplinar y, por ende, poder incorporarse plenamente a las comunidades discursivas profesionales.

En este capítulo nos proponemos realizar una descripción de orden funcional, comunicativo y lexical de tres corpus de discurso especializado escrito, recolectados en tres ámbitos del último grado de la formación diferenciada técnica-profesional secundaria en establecimientos de la ciudad de Valparaíso, Chile (Parodi, 2002b; Cademártori, Gramajo & Parodi, 2003). El objetivo específico que abordaremos en este artículo es determinar y describir los tipos de textos que reciben como lectura obligatoria o de consulta los alumnos de tres ámbitos de la formación diferenciada técnica-profesional en los dos últimos años de su educación formal secundaria. Los ámbitos en cuestión son: Sector Marítimo (Especialidad Operación Portuaria), Sector Metalmecánico (Especialidad Mecánica Industrial) y Sector Administración y Comercio (Especialidad Contabilidad). Con el fin de llegar a una clasificación de los textos recolectados, recurriremos a una metodología de corte inductivo-deductiva, con parámetros funcionales, situacionales y textuales. Nuestra intención es al-

canzar una descripción de los textos recolectados con el mayor grado de profundidad posible y así poder determinar rasgos distintivos y diferenciadores de utilidad práctica para el área en estudio.

En la primera parte de este trabajo, nos concentraremos en una breve revisión de conceptos y criterios fundamentales para enmarcar el trabajo en cuestión, en la segunda, entraremos argumentos y pasos metodológicos; en la tercera, mostramos los resultados cualitativos y cuantitativos de la clasificación de los textos. Se concluye con algunos comentarios y proyecciones.

1. MARCO DE REFERENCIA

1.1 Discurso especializado

Tal como se ha apuntado en el Capítulo 1, adherimos a la tradición que no optado por el uso del término "discurso de especialidad o especializado". Este descriptor ilustra el tipo de discurso en el que centramos el estudio: textos de la ciencia y de la tecnología en su modalidad de divulgación didáctica en contextos institucionales educativa-prácticos secundarios (Martín, Christie & Rothery, 1987; Gläser, 1993; Halliday & Martín, 1993; Rose, 1997; Christie, 1998; Goldman & Bizans, 2002). Nos parece que el término "especializado" revela una gradiente ó continuum, cuestión que se vuelve eje fundamental para aproximarse a textos que tienen diverso grado de especialización y circular, muchos de ellos, en los límites de la transición hacia la diseminación o divulgación propiamente tal.

Diversos autores han apuntado coincidentemente a que el discurso especializado se constituye a partir de un grupo de textos centrados en tópicos prototípicos de un ámbito específico del conocimiento; en este caso, científico y tecnológico. Estos textos presentan una serie de rasgos distintivos que revelan adecuaciones a convenciones lingüísticas, retóricas y estilísticas propias de las tradiciones de una comunidad discursiva determinada y se generan en situaciones comunicativas particulares (Schröder, 1991; Halliday, 1993; Martín, 1993; Cabré, 2002; Crapuscio, 1994, 2003).

En este contexto, nos parece que, inicialmente desde el terreno de los lenguas para propósitos específicos, Schröder (1991) aporta una rica visión de las lenguas especializadas que –entre otros– precisa la idea de orientación hacia la temática, pero sin desconocer la importancia del contexto situacional y de las funciones del lenguaje. Para Schröder (1991:5), estas variedades de lenguaje especializada:

...no son definidas como opuesta a la lengua común; las lenguas para propósitos específicos son sublenguas que pertenecen a un cierto campo de la comunicación orientada hacia un tema; ellas usan los medios lingüísticos y comunicativos de un cierto sistema de lenguaje y de cultura de un modo específico y con una frecuencia específica de ocurrencia, dependiendo del contenido, del propósito y de la situación comunicativa global de un texto o discurso...³

Este autor destaca la concurrencia de rasgos de diversa índole en la definición de este tipo de discurso: lo lingüístico y lo pragmático se conjugan y resultan centrales. La preferencia de Schröder por el término *especial* o *especializado*, lo lleva a sugerir que no se debería utilizar el de "lenguas con propósitos específicos", sino que el de "comunicación

especializada" ya que involucra los factores antes señalados y se plantea como un concepto más abarcador. En nuestra opinión, los textos de especialidad poseen características gramaticales y textuales que, unidas a otros factores de orden no-lingüístico, constituyen en conjunto una matriz de criterios útiles para, por ejemplo, distinguirlos de otros tipos no especializados o para diferenciar clases textuales más específicas al interior de géneros especializados mayores (como lo que se propone en el presente estudio).

Ahora bien, complementariamente y con el propósito de acotar aún más el concepto de discurso de especialidad desde una visión semiótica compleja, nos parece oportuno incorporar la noción de "síndrome", propuesta por Halliday (1993) en el marco de la lingüística sistemática funcional (LSF). De acuerdo a este especialista, a través de la co-ocurrencia de un conjunto de rasgos lingüísticos se puede identificar un cierto registro. Los síndromes, para Halliday (1993), son patrones de co-ocurrencia a partir de rasgos en uno u otro de diversos niveles lingüísticos (de la expresión o del contenido). Estos síndromes son los que caracterizan una variedad de lengua y hacen posible reconocer un registro determinado como tal (por ejemplo, una variedad dialectal o una técnica-científica).

Como se aprecia, en el marco de los estudios que abordan el discurso especializado desde diversas ópticas, reiteramos que para definir un texto como de especialidad la perspectiva lingüística por sí sola no puede dar cuenta de modo radical de los diversos textos que circulan. Se hace evidente que lo lingüístico se debe complementar con rasgos de ámbito funcional y situacional.

2. CORPUS Y METODOLOGÍA

2.1 El Corpus Técnico-Científico PUCV-2003

El Corpus Técnico-Científico (CTC) está compuesto por setenta y cuatro textos con un total de 626.790 palabras, recolectados en establecimientos secundarios técnico-profesionales de la ciudad de Valparaíso, Chile. Como se aprecia en la Tabla 1, no existe una relación directa entre número de textos por ámbito de especialidad y número de palabras. Así, en el ámbito marítimo de operación portuaria se registra la mayor cantidad de textos, pero el menor corpus de palabras. Por el contrario, en el área técnica industrial metalmecánica se recolectó el grupo más reducido de textos (sólo dieciocho), pero ellos conforman la muestra más grande respecto al número de palabras. Por su parte, en el área contabilidad se entregan cifras similares a la anteriormente descrita. En esta área técnica se alcanzó un total de veinte textos con un número considerable de palabras.

Tabla 1: Configuración del CTC

Área Técnica	Número de Textos	Número Total de palabras
Marítimo Operación Portuaria	36	155.160
Mecánica Industrial	18	246.374
Contabilidad	20	225.256
Totales	74	626.790

A continuación, abordaremos la cuestión nuclear que nos ocupa: la clasificación y caracterización de estos textos escritos especializados.

2.2 En busca de una tipología textual: criterios y método

La construcción de tipologías textuales es un asunto relativamente reciente en el ámbito de la lingüística y –sin lugar a dudas– encierra una alta complejidad (Charaudeau, 2004). Ya señalábamos en la introducción de este capítulo que la hegemonía del generativismo chomskiano ocupó durante un periodo de tiempo importante el centro de los estudios lingüísticos y, según algunos (Stubbs, 1996; Leech, 1991; McEnery & Wilson, 1996; Kennedy, 1998), actuó como freno a los desarrollos que se producían en diversas esferas del mundo académico. No menos cierto es que el estructuralismo también fue un obstáculo para las preocupaciones por el texto o discurso y su definición, categorizaciones y clasificaciones se vieron fuertemente marcadas por este movimiento científico.

Ciapuccio (1994, 2003) entrega una exhaustiva revisión del desarrollo de la cuestión de los tipos textuales y sus variadas vertientes. Una constatación interesante es contar hoy en día con aproximaciones más exhaustivas y que incorporan múltiples niveles de análisis. Estamos seguros que estos aproximaciones complejas en que coexisten aspectos cognitivos, lingüísticos y comunicativos pueden brindar descripciones y explicaciones mucho más ricas y poderosas de los fenómenos en estudio. Al mismo tiempo, postulamos que esta perspectiva multiniveles resulta más acorde con nuestro concepto de texto escrito y su directa relación con los procesos de comprensión y producción lingüística (Peroniard, Gómez, Parodi, Núñez, 1998; Parodi, 2003a, 2005a).

De modo más específico y entendiendo que se puede precisar una distinción entre tipo y clase textual (Ciapuccio, 1994), en este trabajo hemos optado por utilizar ambos términos indistintamente debido a que –en la gran mayoría de la bibliografía consultada– estos términos no presentan definiciones taquitas. Siguiendo a Brinker (1988: 124), entendemos que las clases o tipos textuales son:

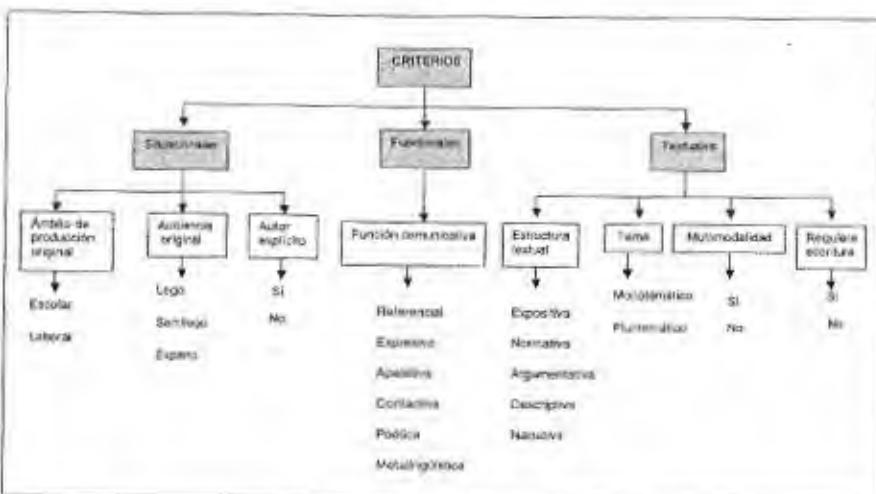
"...modelos complejos y convencionalmente válidos para un conjunto de acciones lingüísticas que pueden ser descritos en cada caso como conexiones de marcas contextuales (situacionales), comunicativo-funcionales y estructurales (gramáticos y temáticos). Ellos se han desarrollado históricamente dentro de la comunidad lingüística y constituyen parte del conocimiento corriente de los hablantes. Poseen un carácter normativo, también facilitan la tarea comunicativa, ya que entregar a los participantes de la comunicación orientaciones más o menos fijas para los procesos de producción y comprensión lingüística".

Ahora bien, en este contexto, para la elaboración de una tipología que dé cuenta de las características fundamentales del Corpus Técnico Científico (CTC), se decidió optar por un enfoque multinivel (Bassols & Torrent, 1997), es decir, se consideró tanto las características internas de los textos como el contexto extra-lingüístico en el cual se producen y circulan. Dicho enfoque tiene sus fundamentos en el carácter eminentemente dialógico del lenguaje (Bajin, 1982) e intenta dar cuenta de la relación lingüística-social entre los participantes de una determinada comunidad discursiva. Los enfoques de varios niveles son actualmente considerados los más adecuados para la elaboración de tipologías válidas

desde un punto de vista teórico (Ciupuscia, 1994, 2003; Bassols & Torrent, 1997). Lo esencial es la elección de los criterios o parámetros adecuados que en su conjunto permitan describir las similitudes y diferencias entre los textos o tipos de textos (Molin, 1993).

La construcción de esta propuesta tipológica obedeció a un enfoque inductivo-deductivo. El paso inductivo implicó "partir de los tipos ya definidos e instituidos tradicionalmente como objetos lingüísticos observables" (De Beaugrande & Dressler, 1981: 251) para, sobre esa base, construir una tipología empíricamente sustentada. Posteriormente, con el objetivo de determinar patrones comunes y divergentes que permitan distinguir y clasificar con mayor precisión los textos técnicos inicialmente clasificados, se elaboró deductivamente una matriz de rasgos más específicos, siguiendo lineamientos tomados de bibliografía especializada. La tipología en cuestión se organiza, de acuerdo al enfoque multirivales, en tres criterios generales de análisis, o saber, situacional, funcional y textual. Estos criterios intentan abarcar aspectos que dicen relación con la interacción de los participantes, las funciones comunicativas, los contextos en que los textos circulan y las estructuras textuales que caracterizan los textos de la muestra recolectada.

A continuación se presenta el Esquema 1 en que se muestran ocho rasgos especificadores a partir de los tres criterios antes mencionados. Posteriormente se describen en detalle cada uno de ellos.



Esquema 1. Criterios y rasgos de clasificación textual

a) Criterio situacional

Si se parte de la premisa de que todo texto aparece siempre en un contexto de situación, será imprescindible determinar qué rasgos permiten describir de mejor manera dicho contexto (Halliday, 1978). En lo que respecta a esta investigación, consideramos tres rasgos específicos para dar cuenta del aspecto situacional: el ámbito de producción original

de los textos, la audiencia original para la cual fueron concebidos y la autoría. Estos rasgos permitirán caracterizar a las clases textuales en lo que respecta a su relación con los participantes de la interacción, es decir, a la relación que se establece entre el texto técnico y el alumno lector. Este punto de partida también lleva a una distinción entre las clases textuales, cuyo ámbito de producción es el ámbito escolar, y otras cuyo ámbito de producción puede situarse en otras comunidades discursivas o en otras esferas del sistema educativo.

a. 1) **Ámbito de producción original:** se refiere al contexto original en el que se produce el texto. Esto es, si el texto se produce en o para el ámbito del trabajo o en o para el ámbito escolar: Escolar/Laboral.

a. 2) **Audiencia original:** se refiere la relación entre escritor y lector en cuanto al conocimiento que se espera tengan estos últimos acerca del tema, éste puede ser de experto a lego, de experto a semilego y/o de experto a experto.

Lego: designamos como lego a la persona que no posee conocimientos acerca del tema del texto.

Semilego: entendemos por semilego a la persona que posee conocimientos mínimos acerca del tema, pero suficientes para seguir aprendiendo.

Experto: es la persona que tiene conocimiento especializado del tema.

En lo que respecta a este estudio, denominamos al alumno lector como "semilego", puesto que hay etapas de inicio de aprendizaje donde el alumno no posee conocimiento acerca de los temas de los textos pero existen otras en que ya cuenta con un acercamiento inicial a los temas especializados. Del mismo modo, pueden existir otras etapas en donde los lectores ya hayan adquirido un conocimiento mayor, el cual les permite una aproximación con mayor facilidad.

a. 3) **Autor explícito:** se refiere a si el nombre del escritor aparece explícito o no en el texto.

b) Criterio funcional

Estos rasgos permiten clasificar a las clases textuales según el propósito comunicativo expresado en el texto con determinados recursos. Lo que Biber (1988) denomina como el propósito del evento comunicativo. Dichos rasgos se corresponden con las funciones comunicativas propuestas por Jakobson (1961).

b. 1) **Función referencial:** apunta a hechos, cosas o ideas; su objetivo es informativo. Está centrada en el contexto, es decir, en los hechos a que se refiere. Se manifiesta, entre otros, en la tercera persona verbal y en una gran cantidad de sustantivos.

b. 2) **Función expresiva:** centrada en el emisor, implica la expresión de sentimientos y emociones. Se manifiesta en la primera persona verbal, las interjecciones y en una gran cantidad de adjetivos.

b. 3) **Función apelativa:** se orienta hacia el destinatario, implica la persuasión y exhibición con la intención de producir una respuesta de éste. Se manifiesta preferentemente en el uso de la segunda persona verbal y en la abundancia de verbos.

b. 4) **Función contactiva:** está orientado hacia el canal de comunicación, tiene como objeto comprobar si el canal funciona correctamente. Se manifiesta a través de frases cliché y expresiones formulaicas.

b. 5) **Función metalingüística:** está orientada hacia el código, su intención esclarar, explicar. Mayoritariamente se manifiesta a través de definiciones y aclaraciones.

b. 6) **Función poética:** está orientada hacia el propio mensaje, se manifiesta en el estilo y las figuras.

Cabe recordar, como bien se sabe, que ninguna de estas funciones existe en estado puro, esto es, ellas se entremezclan en el discurso, aunque suele manifestarse cierta dominancia de una de ellas (Jakobson, 1961).

c) Criterio textual

Estos rasgos caracterizan las clases textuales según el patrón organizativo del texto, al que se ve afectado por el tema, por los elementos gráficos y por las características estructurales propiamente tales.

c. 1) **Estructura organizacional:** Se refiere al patrón de organización de la información que predomina en el texto, ya sea descriptiva, narrativa, expositiva, argumentativa y normativa.

- **Descriptiva:** su función consiste en caracterizar por medio del lenguaje a objetos, personas, situaciones, o procesos, explicando sus partes, cualidades o circunstancias (Bassols & Torrent, 1997). Está condicionada por el contexto comunicativo y el propósito que se pretende alcanzar (Calsamiglia & Tusón, 1999).
- **Narrativa:** supone el deseo de contar acerca de hechos reales o ficticios, pero posibles en un universo textual. Su función es organizar discursivamente las acciones y los acontecimientos en un orden secuencial integrador que muestre la unidad de acción y se oriente hacia un fin; de manera tal que si una de las partes de la acción es desplazada o suprimida el todo se altera (Aristóteles, 1978). Se caracteriza por la presencia de acciones que afectan a personas (van Dijk, 1983), por la transformación de esas acciones (Adam, 1992), por una alta densidad de relaciones causa-efecto, propósito, posibilidad y proximidad temporal.
- **Expositiva:** su finalidad es informar o exponer "objetivamente" un tema para facilitar la comprensión (Bassols & Torrent, 1997). Esto implica la necesidad de que lo entrega de información sea fidedigna, neutra y objetiva (Calsamiglia & Tusón, 1999).
- **Argumentativa:** su finalidad es influir sobre una audiencia determinada. Supone un enunciador que pretende "hacer admitir una conclusión a un destinatario [o destinatarios], ofreciéndole una razón para admitir esa conclusión" (Plantin, 1998: 126).
- **Normativa:** su finalidad es entregar indicaciones e instrucciones precisas acerca del modo en que se deben efectuar determinados procedimientos que tienden a regir la conducta y/o a entregar un estado de situación ideal de cosas y procesos (Silvestri, 1995).
- c. 2) **Tema:** se refiere a si el texto abarca uno o varios temas. Lo que puede dar cuenta de la complejidad textual y de la extensión. Un texto pluritemático puede resultar más complejo, debido a la cantidad de temas tratados y a su extensión. Un texto monotemático poseería una estructura más simple, por lo tanto, podría resultar de más fácil comprensión.
- **Monotemático:** se enfoca en un tema.
- **Pluritemático:** se enfoca en varios temas.

c. 3) **Multimodalidad:** se refiere a la presencia o ausencia de elementos de distintas modalidades (lingüísticos, gráficos). El que un texto tenga elementos multimodales puede facilitar la comprensión, puesto que un mismo concepto es presentado de distintas maneras (Gómez, 2001).

c. 4) **Requiere escritura:** se refiere a la cerrazón del texto al requerir o no completación (escritura).

3. RESULTADOS

3.1 Aplicación de criterios: Hacia una identificación de las clases textuales del CTC

El primer paso -de orden estrictamente inductivo- que se aplicó en el estudio del corpus CTC fue el análisis de los textos con el objetivo de intentar una primera clasificación basada en nociones socialmente instauradas a partir de lo que intuitiva y tradicionalmente distingue a estos objetos lingüísticos. Ello se trabajó grupalmente en el equipo de investigadores y, de este modo, se determinó tentativamente una primera clasificación en doce tipos textuales. Ellos se listan a continuación:

- Artículo Técnico
- Descripción Técnica
- Diagrama
- Formulario
- Glosa Legal
- Glosario
- Guía Didáctica
- Instructivo
- Ley
- Manual Técnico
- Reglamento
- Tabla

El segundo paso -esta vez, de orden deductivo- consistió en la aplicación de la matriz de ocho rasgos caracterizadores a los doce tipos textuales inicialmente detectados. Ello con el objetivo de describirlos de modo más específico, alcanzar definiciones operacionales y lograr establecer comparaciones más enriquecedoras. La Tabla 2 presenta los doce tipos textuales y el resultado del análisis de rasgos, basados en los tres criterios más arriba descritos.

Esta caracterización más fina, presentada en la Tabla 2, permite distinguir algunos rasgos como más diferenciadores de los tipos textuales. Así, desde el punto de vista de las funciones comunicativas, se constata el predominio de las funciones referencial y operativa del lenguaje en la mayoría de los textos del corpus. Ello implica un predominio del foco referencial en el que se destaca la alusión a hechos, cosas y/o ideas presentadas, con especial énfasis en el tema que se quiere comunicar. Comunicativamente hablando, el objetivo central de estos textos es básicamente el de informar, aunque también se evidencia una orientación hacia el destinatario, con la finalidad de producir una respuesta de éste. Como contraparte, es evidente, de acuerdo a estos rasgos, que los textos técnico-científicos del Corpus PUCV-2003 se alejan de todo propósito que implique la expresión de sentimientos y emociones (función expresiva), así como de preocupaciones estilísticas que tienden que ver con el mensaje (función poética).

Tabla 2. Tipos textuales y rasgos caracterizadores

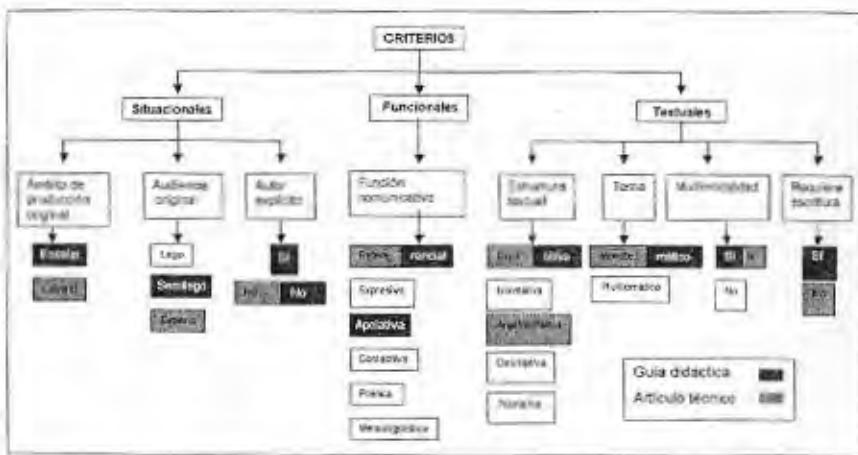
Tipos textuales	Rasgos situacionales	Rasgos funcionales	Rasgos textuales
Ámbito de producción Original	Audiencia original	Autor explícito	Estructura textual predominante
Manual Técnico	Escolar	Semílego	Referencial-apelativa
Guía Didáctica	Escolar	Semílego	Apelativa-referencial
Formulario	Laboral	Semílego	Apelativa
Reglamento	Laboral	Experto	Apelativa
Glosa Legal	Escolar	Semílego	Referencial
Instructivo	Laboral	Experto	Apelativa
Ley	Laboral	Experto	Apelativa
Artículo Técnico	Laboral	Experto	Referencial
Tabla	Escolar	Semílego	Referencial
Descripción Técnica	Escolar	Semílego	Referencial
Diagrama	Escolar	lego	Referencial-mediogénesis
Glosario	Escolar	lego	expresiva

Los datos de la Tabla 2 también permiten señalar que los tipos textuales del CTC se caracterizan mayoritariamente por poseer una estructura organizacional fundamentalmente expositiva y normativa; esto significa que su finalidad es informar o exponer un tema para facilitar su comprensión, lo cual conlleva la necesidad de entregar información neutra y fidedigna. Así como también entregar indicaciones e instrucciones precisas acerca de cómo se deben realizar determinados procedimientos.

La multimodalidad se revela como un rasgo importante entre los tipos de textos técnicos. Según se aprecia, la mayor cantidad de textos se alinea en el tipo multimodal, es decir, donde los recursos clásicamente llamados no verbales (figuras, dibujos, tablas, esquemas, gráficos, etc.) tienen un rol importante en la comunicación de la información; de este modo, se hace evidente que cualquier intento por caracterizar estos textos no puede dejar de lado este rasgo fundamental del discurso especializado.

Por último, se debe poner de relieve el carácter monotemático en la gran mayoría de los tipos textuales del corpus. Este rasgo, unido a los anteriormente descritos (Foco Referencial y Expositivo), brinda una visión de conjunto de las características relevantes de los textos técnicos en estudio. De los doce tipos textuales en análisis, sólo el Manual destaca como una clase de texto pluritemático, clásicamente centrado en una variada gama de saberes.

A partir de la matriz de rasgos caracterizadores, es posible establecer una variedad de comparaciones ya sea atendiendo a algunos rasgos especificadores a través de todos los textos del corpus o en una comparación más exhaustiva pero sólo entre dos o tres tipos de textos. En un primer momento, a modo de exemplificación, se ilustra de manera gráfica - mediante el Esquema 2 - las diferencias y similitudes entre los llamados: Guía Didáctica y Artículo Técnico. Posteriormente, revisaremos - a través del Esquema 3 - las implicancias a partir de dos rasgos situacionales en los doce tipos textuales.



Esquema 2. Comparación de rasgos en dos tipos textuales

Este análisis tipológico comparativo, permite establecer distinciones entre la Guía Didáctica (GD) y el Artículo Técnico (AT) mediante los rasgos situacionales, funcionales y textuales. A partir del criterio situacional, el tipo de texto GD se diferencia del AT en el **ámbito de producción original**: la GD surge en el contexto educativo, creada por y para dicho contexto. El ámbito de producción original del AT, en cambio, es el laboral. Respecto de la **audiencia original** a la cual están dirigidos estos tipos textuales, también se establecen diferencias: mientras la GD está orientada hacia lectores semilegros, el AT está orientado a una audiencia experta. En cuanto a la **autoría**, la GD puede o no llevar el nombre del autor, evidenciando una cierta informalidad en relación con el AT, en el cual el autor siempre se identifica de modo explícito.

El Artículo Técnico y la Guía Didáctica evidencian un énfasis distinto en sus propósitos comunicativos: en el AT se privilegia la **función referencial**, lo que le otorga un carácter más "objetivo" al texto, con un foco informativo. En la GD, por su parte, se enfatiza la **función operativa** por sobre la referencial, es decir existe una mayor orientación hacia el destinatario, tratando de persuadirlo para producir en él una respuesta, hecho que se evidencia en que ella invita a la participación activa del lector pues requiere escritura. Ello se relaciona con los rasgos textuales, específicamente con la **completitud textual** ya que la GD requiere ser completada, a diferencia del AT que no requiere escritura.

Otra diferencia se puede observar en la **estructura textual predominante**: la Guía Didáctica posee una estructura expositiva, esto significa que su finalidad es fundamentalmente la presentación de los temas para ser aprendidos con el objetivo de facilitar su comprensión. El Artículo Técnico, cuya estructura es expositiva-argumentativa, privilegia un foco persuasivo y, por ende, entrega información que apoya la línea que se pretende privilegiar.

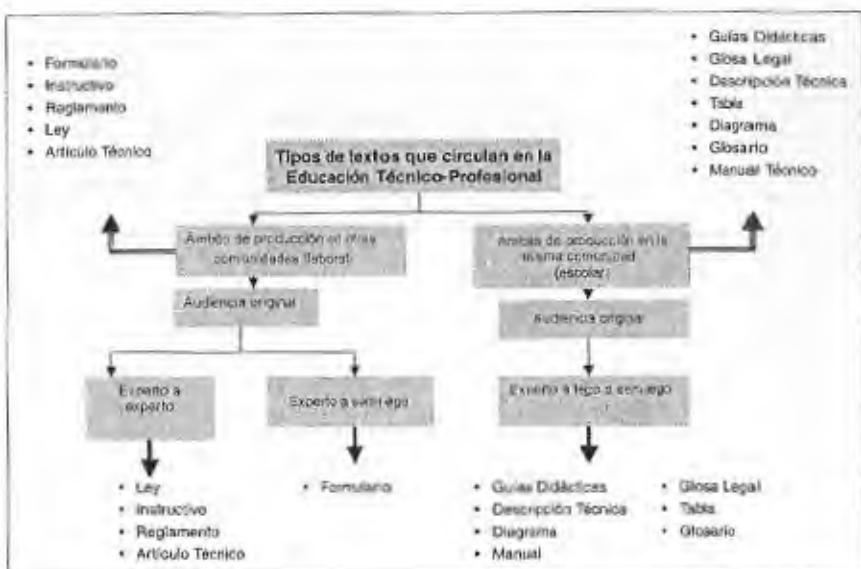
Como contrapartida, sería también posible establecer los rasgos comunes a ambas clases. Asunto que no será abordado en este trabajo.

3.1.2 Ámbito de producción original y audiencia invocada

Tal como se dijo anteriormente, nos ha parecido necesario profundizar en dos de los rasgos situacionales: el **ámbito de producción original** de los tipos de textos y el de la **audiencia invocada** a la que van dirigidos. Ello, por un lado, con el fin de pesquisar la utilidad de dichos rasgos y, por otro, caracterizar el discurso profesional técnico-científico.

En primer lugar, resulta interesante constatar que en la comunidad discursiva técnico-profesional estudiada circulan, por un lado, tipos textuales que son producidos por esa comunidad para sí misma y, por otro, clases textuales que no son generadas en esa comunidad, pero que igualmente son leídas por los miembros de ésta. Estos tipos textuales van dirigidos originalmente a las audiencias específicas para las cuales fueron concebidos y, por tanto, sus recursos lingüísticos, textuales y gráficos están al servicio de éstas. Esta distinción (audiencia original y contexto de producción original) resulta de gran utilidad al establecer la tipología textual, pues nos permite comparar los tipos textuales que son producidos en la comunidad discursiva u otra y, a su vez, comparar las audiencias originales para las cuales fueron concebidos. Ella con el fin de determinar si coinciden con la audiencia lego o semilego a la que debieron ir dirigidos los textos que se utilizan en la comunidad escolar técnica-profesional.

A continuación, el Esquema 3 muestra los tipos de textos que circulan en la comunidad educativa técnico-profesional, según su ámbito de producción.



Esquema 3. Comunidad discursiva escolar técnico-profesional y tipos textuales

El esquema permite visualizar el modo en que estos textos se subdividen, como se dijo anteriormente, en dos grupos, a saber, los tipos textuales cuyo **ámbito de producción** es la propia comunidad escolar y los tipos textuales que son producidos en **otras comunidades**. A su vez, se muestran las audiencias para las cuales fueron creados (legos o semilegros y expertas). Según se aprecia, la mayor parte de los tipos textuales (siete de doce) circulan en la comunidad discursiva escolar de origen: Guía Didáctica, Glosa Legal, Descripción Técnica, Tabla, Diagrama, Glosario y Manual Técnico. Sin embargo, un número significativo de los tipos de textos que leen los alumnos en la comunidad educativa técnico-profesional no son producidos por esa comunidad escolar ni están dirigidos originalmente a ella. Éste es el caso del Formulario, el Instructivo, el Reglamento, la Ley y el Artículo Técnico. De estos tipos textuales sólo uno de ellos, el Formulario, coincide con la audiencia escolar, es decir, va dirigido de expertos a semilegros. Los cuatro tipos de textos restantes van dirigidos a audiencias expertas, lo que indicaría un número considerable de tipos textuales que no van dirigidos a la audiencia semilega de la comunidad escolar técnico-profesional.

Este análisis permite mostrar el modo en que se efectúa –posiblemente– el camino hacia la incorporación paulatina de los futuros profesionales de estos ámbitos técnicos a sus respectivas comunidades discursivas. La coexistencia de textos de alto especialización junto a aquellos de corte más divulgativo y didáctico está mostrando de manera significativa el modo en que la institución escolar formadora de profesionales organiza el grado de especialización del conocimiento y su acceso al mismo.

No se debe entender que esta distribución de tipos textuales según estos dos rasgos constituye una desventaja para los alumnos; muy por el contrario, la utilización de textos con características de alta especialización es justamente lo que permiten su aprendizaje progresivo a la comunidad lingüística experta. Estos son los textos que los futuros alumnos y más adelante profesionales encontrarán en su vida laboral y los otros (más de carácter divulgativo-escuela) cumplen el rol de ayudarlos a desarrollar paulatinamente su conocimiento disciplinar. En efecto, esta complementariedad de tipos textuales es justamente la que caracteriza al ámbito profesional educativo especializado.

Los formadores de estas comunidades -tal vez, intuitivamente- han establecido un adecuado equilibrio entre tipos textuales didáctico-divulgativos y tipos textuales eminentemente altamente especializados en cada uno de los tres ámbitos técnicos en estudio. El orden asignado dentro del currículo escolar es un dato que desconocemos; por ello, este último comentario es altamente especulativo. Habrá que confrontarlo prontamente.

3.1.3 Definiciones de los tipos textuales del corpus técnico

Por último, a partir de la aplicación de la matriz de rasgos a los setenta y cuatro textos del corpus CTC, se definió -de manera más oportada- los doce tipos textuales emergentes. A continuación presentamos una definición para cada uno de ellos.

Artículo Técnico: Texto de carácter didáctico (divulgativo) cuyo fin es explicar diversos aspectos de un objeto, proceso o actividad en una esfera profesional. No tiene estructura fija, aunque suele ir acompañado de ilustraciones, cuadros y tablas. Su función primordial es referencial. Su estructura textual predominante es la expositiva.

Descripción Técnica: Documento explicativo que contiene especificaciones físicas y funcionales, ámbitos de aplicación y otras características de un objeto o proceso técnico. Suele acompañarse de ilustraciones. Su función primordial es referencial. Su estructura textual predominante es la expositiva.

Diagrama: Dibujo geométrico o esquema que describe las partes o el funcionamiento de un objeto o proceso. Su función primordial es referencial. Su estructura textual predominante es la descriptiva.

Formulario: Documento institucional con espacios en blanco para ser completados con datos. Su función primordial es apelativa. Su estructura textual predominante es la normativa.

Glosa Legal: Aclaración o comentario a una ley. Su función primordial es referencial. Su estructura textual predominante es la expositiva.

Glosario: Catálogo o vocabulario donde se definen palabras del ámbito técnico. Está destinado a un grupo no especializado específico del ámbito escolar. Su función primordial es referencial-metalingüística. Su estructura textual predominante es la expositiva.

Guía Didáctica: Texto breve de carácter didáctico que contiene un conjunto de explicaciones e indicaciones acerca de un tema específico, con ejercicios o actividades sugeridas. Su función primordial es apelativa, posee también una función referencial. Su estructura textual predominante es la expositiva-normativa.

Instructivo: Documento, frecuentemente institucional, que proporciona conocimientos útiles para la realización de una actividad. Su función primordial es apelativa. Su estructura textual predominante es la normativa.

Ley: Conjunto de preceptos dictados por la suprema autoridad para ser acatados por todos los gobernados. Conjunto de normas obligatorias para los ciudadanos de un país. Su función primordial es apelativa. Su estructura textual predominante es la normativa.

Manual Técnico: Es un tratado de carácter didáctico enmarcado dentro de una profesión técnica. Rico en ejemplos, tablas y recursos multimodales lo que facilita su comprensión. Su función primordial es la referencial, pudiendo tener secundariamente una función apelativa. Su estructura textual predominante es expositiva-normativa.

Reglamento: Documento institucional que contiene un conjunto de reglas, preceptos o instrucciones para la ejecución de una actividad. Su función primordial es apelativa y su estructura textual predominante es la normativa.

Tabla: Serie ordenada en columnas paralelas de valores numéricos de cualquier clase, de palabras o de signos. Su función primordial es referencial. Su estructura textual predominante es la descriptiva.

3.2 Ocurrencia de los tipos textuales en el corpus CTC

A partir de los doce tipos de textos técnico-científicos detectados en el corpus, según criterios situacionales, funcionales y textuales comentados anteriormente, se entrega una tabla descriptiva con datos numéricos relevantes que dan cuenta detallada de la distribución de los mismos en el corpus CTC.

Tabla 3. Ocurrencia de las clases lexicales en los tres corpus del CTC
[Nota: las celdas numéricas marcadas muestran las principales ocurrencias]

Clase Textual/Área Especialidad	Operación Portuaria (CTCma)		Mecánica Industrial (CTCin)		Contabilidad (CTCcom)		Totales	
	Nº Textos	Nº Palabras	Nº Textos	Nº Palabras	Nº Textos	Nº Palabras	Nº Textos	Nº Palabras
Artículo Técnico	1	9.346	-	-	-	-	1	9.346
Descripción Técnica	4	17.748	1	1.135	3	6.287	8	25.170
Diagrama	1	13	-	-	1	103	2	116
Formulario	-	-	-	-	3	2.287	3	2.287
Glosa Legal	2	4.142	-	-	-	-	2	4.142
Glosario	3	6.518	-	-	1	3.229	4	9.747
Guía Didáctica	12	17.528	-	-	2	2.535	14	20.063
Instructivo	5	8.859	2	3.285	2	5.719	9	17.864
Ley	-	-	-	-	3	67905	3	67905
Manual Técnico	4	84.324	15	241.953	5	137.191	24	463.468
Reglamento	2	4.485	-	-	-	-	2	4.485
Tabla	2	2.197	-	-	-	-	2	2.197
Totales	36	155.160	18	246.374	20	225.256	74	626.790

Lo que primero llama la atención respecto de la distribución de los textos en estas doce categorías es su alto grado de heterogeneidad, detectándose interesantes diferencias entre las tres áreas de especialización (Marítimo, Industrial y Comercial). Aunque cada área puede ser descrita en términos de un perfil de tipos textuales, la clase Manual Técnico resalta como el tipo más frecuente en todo el *corpus*, presente en las tres áreas de especialidad, y el de mayor emergencia respecto al número de palabras en el *corpus* total (da cuenta del 74%). Un análisis más fino de los datos contenidos en la Tabla 5 revela que el mayor número de lexos y, de igual modo, el mayor número de palabras se concentran principalmente en 4 tipos textuales: Manual Técnico (24), Guía Didáctica (14), Instructivo (9) y Descripción Técnica (8).

Según se observa, la mayor diversidad tipológica se registra en el área Marítimo de Operación Portuaria con un total de diez clases textuales; en segundo lugar, se ubica el área Comercial de Contabilidad con un total de ocho tipos. Destaca el área Metalmeccánica Industrial por la escasa variedad tipológica; sólo tres tipos de textos se recolectaron en esta especialidad, de los cuales los manuales explican el 90% de la ocurrencia. Del mismo modo, se aprecia que esta área -con la menor variedad tipológica- es la que muestra la mayor concentración de palabras del total de los tres *corpus* (246.374 palabras). Ella se explica por la alta concentración de Manuales Técnicos como tipo principal de texto, ya que ellos suelen ser de extensión importante, comparados con lo que sucede en los otros tipos de textos que constituyen el *corpus*.

Como una forma de visualizar comparativamente estos cifras, se muestra la ocurrencia de cada tipo textual en el Gráfico 1.

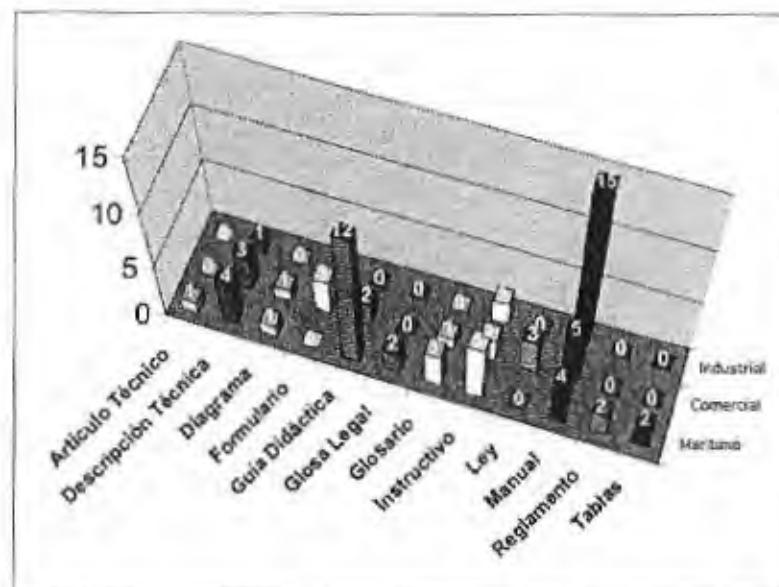


Gráfico 1. Ourrencia de los tipos textuales

A partir de este gráfico, es posible observar que el sector comercial posee ocho clases textuales: Manual, Guía Didáctica, Instructivo, Descripción Técnica, Diagrama, Glosario, Formulario y ley. De esas clases, el Formulario y la Ley son privativos del sector, es decir, sólo se dan en esa área de especialidad. Las clases textuales que no aparecen en este sector son el Reglamento, la Glosa Legal, el Artículo Técnico y la Tabla. El sector marítimo, por su parte, presenta una mayor diversidad de tipos textuales (diez tipos): Manual, Guía Didáctica, Glosa Legal, Instructivo, Descripción Técnica, Diagrama, Glosario, Reglamento, Artículo Técnico y Tabla. Son propias de este sector los reglamentos, la Glosa Legal, el Artículo Técnico y la Tabla. Además, se caracteriza por poseer la mayor cantidad de Guías Didácticas recolectadas (en total 12). Los tipos de textos que no se dan en este sector son el Formulario y la Ley.

El sector industrial, a su vez, posee sólo tres tipos textuales, estos son el Manual, el Instructivo y la Descripción Técnica. No posee un tipo de texto privativo, sin embargo, posee la mayor cantidad de Manuales (en total 15).

Resulta interesante volver a mirar, ahora diagramáticamente, la relación entre el área temática o de especialidad y los tipos de textos que circular, desde aquellos más especializados y propios del ámbito profesional hasta los que se sitúan claramente como típicos del género más divulgativo. Tal como ya se decía más arriba, dos son las áreas que presentan mayor heterogeneidad en los tipos textuales y mayor diversidad en cuanto a la relación entre especializado y divulgativo, a saber, marítima y comercial. En ellas se detecta una - posible - mayor preocupación en la selección de los textos en cuanto a considerar una cierta graduación en los contenidos técnico - científicos, hecho que permitiría en definitiva un mejor acceso al conocimiento profesional, propio del ámbito laboral (Lassen, 2003).

La tensión entre el ámbito de especialización y el continuum de discursos, desde los altamente técnico - científico hasta los de mayor procesamiento con el fin de hacer menor la carga informativa y, por ende, apoyar la transparencia del conocimiento especializado que se quiere comunicar se revela como una relación de extrema importancia en el diseño de materiales divulgativos y didácticos.

En este contexto es que la situación del área industrial aparece como la más preocupante, por cuanto su diversidad tipológica es escasa y se centra en tipos textuales más divulgativos, con escasa aparición de textos técnicos propios del contexto profesional en el que deberán insertarse los actuales alumnos. De cualquier modo, vale la pena enfrentar estos datos con cautela, ya que ellos están tomados de tres instituciones de educación técnico - profesional diferenciada y -obviamente- constituyen una muestra dentro de un universo mucho mayor.

COMENTARIO FINAL Y PROYECCIONES

Los datos aportados en esta investigación, centrada en una clasificación multiniveles con atención a cuestiones lingüísticas, funcionales y situacionales han permitido, entre otros:

- Agrupar los textos del *corpus* según similitudes y diferencias o rasgos compartidos o excluyentes, distinguéndolos unos de otros.

- Definir operacionalmente las clases textuales
- Detectar variabilidad tipológica según criterios y rasgos utilizados.

En cuanto a las tres áreas temáticas de especialización, cabe destacar que el sector marítimo portuario presenta la mayor heterogeneidad de tipos textuales, de los cuales la mayoría corresponde a tipos de textos generados en la comunidad escolar, es decir, textos didácticos propiamente tales dirigidos a audiencias legas y semi-legas. Esto se evidencia en que dicho sector entrega la mayor cantidad de *Guías Didácticas*, *Descripciones Técnicas* y *Glosarios*, todos tipos de textos reconocidos por su carácter didáctico y como propios de la comunidad educativa. Lo anterior se ve complementado con la circulación de variados tipos de textos de especialidad, lo cual nos permite concluir que este sector de especialidad prepararía paulatinamente a sus alumnos en el manejo de un lenguaje progresivamente adecuado a la comunidad profesional.

El sector comercial, si bien comparte con el sector marítimo la heterogeneidad textual, se consigna una menor cantidad de tipos textuales didácticos, adquiriendo relevancia los tipos textuales de especialización. Sin embargo, sigue manteniéndose un equilibrio entre tipos de didácticos y los más especializados. El sector industrial, por su parte, se caracteriza por poseer escasa heterogeneidad textual y –comparativamente– el menor número de documentos recolectados.

Como se aprecia, el ámbito de especialización ha demostrado tener incidencia tanto en la variabilidad de clases textuales detectadas así como en el número de textos que circulan en uno y otro sector. Así, las áreas marítima y comercial evidencian cierta distribución similar en estos terrenos, a diferencia de lo que ocurre con la industrial.

Ahora bien, sin lugar a dudas la aproximación aquí desarrollada apoya la construcción de taxonomías complejas e integradoras en niveles de diversa índole, hecho que resulta en una alternativa más productiva y ayuda a ver datos que de otro modo habría sido imposible o más difícil pescar. Todo ello revela que las propuestas teóricas y aplicadas desde estas vertientes son coherentes y proveen resultados útiles.

En esta línea, a la luz de los datos entregados, es factible obtener un panorama –desde el discurso especializado– acerca de la situación lingüística y didáctica/educativa en que se encuentran las tres áreas técnico-profesionales en cuestión. En nuestra opinión, uno de los aportes valiosos de este estudio lo constituye la heterogeneidad tipológica textual detectada, tanto en lo que respecta a lo que ocurre en cada ámbito de especialidad como en cuanto a la visión de conjunto que la aplicación y análisis de ciertos rasgos específicos otorga. Esta última se aprecia, con especial particularidad, en la determinación de clases textuales originales del ámbito profesional y otras más características de la didáctica escolar. En términos numéricos, esta ocurrencia se aproxima a un equilibrio de ambos tipos de textos según estos rasgos, hecho que resulta altamente satisfactorio. Con ello se supone que estaría siendo cautelada de manera eficiente –al menos en este respecto– la paulatina incorporación de los alumnos a su respectivo campo de especialización profesional.

Estimamos que habría sido inquietante detectar cierto uso excesivo de uno u otro de estos tipos textuales. A partir de ello, es posible hipotetizar que los profesores de estos campos temáticos especializados han realizado –aunque fuera intuitivamente– una cierta

disposición en la selección del material de lectura entregado a sus alumnos. Este equilibrio estaría, idealmente, produciendo un continuum entre tipos de textos dirigidos a un público semi-lega hacia otro con características de expertos propiamente tales. Por supuesto que aún queda por indagar y determinar el modo de empleo y secuencia jerárquica en la lectura y utilización de estos textos, ya que la detección de ambos prototipos textuales [más divulgativo/didáctico y altamente especializado] no implica una entrega y lectura secuenciada de ellos desde lo más divulgativo y didáctico hacia lo más plenamente especializado.

Dado anterior se desprende que se ha privilegiado enfrentar a los alumnos a un conjunto de textos escritos con un grado de mayor transparencia, como son los discursos divulgativo/didáctico. Por ende, se ha procurado –no sabemos si intencionalmente– aligerar la opacidad del lenguaje especializado en virtud de incluir un número importante de materiales escritos más cercanos a alumnos en formación. Con esta equilibración de materiales de lectura se permitiría un mejor acceso a la conceptualización de los contenidos técnico-científicos a que se encuenan los sujetos del campo técnico-profesional y con ello se potenciaría un avance sustancial en la calidad de los aprendizajes. Como bien se sabe, a menudo el proceso de enseñanza-aprendizaje escolar puede volverse más accesible a los lectores si se jerarquizan contenidos y tipos textuales. Esperamos este sea el caso.

Por último, en virtud de este panorama, el programa de investigación en desarrollo en nuestra Universidad debe focalizarse –al menos– en cuatro vías:

- a) Contrastar los tipos textuales detectados por el equipo de investigadores con la opinión de los profesores de las respectivas especialidades técnico-profesionales.
- b) Indagar el modo de entrega o secuencia jerárquica de lectura de estos materiales escolares.
- c) Profundizar en los rasgos lingüísticos de los textos del corpus desde la perspectiva de la lingüística de corpus para alcanzar descripciones más acotadas y de mayor poder explicativo con apoyo de tecnología de vanguardia: como son el uso de los quebradores morfológicos, tematizadores y desambiguadores lingüísticos y estocásticos.
- d) Avanzar en el estudio multirrasgos y multirdimensiones del corpus con el fin de determinar dimensiones o síndromes desde una óptica estadística multivariada.

Capítulo 3

Textos de especialidad y comunidades discursivas técnico- profesionales: Una aproximación a los manuales técnicos

GIOVANNI PARODI

INTRODUCCIÓN

Sostener, hoy en día, que la educación técnico-profesional secundaria está vinculada de manera estrecha al discurso escrito no es una novedad. Sin embargo, afirmar que el desarrollo de las habilidades especializadas en lectura y escritura de textos técnico-científicos constituye un eje fundamental en el acercamiento de estos profesionales a sus comunidades discursivas y, por ende, en su incorporación al mundo laboral es posible que resulte menos obvio. En particular, dado que en Chile aún no alcanzamos los niveles esperados en el manejo de la lengua escrita en el ámbito de la educación general primaria y secundaria, podría suponerse postergable la preocupación por la alfabetización técnico-profesional.

Desde mi punto de vista, tal suposición es evidentemente errónea. Cualquier intento por apoyar la participación como miembros activos de estos profesionales en sus comunidades discursivas está determinado por un manejo eficiente del discurso especializado a través del cual se generan, transmiten, diseminan y adquieren los saberes disciplinarios. El conocimiento acabado de los textos que circulan tanto en los ambientes educativos profesionales como en las organizaciones e instituciones laborales es prioritario para un proyecto de educación técnico-científico secundario, universitario y laboral.

Hoy, existe consenso en cuanto a la abrumadora cantidad de información que nos circunda y ante la cual resulta imposible mantener un grado de actualización en todos los dominios. Sin lugar a dudas debemos ser selectivos. Está claro también que en la educa-

ción formal la aproximación y construcción del conocimiento disciplinar se realiza -inicialmente- a través de la lengua escrita. Unido a esto, es muy cierto que la incorporación a la comunidad discursiva (de manera efectiva y definitiva) se concretará mediante la práctica social comunitaria que implica diversos otros modos de acción, además del discurso escrito.

La llave de la cultura y de la sabiduría, cuestiones mucha más allá de la mera información, no reside en la cantidad de esta última que seguiremos capaces de manejar, sino en el grado de conocimiento significativo que alcancemos y podamos procesar creativamente a partir de ella. Desde esta opción, la construcción de un conocimiento disciplinar no puede ser exigido al profesional en formación de manera automática. Las instituciones formadoras tienen una responsabilidad en ello y deben jugar un papel preponderante: la adecuación de contenidos y metodologías a la luz de los tiempos en que parece primar la "revolución tecnológica" llama a que las instituciones formadoras garanticen una pertinencia en la formación inicial. Es, en este punto, donde la lingüística y sus diversas ramas juegan un rol central: los lingüistas debemos hacernos cargo del conocimiento disponible de tipo científico en determinadas áreas técnico-profesionales y buscar los medios que permitan disponer de la información pertinente para abrir el acceso al discurso especializado de manera efectiva y hacer de este, tal como es, una herramienta poderosa.

Mé parece relevante -en este contexto- insistir que en Chile no conocemos los tipos de textos que se emplean en el sistema educacional secundario técnico-profesional diferenciado. Tampoco tenemos con una descripción -aunque fuera parcial- del perfil de competencia en lengua escrita especializada con que los sujetos egresan de estas instituciones. Tenemos aún menos información acerca de los rasgos lingüísticos y textuales que caracterizan dichos textos especializados escritos. Como se dijo, si la lingüística puede contribuir a ello, una tarea inicial debe ser la recopilación y descripción de los textos que circulan en estos ámbitos y de sus rasgos estructurales y funcionales. Sólo desde allí se puede penetrar en el entramado institucional.

Ahora bien, dentro de este marco y como una forma de contribuir a ello, en la Pontificia Universidad Católica de Valparaíso, hemos decidido llevar a cabo un programa de investigación, inicialmente de corte descriptivo, en el cual interesa conocer, desde diversos ópticos, los tipos de textos escritos que efectivamente se utilizan en el ámbito escolar técnico-profesional en tres diferentes especialidades, a saber, Marítima, Comercial e Industrial. Por una parte, buscamos dar cuenta de los tipos textuales que estos estudiantes enfrentan en su práctica escolar (ver Capítulo 2). Por otra, mediante herramientas computacionales y estadísticas, en el marco de la lingüística de corpus contemporánea, pretendemos llegar a un análisis multirrasgos (AMR) y multidimensiones (AMD) con el fin de identificar (por medio del análisis factorial) los patrones más relevantes de asociación de los rasgos lingüísticos que co-ocurren significativamente en los textos de los corpora recolectados (ver Capítulo 4).

De manera más específica, en este capítulo me concentraré en una extensión de la descripción general del proyecto (ya iniciada en el capítulo anterior), en una caracterización inicial de los corpora recolectados y en una descripción preliminar -a partir de un grupo de rasgos lingüísticos- en un solo tipo textual, el manual técnico. Para ello, en la primera parte del trabajo, presentaré una breve revisión del concepto de texto especializado. También haré referencia a la lingüística de corpus y el uso de grandes muestras de textos con apoyo de computadores. Mostraré de manera sucinta las grandes líneas del programa de

investigación en curso más allá de los límites del presente trabajo. En la segunda parte, entrego una revisión de los pasos metodológicos aplicados. Finalmente, presento los resultados cuantitativos acerca de los tipos textuales detectados en el corpus y efectúo una descripción y análisis de las frecuencias de ocurrencia de sesenta y cinco rasgos lingüísticos en veinticuatro manuales técnicos en tres ámbitos de especialidad. Ciero el capítulo con proyecciones para el área y el proyecto en cuestión.

1. MARCO DE REFERENCIA

1.1 Discurso especializado

Ya destacábamos en el Capítulo 1 que la investigación en torno al llamado discurso especializado, en particular, aquello que se localiza en las diversas modalidades de divulgación científico-tecnológica y en su uso al interior de la sala de clases en el nivel secundario profesional constituye un eje nuclear y estratégico en la planificación lingüística y en el proceso de alfabetización progresiva de las comunidades discursivas en formación.

Desde esta mirada, como la mayoría de los científicos sabemos, hoy en día, existe una gran variedad terminológica en todo ámbito del conocimiento, en particular me refiero a la que ocurre toda vez que requerimos clasificar o tipologizar objetos de estudio. Normalmente, los abordajes son múltiples en razón de supuestos teóricos divergentes. Pero, tal vez, lo más desconcertante sea el hecho de que por razones muy diversas, los especialistas no logramos con facilidad o, por el contrario, no buscamos por decisión coincidir o amalgamar aspectos de las tradiciones. En el caso particular del denominado discurso especial, profesional, técnico, institucional, especializado, etc. alcanzar un relativo orden terminológico y lograr una visión más o menos homogénea tampoco resulta fácil (Cieplausz, 2000; López, 2002). Estoy cierto que no se debe simplificar lo que no corresponde, pero un poco de orden para sobrevivir al caos parece sano.

Tal como se ha adelantado en los capítulos precedentes de este libro, hemos optado por el término discurso de especialidad o especializado, pues nos interesa el discurso de la ciencia y la tecnología en contextos institucionales educativos en donde, hipotéticamente, puede detectarse una gradiente hacia lo más divulgativo. Desde esta óptica, coincidimos con diversos investigadores para quienes el discurso de especialidad se conforma por un conjunto de textos que involucran lóquicos que son prototípicos de un área determinada del conocimiento científico y tecnológico y se ajustan a convenciones particulares de índole lingüística, funcional y situacional, entre otras, propias de las tradiciones de una comunidad discursiva específica.

Ahora bien, desde el marco de la lingüística sistémico funcional (LSF), con el fin de caracterizar un texto como de especialidad, tomo el principio de la diversidad funcional y ocupo a los tres aspectos del contexto situacional a tres variables del registro. Ellos -como bien se sabe- son campo, tenor y modo (Halliday, 1985, 1993; Martín, 1986, 1992, 1997; Halliday & Martín, 1993). En la primera, se considera el espacio institucional o de acción social y se centra en el lóquico o contenido de la actividad junto a la descripción de los participantes que llevan a cabo la comunicación especializada (campo). En la segunda, se considera la estructura de roles sociales, es decir, el tipo de relaciones que se

establecen entre los participantes, en este caso, en relación a la divulgación didáctica (tenor). Y, en el tercero, se focaliza la atención en el rol del lenguaje en la interacción, es decir, en la organización simbólica, ya sea que esto se presente de manera oral o escrita (modo).

Dé manera sucinta, en el caso del discurso especializado, las implicancias lingüísticas caracterizadoras globales serían las siguientes:

- Campo, tecnolectos, procesos descriptivos, taxonomicos
- Tenor, léxico formal, bajo compromiso afectivo
- Modo, escrito (no interactivo, monológico, independiente del contexto)

Con el fin de precisar mayormente la idea de discurso especializado, recurro a la noción de "síndrome" acuñada por Halliday (1993) para destacar que a través de los rasgos co-ocurrentes de variación se puede identificar un tipo de registro determinado.

En palabras de Halliday (1993: 4):

"Cualquier variedad de lengua, sea funcional o dialectal, ocupa un espacio amplio, una región cuyos límites son difusos y dentro de los cuales puede existir considerable variación. Pero puede ser definida y reconocida por ciertos síndromes, es decir, patrones de co-ocurrencia entre rasgos de uno u otro nivel lingüístico - típicamente rasgos de la expresión en el caso del dialecto, rasgos del contenido en el caso de una variedad funcional de 'registro'. Tales síndromes son lo que hacen posible hablar del 'lenguaje de la ciencia'."

Dentro de los discursos de especialidad, los síndromes de rasgos lingüísticos varían en cada caso para satisfacer las necesidades de los hablantes al construir nuevos significados. Estos rasgos pueden ser pesquisados a través de las variables del registro ya comentadas anteriormente, cuyas dimensiones son relevantes ya que ellas, según Halliday (1985, 1993), tienen impacto en el uso lingüístico, en el modo en que usamos el lenguaje, producto de las variaciones en el contexto situacional.

Una aclaración se hace relevante. Aunque adscribimos al paradigma general de la LSF, debemos puntualizar que no coincidimos con la postura de Halliday en cuanto a la "misteriosa entidad llamada mente" (1998: 188). De manera concisa, para Halliday (1994, 1998) la conciencia de orden superior, o sea, la humana, es una conciencia semiótica, que en definitiva también se puede llamar significado. Cabe señalar que Matthiessen (1998) también recurre a la palabra "significado" con el fin de alternar o sustituir la de "pensamiento". En este sentido, el antimentalismo de algunos lingüistas sistémico funcionales no nos permite aceptar de modo radical sus supuestas fundamentales. De manera más profunda, en mi opinión, los procesos semióticos y neurobiológicos, aunque de naturaleza importante en la filogénesis y ontogénesis del hombre no logran -en su conjunto- dar cuenta cabal de la esencia del ser humano como ente autoconsciente y trascendente (Peronard & Gómez Macker, 1985; Gómez Macker, 1998; Parodi, 2003a, 2005a).

LINGÜÍSTICA DE CORPUS COMPUTACIONAL: UN MARCO DE ACCIÓN

Durante un prolongado periodo de tiempo, el enfoque generativo transformacional de Noam Chomsky (1957, 1965) influyó en la postergación de diversos programas de inves-

tigación enfocados al estudio de la variación lingüística y opacó los desarrollos alternativos en otros polos del mundo académico (Francis, 1979; Sinclair, 1982, 1991; Almer & Altenberg, 1991; McEnery & Wilson, 1996). Es bien sabido que el modelo chomskiano al plantear una situación idealizada de comunicación niega la alteridad y, por ende, el individuo pasa a ser un sujeto inexistente. Lo anterior no implica un descensoimiento o falta de reconocimiento al rol que este lingüista desempeñó en el escenario de quebrados paradigmáticos conocidos y, por ejemplo, del nacimiento de interdisciplinas relevantes tal como la floreciente psicolingüística (Peronard, 1994, 1998; Parodi, 1999).

No obstante esto último, es evidente que los postulados chomskianos y su fuerte impacto generaron una discontinuidad en la lingüística de corpus clásica apoyada en la tradición en la lingüística estructural (Fries, 1952). El comienzo de la segunda era de los estudios basados en corpus se puede fijar a comienzos de la década del sesenta, marcada -en parte- por la fuerte irrupción de los computadores en el ámbito lingüístico y el desarrollo de grandes proyectos de investigación en Inglaterra y el norte de Europa, a partir de la construcción de corpus lingüísticos del inglés; ellos constituyen el eje de avanzada de esta nueva repartición. Al respecto, Leech (1991) sostiene que esta nueva lingüística de corpus (de la variedad computacional) extendió su espectro e influencia, de modo que alcanzó hasta el procesamiento de lenguaje natural y se volvió una rama independiente dentro de la lingüística.

Del mismo modo que la hegemonía generativista desestimó inicialmente el estudio del lenguaje natural a través de corpus, también evadió un enfoque de dimensiones probabilísticas. Son varios los investigadores que coinciden en que Chomsky (1957) en su rechazo a un modelo basado en probabilidades y a cuantificaciones de cualquier tipo genera un impacto que influyó negativamente en ciertos desarrollos en lingüística (Leech, 1991; Sinclair, 1991; Stubbs, 1996; McEnery & Wilson, 1996; Kennedy, 1998). Ella debido, en parte, a que se sostendría qué una máquina, a través de un modelo probabilístico, estaría lejos de dar cuenta de un modelo de conocimiento humano. Por supuesto, no asumimos que la máquina puede representar el conocimiento humano ni las habilidades cognitivas de manera definitiva, pero puede ser de gran utilidad para que el científico, a partir de esos datos, alcance mejores interpretaciones de fenómenos en estudio. Los enfoques probabilísticos y basados en corpus no tienen que estar reñidos con visiones cognitivas del lenguaje, por el contrario, los aportes del primero pueden ser de gran utilidad para el segundo y viceversa (Bod, 2003; Jurafsky, 2003).

Ahora bien, todo este movimiento ha conducido a una nueva manera de hacer ciencia en la que se ha enfatizado la preocupación por el uso lingüístico y la variabilidad. En este contexto, es evidente que verímos enfrentando un renacimiento del empirismo sin los residuos de la lingüística estructural de corte behaviorista ni de la psicología conductista importante en los años cincuenta.

Church y Mercer (1993: 21-22) explican que existen tres desarrollos que han galvanizado el renacimiento del empirismo:

1. Los computadores son más poderosos, se encuentran más disponibles y sus costos son menores que en los años setenta y ochenta.
2. Las bases de datos con textos digitalizados ya no están restringidas para usos particulares. Su acceso se ha vuelto más amplio.

3. Debido a cambios políticos y económicos en el mundo, existe un mayor énfasis hoy en día en decisiones y evaluaciones, basadas en el análisis de textos y de grandes conjuntos de datos. Estos análisis han logrado responder exitosamente a las presiones por contar con enormes cantidades de información.

Resulta interesante comprobar que argumentos muy similares a los anteriores son utilizados por otros investigadores para justificar el auge e importancia de la actual lingüística de corpus (Francis, 1979; Conrad & Biber, 1998; Chafe, 1992; Sinclair, 1991; Leech, 1991; Kennedy, 1998; McEnery & Wilson, 1996; Moreno, 1998; Graesser, Gernsbacher & Goldman, 2003). En esta línea, Stubbs (1996) es uno de los fuertes defensores de la lingüística de corpus como teoría. El argumenta que el trabajo con corpus brinda una nueva manera de considerar la relación entre los datos y la teoría, revelando cómo la teoría puede fundarse sobre corpus accesibles de lenguaje natural. Para este investigador, la teoría puede emerger *inductivamente* de los datos, dando así fuerza a una lingüística sustentada en corpus. En palabras de Stubbs (1996: 231):

"La lingüística de corpus todavía cuenta con lineamientos muy preliminares para una teoría que puedan relacionar un texto individual con un corpus de textos, que luego puedan hacer uso de lo que es frecuente en un corpus para identificar lo que es típico en una lengua y puedan utilizar los hallazgos de patrones recurrentes con el fin de construir una teoría de la relación entre el uso voluntario de una lengua y el uso creativo de la misma".

Por su parte, Sinclair (1991) enfatiza que la lingüística de corpus es una técnica cuyo fundamento es el corpus mismo. La clave está en la construcción adecuada de un corpus representativo; de este modo, los resultados generados a partir de dicho corpus tendrán directa relación con la constitución de la base de datos. En la misma línea, Simpson y Swales (2001) argumentan que la lingüística de corpus es esencialmente una tecnología, pero literalmente una tecnología con consecuencias potenciales considerables.

Chafe (1992) parece ser, junto a Stubbs (1996) y Tognini-Bonelli (2001), uno de los más entusiastas respecto a la lingüística de corpus; no obstante, Chafe aboga al igual que Fillmore (1992) por el trabajo mancomunado de técnicas de investigación diversos (tanto cuantitativas como cualitativas), argumentando que las cuantitativas por sí solas no logran revelar los aspectos más profundos del lenguaje y la mente (Chafe, 1994). Su postura es, tal vez, la más interesante y vanguardista en cuanto visualiza que la tarea del lingüista de corpus es tratar de estudiar el lenguaje y, a través de éste, llegar a la mente humana, es decir, indagar la naturaleza del lenguaje como una manifestación de la mente con especial atención a la conciencia humana. No obstante ello, es cauteloso en cuanto a las etiquetas para unir a otro tipo de lingüística y, en definitiva, se inclina por denominaciones más genéricas que no provoquen disputas clásicas: introspección/experimentación (Chafe, 1992, 1994).

Un principio empírico no extraordinariamente novedoso, pero sí muy valioso, que guía estas investigaciones dice que en tanto mayor sea la cantidad de información disponible para el análisis, mayor poder tendrán las conclusiones a las que se arriba. A pesar de que en la actualidad gran parte de las investigaciones acerca del discurso se basan en textos auténticos, muchos de estos trabajos no se apoyan en bases de datos significativas en la línea de la lingüística de corpus. Al respecto, Biber, Conrad y Reppen (1998: 106) opinan que

"La mayoría de las investigaciones no emplean métodos cuantitativos para describir el grado en que las estructuras discursivas se utilizan y relativamente pocos de estos estudios apuntan a producir hallazgos generalizables que se evidencien en los textos".

Aunque en el mundo hispánico se detectan interesantes avances en investigación con textos auténticos la misma que polos de desarrollo tecnológico, en ciertos casos con equipos multidisciplinarios (Santana, Pérez, Carreras Duque, Hernández & Rodríguez, 1997; Raja, 2001; Villaseñor, Montes, Pérez & Valtreydaz, 2002; Ferreira, Campos & Fugger, 1998; Ferreira-Cabrerizo & Atkinson-Abutidy, 2002; Llorente, 2002; Echeverría, 2002; Valiz, 2002; Núñez, Gramajo & Parodi, 2003; Parodi & Gramajo, 2003), es muy cierto que aún la tecnologización de la investigación es escasa y que tanto el desarrollo de programas computacionales como de las tecnologías requeridas no se encuentran accesibles de manera expedita. En mi opinión, la superación de esta barrera metodológica y tecnológica no puede esperar si queremos producir, efectivamente, investigación competitiva y de primer orden, la docencia de pregrado y de postgrado exige que así sea para que –entre otras– la superación de la brecha digital deje de ser una utopía y el acceso al conocimiento especializado esté disponible democráticamente. Por otra parte, es muy cierto que los grandes desarrollos en estas áreas se han producido para la lengua inglesa (entre muchos otros, Francis, 1979; Sinclair, 1982, 1991; Biber *et al.*, 1998; Conrad & Biber, 1998; Simpson & Swales, 2001); en nuestro caso, los costos implicados son de relevancia y los apoyos institucionales y subsidios de investigación no suelen contemplar un gran componente de infraestructura tecnológica.

En suma, en esta investigación adherimos a los principios metodológicos de la lingüística de corpus, cuyo marco de acción resulta muy útil para indagar una lengua en uso y su correspondiente variación, todo ello con apoyo de las tecnologías computacionales de punta. Opino que los avances en este terreno aún hacen difícil pensar en la lingüística de corpus como un paradigma, ya que no se visualiza claramente una teoría lingüística propia y original que la sustente. Es pues, en estos términos, una herramienta metodológica de alta utilidad, dado que los principios que la guían pueden utilizarse, por ejemplo, desde teorías lingüísticas como la LSF de Halliday (1975, 1985). En este sentido, el futuro de la lingüística de corpus es auspicioso y habrá que esperar algún tiempo para ver los resultados y cambios por los que transitará. Al respecto, en un artículo en homenaje a Jan Svartvik, Johansson (1991: 313) concluye así:

"A pesar de los grandes cambios que han ocurrido en menos de tres décadas, desde el primer corpus computarizado, existe un modo en que el rol del corpus en la investigación lingüística no ha cambiado. El corpus permanece como una de las herramientas del lingüista, que es usado junto a la introspección y las técnicas de elicitation. Los sabios lingüistas, como los artesanos de experiencia, afinan sus herramientas y reconocen sus usos apropiados. No es coincidencia que Jan Svartvik haya realizado tan notable trabajo tanto en las áreas de corpus como de la elicitation. Su ejemplo vale la pena mantenerlo fresco en la memoria, cualquiera sean los futuros cambios que se avecinen".

2. PROYECTO EN MARCHA: MARCO GENERAL Y PASOS INICIALES

Como se ha indicado previamente, nos interesa el discurso escrito de especialidad, en su modalidad de divulgación didáctica, como eje de la alfabetización de jóvenes estudiantes de educación secundaria técnico-profesional diferenciada (Glaser, 1982; Halliday & Martin, 1993; Martin, 1992, 1998; Rose, 1997; Veal, 1998; Cameron, 2003; Lasser, 2003). Es un hecho que la realidad escolar y profesional desde esta perspectiva ha sido muy poco indagada en Chile y es escaso el conocimiento del que disponemos acerca de los tipos textuales que circulan en estos ámbitos, a través del que se accede al conocimiento disciplinar. También es escasa o nula la información disponible acerca de la vinculación entre esta formación lingüística especializada y los materiales escritos típicos del ámbito laboral propiamente dicho. Es aún menos lo que se sabe respecto de las prácticas escriturales de estos sujetos en su acercamiento e integración a la correspondiente comunidad discursiva. Otro aspecto interesante de indagar son las concepciones de ciencia y tecnología que ellos poseen, pues es evidente que éstas tendrán relación con su modo de incorporarse a su actividad profesional y de compartir con sus pares. Del mismo modo, nos preocupa el grado de conciencia que posean los profesores de especialidad en cuanto al valor asignado a los textos escritos, por ejemplo, como generadores de la progresiva inscripción en la respectiva comunidad discursiva o como meros objetos reproductores de saberes estéticos y no de competencias comunicativas y progresivamente más expertas.

3. OBJETIVOS Y METODOLOGÍA

3.1 Objetivos

A continuación se entregan los objetivos generales del programa de investigación. Cabe puntualizar que en este capítulo, sólo algunos de ellos serán abordados de manera parcial. En capítulos precedentes y posteriores se exploran otros. Los objetivos en cuestión son:

- Recolectar y tipologizar los textos especializados que se lean en tres ámbitos de Educación Diferenciada Secundaria Técnico-Profesional en algunos establecimientos de Valparaíso, Chile (3º y 4º E.M.)
- Describir, en base a procedimientos de la lingüística de corpus, algunos de los rasgos lingüísticos, textuales y semánticos de esos textos.
- Realizar un perfil de competencia de la habilidad para comprender textos especializados en alumnos de dichos establecimientos educacionales (4º E.M.)

De manera más específica, una vez recolectado el material, se debe:

- Buscar criterios para clasificar los textos según una aproximación multinevles: lingüística, funcional y comunicativa (ver Capítulo 2).
- Establecer una matriz de rasgos lingüístico-textuales, distintivos de variación lingüística y funcional según la bibliografía especializada, con el fin de describir e identificar los rasgos más caracterizadores del corpus técnico-científico (ver diversos capítulos de este libro).
- Comparar el corpus Técnico-científico (CTC) internamente con el fin de detectar diferencias y/o similitudes entre los textos que lo componen (ver Capítulos 2, 5, 6 y 9).

- Comparar las clasificaciones preliminares de tipos de textos en base a criterios textual-comunicativo-situacional con los datos emergentes del análisis factorial a partir de los rasgos lingüísticos en una perspectiva multirrasgos/multidimensiones (ver Capítulo 4).

3.2 Metodología 1: Recolección del corpus

Para llevar a cabo los objetivos, se procedió a identificar tres ámbitos técnico-profesionales relevantes en nuestra región geográfica (Valparaíso, Chile) e iniciar un contacto con estas instituciones formadoras. Tres establecimientos educacionales fueron seleccionados: uno del sector marítimo (Especialidad Operación Portuaria), uno del sector metalmeccánico (Especialidad Mecánica Industrial), y uno del sector de administración y comercio (Especialidad Contabilidad).

Con el fin de acceder y recolectar el material escrito, se decidió entrevistar a los profesores de cada especialidad para pedirles información acerca de los textos que ellos proporcionan a sus alumnos como lectura obligatoria y/o complementaria en sus clases. Optamos por esta alternativa como una forma más directa y eficaz para acceder al material escrito que los alumnos deben enfrentar en su práctica diaria.

Según pudimos comprobar, los medios de acceso para los alumnos son diversos. En muchos casos se recurre a más de una alternativa:

- Se entrega listas de libros, artículos o apuntes existentes en la biblioteca escolar.
- Se entrega apuntes o artículos fotocopiados en la clase.
- Se entrega referencias bibliográficas no existentes en la biblioteca.
- Se indica la forma de acceder a la base de datos en la Intranet del establecimiento escolar, en la que previamente el docente ha incorporado material escrito.
- Se entrega presentaciones impresas de las clases realizadas en programa Power Point.

Una vez conseguida la información preliminar, se procedió a la recolección y posterior organización del corpus. Optamos por recopilar la totalidad de los textos que el docente dice que pone al alcance de los alumnos en cada una de las áreas de especialidad. Aunque inicialmente se había proyectado seleccionar parte del material utilizado, debido a la detección temprana de una muestra limitada, se decidió relevar el universo de los textos indicados por cada profesor. Con ello, se asegura una representatividad del corpus y, en nuestra opinión, su valor adquiere mayores proyecciones.

Todo el material fue catalogado y procesado computacionalmente, es decir, debidamente digitalizado. Para ello, se creó una base de datos a través de la cual el acceso a la interrogación de los corpus es factible con diversos propósitos.

Las tres consideraciones más importantes en el diseño del corpus han sido:

- La representación del universo de textos leídos por los alumnos de las tres especialidades en investigación.
- El tamaño de las muestras (en cuanto al número de textos y número de palabras) para asegurar la utilización de los programas estadísticos requeridos.
- La organización de los documentos de manera que sus divisiones y nomenclaturas apoyen los posteriores pasos metodológicos.

Como se aprecia, una característica esencial del proceso de recolección de información ha sido la recuperación de discurso auténtico en situaciones de uso escolar cotidiano. Cabe destacar, en este aspecto, la colaboración y buena disposición de profesores, bibliotecarios y alumnos de los establecimientos participantes, sin cuya apoyo desinteresado no habría sido posible tener acceso a los textos ni menos lograr los objetivos iniciales en breve plazo.

3.3 Metodología 2: Tipos textuales. En busca de una clasificación

Concluido el relevamiento del corpus, con el objetivo de proponer una primera aproximación tipológica, procedimos al estudio de ciertas características que, -como se dijo- van más allá de lo exclusivamente lingüístico (Schröder, 1991; Cepusio, 2000, 2003; Ciapuccio & Kugel, 2002). La idea era construir una primera matriz de rasgos multinevados (por ausencia o presencia) que nos orientara en la diferenciación y descripción de los tipos de textos recolectados. Buscábamos elaborar una tipología multinevada que incorporara aspectos de la estructura textual, del contexto espacio-temporal, de las relaciones sociales e institucionales y de las funciones comunicativas; en este sentido, nuestro objetivo era una clasificación de base empírica orientada a la complejidad. Esta orientación polifacética implica una postura psicosocial-lingüística cuyos componentes cognoscitivos y pragmáticos logran reflejarse en una taxonomía compleja. Entre otros, este enfoque pretende dar cuenta de nuestra preocupación por el procesamiento del discurso en diversos contextos y por escritores y lectores de índole heterogénea. Cabe señalar que los aspectos lingüístico-gramaticales propiamente tales, fundamentales en una tipología textual según nuestra visión, serán abordados sistemáticamente a través del AMR y del AMD, que se describirán más adelante.

De acuerdo a lo dicho hasta aquí, en la Tabla 1 se entrega un conjunto de rasgos que dan cuenta de los tres criterios nucleares:

Tabla 1. Criterios y rasgos de clasificación textual

CRITERIOS							
Situacional		Funcional	Textual				
Ámbito de producción original	Audencia original	Autor explícito	Funció comunicativa	Estructura textual	Tema	Multimedialidad	Requiere Escritura
Escuela laboral	lego Semilego Experto	Si o No	Referencial Expresiva Apelativa Corativa Poética Metalingüística	Expositiva Argumentativa Narrativa Normativa Narrativa	Mono- temático Plurite- mático	Si o No	Si o No

La aplicación de estos tres criterios y sus rasgos correspondientes proporcionó un agrupamiento de los setenta y cuatro textos del corpus CTC en torno a doce clases textuales. Con el objetivo de visualizar con mayor claridad estas clasificaciones, mediante el análisis cualitativo del material, procedimos a buscar nombres tradicionales del medio técnico pro-

fesional educativo que revelaran la naturaleza de cada uno de ellos. De este modo, recurrimos a denominaciones clásicas para los tipos textuales que esperamos -en la mayoría de los casos- los alumnos y/o profesores del sistema institucional reconocen y posiblemente utilicen intuitivamente. Ellas son las siguientes:

- Artículo Técnico
- Descripción Técnica
- Diagrama
- Formulario
- Glosa Legal
- Glosario
- Guía Didáctica
- Instructivo
- Ley
- Manual Técnico
- Reglamento
- Tabla

Mayores detalles de los pasos investigativos y definiciones operativas se encuentran en el Capítulo 2 de este libro.

3.4 Etiquetaje lingüístico del corpus: Búsqueda de tecnología de punta

En parte, debido a la insuficiente capacidad de manejar y realizar estudios comparativos más abarcadores en base a la totalidad del corpus recolectado con tecnologías de vanguardia así como el deseo de trabajar de manera más profunda las descripciones lingüísticas con apoyo estadístico, decidimos buscar un convenio académico con el Instituto Universitario de Lingüística Aplicada (IULA) de la Universidad Pompeu Fabra (UPF), España. Ello, con el objetivo de acceder a programas computacionales poderosos que permitieran contar con un corpus etiquetado estructuralmente. El procedimiento aplicado a los textos por parte del equipo chileno-catalán consistió en

- Codificación SGML (Standard Generalized Mark Up Language)
- Partidor o separador de oraciones (Splitter o chunker)
- Marco morfológico (Paquete IIC-UFF)
- Desambiguador lingüístico y estocástico

En virtud del convenio firmado con la UPF, los documentos digitalizados del corpus PUCV-2003 estuvieron disponibles a través de una interfaz computacional en Internet. Ello permitió un acceso directo y expedito a la totalidad de los textos recopilados y a una diversidad de alternativas de interrogación con resultados cuantificados y exemplificados en cada caso, limitadas exclusivamente por el tipo de marco de naturaleza morfológica; de modo que el alcance sintáctico y semántico está restringido por el tipo de etiquetas lingüísticas empleadas y por las posibilidades de interrogación.

Paralelamente, el equipo que dirijo en la PUCV desarrolló un programa computacional denominado BUCÓLICO (Buscador de Concordancias Lingüísticas en Corpus), cuyo objetivo central es a partir del corpus PUCV-2003 etiquetado con tecnología del IULA (UPF),

proporcionar alternativas de análisis no contempladas en la interfaz del IIUA, tales como N-gramas, y realizar cálculos y gráficos en base a las ocurrencias y comparaciones detectadas. Este desarrollo tecnológico multidisciplinario (lingüistas, ingenieros computacionales, estadísticos, diseñadores gráficos, entre otros) es, en efecto, una herramienta comparativamente rudimentaria con los avances tecnológicos detectados fuera de Latinoamérica, pero que se constituye en un primer avance en el diseño de instrumentos computacionales elaborados en el marco del programa en desarrollo (ver Capítulo 9).

Dado que el convenio con la UPF estaba restringido a este corpus, también decidimos emprender el diseño de implementar nuestro propio sistema de etiquetaje e interrogación de corpus. Con este fin se adquirió un paquete de alto poder tecnológico para la lengua española (Connexor) y se diseño e implementó una interfaz computacional interactiva a través de Internet. Esta plataforma, denominada *El Grid* (www.elgrid.cl), permite el marcado morfológico de textos y su posterior interrogación. Los productos de las búsquedas pueden ser almacenados temporalmente y los usuarios pueden obtener diversas tipos de información acerca de sus textos.

3.5 Rasgos lingüísticos

Respecto a los rasgos lingüísticos que se indagarán, se decidió elaborar un conjunto inicial de diecisésis categorías representativas de características gramaticales y funcionales del español. Ellos se rastrearon a partir de bibliografía relevante en el tema. El listado de estos diecisésis rasgos se presenta a continuación en la Tabla 2.

Tabla 2. Categorías de base para matriz de rasgos lingüísticos

Matriz de rasgos lingüísticos de base	
A. Marcadores de tiempo verbal	I. Tipos verbales
B. Marcadores de modo verbal	J. Verbos modales
C. Desinencias verbales de persona	K. Marcadores de modalidad
D. Pronombres personales	L. Adverbios
E. Formas nominales	M. Marcadores de subordinación
F. Formas pasivas	N. Frases preposicionales y adjetivas
G. Especificidad léxica	O. Marcadores de coordinación
H. Formas estativas activas	P. Marcadores de negación

A partir de estas diecisésis categorías iniciales, se procedió a construir una matriz más específica de los rasgos caracterizadores del español; de este modo, se llegó a un total de sesenta y cinco rasgos lingüísticos de importancia lingüística y funcional. Paralelamente, la matriz de rasgos debió ser colejada con las posibilidades de marcado e interrogación que brindaban los procedimientos tecnológicos disponibles. A continuación, en la Tabla 3 se presenta el total de estos rasgos, agrupados según las categorías iniciales.

Tabla 3. Rasgos lingüísticos Proyecto Corpus PUCV-2003

Rasgos lingüísticos Proyecto Corpus PUCV-2003	
A. Marcadores de tiempo verbal	I. Tipos verbales
1. Pretérito indefinido (indicativo)	35. Fábulas
2. Pretérito imperativo (indicativo)	36. Paredes
3. Pretérito perfecho (indicativo y subjuntivo)	37. Persuasivos
4. Presente (indicativo y subjuntivo)	38. Perceptivos
5. Futuro (indicativo y subjuntivo)	J. Verbos modales
6. Futuro geotáctico	39. Posibilidad
B. Marcadores de modo verbal	40. Necesidad
7. Indicativo/impérativo	41. Obligación
8. Subjuntivo/impérativo	42. Volición
9. Modo indicativo	K. Marcadores de modalidad
10. Modo subjuntivo	43. Atenudadores
11. Modo imperativo	44. Enfáticos
C. Desinencias verbales de persona	L. Adverbios
12. Primera singular	45. De lugar
13. Segunda singular	46. De tiempo
14. Tercera singular	47. De modo
15. Primera plural	48. De cantidad
16. Segunda plural	M. Marcadores de subordinación
17. Tercera plural	49. Subordinadas sustantivas con "que"
D. Pronombres personales	50. Subordinadas adjetivas pronominal relativo
18. Primera persona singular	51. Subordinadas adverbiales de razón o causa/efecto
19. Primera persona plural	52. Subordinadas adverbiales de concesión
20. Segunda persona singular	53. Subordinadas adverbiales condicionales
21. Segunda persona plural	54. Subordinadas adverbiales de tiempo
22. Tercera persona singular	55. Frases infinitivas en función nominativa
23. Tercera persona plural	N. Frases preposicionales y adjetivas
24. Demostrativo	56. Frases prep. (compl. del nombre)
E. Formas nominales	57. Adjetivos atributivos (calificativas)
25. Nominalizaciones	58. Adjetivos predicativas
26. Sustantivos (comunes y propios)	59. Adjetivos demostrativas
F. Formas Pasivas	60. Participio función adjetiva
27. Pasivas con "se"	G. Marcadores de Coordinación
28. Pasivas con ser sin agente	61. Conjunciones adversativa, aditiva y disyuntiva
29. Pasivas con ser con agente	H. Marcadores de negación
30. Pasivas con estar	62. Adverbios de negación
G. Especificidad léxica	63. Adverbios de tiempo
31. Relación clase/tipo (type/token) por forma	64. Conjunción de negación
32. Relación clase/tipo (type/token) por lema	65. Pronombres de negación
H. Formas estativas activas	
33. Ser	
34. Estar	

4. RESULTADOS

En este apartado, se entregan tres tipos de resultados empíricos, algunos de orden preliminar: (1) descripción global del tamaño del *corpus*; (2) descripción cuantitativa de la ocurrencia de los tipos textuales recolectados antes mencionados, y (3) frecuencia de ocurrencia de los sesenta y cinco rasgos lingüísticos en el total de los textos identificados como *Manual Técnico* en cada uno de los tres ámbitos de especialidad.

4.1 El Corpus Técnico-Científico (CTC)

El Corpus Técnico-Científico (CTC) está compuesto por setenta y cuatro textos con un total de 626.790 palabras, recolectado en establecimientos secundarios técnicos profesionales de la ciudad de Valparaíso, Chile. El detalle de esta información se entrega en la

Tabla 4. Constitución del CTC

Ámbito Técnico-científico	Número de textos	Número de palabras
Operación Portuaria	36 (49%)	155.160 (25%)
Mecánica Industrial	18 (24%)	246.374 (39%)
Contabilidad	20 (27%)	225.256 (36%)
Totales	74 (100%)	626.790 (100%)

Tabla 4:

Los datos de la Tabla 4 muestran que el número de textos por área temática y el correspondiente número de palabras se distribuye diferencialmente. Como se aprecia, en el ámbito Marítimo (Especialidad Operación Portuaria) si bien se cuenta con el mayor número de textos (49% del total), se aprecia el menor número de palabras (25% del total); por el contrario, en el área Metalmecánica (Especialidad Mecánica Industrial) se registra una cantidad más reducida de textos (sólo 18, que alcanzan el 24% de lo recopilado). Sin embargo, estos 18 textos constituyen el grupo más extenso según el número de palabras (39% del total). Por último, el área técnica de Administración y Comercio (Especialidad Contabilidad) cuenta con cifras similares a las anteriormente descritas; un total de 20 textos (27% del total) y con un número importante de palabras (36% del total).

Estas cifras revelan una cierta heterogeneidad respecto a la configuración del *corpus* total, en consideración a cada uno de los tres ámbitos. También muestran que no existe una relación directa entre área de especialidad y porcentaje de textos y de palabras. Explicaciones a estos datos habrá que indagarlas en el estudio más detallado de los tipos de textos que circulan en cada área, pues ello podría tener que ver con la cantidad de material que los alumnos efectivamente leen y con el o los tipos de textos a que se exponen los sujetos en su acercamiento a la disciplina profesional. En todo caso, el número de palabras por ámbito técnico es relevante en términos estadísticos y ello asegura que se podrán efectuar los análisis requeridos para el diseño experimental, en especial, el que dice relación con el análisis factorial.

4.2 Los tipos textuales en el CTC

A partir de los doce tipos de textos técnicos científicos detectados en el *corpus*, según criterios situacionales, funcionales y textuales comentados anteriormente, a continuación se entrega una tabla descriptiva con datos numéricos relevantes que dan cuenta detallada de la distribución de los mismos en el *corpus* CTC.

Tabla 5. Ocurrencia de los doce tipos textuales en los tres *corpus* del CTC
(Nota: las celdas marcadas muestran las principales ocurrencias)

Clase Textual/Área Especialidad	Operación Portuaria (CTCma)		Mecánica Industrial (CTCin)		Contabilidad (CTCcom)		Totales	
	Nº Textos	Nº Palabras	Nº Textos	Nº Palabras	Nº Textos	Nº Palabras	Nº Textos	Nº Palabras
Artículo Técnico	1	9.346	-	-	-	-	1	9.346
Descripción Técnica	4	17.748	7	1.135	3	6.287	8	25.170
Diagrama	1	13	-	-	1	103	2	116
Formulario	-	-	-	-	3	2.287	3	2.287
Glosa Legal	2	4.142	-	-	-	-	2	4.142
Glosario	3	6.518	-	-	1	3.229	4	9.747
Guía Didáctica	12	17.528	-	-	2	2.535	14	20.063
Instructivo	5	8.859	2	3.285	2	5.719	9	17.884
Ley	-	-	-	-	3	6.7905	3	6.7905
Manual Técnico	4	84.324	15	241.953	5	137.191	94	463.468
Reglamento	2	4.485	-	-	-	-	2	4.485
Tabla	2	2.197	-	-	-	-	2	2.197
Totales	36	155.160	18	246.374	20	225.256	74	626.790

Lo que primera llama la atención respecto de la distribución de los textos en estas doce categorías es su alto grado de heterogeneidad, detectándose interesantes diferencias entre los tres ámbitos de especialización (Marítimo, Industrial y Comercial). Aunque cada área puede ser descrita en términos de un perfil de tipos textuales, la clase *Manual Técnico* resulta como el tipo más frecuente en todo el *corpus*, presente en las tres áreas de especialidad, y el de mayor envergadura respecto al número de palabras en el *corpus* total (de cuenta del 74%). Un análisis más fino de los datos contenidos en la Tabla 5 revela que el mayor número de textos y, de igual modo, el mayor número de palabras se concentran principalmente en 4 tipos textuales: *Manual Técnico* (24), *Guía Didáctica* (14), *Instructivo* (9) y *Descripción Técnica* (8).

Según se observa, la mayor heterogeneidad tipológica se registra en el área Marítima de Operación Portuaria con un total de diez tipos textuales, en segundo lugar, se ubica el área Comercial de Contabilidad con un total de ocho tipos. Destaca el área Metalmeccánica Industrial por la escasa variedad tipológica: sólo tres tipos de textos se recopilaron en esta especialidad, de los cuales los manuales explican el 90% de la ocurrencia. Del mismo modo, se aprecia que esta área -con la menor variedad tipológica- es la que muestra la mayor concentración de palabras del total de los tres corpora (246.374 palabras). Ello se explica por la alta concentración de Manuales Técnicos como tipo principal de texto, ya que ellos suelen ser de extensión importante, comparados con lo que sucede en los otros tipos de textos que constituyen el corpus.

4.3 El análisis multirrasgos (AMR) en los manuales del CTC

Con el fin de mostrar -preliminarmente- la utilidad de contar con un número importante de rasgos gramáticos para la indagación profunda de las características de los textos recolectados, a continuación, se entregan tres gráficos que muestran la frecuencia de ocurrencia de las sesenta y cinco variables lingüísticas desagregadamente. Dada que los manuales técnicos concentran el 74% del total de palabras de la muestra de textos, se optó por utilizar este tipo textual por área de especialización en esta aproximación a los resultados.

Por razones de espacio, en los gráficos, las variables lingüísticas se han numerado para su mejor visualización. En la Tabla 3, presentada más arriba, se dio cuenta de la relación rasgo lingüístico y el número asignado.

Con el fin de que la comparación de los datos sea representativa, éstos han sido normalizados a un texto de 1.000 palabras. Este proceso de normalización es crucial dada la variación en la extensión de los textos y, por ende, la diversidad de frecuencias de distribución posibles. Tal como se indicó, para efectos de una presentación visualmente más comprensible, se ha dividido los sesenta y cinco rasgos en tres grupos. Los tres gráficos siguientes entregan las cuantificaciones de las ocurrencias de cada variable lingüística.

El Gráfico 1 da cuenta de los primeros veinte rasgos lingüísticos. Ellos dicen relación (de izquierda a derecha) con rasgos de tiempo verbal (presente, pretéritos y futuros) y de modo verbal, de desinencias verbales de persona (primera, segundo y tercera singular; primera, segunda y tercera plural) y de pronombres personales (de primera persona singular y plural y de segunda persona singular). Una primera observación general permite apreciar una interesante similitud en el comportamiento de las cifras. En la mayoría de los casos, las frecuencias presentan una distribución más o menos homogénea tanto en las ocurrencias menores como en las más destacadas. En principio, tres comentarios resultan pertinentes. Por un lado, los manuales del área marítima se distancian de las otras dos áreas por una mayor frecuencia de ocurrencia del rasgo 4 (presente del indicativo y del subjuntivo), de los rasgos 7 y 9 (indicativo/impérativo y subjuntivo/impérativo, respectivamente) y de los rasgos 14 y 17 (desinencias de tercera persona singular y plural). Por otro, la mayor frecuencia de ocurrencia en estos cinco rasgos se presenta en las tres áreas, aunque el ámbito marítimo es el que presenta siempre el punto más alto. Por último, los manuales técnicos de las áreas industrial y comercial -en lo que a estos veinte rasgos importa- revelan una distribución de frecuencias altamente parecidas, lo que indicaría que

Manuales Técnicos y Rasgos por Área (1-20)
(Tiempo y Modo Verbal - Pronombres y Desinencias de Persona)

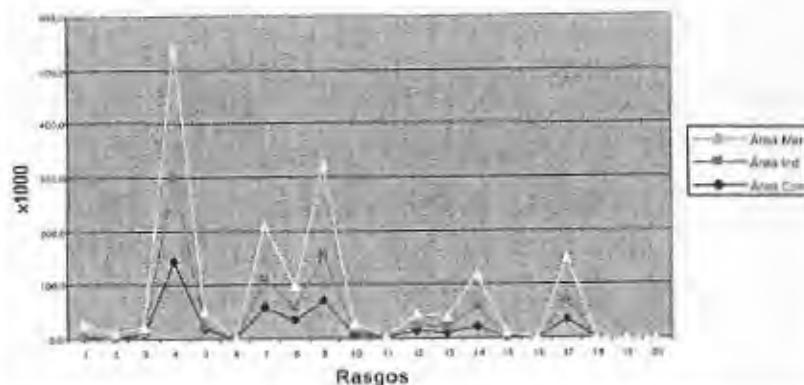


Gráfico 1. Primeros 20 rasgos por área de especialidad

sus organizaciones estructurales podrían ser muy similares en ciertos puntos. No es de extrañar la mayor frecuencia de ocurrencia de rasgos como el presente de indicativo e imperativo, si se piensa que el tipo textual en cuestión -de acuerdo a la propuesta de taxonomía multinivel- está caracterizado con un foco referencial y una posible función secundaria de lódoe y apelativa. Algunos de estos rasgos apuntan también a una estructura textual predominante de tipo expositivo-normativa.

En lo que respecta a los rasgos de desinencia verbal y pronombres personales (rasgos del 12 al 20), los datos graficados muestran mayor homogeneidad, particularmente en las frecuencias de ocurrencia de los pronombres personales de primera (singular y plural) y de segunda (singular). No es extraño que en este tipo de texto sea regularmente escasa la presencia explícita de pronombres personales (rasgos 18, 19 y 20); bien sabido es que en el español escrito académico la tendencia es a su normal elisión.

De manera general, se debe destacar que un número importante de estos veinte primeros rasgos en comparación comparte una frecuencia de distribución en patrones relativamente similares. Veamos, a continuación, los resultados de otros rasgos en el mismo tipo de texto.

Los siguientes veinte rasgos lingüísticos presentados en el Gráfico 2 muestran una frecuencia de ocurrencia altamente más reveladora que en el caso anterior. En este caso, los rasgos son: pronombre personal explicito (de segunda persona plural; de tercera persona singular y plural y, de pronombres demostrativos); nominalizaciones y sustantivos comunes y propios; cuatro formas pasivas (con "se"; con "ser" sin agente; con "ser" con agente; y con "estar"); especificidad léxica (relación tipo/clase de formas y de lemas); formas estativas activas (verbos ser y estar); tipos verbales (públicos, privados, persuasivos, perceptivos) y verbos modales (posibilidad y necesidad).

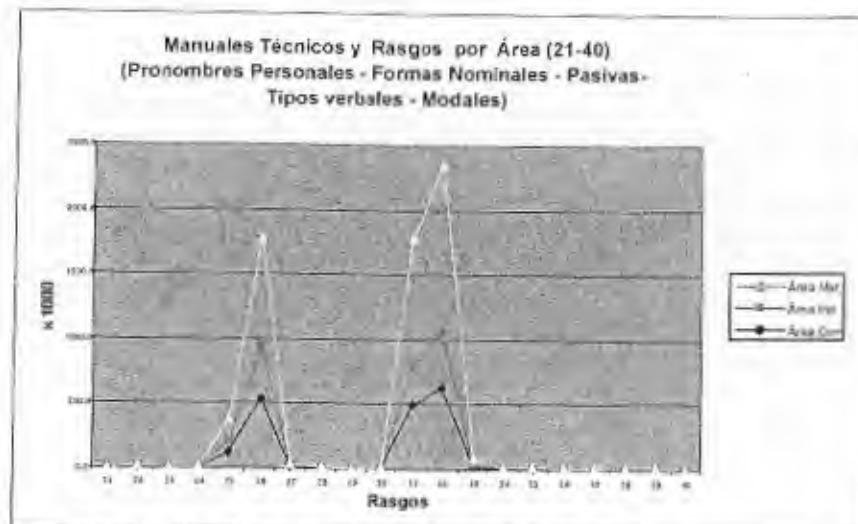


Gráfico 2. Segundo grupo de 20 rasgos por área de especialidad

Como se anticipaba, los datos de este gráfico revelan una distribución bastante sistemática en las frecuencias de los rasgos involucrados, hecho que, por un lado, entrega información relevante acerca del tipo de texto especializado que estamos describiendo y, por otro, permite inferir un impacto mayor en el área técnico-profesional de la clase textual "marítimo". En otras palabras, dado que los Manuales Técnicos dan cuenta (numéricamente hablando) de un alto porcentaje del corpus (74% del total de palabras) y que aparentemente entre ellos no habría un número importante de diferencias significativas en cuanto a la caracterización por ocurrencia de rasgos, es posible pensar en un género (o macrogénero) textual que podría tener un perfil muy similar en algunos niveles y áreas de especialidad (Lossen, 2003). Por supuesto que todo ello habrá de comprobarse, por un lado, con herramientas estadísticas y, por otro, con análisis cualitativos más finos.

Junto a lo anteriormente dicho, no es menor que tanto las oraciones pasivas y las nominalizaciones, así como cierto tipo de verbos parecen estar indicando que su frecuencia de ocurrencia en los manuales técnicos de las tres áreas estudiadas tienden a presentar un patrón de ocurrencia muy homogéneo. Sólo tres rasgos (26, 30 y 32) de los veinte en estudio muestran diferencias importantes, nuevamente a favor del área marítima. Los rasgos en cuestión dicen relación con los sustantivos comunes y propios, las pasivas con verbo estar y la relación clase/lema por lema (Type/Token). Estos tres rasgos en los manuales del área marítima muestran una frecuencia divergente entre las diferentes áreas técnicas de los manuales, revelándose como una posible característica diferenciadora entre ellos. No deja de ser interesante que nuevamente sean los manuales del área marítima los que destaque los más altos niveles de frecuencia de ocurrencia en estos rasgos distintivos. Ello indica que la mayor presencia de este tipo de categorías gramaticales en los manuales del sector marítimo de operación portuaria contribuye a algún tipo de organización textual particular de alta densidad léxica y – posiblemente – alta carga informativa.

Por último, en el Gráfico 3 se entregan los 25 rasgos restantes y sus frecuencias de ocurrencia.

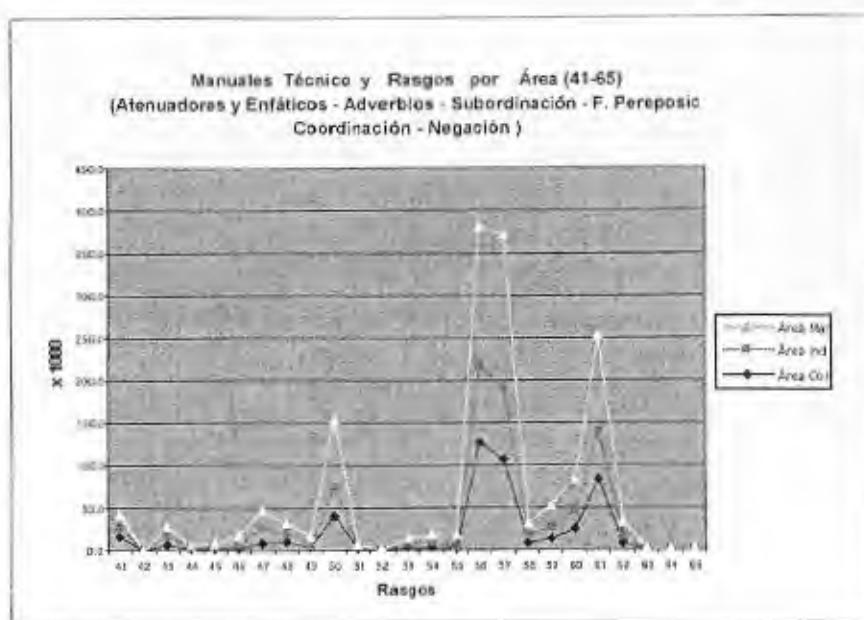


Gráfico 3. Tercer grupo de 25 rasgos por área de especialidad

Este tercer gráfico, en el que se incorporan los últimos veinticinco rasgos de los MT del CTC, los datos se ordenan en siete grandes grupos: verbos modales (obligación y voluntad), marcadores de modalidad (atenuadores y enfatizadores), adverbios (lugar, tiempo, modo y cantidad), marcadores de subordinación (cláusulas nominales, cláusulas relativas, cláusulas adverbiales de causa-efecto, concesión, condición, tiempo, frases infinitivas en función nominal), frases preposicionales y adjetivas (frases preposicionales en función de complemento del nombre, adjetivos atributivos y predicativos, adjetivos demostrativos, participios en función adjetiva), coordinación (conjunctiones aditivas, adversativas y disyuntivas) y marcadores de negación (adverbios de tiempo y negación, pronombres, conjunciones).

La distribución en la frecuencia de ocurrencia de estos veinticinco rasgos restantes guarda estrecha relación con lo dicho respecto de los dos gráficos anteriores. Esto quiere decir que en la mayoría de los rasgos se observa una ocurrencia bastante pareja, sólo en cuatro de ellos se aprecia un aumento en el número de la frecuencia de aparición y un distanciamiento entre las tres áreas de especialidad, nuevamente con predominio del sector marítimo. La posible diferencia significativa entre la ocurrencia de rasgos se materializa en: las subordinadas adjetivas con pronombre relativo, frases preposicionales como complemento del nombre, los adjetivos atributivos y las conjunciones en tres tipos (adversativas, aditivas y disyuntivas). Son ellos los que producen esta distinción entre los ámbitos técnico-profesionales.

les. Cabe enfatizar que la homogeneidad detectada entre la mayoría de los rasgos no debe ser causa para no indagar posteriormente la incidencia de estos rasgos entre los tipos de áreas involucradas.

Ahora bien, este análisis preliminar a partir de los resultados entregados en los tres gráficos precedentes permite señalar que -en base a estos sesenta y cinco rasgos lingüísticos tomados en su conjunto- los manuales de las tres áreas en estudio parecen poseer gran homogeneidad en el modo de estructurar la información. Como se aprecia, no se detectan variaciones importantes entre las ocurrencias de los rasgos, ello quiere decir que la estructura lingüística general de estos textos podría no presentar diferencias determinantes y tender a una organización de los contenidos de modo más o menos similar. En efecto, ello confirma de manera destacable la constitución de un tipo textual por parte de los manuales técnicos. Del mismo modo, estos hallazgos permiten corroborar la tipologización inicial de estos veinticuatro textos dentro de un grupo de características similares en base a criterios funcionales, situacionales y textuales.

Por otro lado y no menos importante, también de acuerdo al estudio comparativo de los sesenta y cinco rasgos lingüísticos de los Gráficos 1, 2 y 3, se puede sostener que la clase textual manual técnico está por sobre el criterio de especialidad; esto es, no se detecta un comportamiento divergente significativo de los rasgos lingüísticos en cuestión entre los ámbitos de especialidad. Aunque no se ha aplicado aún un análisis estadístico, la cuantificación revela gran similitud de ocurrencia en la mayoría de los rasgos, sin que la pertenencia al área marítima, industrial o comercial parezca incidir de manera determinante o mostrar tendencias diferenciadoras.

COMENTARIOS FINALES Y PROYECCIONES

El estudio del discurso de especialidad en sus diversas modalidades está en la actualidad en un momento de auge. Diversos proyectos se están llevando a cabo en varias partes del mundo y Latinoamérica no es una excepción (entre otros, Moyano, 2000; Cubo de Severino, 2002; Ciapuscio & Orelli, 2002; Harvey, 2004; Oyanedel, 2004; Ciapuscio, 2004; Becke, 2005). A pesar de ello, tal como puntualicé en la introducción a este capítulo, se cuenta con un conocimiento escaso o nulo acerca de aspectos lingüístico-discursivos de los materiales escolares y de las prácticas lingüísticas de los alumnos en preparación para el mundo técnico-profesional; de modo más general, tampoco se maneja información acerca de los procesos de alfabetización técnico-científica en estas comunidades discursivas.

Con las ideas anteriores en mente, espero que la información aportada en este trabajo contribuya a la elaboración de un perfil lingüístico acerca de los alumnos de áreas técnico-profesionales en la educación diferenciada secundaria. Por cierto, en este estudio, he aportado solo información parcial basada en solo algunos establecimientos educativos y sólo tres de las muchas áreas técnicas existentes en Chile. También es evidente que la realidad geográfica de nuestro país impone temáticas, ritmos y urgencias muy divergentes y que cualquier intento homogenizador es peligroso. En este sentido, espero que los resultados preliminares que entrego constituyan un aporte inicial para visualizar mejor el papel de las prácticas discursivas en la alfabetización de los sectores técnico-profesionales de la educación secundaria en Chile y, de modo más particular, se afiance el rol que el discurso escrito

[en este caso] puede ostentar como eje fundante de las relaciones en comprensión y producción lingüística. Sin lugar a dudas, las decisiones informadas académicamente aportarán al mejoramiento de ciertas políticas educativas. De manera más específica, se espera contribuya a la revisión y redirección de didácticas escolares de divulgación del conocimiento especializado.

La metodología utilizada en la determinación de las clases textuales del corpus PUCV-2003 desde una óptica compleja multiniveles -por una parte- ha mostrado ser una herramienta útil que distingue doce clases de textos al interior de los setenta y cuatro documentos recogidos en las tres áreas de especialización (ver Capítulo 2). Por otra, la aproximación multirrasgos -aplicada de manera preliminar en este trabajo- al total de los veinticuatro Manuales Técnicos del corpus ha revelado que estos textos -en base a los sesenta y cinco rasgos lingüísticos- tienden a agruparse con un comportamiento relativamente homogéneo. Ello prueba que constituyen lo que inicialmente y, en base a otros criterios, se definió como un tipo textual particular.

Aunque son un número menor en relación al total de los sesenta y cinco rasgos lingüísticos explorados, existen diez rasgos que tendrían relevancia en la detección de la variabilidad emergente del área marítima, en comparación con las otras dos áreas de especialización (comercial e industrial), cuyo comportamiento se deriva como mucho más homogéneo. Su estudio requiere mayor atención desde diversas ópticas. La mayor o menor incidencia definitiva de este hecho habrá de ser comprobada empíricamente a posteriori por medio de los correspondientes análisis estadísticos y de un estudio cualitativo comparativo detallado.

Sin lugar a dudas, la incorporación de tecnología computacional avanzada así como la de los principios metodológicos de la lingüística de corpus computacional permitieron alcanzar conclusiones relevantes, hecho que habría sido imposible de manejar en los tiempos empleados considerando el número de Manuales Técnicos (24), el de palabras involucradas (463.468) y el de rasgos lingüísticos representativos (65). Es importante destacar que conclusiones basadas en grandes corpus de textos naturales y combinados con múltiples variables trabajados simultáneamente -como ha sido el caso de esta investigación- brindan una mayor confiabilidad y robustez a las conclusiones prouestas y -en definitivo- a los estudios lingüísticos y discursivos en general. La posibilidad de explorar la variabilidad inherente a gran escala es también una fortaleza no menor.

Ahora bien, el desafío de enfrentar un estudio de dimensiones como el aquí esbozado conlleva de siyá una serie de riesgos importantes. Si a ello sumamos la escasa bibliografía hispánica producida en una línea funcional/comunicativa desde la gramática y el discurso, las complicaciones aumentan. No obstante, los logros alcanzados hasta este momento son altamente satisfactorios.

Sin lugar a dudas, es necesario enfatizar que el estudio de orden cuantitativo descriptivo como el aquí realizado no es suficiente para obtener una panorámica de los textos de especialidad, como se indicó al inicio del trabajo, esa es sólo una parte de la mirada y se debe complementar con los análisis estadísticos pormenorizados junto a los análisis cualitativos más enriquecedores, con el fin de observar en detalle el comportamiento de estos rasgos en las tramas del discurso especializado (ver Capítulos 7 y 8).

Finalmente quiero destacar que un objetivo emergente a partir de este artículo lo com-

lityó la ampliación de los corpus del proyecto original. Con el fin de alcanzar una descripción exhaustiva del corpus técnico-científico (CTC), se optó por recolectar y etiquetar otros dos corpus no técnicos, cuyo procesamiento ya entró en la fase final para los cálculos estadísticos requeridos. Ellos se componen, por un lado, de un grupo de textos orales interactivos a base de entrevistas (CEO) a alumnos de las mismas edades y, en muchos casos, de los mismos colegios de la muestra en estudio; por otro, disponemos de un tercer corpus a base de libros de literatura latinoamericana (CLL), textos que los mismos alumnos de los establecimientos técnico-profesional reciben como lecturas obligatorias en la asignatura de lengua castellana y comunicación. Ellos permitirán, por una parte, efectuar comparaciones profundas de los rasgos lingüísticos caracterizadores entre estos tres diferentes registros (técnicos, no técnicos; orales y escritos), hecho que entre otros determinará la real incidencia y representatividad del CTC y, en particular, del Manual Técnico. También aportarán de manera certeza a la comparación estadística a realizarse en el Análisis Multirrasgos y el de Multidimensiones (ver Capítulo 4).

Capítulo 4

Lingüística de corpus y análisis multidimensional: Exploración de la variación en el Corpus PUCV-2003

GIOVANNI PARODI

INTRODUCCIÓN

Los corpus de textos naturales computarizados han mostrado un importante impacto en los análisis lingüísticos en las últimas dos o tres décadas. En particular, las investigaciones acerca de la lengua inglesa, algunas europeas y ciertas asiáticas han revelado que los estudios lingüísticos basados en grandes muestras de textos digitales no siempre corroboran las intuiciones iniciales de los investigadores. El uso de corpus con sustento digital, así como la disponibilidad de programas computacionales que permiten su tratamiento y posterior análisis han impulsado las indagaciones lingüísticas hacia fronteras antes imposibles.

El empleo conjunto de aproximaciones lingüísticas basadas en corpus y de enfoques computacionales no ha sido una combinación habitual en el ámbito de las ciencias del lenguaje ni en el tratamiento del lenguaje natural. Desafortunadamente, los unos han tendido a operar independientemente de los otros. Esto ha conducido a que un reducido número de textos sea analizado manualmente sin aprovechar los beneficios de la digitalización de los corpus y los robustos programas computacionales de análisis automático; además, muchas veces los estudios se focalizan en solo unas pocas oraciones sin incorporar los avances de la lingüística del discurso con un nuevo objeto de ciencia, potencialmente mucho más productivo.

Se hace necesario buscar la complementación entre ambas polurias (lingüística e informática), tendiendo al equilibrio y evitando el

aislamiento perjudicial. De no ser así, las investigaciones deben asumir limitaciones a su potencial; relativamente pequeñas muestras de texto son analizadas y los resultados pierden robustez. Como es de esperar, algunos investigadores en ambas áreas científicas han combinado los recursos disponibles y han capitalizado los esfuerzos implementados. Diversos proyectos, en particular para la lengua inglesa, desde hace ya más de treinta años han explotado el uso de grandes muestras de textos digitalizados y de programas computacionales automáticos para el etiquetamiento morfológico y sintáctico de los documentos. En estas investigaciones se ha explorado una amplia gama de rasgos lingüísticos a través de enormes cantidades de textos pertenecientes a variadas clases textuales.

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 localiza el llamado discurso divulgativo de especialidad (Ciapuccio, 2000, 2003; Ciapuccio & Kugel 2002; Calsamiglia, 2000; Cassany, López & Martí, 2000; López, 2002; Cademartori, 2003) o aborda los marcadores discursivos (Portolés 1998, Martín Zorraquino & Portolés, 1999; Montolio, 2001); otros, se concentran en algunos pocos rasgos lingüísticos (Ciapuccio, 1992; Harvey, 2002); los menos, estudian la terminología especializada (Cobré, 1999, 2000, 2002; Lorente, 2002). Resulta aún mucho más difícil encontrar investigaciones acerca 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 & 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 capítulo 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 multirastros 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 *corpus*: 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 las áreas de especialización del subcorpus técnico-científico.

En la primera parte del capítulo se entregan antecedentes relevantes al tema del análisis multirastros y multidimensiones. En el marco metodológico de la investigación, se detallan los *corpus*, los rasgos lingüísticos explorados, los procedimientos de etiquetaje

morfológico automático de los textos y de técnicas estadísticas ejecutadas. Posteriormente, se presentan los resultados, entre los que destacan, los cinco factores de agrupamientos de rasgos, interpretados en cinco dimensiones relevantes y factibles de análisis funcional. La dimensión denominada "Foco Informacional" resulta ser la que -estadísticamente- mejor da cuenta del *corpus* técnico-científico escrito. Por último, se adjuntan dos anexos que juzgamos de alta relevancia ya que ellos presentan información funcional acerca de los setenta y cinco rasgos lingüísticos e información descriptiva y explicativa de las cinco dimensiones analizadas.

1. MARCO DE REFERENCIA

Biber (1994) defiende dos ideas significativas respecto a la representatividad de los estudios lingüísticos basados en *corpus*: 1) todo *corpus* debe tener una amplitud importante, y 2) un *corpus* debe contener registros o categorías textuales diversificadas. Ello confiere mayor validez a 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 diversos 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 validados 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). Las investigaciones en que se ha aplicado el análisis multivariado han revelado que diferentes dimensiones se construyen a partir de conjuntos diferentes de rasgos lingüísticos co-ocurrentes, reflejando así diversas interpretaciones funcionales subyacentes (por ejemplo: objetividad, abstracción de información, modalización). Del mismo modo, las tradicionales distinciones de índole más dicotómica (interactiva/no interactiva), se ven desafiadas por los estudios con análisis multidimensional ya que se ha llegado a demostrar que existe un continuum de variación lingüística a lo largo de los registros. Por supuesto, esto último es concordante con las investigaciones que adscriben a la idea de categorías de límites difusos (fuzzy categories) y que hoy en día tienen gran aceptación entre la comunidad científica (Takoff, 1972).

Un supuesto teórico fundamental del enfoque multivariado lo constituye el principio de que la co-ocurrencia de rasgos lingüísticos (determinada mediante procedimientos estadísticos) refleja funciones comunicativas compartidas, es decir, que estos patrones de co-ocu-

riencia de rasgos se interpretan en términos de funciones situacionales, sociales y cognitivas comunes. En otras palabras, los rasgos lingüísticos co-ocurren en determinados textos porque ellos muestran funciones compartidas específicas. Por ejemplo, las oraciones pasivas, las nominalizaciones y los verboides se relacionan todos con *informatividad* y *focalización del objeto*. Del mismo modo que los pronombres de primera y segunda persona singular, el tiempo presente, el modo imperativo y los adverbios de lugar y de tiempo están ligados a la *interactividad*.

En este sentido, se asume que un grupo de rasgos co-ocurren frecuentemente en ciertos textos porque ellos son usados para expresar un conjunto determinado de funciones comunicativas; esto es, no puede determinarse de manera *a priori* estas funciones. Por ello, resulta crucial –inicialmente– contar con un análisis individual de rasgos en términos funcionales, pues desde allí se establecen los fundamentos para determinar las interpretaciones funcionales subyacentes al conjunto de rasgos co-ocurrentes.

Tal como se adelantó en la introducción, en esta investigación seguimos –en líneas generales– la propuesta metodológica de Biber (1988) respecto del Análisis Multirrasgos (AMR) y Multidimensiones (AMD). Ella, dada que tal enfoque fue creado originalmente como un método analítico para el estudio detallado de las variaciones entre registros, como se aprecia, en nuestro caso, nos resulta de utilidad para describir y comparar los textos de los tres corpora. Este enfoque metodológico fue inicialmente desarrollado, según Biber (1988, 1994) para (1) determinar los patrones lingüísticos sobresalientes y en co-ocurrencia en una lengua, desde una perspectiva empírica cuantitativa; y (2) comparar registros orales y escritos en un espacio lingüístico definido por aquellos patrones en co-ocurrencia.

El AMR y AMD unifican las herramientas metodológicas de la lingüística de corpus según las últimas tendencias (Sinclair, 1991; Leech, 1991; Svartvik, 1992), esto es, requiere de un diseño de corpus representativo, programas computacionales automáticos para etiquetar morfosintácticamente los textos y programas para analizar y cuantificar la ocurrencia de los rasgos lingüísticos previamente marcados en el corpus (Biber, Reppen, Clark & Walker, 2001).

2. METODOLOGÍA

2.1 Los objetivos de la investigación y los corpus

Como se indicó, se buscó 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 recolectados, según las dimensiones a determinar e interpretar funcionalmente. 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. También se utilizan técnicas estadísticas para definir los factores y comparar los resultados estadísticos.

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). Inicialmente se recolectó el CTC y, posteriormente, con el objetivo de llevar a cabo procedimientos comparativos entre diversos registros que brinden una profunda y certeza descripción del CTC y cumpliendo procedimientos de rigurosidad en el marco de la lingüística de corpus, se recolectó otros dos corpora, a saber, el CLL y el CEO. La siguiente tabla muestra su distribución por número de textos y palabras.

Tabla 1. Constitución Global del Corpus PUCV-2003

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%)

2.1.1 Corpus de textos Técnico-Científicos (CTC)

Tal como se muestra en la Tabla 2, el Corpus Técnico-Científico (CTC) está compuesto por setenta y cuatro textos con un total de 626.790 palabras, recolectado en establecimientos secundarios técnico-profesionales de la ciudad de Valparaíso, Chile, en tres áreas técnicas. Estas tres diferentes áreas del conocimiento técnico especializado dicen relación con la formación de tres diferentes profesionales técnicos, a saber, sector marítimo (Especialidad Operación Portuaria), sector metalmeccánico (Especialidad Mecánica Industrial), y sector de administración y comercio (Especialidad Contabilidad). Los textos recolectados corresponden a aquellos que se entregan a los alumnos como parte de lecturas obligatorias o complementarias en cada área técnica, esto es, son parte importante del acceso de estos estudiantes al conocimiento especializado.

El desglose de esta información se entrega en la Tabla 2:

Tabla 2. Constitución del CTC

Área Técnica CTC	Número de textos	Número de palabras
Marítimo (Operación Portuaria)	36 (49%)	155.160 (25%)
Industrial (Mecánica)	18 (24%)	246.372 (39%)
Administración y Comercio (Contabilidad)	20 (27%)	225.256 (36%)
Totales	74 (100%)	626.790 (100%)

Como se aprecia, no existe una relación directa entre número de textos por ámbito de especialidad y número de palabras. Así, en el ámbito marítimo de operación portuaria se registra la mayor cantidad de textos (49% del total), pero el menor corpus de palabras (25% del total). Por el contrario, y de manera interesante, en el área técnica de mecánica industrial se recolectó el grupo más reducido de textos (24%), pero ellos conforman la muestra más grande respecto al número de palabras (39%). Por su parte, el área de administración y comercio (Contabilidad) arroja cifras similares a la anteriormente descrita. En

ella se obtuvo un total de 20 textos (27% del total) y un número elevado de palabras (36% del total). Estas cifras revelan una cierta heterogeneidad respecto a la configuración del corpus de acuerdo a cada ámbito de especialización y también muestran que no existe una relación directa entre área técnica y porcentaje de textos y palabras. En todo caso, el número de palabras por ámbito demuestra no ser relevante en términos estadísticos.

Mediante un análisis multinevral se detectó doce tipos textuales diferentes en el CTC, los cuales fueron rastreados cualitativamente a partir de los corpus (para un estudio detallado de esta determinación tipológica, ver Capítulo 2).

2.1.2 Corpus de textos de Literatura Latinoamericana (CLL)

La selección de los doce textos que componen el Corpus de Literatura Latinoamericana escrito se ejecutó basándose en entrevistas con los profesores de la asignatura de Lengua Castellana y Comunicación de los tres establecimientos técnico-profesionales. En ellas se les solicitó un listado de obras literarias que ellos dieran como lectura a sus alumnos de 4º año de Enseñanza Media en las tres áreas de especialización de las cuales se recogió el CTC. Luego de una comparación de los listados de textos obtenidos, se decidió estrictamente en base a un criterio de homogeneidad, constituir este corpus según las obras literarias que coincidían entre los tres establecimientos educacionales. Es decir, el corpus se conformó a partir de las obras a que todos los alumnos estaban expuestos. Ello derivó en este grupo de autores y las correspondientes obras. En la Tabla 3, se presenta los detalles descriptivos de cada texto junto al número parcial y total de palabras.

Tabla 3. Constitución del CLL

Obra literaria (CLL)	Número de palabras
CLL 1 (PUCV 75)	27.853 (6,5%)
CLL 2 (PUCV 76)	1.414 (0,5%)
CLL 3 (PUCV 77)	30.797 (7%)
CLL 4 (PUCV 78)	56.491 (12%)
CLL 5 (PUCV 79)	47.173 (10%)
CLL 6 (PUCV 80)	33.405 (7%)
CLL 7 (PUCV 81)	94.779 (21%)
CLL 8 (PUCV 82)	50.704 (11%)
CLL 9 (PUCV 83)	12.974 (3%)
CLL 10 (PUCV 84)	24.467 (5%)
CLL 11 (PUCV 85)	4.628 (1%)
CLL 12 (PUCV 86)	75.175 (16%)
Totales	459.860 (100%)

La cantidad de palabras de este corpus presenta un número inferior en cerca de ciento setenta mil palabras al del CTC. Este hecho no constituye en sí un problema para comparaciones ya que las cifras se utilizan normalizadas en textos de 1.000 palabras. Además, a pesar de ser un número inferior al otro corpus, su cantidad es significativa para los estándares empleados actualmente en lingüística de corpus y permite sin dificultades la aplicación de programas computacionales estadísticos como los que se requieren para el análisis factorial.

2.1.3 Corpus de textos de Entrevistas Orales (CEO)

El tercer corpus está formado por dos entrevistas orales realizadas a un total de setenta y cinco alumnos de 4º año de Enseñanza Media de establecimientos técnicos y no técnicos-profesionales de la ciudad de Valparaíso (educación diferenciada y no-diferenciada). La primera entrevista, de tipo entrevista en profundidad semi-dirigida, consistió en una conversación acerca de técnicas de estudio y estrategias de lectura y comprensión. La segunda entrevista, se estructuró según algunas de las temáticas abordadas en la primera conversación y tuvo un carácter más abierto y menos dirigido que la primera. Las entrevistas se realizaron por alumnos y alumnas de último año de la carrera de Pedagogía en Castellano de la PUCV. Se recurrió, en parte, a este perfil de entrevistador con el objetivo de crear un ambiente de confianza y distensión en la conversación.

Solo por razones de organización interna y mejor acceso de procesamiento técnico, se decidió dividir las ciento cincuenta entrevistas en cuatro archivos computacionales. Por ello, la distribución y cuantificación de este corpus oral dialógico se presenta del siguiente modo en la Tabla 4.

Tabla 4. Constitución del CEO

Número Archivo (CEO)	Número total de palabras
CEO 1 (PUCV 87)	86.616 (22%)
CEO 2 (PUCV 88)	89.199 (24%)
CEO 3 (PUCV 89)	102.092 (27%)
CEO 4 (PUCV 90)	102.187 (27%)
Totales	380.094 (100%)

2.2 Análisis Multirrasgos y Multidimensiones: Pasos metodológicos

La realización acuciosa de un completo análisis MR y MD implica una serie de decisiones y etapas rigurosas y, en algunos casos, conlleva alta complejidad técnica computacional y dominio estadístico especializado. Bibel (1988), Bibel *et al.* (1998) y Conrad y Bibel (2001) entregan una serie de pasos metodológicos para su aplicación. Ellos pueden ser resumidos en los siguientes doce puntos nucleares:

1. Diseño, recolección, organización y digitalización del corpus.
2. Selección de un conjunto de rasgos lingüísticos sobre la base de bibliografía especializada de acuerdo a los registros inválidos que serán considerados en el análisis.
3. Caracterización funcional de los rasgos lingüísticos seleccionados.

4. Disponibilidad de programas computacionales capaces de analizar automáticamente los textos en formato plano (ASCII o txt).
5. Marco estructural o etiquetado morfológico y/o sintáctico de los textos del *corpus*.
6. Interrogación manual o (semi)automática de cada uno de los rasgos a partir de los rasgos en estudio para determinar su ocurrencia.
7. Construcción de bases de datos normalizadas, dado el número de palabras divergente entre los textos.
8. Aplicación, con asistencia de programas computacionales, del análisis factorial a las frecuencias de ocurrencia de los rasgos. Ello con el fin de reducir las variables involucradas y determinar patrones de co-ocurrencia entre los rasgos lingüísticos.
9. Establecimiento de un conjunto de factores (cada factor queda conformado por un conjunto de rasgos lingüísticos) mediante el análisis factorial, con algún tipo de rotación (Varimax, Cuantimax, Oblimin, etc.).
10. Interpretación funcional de los factores, producto del análisis factorial, a partir de la co-ocurrencia de rasgos, constituyendo así una dimensión subyacente de variación.
11. Confirmación o refutación de la interpretación de los factores mediante el cálculo de los puntajes factoriales.
12. Cálculo de los puntajes de dimensión para cada texto respecto de cada uno de las dimensiones. En esta fase, se comparan los puntajes de cada registro en cada dimensión y se estudian similitudes y/o diferencias lingüísticas y funcionales.

2.3 Rasgos lingüísticos

Respecto a los rasgos lingüísticos a Indagor, se llevó a cabo inicialmente un rastreo bibliográfico con el fin de identificar categorías gramaticales representativas que mostraran relevancia funcional en español. Con esta información disponible, que reveló un cierto grado de dificultad en su detección, se construyó una matriz con un total de sesenta y cinco rasgos lingüísticos caracterizadores del español. En la Tabla 5 se reiteran los rasgos ya presentados en el Capítulo 3, agrupados en torno a las quince categorías más generales.

Tabla 5. Rasgos lingüísticos Proyecto Corpus PUCV-2003

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)	
4. Presente (indicativo y subjuntivo)	I. Tipos verbales
5. Futuro (indicativo y subjuntivo)	35. Pùblicos
6. Futuro perfectivo	36. Privados
B. Marcadores de modo verbal	37. Persuasivos
7. Indicativo/impérativo	38. Perceptivos
8. Subjuntivo/impérativo	
9. Modo: indicativo	J. Verbos modales
10. Modo: subjuntivo	39. Posibilidad
11. Modo: modalivo	40. Necesidad
	41. Obligación
	42. Volición

C. Desinencias verbales de persona	J. Verbos modales
12. Primera singular	39. Posibilidad
13. Segunda singular	40. Necesidad
14. Tercera singular	41. Obligación
15. Primero plural	42. Volición
16. Segundo plural	
17. Tercero plural	K. Marcadores de modalidad
	43. Atenuadores
	44. Enfáticos
D. Pronombres personales	L. Adverbios
18. Primera persona singular	45. De lugar
19. Primera persona plural	46. De tiempo
20. Segunda persona singular	47. De modo
21. Segunda persona plural	48. De cantidad
22. Tercera persona singular	
23. Tercera persona plural	M. Marcadores de subordinación
24. Demasiativos	49. Subordinadas sustitutivas con "que"
	50. Subordinadas adjetivas con relativo
E. Formas nominales	51. Subordinadas adverbiales de razón o cuya
25. Nominalizaciones	52. Subordinadas adverbiales de concesión
26. Sustantivos (comunes y propios)	53. Subordinadas adverbiales condicionales
F. Formas Pasivas	54. Subordinadas adverbiales de tiempo
27. Pasivas con "se"	55. Frases infinitivas en función nominal
28. Pasivas con ser sin agente	
29. Pasivas con ser con agente	N. Frases preposicionales y adjetivos
30. Pasivas con estar	56. Frases prep. (compl. del nombre)
G. Especificidad léxica	57. Adjetivos atributivo/sustitutivo
31. Relación tipo/token por forma	58. Adjetivos predicativos
32. Relación tipo/token por forma	59. Adjetivos demostrativos
H. Formas estativas activas	60. Participios función adjetiva
33. Ser	
34. Estar	N. Marcadores de Coordinación
I. Tipos verbales	61. Conjunciones aditivas, odi y disy.
35. Pùblicos	
36. Privados	O. Marcadores de negación
37. Persuasivos	62. Adverbio de negación
38. Perceptivos	63. Adverbios de negación temporal
	64. Conjunción de negación
	65. Pronombres de negación

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 el Anexo 1) apartado desarrollado por el equipo de investigadores de la Pontificia Universidad Católica de Valparaíso (PUCV). Estamos ciertos que este Anexo constituye un aporte en sí mismo al estudio de categorías lingüísticas del español que presentan importancia funcional.

2.4 Etiquetaje lingüístico automático

En un intento por buscar el mencionado equilibrio entre lo lingüístico y lo computacional y debido a la insuficiente capacidad de manejar y realizar estudios comparativos más abarcadores a base de la totalidad del *corpus* recolectado con tecnologías de vanguardia, tal como se explicó en el capítulo anterior, establecimos un convenio académico con

la Universidad Pompeu Fabra (UPF) de España. A través de éste, se logró el acceso a programas computacionales que nos brindó la posibilidad de contar con un corpus etiquetado estructuralmente y el acceso a su interrogación por medio de una interfaz gráfica. El procedimiento aplicado a los textos por el equipo chileno-catalán consistió en:

- Codificación SGML (*Standard Generalized Markup Language*)
- Partidor o separador de oraciones (*Splitter o chunker*)
- Marcaje morfológico (Paquete LIC-UPF)
- Desambiguador lingüístico y estocástico

Una vez concluidos estos procedimientos, se procedió a la interrogación de los textos a través de la interfaz computacional con el fin de obtener la ocurrencia de los sesenta y cinco rasgos. Para ello se utilizó diversos accesos a la interfaz, ya sea la interrogación directa a través de búsquedas de ocurrencias simples o a través de comandos en CQP (*Corpus Query Program*) para aquellas rasgos de alcance sintáctico y no solo morfológico. Para procesar estadísticamente las frecuencias de cada rasgo lingüístico en cada uno de los 90 textos, los cómputos de frecuencias fueron sometidos a un proceso de normalización.

Posteriormente, diseñamos e implementamos nuestro propio sistema de etiquetaje e interrogación de corpus. Para ello, inicialmente, adquirimos un conjunto de programas computacionales para la lengua española (Connexor). Luego se constituyó un equipo interdisciplinario que diseñó y dio forma a una plataforma tecnológica interactiva, que funciona a través de Internet. Esta interfaz de consulta lingüística, denominada El Gral (www.elgral.cl), permite el marcaje morfosintáctico de textos de modo remoto como también su posterior interrogación. Los productos de las búsquedas pueden ser almacenados temporalmente y los usuarios pueden obtener diversos tipos de información acerca de sus textos.

2.5 Análisis factorial

Como se sabe, el análisis factorial es un procedimiento estadístico que permite identificar agrupamientos de rasgos lingüísticos que co-ocurren frecuentemente en los textos. Este análisis identifica correlaciones entre un número amplio de variables (sesenta y cinco rasgos lingüísticos) y aquellas que se distribuyen de modo similar. La estructura factorial en que se agrupa conjuntamente a las variables que tienden a co-ocurrir es producto de una matriz correlacional de todas las variables inicialmente involucradas. Cada grupo de variables co-ocurrentes resulta ser un factor, el cual es posteriormente interpretado en términos de categorías funcionales como una dimensión de variación. Este procedimiento tiende -entre otros- a la reducción de las variables involucradas en virtud de su co-ocurrencia significativa (Oakes, 1998; Hair, Anderson, Tatham & Black, 1999).

En el caso de esta investigación, una vez que se realizó el análisis factorial (Factores Principales), se determinaron siete factores principales con una rotación tipo Oblítnia (Oakes, 1998; Hair *et al.*, 1999). Estos factores fueron confirmados a través de los puntuajes factoriales. Se obtuvo así siete posibles dimensiones de las cuales solo fue factible interpretar funcionalmente cinco; por tanto, dos factores no permiten esbozar una dimensión consistente debido tanto al reducido número de rasgos constitutivos como a su naturaleza heterogénea. En general, se requiere al menos cinco pesos relevantes para alcanzar una interpretación significativa del constructo subyacente al factor (Biber, 1988).

Dentro de cada factor, se prelatura el listado de rasgo lingüístico determinados y frente a ellos aparece -en cada caso- un número, normalmente, con decimales. Esta cifra es el peso factorial que indica una medida de fuerza de la relación entre el rasgo en cuestión y el factor como un todo. En otras palabras, este número muestra cuán representativo es el rasgo lingüístico del constructo funcional que subyace al factor. Como se sabe, las cargas o pesos factoriales fluctúan entre +1 y -1. Un valor más cercano a +1 refleja mayor representatividad del rasgo dentro del factor.

3. RESULTADOS

En este apartado, se entregan dos tipos de resultados empíricos a partir de los siguientes procedimientos: (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, CII y CEO) y de los tres ámbitos (mártico-científico del CTC (mártico, industrial y comercial). En otras palabras, se procede a interpretar funcionalmente los parámetros estadísticos encontrados y se estudia su incidencia en cada uno de los registros y áreas de especialización profesional con el fin de identificar relaciones en el nivel lingüístico y funcional.

3.1 Factores y dimensiones

Tal como se verá a continuación, la solución final al análisis factorial concluyó con cinco factores óptimos. En ellos se excluyó todos aquellos rasgos que puntuaban con un valor absoluto menor a 0,40, dado que normalmente en este tipo de estudios se estiman sin importancia relativa frente a la interpretación, incluso si fueran estadísticamente significativos. Otras investigaciones utilizan como valor de referencia la cifra de 0,30 (Biber, 1988), no obstante ello, en este estudio se decidió usar un punto de corte superior.

Solo los rasgos destacados o importantes deberían ser interpretados como parte de cada factor. Una carga negativa o positiva no influye la relevancia de un peso, sino más bien, releva grupos de rasgos que se encuentran distribuidos en los textos de un modo complementario. Así, en los cinco factores siguientes, ciertos rasgos con valor positivo co-ocurren con una alta frecuencia en los textos del corpus, mientras que otro grupo de características co-ocurrentes pero con pesos negativos generan fuertes lazos entre ellos. Ambos grupos de rasgos presentan una especial relación. Ellos se distribuyen en un patrón complementario de ocurrencia, es decir, algunos rasgos (con peso positivo) con alta frecuencia en un texto tienden a denotar la ausencia de ciertos rasgos (con peso negativo) en los mismos textos y viceversa (distribución complementaria). Una explicación pormenorizada de cada uno de las cinco dimensiones se adjunta en el Anexo 2.

Como se adelantó, los rasgos lingüísticos, agrupados en cada factor, son interpretados funcionalmente en una dimensión textual a través de un análisis evaluativo de las funciones comunicativas más ampliamente compartidas por esos rasgos en cuestión. Los resultados, presentados en las Tablas siguientes, muestran los rasgos y los correspondientes pesos positivos y negativos en cada una de las cinco dimensiones.

FACTOR 1**Dimensión 1: Foco Contextual e Interactivo**

Subordinadas adverbiales de causa efecto	0,945
Adverbios de tiempo	0,934
Adverbio de negación	0,928
Pronombres segunda persona singular	0,911
Pronombres primera persona singular	0,823
Desinencias de segunda persona singular	0,813
Pronombre de negación	0,731
Adverbios de lugar	0,723
Modo indicativo	0,693
Desinencias primera persona singular	0,668
Futuro perifrásico	0,662
Enfatizadores	0,652
Formas activas "ser"	0,637
Verbos modales de volición	0,630
Pronombres destratificativos	0,592
Pronombres de segunda persona singular	0,531
Subordinadas adverbiales condicionales	0,523
Adverbios de negación temporal	0,503
Subordinadas sustantivas	0,497
Subordinadas adverbiales de tiempo	0,487
Verbos privados	0,474
Frases infinitivas en función nominal	0,466
Presente	0,424
Frases preposicionales complemento del nombre	-0,545
Nominalizaciones	-0,479
Sustantivos	-0,443
Participios en función adjetiva	-0,437

los rasgos que constituyen el Factor 1 son los más numerosos y ostentan los pesos estadísticos más altos. De los 23 rasgos iniciales con peso estadístico sobre 0,40, se destacan quince rasgos con pesos superiores a 0,60. Cabe señalar que los rasgos positivos que co-ocurren en esta dimensión tienen los pesos más altos en comparación con todos los rasgos que caracterizan a las otras seis dimensiones. Los rasgos con pesos negativos son relativamente menores en número y sólo cuatro presentan pesos sobre 0,40. Los rasgos que se reúnen en este factor denotan una gran relación funcional entre la mayoría de ellos. Su interpretación no resulta compleja. De este modo, la estructura del Factor 1 no constituye un producto exclusivo de la técnica de extracción factorial, sino que se basa en cuestiones lingüísticas y comunicativas subyacentes. Ésta resulta ser un factor poderoso que representa un patrón de variación sustancial entre los textos orales y escritos, especializados y no especializados del español.

Para interpretar esta dimensión se debe evaluar sistemáticamente los posibles funciones comunicativas compartidas por los rasgos co-ocurrentes. Por todo ello se decidió nombrar a este primer factor como Dimensión 1 "Foco Contextual e Interactivo". En este caso, comparativamente, los rasgos con peso negativo son menos e indicar una clara interpretación (nominalizaciones, frases preposicionales 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 informativa. Por su parte, los rasgos con altos puntajes de ocurrencias positivas revelan una referencia directa al contexto lúco y temporal, determinan marcos de orden de la sucesión de hechos, establecen una vinculación con la acción y expresan motivos y consecuencias. También, a través de estos rasgos, se expresa referencia directa a los participantes y existen suficientes evidencias de que la marca de la situación está presente. Esta dimensión se concreta en la acción, en la sucesión de acontecimientos y en las relaciones interpersonales de tipo dialógico. Los rasgos lingüísticos involucrados, en su conjunto, permiten suponer que los textos caracterizados por esta dimensión no contienen información altamente abstracta; sino por el contrario, la alta frecuencia de ocurrencia de rasgos específicos y estadísticamente positivos se asocian con un foco en la explícita y dependencia del contexto y en la activa participación de los interlocutores, rasgos clásicos del discurso oral y dialógico.

Esta Dimensión 1 representa un parámetro fundamentalmente importante de variación, compuesto por un conjunto amplio de rasgos lingüísticos y definitorios de una clara distinción entre dos polos opuestos como son, por un lado, lo contextual e interactivo (oralidad: conversaciones) y, por otro, la informativa, altamente planificado y cohesionada (escritura: exposición).

FACTOR 2**Dimensión 2: Foco Narrativo**

Pronombres segunda persona plural	0,842
Pronombres primera persona singular	0,828
Futuro perifrásico	0,823
Preterito imperfecto	0,820
Pronombre tercero persona plural	0,708
Modo indicativo	0,686
Desinencias primera persona plural	0,667
Verbos modales de volición	0,651
Preterito indefinido	0,614
Pronombre de negación	0,590
Verbos privados	0,577
Adverbios de lugar	0,533
Pronombres segunda persona singular	0,529
Verbos percepcionales	0,496
Adverbio de negación	0,493
Adverbios de negación temporal	0,482
Formas activas "estar"	0,460

Verbos públicos	0,445
Pronombres primera persona singular	0,431
Desinencias tercera persona singular	0,423
Conjunciones adver., disy. y aditivas	0,411
Frases infinitivas en función nominal	0,405
Conjunción "y"	0,402
 Nominalizadores	
Frases preposicionales complemento del nombre	-0,581
Adjetivos atributivos	-0,562
	-0,442

En oposición, el Factor 2 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 del Factor 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. Estas características co-ocurrentes llevaron a denominar el factor como Dimensión 2 "Foco Narrativo". Ella permite identificar textos literarios orales o escritos, a diferencia de textos altamente especializados.

FACTOR 3

Dimensión 3: Foco Compromiso

Verbos privados	0,824
Pronombres primera persona singular	0,789
Pretérito indefinido	0,705
Verbos modales de volición	0,655
Desinencias primera persona singular	0,640
Modo indicativo	0,630
Pretérito imperfecto	0,604
Pronombre de negación	0,569
Pronombres segunda persona singular	0,563
Desinencias segunda persona plural	0,562
Frases infinitivas en función nominal	0,518
Subordinadas sustantivas	0,467
Subordinadas adverbiales de concesión	0,452
Formas activas estar	0,435
Pronombres segunda persona singular	0,427
Adverbio de negación temporal	0,411
Pronombres primera persona singular	0,402
 Frases preposicionales complemento del nombre	
	-0,457

El tercer Factor ha sido interpretado, a partir del conjunto de rasgos co-ocurrentes sistemáticamente bajo en nombre de Dimensión 3, cuya nombre más específico resulta ser "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". La clara identidad de quién escribe o habla queda manifiesta en el texto de manera explícita y quien participa se compromete e involucra con lo que dice y hace. Esto compromiso con el discurso y su contenido revela los afectos y los propósitos del escritor/hablanente. Esta dimensión está asociada a textos en los que sobresele 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. Dicho de otro modo, esta dimensión caracteriza a textos en los que aparecen participantes reales que expresan intenciones y actitudes proposicionales frente a lo dicho.

FACTOR 4

Dimensión 4: Foco Modalizador

Formas activas "ser"	0,671
Atenudadores	0,656
Verbos modales de posibilidad	0,641
Adverbios de modo	0,606
Adjetivos predicativos	0,565
Desinencias tercera persona plural	0,549
Subordinadas adjetivas	0,514
Desinencias tercera persona singular	0,405

Sustantivos	-0,494
-------------	--------

Para construir la interpretación del cuarto factor, se constituye la Dimensión 4, llamada "Foco Modalizador". Para ello se atendió de manera especial a la alta co-ocurrencia significativa de atenuadores (parecer que, creer, tal vez, a lo mejor, quizás, quizál), 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 sistemática de estas marcas lingüísticas tiende a darse en textos con énfasis en cómo (modus) se dicen las cosas, más que en lo dicho (dicunt). Esta distribución de rasgos presenta el contenido de un discurso como incierto y abierto a la verificación; por el contrario, se aleja de la supuesta "objetividad" que otros grupos de rasgos típicamente tienden a representar, tales como aquellos que enfatizan lo referencial. En estos últimos, las restricciones impuestas llevan a textos con alta precisión léxica y densidad informativa.

Así, la Dimensión 4 "Foco Modalizador" agrupa rasgos característicos de textos orales y escritos, clásicamente ligados a la narración y descripción. No obstante ello, también puede marcar textos expositivos y argumentativos, más representativos de la alta especialización.

FACTOR 5

Dimensión 5: Foco Informacional

Verbos modales de obligación	0,496
Modo subjuntivo	0,494
Nominalizaciones	0,456
Participios en función adjetiva	0,413
Frases preposicionales complemento del nombre	0,413
Desinencias tercera persona singular	-0,632
Pretérito indefinido	-0,630
Forma estativa activa "estar"	-0,595
Verbos privados	-0,575
Pronombre de negación	-0,572
Verbos modales de volición	-0,503

Finalmente, el último y quinto factor ha sido interpretado como Dimensión 5: "Foco Informacional" y 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 (comunes y propios) y frases preposicionales son todos rasgos indicadores de integración y compactación de información altamente abstracta, típico 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. Por otra parte, los seis rasgos negativos aquí presentes (tercera persona singular, pretérito indefinido, verbo estar estativo activo, verbos privados, pronombres de negación y verbos modales de volición) apuntan, preferentemente, hacia una contextualización de eventos señalados en el discurso, tendencia atenuada en textos de alta informatividad.

Ahora bien, cada una de estas cinco dimensiones emergentes son producto de un conjunto distintivo de rasgos lingüísticos co-ocurrentes; cada una de ellas define potencialmente un grupo divergente de similitudes y diferencias entre registros y áreas de especialización. Cabe señalar que producto del tipo de rotación seleccionada (Oblimin), la cual resulta más apropiada en el manejo de datos lingüísticos y arroja una mayor certeza de los hechos analizados, la conformación de un factor puede presentar rasgos repetidos con otro factor, ya que estos no son eliminados del análisis una vez incluidos en un primer factor. Esto implica posibles complejidades llegado el momento de la interpretación funcional; no obstante, su conformación resulta ser más realista en cuanto al lenguaje humano y su relación con una determinada dimensión. Como se sabe, en la lengua un tipo de rasgo caracterizador es fácilmente detectable como parte constitutiva en más de una función comunicativa; es el conjunto de rasgos en co-ocurrencia sistemática lo que revela un patrón de variación singular con posibilidad de interpretación comunicativa.

Las dimensiones 1, 2 y 5 ("Foco Contextual e Interactivo", "Foco Narrativo" y "Foco Informacional") se muestran bastante distintivas, dado que la mayoría de sus rasgos positivos son diferentes. Por ello, es posible establecer una clara separación entre las funciones que buscan representar y los tipos de textos que distinguen. Esta distinción agudo y fluida está fundamentada en que muchos de los rasgos negativos presentan también heterogeneidad y por ello se apunta a registros o tipos de textos muy estereotipados con escasos entrecruzamientos. No obstante, las dimensiones 3 y 4 ("Foco Compromiso" y "Foco Modalizador") no parecen apuntar a categorías tan firmemente establecidas y es evidente que tienden a resultar similares en algunos aspectos. Este hecho se explica en que, por una parte, comparten un número de rasgos lingüísticos cuyas funciones subyacentes pueden ser muy similares y, por otro, en que algunos de sus rasgos aunque no idénticos tienden a una interpretación funcional similar. Ello no ha de extrañar ya que —tal como se advirtió anteriormente— la rotación tipo Oblimin de los datos conlleva este tipo de resultados, resultados que normalmente dan cuenta de lenguas naturales.

3.2 Variabilidad en el Corpus PUCV-2003

En este apartado, se estudian las similitudes y diferencias entre los registros y áreas de especialización técnico-científicas con respecto a cada una de las cinco dimensiones. Los tres registros en estudio pueden ser similares en algunos aspectos del *conjunto* de los patrones especificados, pero también pueden variar grandemente entre una dimensión y otra; este tipo de cuestiones nos interesa pesquisar empíricamente.

En un primer momento, nos interesa indagar el impacto de las dimensiones detectadas en la distinción entre oralidad y escritura. Como bien se sabe, esta distinción ha resultado polémica llegado el momento de determinar rasgos específicadores y, en algunos casos, los autores no han alcanzado acuerdo (Halliday, 1985; Chafe & Danilewicz, 1987; Biber, 1988; Louwse, McCarthy, McNamara & Graeser, 2004). El Gráfico 1 muestra los resultados obtenidos en este punto.

Compor un corpus de registro escrito conformado por textos técnico-científicos de divulgación didáctica y textos de literatura latinoamericana con otro compuesto por entrevistas orales semi-dirigidas a estudiantes secundarios constituye un primer acercamiento a cuestiones interesantes. En primer lugar, cabe destacar que las dimensiones construidas a partir de los 65 rasgos lingüísticos están revelando un impacto diferenciador, hecho que en sí mismo es un hallazgo. El registro oral aquí estudiado muestra su comportamiento muy variado a lo largo de las 5 dimensiones, pero relativamente fácil de interpretar; la alta variación que se observa entre la Dimensión 1 y la Dimensión 5 son prueba inequívoca que los rasgos constitutivos de una y otra se oponen en la identificación de textos prototípicamente orales de otros evidentemente escritos. En este sentido, la mayor ocurrencia del modo subjuntivo, verbos modales de obligación, nominalizaciones, participios en función adjetiva y frases preposicionales como complemento del nombre no aparecen como rasgos que tienden a co-ocurrir sistemáticamente en los textos orales aquí estudiados; ellos adquieren —de este modo— un carácter identificador de modalidades escritas de la lengua y se distancian prototípicamente de la oralidad.

De acuerdo a las cifras del Gráfico 1, de modo muy claro se puede caracterizar la

oralidad a partir de las 4 primeras dimensiones, pues todas ellas presentan índices positivos. Al mismo tiempo, se hace evidente que la interactividad y contextualización (Dimensión 1) son los rasgos identificadores y más representativos de un discurso oral dialógico como el que constituye este corpus. Los rasgos de la narratividad y del compromiso aparecen también como caracterizadores del tipo de textos orales aquí descritos. No es una casualidad que la Dimensión 5 "Foco Informacional" sea la única que registre un peso negativo y una diferencia estadísticamente significativa con el resto de las otras dimensiones y, muy en particular, con la Dimensión 1 ("Foco Contextual e Interactivo").

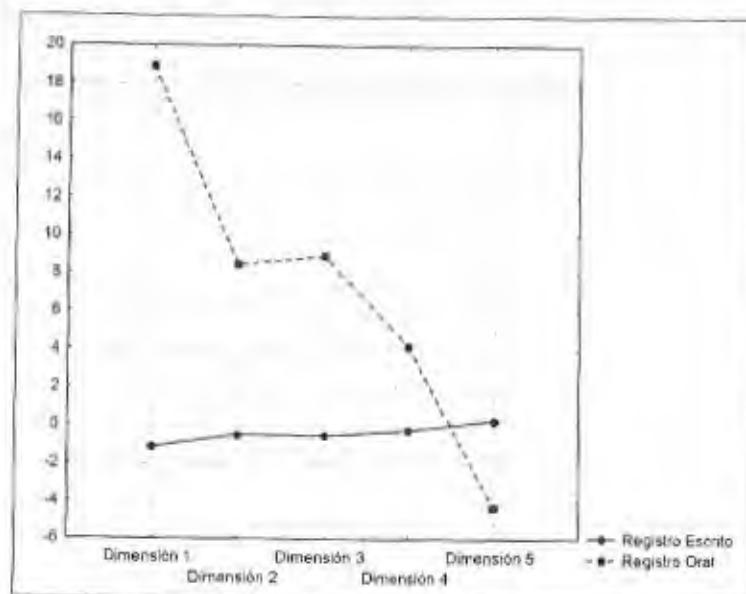


Gráfico 1. Comparación entre oralidad y escritura

En cuanto al registro escrito, el Gráfico 1 revela un comportamiento relativamente homogéneo de las cifras a lo largo de las 5 dimensiones. Coincidientemente con lo dicho más arriba respecto de la oralidad, las cuatro primeras dimensiones presentan ocurrencias negativas y sólo la Dimensión 5 ("Foco Informacional") alcanza un peso positivo, distanciándose de las primeras y constituyéndose en la principal caracterizadora de la escritura. llama la atención, eso sí, la escasa variabilidad entre dimensiones en el registro escrito. Si bien es claramente explicable para las dimensiones 1, 2 y 3, habría sido esperable que la Dimensión 5 mostrara mayor diferencia con el resto y posicionara a la escritura de modo más distante de las otras.

Ahora bien, si para el idioma inglés Biber (1988) identificó un conjunto de rasgos y dimensiones que logran capturar la diferencia entre diversos tipos de registros a lo largo del continuo oralidad/escritura, Louwse, McCorffy, McNamara y Graesser (2004) sostienen que el estudio de Biber se concentra en el nivel de la palabra y que no alcanza un

impacto que involucra al texto como una unidad mayor. Louwse *et al.* (2004) dan cuenta de una investigación que siguió los mismos pasos metodológicos y se apoyó en el mismo corpus que Biber (1988), pero utilizó un conjunto mucho más amplio de categorías e incluyó aspectos de cohesión textual en varios niveles. En esta nueva indagación para la lengua inglesa, se logra determinar una dimensión específica que distingue oralidad de escritura y que se apoya en un amplio grupo de variables que intentan capturar relaciones textuales mucho más allá del horizonte de la palabra.

Hacemos referencia a estos estudios de la lengua inglesa con el fin de precisar el alcance de nuestros datos. Es bien sabido que para el español no contamos con antecedentes en esta línea de investigación; por lo tanto, no disponemos de otros estudios como el aquí presentado que nos permita comparar nuestros hallazgos. Si bien hemos seguido de modo general los pasos metodológicos propuestos por Biber (1988), nuestras categorías indagadas (aunque no ostentan la productividad de las empleadas por Louwse *et al.* 2004) buscan de modo certeza -desde el plano de la palabra y la oración- una proyección en la dimensión textual. Desde esta mirada, los datos aportados en el Gráfico 1 constituyen un indicador importante para el español y han de considerarse como un referente, aunque queda claro que la cantidad de textos y los tipos constitutivos en cada modalidad de lengua no son un conjunto suficientemente variado y amplio que permitan una mayor distinción intragrupal.

El siguiente gráfico entrega datos relativos al registro especializado (escrito) y a los registros no-especializados, tomados estos últimos en su conjunto (orales y escritos). Para ello, se han calculado los puntajes promedio por dimensión (determinados por medio del puntaje factorial o *factor score*).

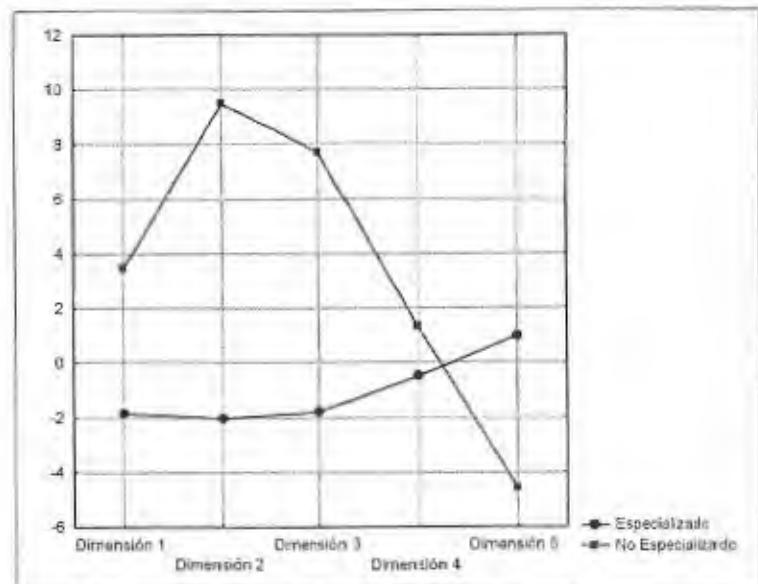


Gráfico 2. Comparación entre registros especializados y no-especializados

Como se aprecia en la gráfica, la comparación de los pesos factoriales entre las cinco dimensiones con base en los textos de especialidad y los considerados aquí como no especializados (orales y literarios) muestra una evidente diferenciación: los textos no especializados destacan por pesos positivos en las cuatro primeras dimensiones en donde la narratividad aparece como la dimensión más identificadora junto al Foco Compromiso. Son los marcos de persona singular y plural (pronombres y desinencias verbales), el pretérito imperfecto y el indefinido, el modo indicativo y los verbos modales de volición, entre otros, los que identifican los textos de literatura latinoamericana escrita y el corpus de textos orales, es decir, los que representan la ausencia de especialización temática. Junto a ellos las rasgos característicos de la interactividad, contextualización temporal y espacial y modalización también se constituyen en prototípicos de los discursos no especializados. Contrariamente, la Dimensión 5 se destaca por un peso altamente negativo y estadísticamente significativo respecto de las otras cuatro dimensiones, hecho que indica que la compactación de información y alto grado de referencialidad no se evidencian como característica de estos textos no especializados.

Por su parte, los textos del corpus técnico-científico (CTC), considerados como especializados, alcanzan pesos totalmente opuestos a los obtenidos por el corpus no especializado. En este caso, los datos estadísticos son muy similares a los del corpus escrito, entregados en el Gráfico 1. En las cuatro primeras dimensiones se registran pesos negativos y solo emerge la Dimensión 5 "Foco Informacional" como la distintiva para este grupo de textos. Estos datos son congruentes con la teoría de corte lingüística y funcional que hemos venido sustentando para diferenciar e identificar a los textos prototípicos del discurso especializado, de modo que los hallazgos entregados se constituyen en una corroboración empírica relevante que pone de manifiesto las diferencias entre un determinado registro y otro. Estos valores permiten avalar empíricamente la postura a favor del denominado discurso de especialidad y muestran que las 65 categorías exploradas son un conjunto de rasgos valiosos como indicadores de variabilidad.

Un hallazgo importante lo constituye el que, según las cifras entregadas en los dos primeros gráficos, la especialización se detecta como estrechamente vinculada a la modalidad escrita de la lengua. Tal como lo ha mostrado el punto factorial promedio respecto de la Dimensión 5 "Foco Informacional", es esta dimensión y el conjunto de rasgos lingüísticos y funcionales asociados a ella, lo que establece la distinción fundamental entre, por una parte, lo que está más cerca de la escritura que de la oralidad y, por otra, lo que resulta más prototípicamente especializado o no. La explicación que se desprende de estos datos apunta claramente a que las características típicas de un discurso especializado coinciden con los rasgos clásicos de la escritura en su versión más prototípica. De este modo, queda claro que la supuesta especialización a través de la modalidad oral de la lengua no encuentra apoyo en nuestros datos y cabría suponer, entonces, que cuando se habla de oralidad especializada, a lo que se tiende mayoritariamente, es a un tipo de discurso escrito que se dispone de manera oral (el ejemplo clásico es una conferencia académica que se dispone oralmente pero que se produce de modo escrito).

También vale la pena llamar la atención sobre el comportamiento estadístico de la Dimensión 5. Tal como se observa en los valores hasta ahora disponibles, es la función informativa de la lengua, manifiesta a través de diversos mecanismos, la que se hace

presente como rasgo distintivo entre las diversas aproximaciones exploradas. Es la Dimensión 5 la que obtiene punto negativo para el registro oral y la que puntúa positivamente para la escritura, mostrando una diferencia estadísticamente significativa entre ambos valores. Es, del mismo modo, la Dimensión 5 la que arroja puntos factoriales positivos para el corpus especializado y la que muestra puntos negativos elevados para el corpus no especializado, siendo la diferencia entre ambos estadísticamente significativa. Todo ello implica, desde la perspectiva de esta investigación interesada en el discurso especializado escrito, que la Dimensión 5 "Foco Informacional" juega un rol central en la descripción de este tipo de discurso y que sus rasgos lingüísticos y funcionales constituyen un conjunto de características a profundizar y destacar. Así, entre otros, la monorreferencialidad, la precisión, la compactación, la carencia de emotividad, son rasgos relevantes que se desprenden como prototípicos del discurso focalizado en este libro.

A continuación, en el Gráfico 3, se presentan los puntos promedio por dimensión para cada uno de los tres registros del Corpus PUCV-2003 (técnicos, literarios y orales), con respecto a las cinco dimensiones en cuestión. Recordemos que con ello se busca —a través de las dimensiones ya determinadas— comparar los registros esta vez de modo más detallado con el fin de determinar relaciones específicas entre el español escrito especializado y el resto de los registros; de modo más preciso, se intenta identificar posibles diferencias y/o similitudes lingüísticas y funcionales entre los textos en estudio.

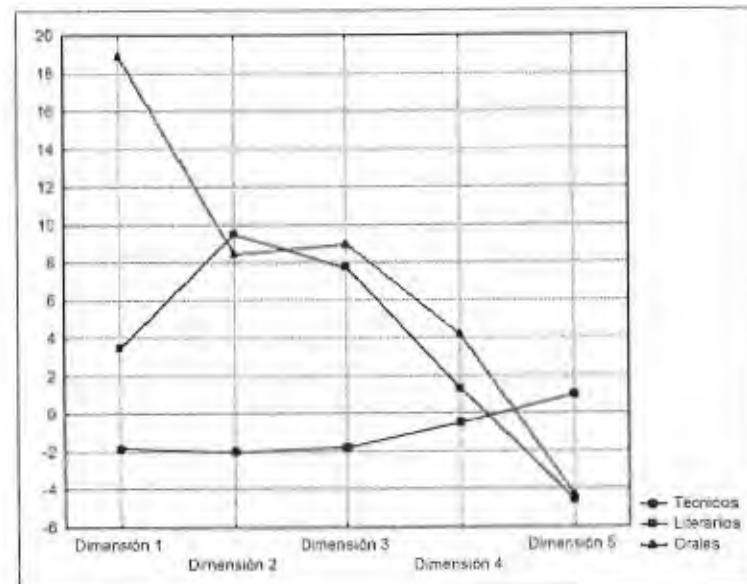


Gráfico 3. Dimensiones y registros

Resulta interesante comprobar que los textos técnicos 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 sistemática 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 informacional, el cual se aleja claramente de contenidos interpersonales y afectivos. Al mismo tiempo, como se observa, los textos de literatura latinoamericana y los textos basados en entrevistas orales poco planificadas, con un grado importante de espontaneidad, muestran puntos 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 pasado, pronombres y adjetivos demostrativos se constituyen en prototípicos.

Este análisis comparativo arroja similitudes y diferencias muy elocuentes. Permite identificar las áreas de intersección entre la oralidad de las entrevistas y la narratividad de la literatura latinoamericana. 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.

Como se aprecia, la variación entre estos registros enfrenta un *continuum*, identificado en este caso por medio de las dimensiones y los rasgos lingüísticos que ellas capturan y permite identificarlos y caracterizarlos prototípicamente. Las implicancias derivadas de ello son múltiples, en particular, en lo que respecta al discurso especializado de divulgación didáctica, foco de interés de esta investigación. El diseño de materiales didácticos que aborden el discurso especializado en español escrito y quienes se concentren en su enseñanza deben capitalizar estas descripciones.

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

Es interesante observar que también en el Gráfico 4 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-científicos. Según se aprecia, el ámbito comercial presenta el mayor puntaje promedio positivo en esta dimensión (3,8), hecho que muestra su mayor carga informativa a través de alta densidad léxica y complejidad sintáctica.

Parodi (ver Capítulo 3) ya había detectado, a través de un estudio descriptivo simple, cierta variabilidad interna en el comportamiento de los sesenta y cinco rasgos lingüísticos entre las tres áreas técnicas. Esta investigación preliminar muestra un patrón distintivo entre el área marítima y las otras dos. El mismo dato encuentra apoyo certero en estos análisis dimensionales con técnicas tipo factor score. Según se comprueba, el mayor puntaje negativo en cuanto a la carga informativa lo obtiene el ámbito de especialización marítima (-0,1), el mismo que aparecería distante de los otros dos en el trabajo que comentamos. Considerando este resultado, la Dimensión 5 parece distinguir entre textos de un ámbito y otro. Será necesario un estudio cualitativo profundo para dar cuenta con mayor detalle de qué clases textuales del área comercial se alinean en torno a este patrón dimensional.

Seguidamente en la jerarquía, la Dimensión 4 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.

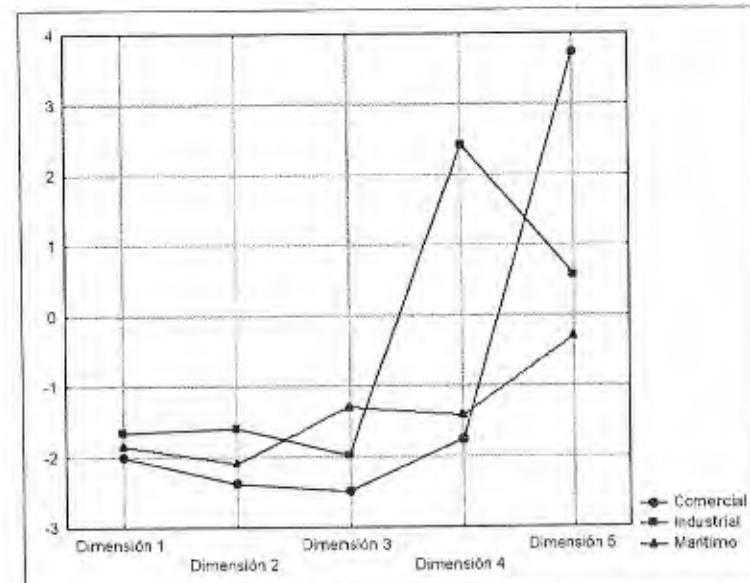


Gráfico 4. Dimensiones y ámbitos de especialización

En el resto de las dimensiones (1, 2 y 3), las tres áreas del CTC presentan ciertos negativos y relativamente similares, hechos que indican que estas dimensiones no aportarían gran explicación diferenciadora en la descripción del discurso técnico-científico de divulgación didáctica escrita. 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. Todo ello es congruente con otros antecedentes disponibles acerca de este tipo de textos.

CONCLUSIONES

En este Capítulo se combinó los beneficios de los recursos aportados por la lingüística de corpus y los avances en el desarrollo de programas tecnológicos de vanguardia para el español en la construcción e interrogación de una base de datos digital con etiquetaje lingüístico automático. Para ello se buscaba incorporar las tecnologías emergentes a la investigación lingüística así como explorar la distribución de un grupo de rasgos lingüísticos relevantes funcionalmente en la descripción de tres corpora de lenguaje natural a través de registros diversificados. También -de modo más específico- se pretendía llevar a cabo una descripción inicial del corpus técnico-científico desde la perspectiva del análisis multidimensional.

Aunque esta podría ser considerada una investigación exploratoria, basada en un número relativamente pequeño de textos y con escasa variabilidad de registros, los datos aportados permiten sostener que su carácter es poderoso y que son claramente interpretables en varios sentidos: estadística, lingüística y funcionalmente.

El desafío de avanzar hacia un campo inicialmente desconocido para el equipo chileno ha sido paulatinamente enfrentado con éxito. Se hace evidente que para proseguir en este terreno multidisciplinario debemos implementar, en nuestra Universidad, herramientas tecnológicas que nos permitan mayor autonomía en el proceso de marcaje y etiquetaje de los textos, así como intentar desarrollar medios de interrogación amigables para los que no necesariamente se interesan en aprender lenguajes computacionales tales como los comandos en CQP. Para ello, en este momento -gracias al apoyo institucional y gubernamental- se está instalando un laboratorio computacional con equipamiento (máquinas y programas computacionales para el español) y equipo humano de apoyo especializado. Este otro desafío nos espera ahora.

En cuanto a la determinación e identificación de rasgos relevantes y dimensiones lingüísticas, desde una óptica general, el análisis multidimensional 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, sino que la determinación de regularidades sistemáticas que devuelven 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 multirrasgos y la muestra relativamente grande de textos (comparado con análisis de textos ejemplares) permite generar conclusiones robustas y con mayores implicancias. Estos aspectos llevan a la lingüística en esta línea- no solo a contar con una alternativa metodológica de investigación, sino a delinear un paradigma que brinda renovados bríos a las indagaciones venideras para el español.

El hallazgo de las cinco dimensiones, identificadas a partir de un análisis cuantitativo de la distribución de los sesenta y cinco rasgos lingüísticos en los noventa textos, resultó en un recurso de extraordinario potencial descriptivo. Los aportes, a partir de la visión multidimensional, integran una distinción fundamental entre los textos estudiados de la modalidad oral y escrita: los primeros aparecen incluidos en la interactividad, contextualización y las relaciones interpersonales; esto es, fuertemente descielos por la Dimensión 1 "Foco Contextual e Interactivo"; al mismo tiempo, la oralidad aquí descrita (entrevistas) no se relaciona directamente con una prosa informativamente 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 AMD también ha permitido identificar registros más expositivos y técnicos escritos de otros claramente narrativos literarios escritos.

Un punto de relevancia lo constituye la distinción que atañe a la especialización y su directa vinculación con la modalidad escrita de la lengua. Los rasgos lingüísticos y funcionales asociados a la Dimensión 5 "Foco Informativo", detectados empíricamente, son un dato a ser considerado y explorado en profundidad en posteriores investigaciones.

Otro posible hallazgo destacable 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; Bibel, 1994; Aljmer, 2002; Reppen, Fitzmaurice & Bibel, 2002). El español, en términos generales, al igual que el inglés, el sámi y el coreano presentan múltiples dimensiones que reflejan al menos la interactividad, el foco informativo y el compromiso. En todas estas lenguas -gracias al AMD- se ha podido establecer distinciones entre oralidad y escritura y diversos registros específicos en cada modalidad de lengua. 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.

La descripción del CTC se desarrolló parcialmente en este trabajo. Es necesario abordar un análisis más pormenorizado de las distinciones entre las tres áreas de especialización del CTC (comercial, marítimo e industrial) y se deberá continuar con el cálculo de los puntuajes factoriales entre las diversas clases textuales que componen este corpus con el fin de indagar el aporte del AMD a la diferenciación y descripción de las mismas. Los antecedentes empíricos hasta ahora aportados dan cuenta de una interesante homogeneidad a lo largo de las Dimensiones 1, 2 y 3 entre las tres áreas técnicas-científicas: estos textos no parecen tener una fuerte marca de narratividad, de involucramiento de los participantes ni de interactividad. Por el contrario, los textos del área industrial y marítima sí presentan diferencias significativas en las Dimensiones "Foco Modalizador" y "Foco Informativo". El dominio industrial muestra una tendencia hacia la modalización y el dominio marítimo se distingue de los textos de las otras dos áreas en un grado positivo mayor de densidad informativa.

Cabe destacar que como una forma de facilitar la interrogación misma y de contar con otras herramientas (tal como N-Gramas) y facilidad de graficación de los datos obtenidos, se construyó el programa BUCÓLICO (Buscador de Concordancias Lingüísticas en Corpus); éste, no solo cuenta con los noventa textos del Corpus PUCV-2003, sino que permite la incorporación de nuevos textos para su estudio (al respecto, ver Capítulo 9).

Por último, estimo que esta investigación contiene poderosas implicancias para el diseño de pruebas de diverso índoles tales como de evaluación de contenidos técnicos y de comprensión textual propiamente dicho. También para la elaboración de material didáctico. Ello, ya que los análisis de los registros describen las características del tipo de uso lingüístico empleado en el material al que se exponen estos alumnos de establecimientos técnico-profesionales en las áreas mencionadas. De acuerdo a nuestros resultados, estos estudiantes requieren desarrollar la habilidad para manejar una variedad del español especializado escrito muy particular: prosa densa académicamente en términos léxicos, morfológicos y sintácticos pero también textos con marcas de modalización.

Por último, los hallazgos aportados muestran la relevancia del estudio del uso lingüístico en entornos técnico-científicos escolares del ámbito profesional y confirmar la necesidad de crear materiales especializados de instrucción sobre la base de descripciones empíricas de los registros en estudio. Esta evidencia también sugiere que existen razones significativas para atender diferencialmente al análisis, comprensión y producción del registro técnico-científico que circula y se genera tanto en el ámbito educativo como en el ambiente laboral. Todo estudiante debería practicar con una amplia variedad de registros ya que en su vida profesional no solo encontrará los técnicos especializados. Tampoco parece recomendable que alumnos de estos liceos técnico-profesionales secundarios se enfrenten de manera automática a este tipo de textos marcados por una prosa informativa extremadamente compleja. El material debería contener una organización progresiva desde lo más divulgativo didáctico hasta aquellas textos más típicos del ámbito profesional y laboral.

ANEXOS

ANEXO 1: RASGOS LINGÜÍSTICOS CORPUS PUCV-2003

A. MARCADORES DE TIEMPO VERBAL

1. Pretérito indefinido (indicativo) [PRET.IND]

Expresa anterioridad del verbo al acto de la palabra. Es acción que ya no se puede volver a realizar (Contreras, 1984). Funcionalmente, ha sido definido como tiempo representativo de la secuencia narrativa, porque es el tiempo en el que se explica la complicación que perturba el estado inicial (Bassols & Torrent, 1997). Se le considera característico de la lengua escrita y de la prosa narrativa (De Kock & Gómez, 2002).

2. Pretérito imperfecto (indicativo) [PRET.IMP]

Expresa una acción durativa, o bien, reiterada o habitual, simultánea en todo o en parte con otra también pretérita, durativa o instantánea. Por esto se dice que es un tiempo relativo (Alvar, 2000; Moliner, 1986). Específicamente, señala simultaneidad respecto a un momento anterior al punto central (Alvar, 2000). Ha sido considerado característico de la lengua escrita, en particular de la prosa narrativa (De Kock & Gómez, 2002). Bassols y Torrent (1997) señalan que es un tiempo con el que se describen los estados iniciales en la secuencia narrativa, como también los incisos descriptivos insertos en el relato. También consignan su uso en construcciones argumentativas, tal vez porque como anota Moliner (1986), el pretérito imperfecto en los verbos modales [deber, poder, tener que] expresa opinión acerca de la conveniencia o procedencia de una cosa.

3. Pretérito perfecto (indicativo y subjuntivo) [PRET.PER]

Expresa una acción pasada que tiene incidencia en el momento de la enunciación, que dura hasta el momento mismo en el que se habla (Moliner, 1986). Gili Goya (1980) constata que aunque el español moderno establece la diferencia entre el pretérito indefinido y el perfecto, extensas zonas de España e Hispanoamérica han preferido una forma por sobre la otra, debido a que en ambas predomina el aspecto perfectivo.

4. Presente (indicativo y subjuntivo) [PRES]

Expresa las acciones que coexisten con el acto de la palabra (Gili Goya, 1980). Es usado para todo lo universal. Habitual en máximas y sentencias. Cuando la acción se refiere al momento del habla, se trata del presente actual. Pero además, con el presente se enuncian las verdades intemporales y las acciones habituales. Puede expresar acciones pasadas, en el llamado presente histórico (Moliner, 1986). Asimismo, cuando la acción se siente de segura realización, puede tener valor de futuro (Moliner, 1986).

5. Futuro (indicativo y subjuntivo) [FUT]

Expresa acción venidera, independientemente de cualquier otra acción. Segundo Gili Goya (1980) el empleo del futuro supone una cierta capacidad de abstracción por parte del hablante, debido a esto aparece tardíamente y es poco frecuente en el habla infantil. Es preferido en

ciertas modalidades, como la lengua escrita, su uso en la lengua oral asienta conforme el grado de formalidad del habla (Arroyo, 2000). El futuro del subjuntivo aparece solo en ciertos casos, en oraciones con "se" y en las "oraciones de relativo pronominal o adverbial, y esto solamente en lenguaje anticuado y en algunas frases estereotipadas, propias, por ejemplo, del lenguaje judicial o de las leyes..." (Moliner, 1986: 1476).

6. Futuro perifrásico [FUT.PERI]

Tiempo que corresponde al mundo comentado, en oposición al imperfecto, tiempo de lo narrado (Weinrich, 1974). Es más común en los registros coloquiales y descuidados (Arroyo, 2000; Contreras, 2000).

B. MARCADORES DE MODO VERBAL

7. Indicativo/impérativo [INDIC.IMP]

Se da una coincidencia de forma entre la segunda persona singular del impérativo y la tercera persona singular del indicativo.

8. Subjuntivo/impérativo [SUBJ.IMP]

Existe un sincretismo entre las formas del subjuntivo y del impérativo debido a que las únicas formas específicas del impérativo son las segundas personas del singular (tú) y el plural (vosotros). Las otras personas se toman del presente del subjuntivo. En las frases negativas, las segundas personas se sustituyen por las del subjuntivo.

9. Modo indicativo [MOD.IND]

Expresa la modalidad declarativa/experimental (Cepeda, 2002), como también estados o acciones consideradas reales (Gómez Macker & Péronard, 1988). El modo indicativo es propio del intercambio discursivo oral (Cepeda, 2002). Este modo hace referencia a hechos reales localizados en un tiempo verdadero (Criado de Val, 1962; Gómez Macker & Péronard, 1988). Generalmente se refiere a un acontecimiento localizado a partir de un pretérito. Según Alcoba (1999), el modo indicativo se utiliza cuando hay aserción. En este modo predomina la función representativa y su modalidad distintiva es la lógica o declarativa: "Se define por la reflexión objetiva entre hablante y mensaje y la modalidad declarativa [...] y como la modalidad declarativa es la forma no marcada [...] el modo indicativo es siempre el más extenso" (Hernández, 1996: 473).

10. Modo subjuntivo [MOD.SUBJ]

Permite especular sobre hechos inciertos, sujetos a una apreciación subjetiva (Criado de Val, 1962). En este modo se expresa la subjetividad del hablante o del sujeto de la comunicación ante el enunciado, es el modo de la hipótesis y de la virtualidad. En toda expresión con subjuntivo se percibe la presencia del hablante y del enunciado. El hablante puede aparecer, además, "como agente de la enunciación, como sujeto del enunciado, hecho que no podía darse en el llamado impérativo" (Hernández, 1996: 375). Sin embargo, además de estas características, se debe tener presente que el subjuntivo aparece normalmente como oración subordinada, y que ésta remite a textos de mayor complejidad sintáctica (Gili Goya, 1980), cuya función es enmarcar la información del discurso de diversas maneras (Biber, 1988).

11. Modo impérativo [MOD.IMP]

Su misión propia es el mandato directo. Se limita casi exclusivamente al lenguaje hablado, locutancial (Criado de Val, 1962). Estas formas verbales se oponen a las demás formas personales porque están especializadas en las expresiones de la modalidad deontica apelativa de mandato. Las formas de impérativo no transportan otro significado que el de mandato, en cambio, es posible formular mandato mediante otras formas lingüísticas como el futuro o el presente de indicativo, presente del subjuntivo, etc. (Gili Goya, 1980).

C. DESINENCIAS VERBALES DE PERSONA

12. Primera singular [DES.1S]

Refleja el carácter egocéntrico de un texto, implica una necesidad de comunicación directa. Propio de un estilo directo, característico de la lengua escrita y de la prosa narrativa (De Kock & Gómez, 2002). En el lenguaje científico y en el de divulgación de la ciencia se constata la tendencia común a evitar la referencia a la primera persona y a usar otros procedimientos de presentación de la persona del autor (Clapuscio, 1992).

13. Segundo singular [DES.2S]

Se utiliza la segunda persona con el objeto de producir un efecto determinado: "generalizar la experiencia enunciada e incluir al interlocutor de una manera personal y afectiva" (Calzamiglio & Tusón, 1999: 139). Se asocia al lenguaje coloquial.

14. Tercera singular [DES.3S]

Aporta significación de tiempo, modalidad, persona y número del enunciado donde aparece la forma verbal (Alcoba, 1999). Esta desinencia es característica en la lengua escrita, en particular, en el artículo científico (Koset, 2002).

15. Primera plural [DES.1P]

La identificación de la persona que habla con la primera persona del plural incorpora al locutor a un grupo (Calzamiglio & Tusón, 1999). El llamado "nosotros de modestia" reemplaza al yo porque, en ocasiones, se considera inapropiado utilizarlo en público. Otra función de la primera persona plural es incluir al interlocutor, involucrándolo y suavizando así los mandatos o las peticiones. Clapuscio (1992) ha constatado que este uso es frecuente en el discurso de divulgación dirigido al gran público. Asimismo la autora destaca en este tipo de textos y también en aquellos de divulgación destinado a un público más restringido, el uso de nosotros que incluye de modo aparente al interlocutor y cuya función es "ordenar la interacción y asegurarla" (Clapuscio, 1992: 198).

16. Segundo plural [DES.2P]

Al igual que en los casos de primera y segunda persona singular, esta desinencia verbal manifiesta el carácter egocéntrico y la necesidad de comunicación directa.

17. Tercera plural [DES.3P]

Esta desinencia es rasgo característico de la lengua escrita y del artículo científico (Koset, 2002) debido a su uso para constituir oraciones impersonales, en las que el sujeto se calla intencionalmente, carece de interés o es imposible de determinar (Gili Goya, 1980).

D. PRONOMBRES PERSONALES

18. Primera persona singular [PRON.1S]

Debido a que en español los verbos conjugados llevan la marca de persona en la desinencia, el pronombre sujeto es innecesario casi siempre. En primera y segunda persona su aparición es enfática y "significa insistencia particular en hacer resaltar el sujeto" (Gili Goya, 1980: 172). Refiere directamente a los participantes, marcadores de la presencia del yo (Biber, 1988). En la narración, distingue al testigo-participante o al protagonista (Bassols & Torret, 1997). Abunden cuando el contexto no aclara en forma suficiente la persona verbal. Es propio de los entrevistas orales (Castellano, 2000). En general, los pronombres personales de la primera y segunda persona -denominados deicticos- remiten a los participantes en el acto de comunicación, función que les es propia. La propiedad semántica esencial de los pronombres personales es que no permiten asignar valores de verdad a los enunciados independientes del contexto (Fernández, 1999). Otro aspecto importante es que los pronombres de primera y segunda persona son reversibles, en el sentido de que el "yo" de turno no puede menos que ordenar su

derecho o un "tú", si quiere tener un interlocutor válido, aunque no son imprescindibles para expresar el concepto de persona gramatical (Fernández, 1999).

19. Primera persona plural [PRON.1P]

Con la forma nosotros se evita utilizar el pronombre "yo", sin que ello implique, en principio, importantes diferencias semánticas. Este uso se considera más cortés y es por esta razón que está muy extendido en el género académico (Ucedo, 1995). Este pronombre puede tener varios tipos de referentes, en lo que se llama "plurales ficticios" (Alcina & Blecua, 1975).

20. Segunda persona singular [PRON.2S]

Las formas "tú" y "usted" son características de la lengua escrita, específicamente de la prosa narrativa y la poesía (De Kock & Gómez, 2002). Requiere de un destinatario específico e indican un alto grado de interacción y acción (De Kock & Gómez, 2002). En los textos narrativos, puede aparecer la narración en segunda persona singular, en caso de que los propios personajes adopten una perspectiva exterior que les dé mayor lucidez para el autoanálisis (Genette, 1972).

21. Segunda persona plural [PRON.2P]

Es válido para este pronombre cuánto se ha dicho para la segunda persona singular.

22. Tercera persona singular [PRON.3S]

Los pronombres él/ella aparecen cuando existe el riesgo de malinterpretar la referencia porque la información que aporta el contexto falla. Se asocia normalmente con entrevistas orales (Castellano, 2000). Él/ella tienen un uso deictico que los asemeja a los demostrativos, pero también un uso referencial, esto es, tienen la capacidad de retomar los rasgos de un individuo presente en el contexto lingüístico (Fontanella, 1999). En el caso de los textos narrativos, la tercera persona corresponde al narrador omnisciente (Bassols & Torrent, 1997). A diferencia de los demás pronombres personales, la tercera persona singular es la "no persona", es decir, está excluida de la instancia de comunicación y remite no al enunciado, sino a una situación "objetivo". En este sentido se trata del término no marcado y de hecho no existe en todas las lenguas (Fernández, 1999).

23. Tercera persona plural [PRON.3P]

Al igual que para los pronombres de tercera persona singular, el pronombre sujeto aparece cuando puede haber ambigüedad, puesto que son muchas las posibles terceras personas.

24. Pronombre demostrativo [PRON.DEM]

Al igual que otros elementos lingüísticos agrupados bajo la denominación de deicticos, los demostrativos señalan, seleccionándolos, algunos elementos del entorno contextual (Calsamiglia & Tusón, 1999), adquiriendo su sentido pleno solo en el contexto en el que se emiten. Son usuales en el lenguaje oral.

E. FORMAS NOMINALES

25. Nominalizaciones [NOMINAL]

Este término designa a los nombres derivados de bases verbales y adjetivales, así como al proceso de su formación. Los nombres derivados pueden tener como referente un evento o un proceso, un objeto, un estado, una propiedad o un producto resultante de un acontecimiento o proceso, recurso típico del lenguaje técnico (académico) para expresar significados complejos y abstractos (Picallo, 1999). Permiten integrar información en pocos palabras (Chafe, 1982, 1985) y reducir oraciones completas en series de frases nominales más compactas y eficientes (Janda, 1985), además tienen la función de transportar información altamente abstracta (Biber, 1988). Ciapucio (1992) incluye la nominalización dentro de los procedimientos de desagentivación, característica de la comunicación científica.

26. Sustantivos (comunes y propios) [SUST]

Según Bibet (1988), los sustantivos son los principales portadores del significado referencial del texto. La ocurrencia de algunos tipos de sustantivos [los más largos y los derivacionales] se asocia con discursos cuya función es altamente informativa, caracterizados por una cuidadosa integración de la información. El nombre significa de manera independiente y autónoma; en su significado suele distinguirse un núcleo denotativo y una serie de connotaciones, que se incrementan en el contexto (Hernández, 1996).

F. FORMAS PASIVAS

27. Pasivas con "se" [PAS.SE]

Se especializan en acciones de carácter habitual, general, con objetos internos y sujetos implícitos poco delimitados, aunque pueden aparecer en todo tipo de contextos. Tiende a usarse con cuando el sujeto de la frase es nombre de cosa. En general, es más frecuente que la pasiva con ser (Martín, 1980). Aparece tanto en la lengua hablada como en la escrita. Se ha notado un incremento en el uso de esta construcción en el lenguaje divulgativo informativo (Mendikoetxea, 1999a y b). Ciapucio (1992) registra su frecuencia en textos de divulgación científica con lo que se acentúa el carácter impersonal del texto. Según la autora, esta tendencia a la omisión del agente es una característica del lenguaje de la ciencia que se mantiene en la divulgación.

28. Pasivas con "ser" sin agente [PAS.SER-a]

Se especializan en acciones de carácter puntual, con objetos externos y un marcado carácter intencional que denota la existencia de un sujeto implícito delimitado. Son de mayor uso en la lengua escrita. La ausencia del agente se ha atribuido a la intención de callar o ocultar el sujeto nociional (Mendikoetxea, 1999a y b).

29. Pasivas con "ser" con agente [PAS.SER+a]

Con estas pasivas parece ser siempre gramaticalmente posible la expresión del sujeto nociional, independientemente de su papel semántico: agente, destinatario, experimentante o fuente. La expresión del agente se da fundamentalmente en textos escritos de carácter periodístico (Hernández, 2000a).

30. Pasivas con "estar" [PAS.ESTAR]

Los gramáticos no se han puesto de acuerdo acerca del carácter pasivo de estas construcciones. Los que las rechazan, argumentan que las pasivas con ser expresan una acción que recae en el sujeto paciente, en cambio, las construcciones con estar expresan el resultado de esa acción. Según Mendikoetxea (1999b) ambas construcciones tienen un significado pasivo.

G. ESPECIFICIDAD LÉXICA

31. Relación clase/tipo (type/token) por forma [TYP.TOK.form]

Es una relación porcentual que expresa el rango de variabilidad de unidades dentro un universo. En este caso se refiere a la especificidad léxica de las formas que aparecen en un texto. Una tasa alta se relaciona con el lenguaje escrito debido al carácter planeado de su producción. Por el contrario, una tasa baja se asocia con el lenguaje oral por el carácter espontáneo de la oralidad (Horowitz & Samuels, 1987).

32. Relación clase/tipo (type/token) por lema [TYP.TOK.lem]

Se refiere a un tipo especial de relación que expresa la variabilidad o especificidad de los lemas de un texto, entendidos estos como las raíces léxicas de las formas que aparecen en un texto determinado.

H. FORMAS ESTATIVAS ACTIVAS

33. Ser [ACT.SER]

Los verbos conectores expresan básicamente relaciones de equivalencia, igualdad, similitud, dependencia o atribuyen cualidades o valores, como por ejemplo, *ser*, *parecer*, *equivaler*. Se considera que *ser* atribuye cualidades consideradas como permanentes. Este tipo de verbos, mayoritariamente copulativos o pseudocopulativos, aunque no se los identifique como unidades portadoras de conocimiento especializado, forman parte de la expresión de este conocimiento, es decir, no tienen valor especializado, pero forman parte de lo que se ha llamado unidades de conocimiento especializado (Llorente, 2002).

Destacan por su frecuencia en la descripción (Bassols & Torrent, 1997). Con "ser", se formulan juicios que no dependen de la experiencia inmediata (Gili Gaya, 1980). Su empleo en las oraciones copulativas se reduce a servir de nexo entre el sujeto y el predicado y permiten la expresión temporal (Gili Gaya, 1980).

34. Estar [ACT.ESTAR]

En tanto verbo conectar cumple las mismas funciones antes mencionadas para el verbo *ser* pero a diferencia de este, con el verbo *estar* las cualidades se consideran transitorias o accidentales y además se perciben como resultado de un cambio o transformación (Gili Gaya, 1980). En tanto expresa estados, está léxicamente incapacitado para expresar un cambio o progreso durante el periodo de tiempo en que transcurre (De Miguel, 1999).

I. TIPOS VERBALES

35. Públicos [V.PUBLIC] (Decir, explicar, admitir, estar de acuerdo, declarar, quejarse, remarcar, replicar, sugerir)

Reportan acciones propias de la actividad científica (Harvey, 2002). Son verbos que denotan acciones que pueden ser observadas públicamente; son actos de habla primarios tales como decir y explicar y se usan comúnmente para introducir aseveraciones indirectas (Biber, 1988).

36. Privados [V.PRIVAD] (Descubrir, creer, concluir, decidir, adivinar, encontrar, sentir, temer, determinar, demostrar, estimar, reconocer)

Se refieren a actividades perceptibles sólo por el propio hablante y se utilizan también como atenuadores (Palmer, 1974; Weber & Benivoglio, 1991). Los verbos privados se distinguen por expresar estados intelectuales o actos intelectuales no observables (Biber, 1988). Corresponden a los procesos mentales de Halliday (1994) y a los verbos epistémicos léxicos de Hyland (1998).

37. Persuasivos [V.PERSUA] (Acordar, arreglar, preguntar, rogar, insistir, instruir, estipular, sugerir)

Implican intenciones que suponen un cambio en el futuro (acordar, arreglar, preguntar, rogar, comandar, decidir, demandar, garantizar, insistir, instruir, ordenar, comprometer, pronunciar, proponer, recomendar, requerir, estipular, sugerir, urgir) (Biber, 1988). Según Bosari (2000) aluden a acciones actitudinales que podemos realizar para sobre el contenido proposicional de lo dicho.

38. Perceptivos [V.PERCEP] (Sentir, ver, oír, oler).

Expresan estados intelectuales cuando el objeto sobre el que recae la acción es un sustantivo abstracto. Tienen un significado más concreto cuando los objetos en los que recae la acción son sustantivos concretos. Corresponden a un subtipo de los procesos mentales de Halliday (Biber, 1988). Revelan focalización interna en la narración (Arlanza, 2001).

J. VERBOS MODALES

39. Posibilidad [V.MOD.POS]

Expresan la opinión y la actitud del hablante/escritor respecto al contenido declarado, atenuando generalmente la fuerza de los enunciados que se omiten. Son frecuentes en los artículos de investigación científica (Hyland, 1998) como una forma de adelantarse a las posibles objeciones y permite que el autor no aparezca tan taxativo.

40. Necesidad [V.MOD.NEC]

Expresan compromiso del escritor con lo dicho en el artículo de investigación científica (Hyland, 1998), aumentando la fuerza de lo aseverado.

41. Obligación [V.MOD.OBL]

Expresan el punto de vista del escritor que juzga la verdad de lo dicho en términos de certeza (Hyland, 1998). Además, la modalidad deónica se apoya de la necesidad o la posibilidad de los actos ejecutados por agentes moralmente responsables (Ortiz, 2000).

42. Volición [V.MOD.VOL]

Para Langacker (1990) los modales de capacidad y de volición, o sea, los modales dinámicos, se anclan en la realidad, mientras que los otros modales disponen de un mecanismo de proyección y parece que se anclan en la predicción. Según Olbeitz (1998), los modales dinámicos (de capacidad y volición) corresponden a la modalidad orientada al participante.

K. MARCADORES DE MODALIDAD

43. Atenuadores [ATENUAD]

Son marcadores informales y menos específicos de probabilidad e incertidumbre. Marcán el contenido de una proposición como incierta. Co-ocurren con marcadores de interactividad como los pronombres de segunda persona y los signos de interrogación. Aparecen en mayor medida en textos científicos de ciencias sociales (Markkanen & Schröder, 2000). Son propios del discurso académico, pero varían notablemente su uso según el tipo de texto. Son escasos en los manuales o en los materiales de apoyo para un curso. En estos tipos de texto se transmite un saber decentrado, sobre el que hay un consenso amplio: se enseña lo que la comunidad académica en cuestión tiene por seguro, y al enseñar, se simplifica ignorando las alternativas que pueden hacer avanzar el conocimiento (Vásquez, 2001). Son recursos que se ponen al servicio de la eficacia o eficiencia pragmática, es decir, se emplean para lograr efectos comunicativos que van más allá de la mera transmisión de información (Narbóna, 2000).

44. Enfáticos [ENFATIZ]

Acentúan el valor de los verbos. Se usan para indicar en términos positivos la confiabilidad de las proposiciones. Pueden ser usados en funciones no proposicionales para señalar solidaridad con el interlocutor. Característico de la entrevista oral (Cepeda, 2002).

L. ADVERBIOS

45. De lugar [ADV.LUG]

Sitúan la significación del verbo en coordenadas espaciales y añaden información que completa la estructura argumental del predicado (Bosque, 1990).

46. De tiempo [ADV.TIEMP]

Debido a su función delictiva establecen marcos de orden de la sucesión de hechos o pistas contextuales para la interpretación de lo dicho (Kovacs, 1999). Son circunstanciales en posición posverbal. Actúan como circunstanciales si están en el ámbito de la interrogación y de la negación.

47. De modo [ADV.MOD]

Según Bassols y Torrent (1997) son propios de la descripción. En principio denotan la manera en que se presentan los acontecimientos o se realizan las acciones (Bosque, 1990).

48. De cantidad [ADV.CANT]

Expresan cantidad o intensidad y afectan principalmente al verbo (Pérez Rioja, 1971). Como permiten expresar cualidades o determinaciones, contribuyen a imprimir mayor expresividad a las descripciones o ampliar una explicación.

M. MARCADORES DE SUBORDINACIÓN

49. Subordinadas sustantivas con "que" [SUB.SUST]

Las oraciones subordinadas sustantivas introducidas por "que" desempeñan las funciones propias de los sustantivos, aunque no se corresponde en todos los casos con sintagmas nominales. Por su parte, el empleo del determinante está semánticamente establecido: permite designar eventos o procesos que no se contemplan en su realización sino en su resultado (Delbecque & Lamiroy, 1999; Demonte, 1997).

50. Subordinadas adjetivas de pronombre relativo [SUB.ADJ]

El empleo de las subordinadas de relativo (adjetivas) "permite atribuir al sustantivo cualidades muy complejas para las cuales no tiene el idioma adjetivos o participios léxicos" (Gili Gaya, 1980: 301).

51. Subordinadas adverbiales de razón o causa/efecto. [SUB.ADV.c-e]

Señalan la razón del hecho mencionado en la oración principal. Permiten expresar la causa o la consecuencia (Gili Gaya, 1980). La causa o la consecuencia van normalmente introducidos por un conector causal, que indica que los enunciados que lo siguen explican o dan razón de los enunciados antecedentes; o consecutivo, que indican que los enunciados que lo siguen son efecto de los razonamientos antecedentes (Calsamiglia & Tusón, 1999).

52. Subordinadas adverbiales de concesión [SUB.ADV.conc]

Expresan una reserva que carece de eficacia para que se efectúe lo expuesto en la oración principal, cuya realización se cumple a pesar del obstáculo (Bassols & Torrent, 1997). Contribuyen a la estructuración del discurso argumentativo.

53. Subordinadas adverbiales condicionales [SUB.ADV.cond]

Indican la condición que debe cumplirse para que se realice lo expresado en la oración principal (Gili Gaya, 1980). Contribuyen a hacer explícitos los razonamientos, por tanto, son propios de los discursos expositivos y argumentativos.

54. Subordinadas adverbiales de tiempo [SUB.ADV.tiem]

Expresan el tiempo de la acción contenida en la oración principal, determinando una relación de anterioridad, simultaneidad o posterioridad (Pérez-Rioja, 1971). Su función es ubicar temporalmente la acción principal en relación con la subordinada. Con ayuda de estos adverbios es posible expresar los matices temporales para los que la sola presencia del verbo no es suficiente (Gili Gaya, 1980).

55. Frases infinitivas en función nominal [FRA.INF.nom]

Indican un intento por condensar información, omitiendo los participantes (Biber, 1988; Chafe, 1982, 1985). Tienen un carácter eventivo relacionado con la descripción de procesos o de actividades alécticas (Demonte & Varela, 1997). Desde un punto de vista lógico, son oraciones abiertas, puesto que el sujeto del infinitivo es una variable que se extrae de los sintagmas nominales que aparecen en el mismo contexto lingüístico (Gómez, 1999). Según Biber (1988), estas estructuras no cumplen funciones prototípicas ni fijas.

N. FRASES PREPOSICIONALES Y ADJETIVOS

56. Frases preposicionales (complemento del nombre) [FRA.PREP.cn]

Sirven para integrar grandes cantidades de información en un texto (Biber, 1988; Chafe, 1982, 1985). Por otra parte, expresan distintos tipos de relación entre el nombre y el complemento (relaciones de pertenencia, de asociación, de origen, de materia, de medida, peso o edad) (Picallo, 1999).

57. Adjetivos atributivos (calificativos) [ADJ.ATRIB]

El adjetivo puede cumplir la función de predicado de una oración copulativa, caracterizadora o como complemento predicativo, bien obligatoriamente escogido por el verbo, bien opcional. La construcción modificadora o atributiva tiene una estrecha relación con la construcción del predicado nominal ya que casi todos los adjetivos que funcionan como predicados en oraciones copulativas caracterizadoras pueden ser también modificadores. Los adjetivos son palabras que se aplican a otras palabras que nombran objetos físicos o mentales, por medio de los adjetivos se adscribe a esos objetos una propiedad o un conjunto de propiedades. Más estrechamente, un adjetivo modificador adscribe propiedades cuya especificación sirve para definir o delinear con mayor precisión a la entidad mentada para caracterizarla e identificarla entre varias similares para clasificarla o establecer taxonomías culturales y científicas para indicar relaciones genéticas o metonímicas (relaciones parte-todo). La característica fundamental de los adjetivos (atributivos y predicativos), lo que lo diferencia de los nombres, es que son términos generales y por ello pueden aplicarse a múltiples objetos y son de carácter gradual (Picallo, 1999).

58. Adjetivos predicativos [ADJ.PRED]

Son adjetivos que expresan cualidades del sujeto por medio de las oraciones atributivas (Gili Gaya, 1980), dicho de otro modo, su dependencia con el sustantivo se hace de forma indirecta a través de un índice verbal (Marcos Marín, 1975). Hay adjetivos que se combinan con el verbo ser: pertenecen a este grupo los adjetivos de relación y procedencia. Otros adjetivos sólo se utilizan con el verbo estar. En general, son todos aquellos que indican un resultado. Por eso, los participios -salvo en construcciones pasivas- se combinan siempre con el verbo estar. Debe tenerse en cuenta que hay ocasiones en que los participios se combinan con el verbo ser porque de esta combinación surge un significado diferente. Otros adjetivos que sólo se combinan con el verbo estar son: contento, satisfecho. Lo mismo sucede con adjetivos que poseen significado de participios, indicando el resultado de algo. Otros adjetivos se combinan con el verbo ser y estar, el significado varía en cada caso. Con el verbo ser, el adjetivo designa: a) una cualidad inherente a lo designado por el sujeto; b) una característica más o menos permanente del sujeto, que pertenece a su descripción o que lo clasifica (lo introduce en una clase de entes). Con el verbo estar, designan cualidad adquirida, producto de un cambio real o pensado.

59. Adjetivos demostrativos [ADJ.DEM]

Los adjetivos demostrativos actualizan al sustantivo. Pertenece a la categoría pragmática de los definitivos por lo tanto su interpretación está estrechamente ligada tanto al contexto textual como extra-textual. Los determinantes (los artículos, posesivos y demostrativos) y los cuantificadores (los numerales e indefinidos) constituyen clases cerradas con un número fijo de miembros desprovistos de significado léxico. A estos términos, a los que no es posible darle una definición de diccionario (no corresponden en verdad a ningún campo nocional), no les falta, sin embargo, significado gramatical o funcional: los artículos marcan género y número, sirven para introducir un nombre en el discurso o para establecer una relación anafórica, expresan unicidad o presuposición de existencia (Picallo, 1999).

60. Participio función adjetiva [PARTICI.adj]

Se sitúa preferentemente en el discurso escrito más que en el oral y su interpretación usual es que se usa para la integración y las elaboraciones estructurales (Biber, 1988). Janda (1985) establece que se utilizan en la forma de apuntes porque son más compactos e integrados y por ello sirven para la producción de un discurso altamente informacional cuando el tiempo es limitado. Clapuscio (1992) destaca el uso de los verboides, entre ellos el participio, como recurso de desagentivación, procedimiento de ocultación del agente que permite condensar la información. La autora ha constatado su frecuencia en textos de divulgación científica y en textos científicos.

N. MARCADORES DE COORDINACIÓN

61. Conjunciones adversativas, aditivas y disyuntivas [CONJ.dis.adv.ad]

Son características de la lengua escrita y de la prosa narrativa (De Kock & Gómez, 2002). La coordinación es el procedimiento gramatical que se usa para asociar constituyentes sintácticos sin establecer una jerarquía gramatical entre ellos (Camacho, 1999). El uso de estas conjunciones es frecuente como índice de simplicidad o sencillez de la lengua hablada. Según Ávila (2000), el habla utiliza con mayor frecuencia la coordinación antes que la subordinación, especialmente en las conversaciones cara a cara y a distancia.

O. MARCADORES DE NEGACIÓN

62. Adverbio de negación [ADV.NEG]

Importan información más coloquial y fragmentada (Biber, 1988). Son indicadores de la modalidad negativa en el sentido que hacen referencia a la actitud del emisor con respecto al receptor y al propio mensaje.

63. Adverbios de negación temporal [ADV.NEG.tiem]

Los adverbios de negación temporales (nunca, jamás, tampoco) son capaces de expresar negación cuando preceden al verbo (Kovacci, 1999).

64. Conjunción de negación [CONJ.ni]

La conjunción no presenta una caracterización funcional particular, adquiriendo una orientación comunicativa determinada por el rol de los elementos adyacentes que pone en relación (Hernández, 1996). Tal es el caso de la conjunción negativa "ni".

65. Pronombre de negación [PRON.NEG]

Es de uso más coloquial (Tottie, 1983). Son pronombres invariables irrefinables cuantificadores existenciales en el ámbito de la negación (Sánchez, 1999). Puesto que funcionan como sustantivos, admiten la presencia de un adyacente adjetivo, que adopta invariablemente el significante propio de los morfemas de masculino y singular (Alarcos Llorach, 1999).

ANEXO 2: EXPLICACIÓN DE LAS CINCO DIMENSIONES

Dimensión 1: Foco Contextual e Interactivo

Para interpretar el Factor 1 y lograr determinar la dimensión subyacente, se deben evaluar las funciones compartidas por la mayoría de los rasgos que co-ocurren. De los 23 rasgos iniciales con peso estadístico sobre 0,40, se desatan quince rasgos con pesos superiores a 0,60. Cabe señalar que los rasgos positivos que co-ocurren en este factor tienen los pesos más altos en comparación con todos los rasgos que caracterizan a los otros seis factores. Por su parte, los rasgos con pesos negativos en este factor son relativamente menores en número y sólo cuatro presentan pesos sobre 0,40.

Listados en un orden decreciente, es decir, de mayor a menor peso, los rasgos más representativos son: subordinadas adverbiales de causa/efecto (0,945), adverbios de tiempo (0,934), adverbio de negación (0,928), pronombres de segunda persona (0,911) y de primera singular (0,823), desinencias verbales de segunda persona singular (0,813), pronombres de negación (0,731), adverbios de lugar (0,723), modo indicativo (0,693), desinencias de primera persona singular (0,668), pronombres demostrativos (0,592), subordinadas adverbiales de tiempo (0,487), y tiempo presente (0,420). Este factor es, además, uno de los que presenta mayor número de rasgos negativos, entre los cuales se cuenta: frases preposicionales (-0,545), nominalizaciones (-0,479), sustantivos (-0,443), participios en función adjetiva (-0,437).

Los rasgos que se reúnen en este factor denotan una gran relación funcional entre la mayoría de ellos. Su interpretación no resulta compleja y la bibliografía pertinente entrega apoyo suficiente. De este modo, en principio, se puede sostener que altas frecuencias de co-ocurrencia de adverbios de tiempo y de lugar, subordinadas adverbiales de causa/efecto y de tiempo y de pronombres demostrativos 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 (lo que explica la sucesión de los acontecimientos). También ellos implican un compromiso del autor del texto y condensan la marca de lo situado (RAE, 1973; Pérez-Rioja, 1971; Di Tullio, 1997; Kovacci, 1999). Junto a estos rasgos de situacionalidad y contexto, el modo indicativo y el tiempo presente apuntan a una modalidad declarativa/experiencial (Hernández, 1996; Bassols & Torrent, 1997; Ávila, 2000; Criado de Val, 1962; Gómez Macker & Peronard, 1988), característica de la entrevista oral (Cepeda, 2002). Los pronombres de primera y segunda persona singular y las desinencias verbales de primera y segunda persona singular reflejan directamente a los participantes. También son clásicamente considerados como marcadores de la presencia del sujeto en el texto e indican un foco interpersonal y un estilo más comprometido. Estos últimos rasgos normalmente involucran a un destinatario específico e indican un alto grado de acción, ya que aluden a una comunicación directa, un estilo directo (De Kock & Gómez, 2002); por ende, son representativos de la interacción dialógica.

Por su parte, los rasgos con pesos estadísticos negativos tales como las nominalizaciones y los sustantivos son clásicamente considerados portadores de la carga referencial del texto y una alta frecuencia de ellos señala una fuerte densidad informacional (Piccallo, 1999; Di Tullio, 1997; Clapuscio, 1992; Chafe, 1982, 1985; Chafe & Danielewicz, 1987; Halliday, 1993; Biber & Finnegan, 1986; Biber, 1988). Del mismo modo, las estructuras adjetivas (frases de participio en función adjetiva y frases preposicionales) actúan como expansiones en el proceso de textualización, es decir, ellas permiten la integración y precisión de grandes cantidades de información en un texto (Clapuscio, 1992; Harvey, 2002). Como se comprende, los textos caracterizados por estos tipos de rasgos suelen ser altamente abstractos, condensar cantidad-

des de información importante y expresar significados complejos; por lo tanto, constituyen un recurso típico del lenguaje técnico y científico (Hernanz, 1999; Biber, 1988; Burdach, 2000; Zarzalejos, 2001; Moyano, 2000).

En resumen, la dimensión que se infiere a partir de este factor se concreta en la acción, en la sucesión de los acontecimientos y en las relaciones interpersonales tipo dialógico; lo que también se destaca a través de la deixis temporal, espacial y demostrativa. Todos los rasgos lingüísticos involucrados, en su conjunto, permiten suponer que los textos caracterizados por esta dimensión no contienen información altamente abstracta como tampoco evidencian una integración concisa y precisa de la información, sino por el contrario, la alta frecuencia de ocurrencia de estos rasgos tipificadores y estadísticamente positivos se asocian con un foco en la explicitud y dependencia del contexto y en la activa participación de los interlocutores, rasgos clásicos del discurso oral y dialógico. Esta dimensión emergente caracteriza los discursos en que se explicita las marcas espacio-temporal y donde existe colaboración mutua entre los participantes, lo que los torna más auténticos y menos planificados; la espontaneidad e interactividad son típicas de estos discursos. En ellos pueden coexistir múltiples organizaciones discursivas, tales como: narración, exposición, argumentación y descripción. Por todo ello, como un modo de reflejar las funciones primordiales detectadas en conjunto, hemos nombrado esta dimensión con la etiqueta "Foco Contextual e Interactivo".

Puntajes factoriales y relaciones entre corpus (literario, oral y técnico)

Dado que nos interesa mostrar las relaciones entre los tres registros (técnico-científico, CTC; literatura latinoamericana, CLL; y entrevistas orales semi-estructuradas, CEO) y -más específicamente- aislar variables identificatorias del discurso técnico-científico en contextos didácticos, se comparan estos tres registros a través de cada una de las dimensiones interpretadas. Para ello se realiza un estudio estadístico a partir de los puntajes factoriales que resumen la frecuencia de cada uno de los rasgos en cada factor para cada uno de los textos. Los puntajes factoriales para cada texto se promedian con los puntos de todos los textos de un corpus específico (CTC, CLL y CEO) y, de este modo, se obtiene un promedio del punto factorial para cada dimensión. Estos puntajes promedio por corpus o registro se comparan con el fin de determinar los tipos de relaciones (similaridad o diferencia) existentes (Hair, Anderson, Tathan & Black, 1999).

De la aplicación de estos procedimientos estadísticos se deduce que, en la Dimensión 1 Foco Contextual e Interactivo, las entrevistas orales (CEO) evidencian el mayor peso (26,16), los textos del corpus literario (CLL), por su parte, ostentan un peso mucho menor (7,12) y, en tercer lugar, se ubican los textos del corpus técnico-científico (CTC) con un peso negativo (-2,14). Estos datos estadísticos permiten confirmar que la Dimensión 1 aparece fuertemente ligada, en primer lugar, a los textos orales en su modalidad dialógica tipo entrevista, seguida por los textos literarios escritos; por tanto, se hace evidente que la dimensión Foco Contextual e Interactivo no es prototípica de los textos técnicos.

Dimensión 2: Foco Narrativo

El Factor 2 también presenta 23 rasgos co-ocurrentes que se distribuyen entre los puntajes 0,84 y 0,40. En orden decreciente, los rasgos son: pronombres de tercera persona singular (0,842), pronombres de primera persona plural (0,828), futuro perifrásico (0,823), pretérito imperfecto (0,820), pronombres de tercera persona plural (0,708), modo indicativo (0,686), desinencias de primera persona plural (0,667), verbos modales de volición (0,651), pretérito indefinido (0,614), pronombres de negación (0,590), verbos privados (0,577), adverbios de lugar (0,533), pronombres de segunda persona singular (0,529), verbos perceptivos (0,496),

adverbio de negación (0,493), adverbios de tiempo (0,482), verbo estar estativo activo (0,460), verbos públicos (0,445), pronombres de primera persona singular (0,431), desinencias de tercera persona singular (0,423), conjunciones aditivas, disyuntivas o adversativas (0,411), frases infinitivas en función nominal (0,405) y conjunción de negación "ni" (0,402). Por otra parte, con peso negativo, están los rasgos relacionados con nominalizaciones (-0,581), frases preposicionales en función del complemento del nombre (0,562) y adjetivos atributivos (-0,442).

La presencia significativa de los pronombres personales, especialmente los de tercera persona singular y primera persona plural, sujetos humanos, protagonistas de historias (Longacre, 1983), al igual que las respectivas desinencias, muestra un marcado acento en la identificación de las personas del discurso, las presentes en el momento de la enunciación y las ausentes en relación a aquellas (Calsamiglia & Tusón, 1999). Congruente con lo anterior, surgen en co-ocurrencia los rasgos asociados a los tiempos verbales del pasado, el pretérito imperfecto, que describe situaciones y circunstancias y, su contraparte, el pretérito indefinido, que señala los sucesos, y, por tanto, el dinamismo de las acciones (Kovacci, 1993), denotando así una referencia directa al mundo narrado (Weinrich, 1974; Arroyo, 2000; De Kock & Gómez, 2002). No menos importante es la co-aparición del futuro perifrásico, demostrándose que si bien los hechos narrados se sitúan generalmente en el pasado, otros transcurren también en la actualidad de los personajes (Contreras, 2000). A estos tiempos verbales se asocia el uso del modo indicativo, que permite a los protagonistas expresar estados o acciones como reales (Gómez Macker & Peronard, 1988). Todo esto se complementa con el uso de los verbos modales de volición, verbos privados, verbos perceptivos y verbos públicos que se orientan a dar cuenta de las actitudes subjetivas del sujeto hablante (Aranzén, 2001). De hecho, la notoria de los verbos de actividad mental -empleados sobre todo en primera persona singular- no es sólo su uso, según lo que Palmer (1974) denomina verbos privados y que se refieren a actividades perceptibles sólo por el propio hablante, sino que conjuntamente ellos pueden cumplir también una función atenuadora (Weber & Bentivoglio, 1991).

Ahora bien, la trama personal y temporal hasta aquí conformada encuentra su anclaje en los adverbios de lugar y de tiempo que sitúan la significación del verbo en unas coordenadas espaciales o temporales (Bosque, 1990). Asimismo, el adverbio de negación, la conjunción "ni" y las conjunciones aditivas, adversativas o disyuntivas sirven de apoyo a la sucesión de hechos en el pasado (Biber, 1988; De Kock & Gómez, 2002; RAE, 1973; Pérez-Rioja, 1971). Por último, el verbo estar estativo activo destaca por su frecuencia en secuencias descriptivas (Bassols & Torrent, 1997; Loriente, 2002) y las frases infinitivas en función nominal se explican por su incidencia en construcciones en que aparecen los verbos modales de volición y los privados, los que remiten a los participantes en un evento comunicativo real o ficcional (Hernanz, 1999).

Los rasgos con peso negativo, tales como las nominalizaciones, las frases preposicionales y los adjetivos atributivos se complementan en la integración y densidad de la información (Chafe, 1982, 1985, 1994; Janda, 1985; Clapuscio, 1992; Halliday & Martín, 1993; Hernanz, 1999; Burdach, 2000; Moyano, 2000; Zarzalejos, 2001). La dimensión que se perfila a partir del Factor 2 se identifica con una sucesión cronológica de eventos principalmente en el pasado y descripción de todo lo que rodea a dichos eventos. Esto último se complementa con los indicadores de tiempo y lugar. Además, la fuerte incidencia de la deixis personal permite expresar la presencia de los protagonistas, ya sea por medio de la presencia de puntos de vista internos mediados por la conciencia del hablante o a través de puntos de vista externos, situados fuera de esta conciencia. Al primer punto de vista, corresponden estados internos, perceptivos, privados y, al segundo, aquellos más públicos. En síntesis, este factor se asocia preferentemente con una sucesión de acontecimientos, que implica circunstancias de tiempo y

Iugar, como también la participación de las personas del discurso. Por tanto, el Factor 2 permite conformar una dimensión funcional que hemos denominado Foco Narrativo.

La aplicación de los correspondientes procedimientos estadísticos entre registros revela que la Dimensión 2 Foco Narrativo presenta el mayor puntaje factorial (29,32) en el registro literario (CLL), a este le siguen los textos del corpus de entrevistas orales (CEO) con un peso positivo de 21,22 y, por último, se ubican los textos del corpus técnico-científico (CTC) con -comparativamente- el menor peso positivo: 1,12. El resultado obtenido por el registro literario (CLL) es congruente con los rasgos co-ocurrentes y que tipifican a dicho discurso, esto es, la estructura narrativa y el mundo ficticio que instaura. La primera consiste en el ordenamiento de los eventos relatados en el pasado. Lo segundo, en cambio, en la serie de objetos, personajes, lugares, tiempos, eventos y leyes que rigen las relaciones entre ellos.

Dimensión 3: Foco Compromiso

Para interpretar el Factor 3 se han considerado 17 rasgos que co-ocurren con valores superiores a 0,40. Listados en un orden de mayor a menor peso, los rasgos considerados son: verbos privados (0,824), pronombres de primera persona singular (0,789), pretérito indefinido (0,705), verbos de volición (0,655), desinencias de primera persona singular (0,640), modo indicativo (0,630), pretérito imperfecto (0,604), pronombres de negación (0,569), pronombres de tercera persona singular (0,563), desinencias de segunda persona plural (0,562), frases infinitivas (0,518), subordinadas sustantivas (0,467), subordinadas adverbiales de concesión (0,452), verbo estar estativo activo (0,435), pronombres de segunda persona singular (0,427), adverbios de tiempo negación (0,411), pronombres de primera persona plural (0,402).

Los verbos privados (Biber, 1988; Weber & Berlivioglio, 1991), los verbos de volición (Gómez, 1999) y los pronombres de primera y segunda persona singular remiten a los participantes de un acto de comunicación (Fernández, 1999), específicamente, a personas que manifiestan sus intenciones y actitudes. De igual modo, los pronombres de primera persona plural y las desinencias de primera persona singular y de segunda persona plural confirman como característica esencial de este factor la manifestación explícita del emisor del discurso (Calsamiglia & Tusón, 1999; Crismore, 1989). La presencia de marcas de pretérito indefinido sugiere que los verbos descritos anteriormente se refieren a acciones pasadas con un final temporal determinado, es decir, a construcciones que marcan el resultado de la acción que el verbo expresa. El modo indicativo, por su parte, hace referencia a hechos reales localizados en un tiempo verdadero (Criado de Val, 1962; Contreras, 1984) y expresa la modalidad declarativa/experiencial característica del intercambio discursivo oral (Cepeda, 2002). En relación a lo anterior, se postula también que el modo indicativo es un rasgo a través del cual se expresan estados o acciones como reales (Gómez Macker & Peronard, 1988), es decir, este rasgo caracteriza intercambios lingüísticos cuyos referentes son hechos concretos en un aquí y un ahora determinados. En las oraciones sustantivas subordinadas, el empleo de la subjunción que está semánticamente condicionada permite designar eventos o procesos que no se contemplan en su realización sino en su resultado, a sea, como hechos ya establecidos, que se conciben como algo previo a la enunciación (Delbecque & Lamiroy, 1999). La presencia de este rasgo sugiere que los interlocutores en el discurso, destacados anteriormente, se manifiestan directamente sobre hechos específicos con un final temporal determinado. La aparición de frases infinitivas puede ser explicada a partir de su recurrencia en construcciones con los tipos de verbos que hemos presentado anteriormente: a saber, verbos privados y de volición. Aunque la función característica de estas construcciones es condensar información (Halliday, 1994; Burdach, 2000), en este caso esta función está subordinada a otros rasgos que aparecen con más peso tales

como verbos privados y de volición, además de las desinencias y los pronombres personales en general. Por último, las frases preposicionales (Hernanz, 1999; Zarzalejos, 2001; Harvey, 2002), único rasgo que aparece con un valor negativo (-0,457) se asocian con una alta carga informativa ya que expanden construcciones de mayor abstracción.

La interpretación de los rasgos de alta ocurrencia tales como verbos privados, verbos de volición y pronombres y desinencias verbales de primera persona en este factor constituyen la Dimensión 3 Foco Compromiso. Esta dimensión está asociada a textos en los que la intención y la actitud del emisor tienen mayor relevancia que el mensaje mismo; dicha de otra forma, esta dimensión caracteriza a textos en los que aparecen participantes reales que expresan intenciones y actitudes proposicionales frente a lo dicho. La clara presencia del "yo" a través de los pronombres y desinencias de primera persona es una evidencia clara de la voluntad de involucrarse en el discurso de manera explícita y asumir un rol preponderante.

Por último, podemos observar que el comportamiento de esta dimensión presenta su valor más alto en el CLL (9,52). En segundo lugar, se ubica el registro CEO con un valor de 7,7. A diferencia de estos puntos, en cambio, aparece con un valor negativo (-1,99) el registro CTC. Todo ello permite señalar que la Dimensión 3 Foco Compromiso caracteriza más a los registros en los que aparecen participantes reales que expresan intenciones y actitudes frente al mensaje, a saber, el CEO y el CLL. El valor negativo del CTC en esta dimensión se explica por la naturaleza del discurso técnico-científico, el cual se focaliza en el contenido informativo del mensaje lo que no exige la identificación de los interlocutores en el discurso; razón por la cual, los rasgos que expresan compromiso comparativamente con los otros registros, son menos frecuentes.

Dimensión 4: Foco Modalizador

Para el Factor 4 se han considerado 8 rasgos co-ocurrentes con peso positivo y un rasgo con peso negativo. Tal como en los casos anteriores, se excluyen los rasgos con peso inferior a 0,40. Los rasgos en cuestión son: formas activas con ser (0,671), atenuadores (0,556), verbos modales de posibilidad (0,641), adverbios modales (0,606), adjetivos predicativos (0,565), desinencias de tercera persona plural (0,549), subordinadas adjetivas (0,514) y desinencias de tercera persona singular (0,405). Además, se constata la presencia de sustantivos con peso negativo (0,494).

La co-ocurrencia de adjetivos predicativos, cuya función calificativa-atributiva se realiza de modo indirecto o no adjunto (González, 2000), unido a formas activas con "ser" sugieren una relación con una función descriptiva (Gili Goya, 1980; Bosque, 1990, 1999; Bassols & Torrent, 1997). La co-ocurrencia de los rasgos anteriores se vuelve relevante al aparecer junto a subordinadas adjetivas que permiten atribuir al sustantivo cualidades muy complejas para las cuales no tiene el idioma adjetivos o participios léxicos (Gili Goya, 1980: 301) y a adverbios de modo que denotan la manera en que se presentan los acontecimientos o se realizan las acciones (Bosque, 1990). Se puede decir que ambos rasgos (subordinadas adjetivas y adverbios de modo) remiten a procesos de "incidencia de un segmento lingüístico sobre otro para configurar una unidad superior" (Hernández 2000b: 391); del mismo modo, su presencia permite la articulación de una sintaxis más expandida y de menor concisión. Por su parte, los atenuadores -marcadores informales y menos específicos de probabilidad e incertidumbre- presentan el contenido de una proposición como incierto (Markkanen & Schröder, 2000). Su co-ocurrencia junto a adverbios de modo y verbos modales de posibilidad caracterizan textos más modalizados, en los que se marca la expresión de la actitud del hablante/escritor respecto al contenido declarado; en otras palabras, se destaca la visión particular del locutor respecto al contenido

de los enunciados que emite (Calzomiglio & Tusón, 1999; Hyland, 1998). En síntesis, la conjunción sistemática de estos rasgos apunta a discursos con énfasis en *cómo* (*modus*) se dicen las cosas en relación al contenido mismo, más que a *lo dicho* (*dicunt*).

El sustantivo ha sido clásicamente asociado a la función referencial. El peso negativo de este rasgo (-0,494) y los pesos positivos de los rasgos ya mencionados sugieren fuertemente la relación del Factor 4 con una función más descriptiva que informativa, más expresiva que referencial. Por todo lo dicho anteriormente, cómo un modo de reflejar las funciones principales que aparecen conjuntamente, hemos llamado a esta dimensión como Foco Modalizador, asociada principalmente a discursos con marcas actitudinales explícitas.

De la aplicación de los procedimientos estadísticos en base a los puntos factoriales, se desprende que los textos del corpus con más alto puntaje en esta dimensión corresponden al registro entrevistas orales (CEO), con 4,17. Los textos del corpus literario (CLL) presentan un punto bajo, de 1,31 en la dimensión Foco Modalizador (algunos de ellos tienen puntaje negativo). En cuanto al corpus técnico científico (CTC), la mayoría de los textos muestra un punto negativo promedio de -0,49. Los puntos obtenidos por CEO en la Dimensión 4, Foco Modalizador, son congruentes con algunas de las características comúnmente aceptadas de la oralidad. Entre estas, Norbona (2000) destaca el propósito socializador de la interacción cara a cara, que privilegio el uso de estrategias como la atenuación o la intensificación; así, los hablantes tratan de hacerse comprender y, al mismo tiempo, buscan influir sobre el oyente.

Dimensión 5: Foco Informativo

El Factor 5 está conformado por la correlación más significativa de los siguientes rasgos: verbos modales de obligación (0,496), modo subjuntivo (0,494), nominalizaciones (0,456), participios en función adjetiva (0,413), frases preposicionales (0,413), desinencias de tercera persona singular (0,632), pretérito indefinido (-0,630), verbo estar estativo activo (0,595), verbos privados (0,575), pronombres de negación (-0,572), y verbos modales de volición (-0,503).

La presencia de los rasgos positivos co-ocurrentes como verbos modales de obligación y modo subjuntivo da cuenta de la necesidad y certeza de los juicios expresados, correspondiendo fundamentalmente a una modalidad deontica (Hyland, 1998; Osorno, 2000). Además, en el caso del subjuntivo, éste remite a organizaciones de mayor complejidad sintáctica (Gili Goya, 1980) y a subordinación, cuya función es enmarcar la información del discurso (Delbeque & Lamiray, 1999; Galán, 1999). Por otra parte, este mismo rasgo es usado para la expresión de la especulación subjetiva y el mandato (Criado de Val, 1962; Gómez Macker & Peronard, 1988). La co-ocurrencia de nominalizaciones junto a participios en función de adjetivo y frases preposicionales se presenta como indicador de integración y compactación de información altamente abstracta, propia de un discurso académico nominal (Burdach, 2000; Picallo, 1999; Chafe, 1982, 1985; Bibel, 1986, 1988).

Los rasgos negativos que presentan un mayor peso en este factor son: las desinencias de tercera persona singular, las cuales se utilizan cuando existe el riesgo de malinterpretar la referencia porque la información que aporta el contexto falla (Castellano, 2000). El pretérito indefinido, el cual sitúa la acción de un suceso o evento en un espacio temporal acabado, es decir, es acción que ya no se puede volver a realizar (Contreras, 1984). El verbo estar estativo activo, aunque no tiene valor especializado (Lorente, 2002), forma parte de las llamadas unidades de conocimiento. Los verbos privados, que se destacan por expresar estados intelectuales o actos intelectuales no observables (Weber & Bentivoglio, 1991; Bibel, 1988). Los pronombres de negación, cuyo uso es preferentemente coloquial (Sánchez, 1999; Taitie, 1983). Y, por último, los verbos modales de volición, que dan cuenta de una modalidad dinámica arraigada en la

realidad (Langacker, 1990) o de una modalidad orientada al participante, es decir, una modalidad que pertenece al estatus del sujeto (Olbertz, 1998). Cabe destacar que todos los rasgos positivos constituyentes de este factor se presentan preferentemente en discursos escritos y, en particular, prototípicamente en el artículo de investigación científica (Hyland, 1998; Cornillie, 2003; Criado de Val, 1962; Harvey, 2002; Burdach, 2000).

En síntesis, los rasgos positivos agrupados en este factor están fundamentalmente orientados hacia la informatividad, entendida ésta como la concentración de información en unidades y estructuras lingüísticas más pequeñas, que presentan los datos lo más concisa y precisamente posible (Biber, 1988; Halliday & Martín, 1993; Burdach, 2000). Observamos así un conjunto de rasgos que dan cuenta de una alta carga informativa, asociada a una entrega regulada y compactada de la información, fundamentalmente referencial. Como se aprecia, los discursos caracterizados por estos tipos de rasgos suelen ser altamente abstractos, condensar cantidades relevantes de datos y expresar significados complejos. Por otro parte, los rasgos negativos de este factor se orientan hacia una contextualización de los estados o actos intelectuales no observados de un sujeto, en un momento determinado (Capusso, 1992). La presencia de los rasgos antes mencionados permite distinguir claramente entre discursos con alta cantidad de información y, por ende, un mayor grado de abstracción de los que presentan menor cantidad de información, por ello es posible denominar a este Factor 5 como Dimensión Foco Informativo.

En relación al estudio estadístico comparativo, observamos que esta dimensión nos permite diferenciar significativamente el registro CTC (0,90) del CEO (-3,58) y del CLL (-4,28). De esta manera observamos que la configuración de los datos apoya la idea de que el registro CTC está conformado por textos cuyos rasgos lingüísticos dan cuenta de una alta cantidad de información, en tanto que los registros CLL y CEO presentan rasgos lingüísticos escasamente relacionados con la compactación y abstracción de la información.

Capítulo 5

Hacia una identificación automatizada de rasgos argumentativos en corpus

RENÉ VENEGAS V.

INTRODUCCIÓN

La argumentación ha sido objeto de estudio desde épocas remotas. Los primeros atisbos de argumentación se reconocen en la Grecia antigua a través de la retórica sofística, lo cual emerge producto, principalmente, del proceso de recuperación de tareas llevado a cabo después de las tiranías de Gelon y Hieron. Es decir, la argumentación nace como respuesta a una catástrofe cultural y política (Platón, 1998).

Platón (1985 b) fue enemigo acérrimo de la retórica de los sofistas, a los que reprimió el dar preeminencia a la opinión frente a la verdad. De hecho, Platón saca a la palestra la relación entre retórica y filosofía, las que para él están totalmente separadas (cfr. Gorgias). A pesar de su postura antisofística su actitud en relación a la retórica no es del todo negativa. En el Fedro, Platón (1985 a) diferencia entre dos retóricas: una, la de los sofistas, con connotaciones negativas, porque trata de persuadir a cualquier precio sin tener ninguna consideración por la honestidad intelectual. Por otro lado, existe una retórica positiva interesada por la dialéctica y por la búsqueda de la verdad que ayuda a la formación de los espíritus.

Más tarde, aparece una figura clave en el desarrollo de la retórica: Aristóteles, quien distingue entre dos ámbitos bien diferenciados: por un lado la ciencia donde las demostraciones se han de basar en la certeza y la verdad, y por otro, el discurso persuasivo que argumenta sobre lo probable, sobre lo verosímil. La primera, al

basarse en la certeza trata de convencer a un auditorio universal con los mismos razonamientos, mientras que el segundo utiliza pruebas para persuadir no a todos de la misma manera, sino en relación al tipo de público. De aquí surgen los diferentes géneros de la retórica: el judicial, el epidídico y el deliberativo.

A través de la civilización griega la retórica llega a los romanos. La subsistencia de este arte en el Imperio Romano está conectado con las diferentes formas de gobierno. Aflora con la república, pero se cierra en sí misma cuando ésta cae.

La retórica durante los siglos XVIII y XIX pierde su importancia y es considerada solamente desde un punto de vista estilístico, convirtiéndose a principios del siglo XX en sólo un listado de figuras retóricas, por último, con el advenimiento del positivismo lógico adquiere una franca connotación peyorativa.

No es sino hasta final de los años cincuenta del siglo XX, que resurgen los estudios sobre la argumentación. El año 1958 es crucial, pues se publican dos obras fundamentales para los estudios de la argumentación. Por una parte, aparece el *Tratado de la Argumentación*, la *Nueva Retórica* de Perelman y Olbrecht-Tyteca y, por otra parte, el libro *The Uses of Argument* de Toulmin, obras provenientes de "horizontes diversos y escritas en estilos totalmente diferentes" [Plantin, 1998:17]. Ambas obras coinciden en una referencia común a la práctica jurídica y en la búsqueda de un pensamiento argumentativo orientado a los asuntos humanos. La gran diferencia existente entre estas obras, sin duda, está en la tendencia epistemológica, para la primera se reconoce una tendencia empírica-inductiva, en tanto que para la segunda es transparente su tendencia racionalista-deductiva. Ambas obras constituyen hoy clásicos de los estudios en argumentación.

Valga esta brevísima síntesis del desarrollo de la argumentación como punto de partida de un intento por descubrir los aspectos que la configuran como una clase de textos, una base textual o como hoy en día se prefiere, siguiendo las ideas de Adam [1992], una secuencia textual.

En este artículo nos interesa presentar los rasgos lingüísticos propios de la argumentación. Esta tarea se llevará a cabo planteando muy someramente la conceptualización de argumentación y los rasgos que la identifican en los textos en autores ya clásicos en esta área de investigación. En primer lugar, revisamos las ideas de Perelman y Olbrecht-Tyteca (1994), para luego adentrarnos en las de Toulmin (1958), fundadores de la hoy conocida Teoría de la Argumentación. Más adelante, se presentan dos visiones respecto del texto argumentativo, la primera de van Dijk (1987) y la segunda de Adam (1992). Finalmente, mostramos las posibilidades de reconocimiento de la argumentatividad a partir de los rasgos lingüísticos que nos presentan explícitamente, y en ocasiones implícitamente, los autores mencionados, así como también otros reconocidos autores que se han preocupado por el tema de la argumentación.

Lo anterior nos permitirá establecer una categorización de los rasgos lingüísticos relevantes de la argumentatividad, a partir de los cuales se desarrollará una investigación de corte descriptivo-cuantitativa que pretende indagar a partir de un corpus seleccionado la frecuencia de aparición de estos rasgos en distintos tipos de textos.

De esta manera, presentamos aquí en un primer momento el estudio bibliográfico orientado a la identificación y sistematización de los rasgos lingüísticos asociados a la argumen-

tación mencionados por los autores estudiados, para luego, en un segundo momento, dar cuenta de la búsqueda automatizada de estos rasgos en el *corpus textual PUCV-2003*, desarrollado por el equipo de lingüística de la *Pontificia Universidad Católica de Valparaíso* y constituido por 90 textos, organizados en tres grandes corpus, a saber: *Corpus Técnico-Científico* (CTC), *Corpus Literario* (CLL) y *Corpus de Entrevistas Orales* (CEO), con un total de 1.466.654 palabras. Los resultados cuantitativos obtenidos permiten diferenciar claramente los tres corpus en términos de sus rasgos de argumentatividad, siendo el *Corpus de Entrevistas Orales* el que más alta frecuencia de rasgos presenta, seguido por el *Corpus Técnico-Científico* y luego por el *Corpus Literario*.

1. MARCO TEÓRICO

1.1 Rasgos lingüísticos de la Argumentación

En este apartado se llevará a cabo una revisión bibliográfica, enfocada específicamente, a la conformación de un listado de rasgos lingüísticos caracterizadores de la argumentatividad. De esta manera intentaremos encontrar, en autores ya clásicos, algunos rasgos léxicos o morfosintácticos que nos permitan distinguir, en un texto cualquiera, la presencia de argumentación.

Iniciaremos esta búsqueda con dos de los acercamientos más clásicos de la teoría de la argumentación, esto es, revisaremos la obra de Perelman y Olbrecht-Tyteca (1958/1994), para luego revisar la obra de Toulmin (1958).

Para Perelman y Olbrecht-Tyteca (1958/1994) la teoría de la argumentación responde al objetivo de provocar o aumentar la adhesión de las personas a las tesis presentadas, por medio de técnicas discursivas. Estas técnicas discursivas se plasman en ciertos rasgos lingüísticos, los cuales son mencionados en el apartado § 38 *Formas verbales y argumentación*. En este apartado plantean que la calidad argumentativa se relaciona con la elección léxica, en ese sentido establecen que un término puede servir de indicio (de distinción, de familiaridad o de sencillez) o también, para situar el objeto del discurso dentro de una categoría mejor de lo que lo haría el uso del sinónimo. Dan cuenta también de la elección de términos alejados del lenguaje habitual como término indicadores de argumentación. Los autores determinan que el uso de paráfrasis sirve para desvalorizar un término determinado y presentarlo como más objetivo ("mentiroso" por "persona que induce a error"). Por otra parte, en el apartado § 39 se refieren a las modalidades, de aquí podemos extraer el uso argumentativo de la negación, además establecen que el uso de la subordinación y el de expresiones como: "por el amor de", "en consideración a", "a causa de", "aunque", "a pesar de", "sin duda" son posibilidades del lenguaje que pueden ser usadas en la argumentación.

Explicitan categóricamente que la modalidad asertiva conviene a la argumentación. Así como también dicen que la modalidad interrogativa es importante para la retórica. En cuanto a los tiempos, en una función argumentativa, promueven un estudio detallado por cada lengua y explicitan algunas posibilidades para el francés. Finalizan el apartado con la revisión de los pronombres, diciendo que el impersonal "se" a cambio de "yo" permite objetivizar, disminuir responsabilidad, así como lo sería cualquier uso de tercera persona en vez de la primera. Otros elementos mencionados por los autores son: el empleo

argumentativo del pronombre o adjetivo indefinido (cierto/a/s, uno/a/s), el uso del artículo determinado, del singular y plural y del demostrativo.

Por su parte, Toulmin (1958) considera como argumento todo aquello que se ofrece, o todo lo que es utilizado, para justificar o para refutar una proposición. Aquello puede ser el enunciado de un hecho, un resultado de la experiencia, a veces simplemente un ejemplo, una definición, el recuerdo de una regla, una creencia comúnmente compartida, o incluso la explicitación de una contradicción.

Este autor lamentablemente no ofrece un listado explícito de los rasgos lingüísticos de los argumentos, pero es posible determinarlos inferencialmente siguiendo los elementos funcionales de su modelo. Así tenemos:

Dato:

- a) Dato ejemplo: por ejemplo, verbigracia
- b) Dato testimonio: pronombre de primera persona, verbos en pasado, según
- c) Dato estadístico: datos cuantitativos

Tesis:

Uso de ciertos verbos como: pensar + que, creer + que

Garantía:

- a) Garantía de autoridad: según, como dice, se plantea que
- b) Garantía motivacional: léxico en relación a emociones y valores
- c) Garantía sustantiva: correctores de causa y efecto, cuantificadores universales y generalizaciones.

Respaldo:

- a) Respaldo estadístico: datos cuantitativos
- b) Respaldo de exemplificación: uso preferente del indicativo
- c) Respaldo de testimonio: verbos de opinión

Calificador:

Enfatizadores y atenuadores léxicos: probablemente, quizás, por lo que parece, presumiblemente, etc.

Refutación:

Negación, verbo dudar, a excepción de

Insistimos en que los rasgos lingüísticos presentados aquí son en su mayoría producto de inferencias realizadas a partir de las funciones de los componentes del argumento, según Toulmin (1958).

Estos autores, pilares teóricos de la argumentación, si bien nos ofrecen un marco teórico de la argumentación coherente con las tradiciones epistemológicas de cada cual, empíricista-inductiva para el primero y racionalista-deductiva para el segundo, no nos ofrecen de manera precisa y acabada un marco específico de rasgos lingüísticos propios de la argumentatividad que pueda ser utilizado por los analistas textuales para el reconocimiento de la argumentación en los textos.

Es por ello que recurriremos a dos autores, ya consagrados en el análisis de textos,

para intentar conformar un grupo de rasgos lingüísticos que permitan identificar cuándo un texto presenta argumentación.

El primero de ellos es van Dijk (1987) quien sostiene que los textos argumentativos están constituidos por un esquema básico muy conocido, a saber, hipótesis (premisa)-conclusión.

Para van Dijk (1987:158):

"La tarea [de la argumentación] consiste en convencer al oyente de la corrección o la verdad de la aseveración, aduciendo suposiciones que la confirmen y la hagan plausible, o bien suposiciones a partir de las que pueda deducirse la aseveración".

Van Dijk (1987) propone que la estructura del texto argumentativo puede ser descompuesta más allá de las categorías convencionales de hipótesis y conclusión. Estos dos elementos son básicos, pero por sí mismos no son suficientes, ya que hay que establecer la legitimidad de la derivación de la conclusión a partir de la premisa o hipótesis. Es lo que van Dijk (1987) llama "legitimidad o justificación", que establece la garantía del proceso deductivo y Toulmin (1958) denomina, regla general. A estas categorías se pueden añadir la especificación de lugar y tiempo en que esa regla general se aplica; especificación denominada *marco* por van Dijk (1987). Este se refiere tanto al lugar en que se produce (calle, casa, etc.) como al marco conceptual, es decir, el conocimiento compartido por los participantes de la argumentación. A veces es necesario especificar las circunstancias en que se produce la argumentación, enumerando los hechos, los elementos constatables que acompañan a la argumentación. Finalmente, en la argumentación cotidiana casi siempre se incluye una cláusula de pretexto, que establece la probabilidad de la secuencia hipótesis-conclusión, donde es muy posible que existan excepciones.

De las explicaciones que nos aporta sobre lo que es la argumentación sólo es posible inferir que para reconocer la argumentación desde sus rasgos lingüísticos se debe prestar atención a adverbios de lugar y tiempo, a locuciones conclusivas, como *a pesar de*, locuciones de limitación, como *sólo en el caso de que*, locuciones de suposición como *en el caso de que* y la conjunción condicional o de suposición *si*.

A continuación presentaremos muy someramente el modelo de Adam (1992). Concretamente, en este modelo la noción de prototipos de texto ha sido útil, para redefinir la noción de superestructura de van Dijk (1987) y para distinguir dos dimensiones: la pragmática, referida a la intencionalidad, al marco de referencia de los enunciados y a la cohesión semántica global -equivalente a la macroestructura de van Dijk (1987)- y la secuencial, referida a la gramática del texto, -equivalente a la microestructura de van Dijk (1987)- y a la organización de las diferentes secuencias del texto.

Según Adam (1992), un texto puede estar estructurado en diferentes secuencias de base, dado que la posibilidad de que se estructure de manera única es muy baja. En todo caso, siempre hay un tipo de secuencia que destaca y que define la estructura dominante del texto. Por ejemplo, en una argumentación tienen cabida secuencias introductorias descriptivas, narrativas o de otros tipos, pero globalmente la secuencia que predomina es la argumentativa, con sus propias características y éste es el mensaje que le llega al lector oyente. Un texto determinado, pues, se puede considerar como argumentativo si se aproxima a este prototípico. Esta noción es importante para no caer en la rigidez de la estructura

del texto -hecho que podría suceder si se utiliza el modelo touminiano de forma mecanicista- con el fin de no eliminar su creatividad y para analizarlo de una forma más flexible. En palabras de Adam (1992: 19):

"chaque texte est une réalité beaucoup trop hétérogène pour qu'il soit possible de l'enfermer dans les limites d'une définition stricte".

Adam (1992) toma el modelo de Toumin como base de la estructura argumentativa, pero analiza los textos como secuencias argumentativas encadenadas en las que se puede producir el caso de que la conclusión de una secuencia sea la premisa de la siguiente. En la Figura 1 se esquematiza la secuencia argumentativa prototípica.

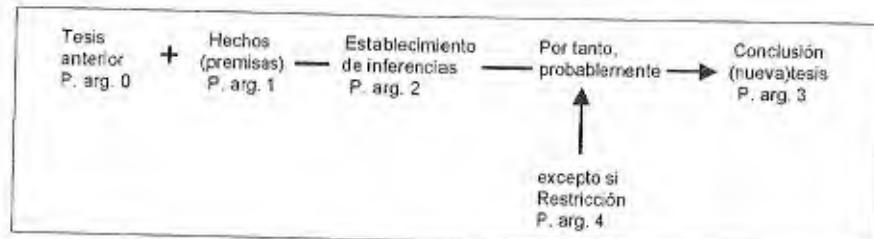


Figura 1. Secuencia argumentativa según Adam (1992)

Finalmente, este autor plantea que la función del texto argumentativo es la de intentar convencer a otro de una afirmación u opinión [tesis] que normalmente conlleva polémica y que admite opiniones o hipótesis diferentes. En su estructura se desarrollan argumentos y contraargumentos. En su construcción se recurre a estructuras oracionales extensas, con abundante coordinación [adversativas, copulativas] y subordinación [causales, consecutivas y finales]. Es un modo de estructuración textual propio del ensayo, de la publicidad, de columnas de opinión y editoriales, de debates, mesas redondas, etc.

1.2 Rasgos lingüísticos de la Argumentación: Otros autores

Dada que hasta este momento los rasgos aportados por los autores seleccionados en esta primera revisión [Perelman & Olbrechts-Tyteca, 1958/1994; Toulmin, 1958; van Dijk, 1987 y Adam, 1992] no ha sido tan fructífera como se esperaba, incorporaremos las propuestas de otros autores de modo que podamos llevar a cabo un acusoso repertorio de los rasgos lingüísticos que convergen y que puedan ser indicadores de la presencia de argumentación en un texto cualquiera.

Así, Bassols y Torrent (1998) proponen una estructura básica de la argumentación, la cual sería: *No + SER/presente}+SN* (ej. *la solución del problema no es sencilla*).

En cuanto a rasgos más específicos, siguiendo en gran parte a Perelman y Olbrechts-Tyteca (1958/1994), dicen que los verbos que dominan esta base textual suelen ser verbos relacionados con la causalidad y la consecuencia: *causar, hacer, originar, generar, activar, ocasionar, producir, motivar, suscitar, determinar*. O bien con las operaciones argumentativas fundamentales: *resultar, deducir, concluir, suponer, conjutar, inferir*.

Y los verbos "dicendi": *decir, suscribir, afirmar, declarar, estimar, considerar, describir, expresarse, arbitrar, implicar, alegar, admitir, asentir, aludir, asegurar*.

Los tiempos verbales varían según la naturaleza de los hechos o de las pruebas aducidas. Se utiliza el pasado para todo aquello hecho y terminado. El presente, para todo lo universal, normal, para las máximas las sentencias, el presente siempre es actual. El imperfecto sirve para todo lo transitorio: ej. *defendía la dictadura del proletariado*.

Plantean que cuando queremos reforzar la presencia de un elemento, usamos el artículo definido, que consigue persuasivamente una unificación del punto de vista. Afirman que cuando nos interesa transformar lo subjetivo en normal recurrimos a la construcción con el pronombre reflexivo *se*, ya que disminuye la responsabilidad del hablante y objetiva el enunciado. Además como se conserva un alto grado de ambigüedad, este pronombre posee un gran ámbito de aplicación argumentativa. Los enunciados demasiado precisos son peligrosos, se tienen que evitar a toda costa.

Las modalidades que convienen más a la base textual argumentativa son la assertiva y la interrogativa. Esta última es una forma hábil de desestimar razonamientos, ya que presenta alternativas, provoca divisiones, genera complicidad y arrastra muchos implicitos útiles para la argumentación.

Las autoras plantean que, en cuanto a cohesión y coherencia, los textos argumentativos utilizan una serie de recursos muy interesantes. Algunos son:

a) Las marcas de orden que introducen párrafos: *Ordinales: primero, segundo, tercero, etc.* Locuciones de orden: *En primer lugar, por un lado, por otra parte, finalmente, para acabar*.

b) Las comillas y las citas: Para citar palabras que refuerzan nuestra intervención, para lo cual normalmente recurrimos a autoridades y personajes. Para dar un tono especial a determinadas palabras y para recordar la máxima de un pensador.

c) La interrogación retórica: Para atraer el interés del receptor, para implicarlo directamente.

d) Los guiones y las rayas: Para poder marcar niveles de pertinencia entre las informaciones, los de los guiones quedan en segundo término, para pasar del tono objetivo al subjetivo.

e) Los nexos que expresan causa o consecuencia: conjunciones: *ya que, porque, pues, luego, así pues*. Locuciones causales: *con motivo de, a causa de, gracias a, con ocasión de, por culpa de, visto que, en vista de que, como que, puesto que, dado que, considerando que*. Locuciones consecutivas de modo: *que, de suerte que, de manera que, así que, en conclusión, por lo tanto*.

f) Adverbios y locuciones adverbiales: *consecuentemente, en consecuencia, por consiguiente*.

Desde el punto de vista pragmático, estiman que la argumentación es un conjunto de estrategias discursivas dirigidas a la demostración de una opinión. Argumentar, dicen, es dar razones para tal o cual conclusión, ligar argumentos y conclusiones.

Siguiendo la teoría de los actos de habla, observan que en las argumentaciones que se dan en las discusiones racionales, los actos que más abundan son:

a) Representativos: *afirmar, suponer, aseverar, enunciar con certeza, afirmar confidencialmente*.

b) Directivos: *ordenar, prohibir, solicitar, rogar, invitar, desafiar, reñir, preguntar*. Aquí estarían las preguntas retóricas.

c) Comisarios: *prometer, jurar, ofrecer*.

d) Declarativos: que sirven para una definición, una ampliación, una precisión o una explicación.

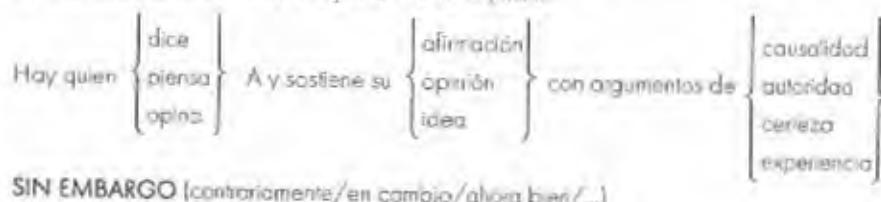
En cuanto a los operadores argumentativos, establecen que son particulares (casi, sólo, ya, de sobra) que imponen al receptor un procedimiento preciso sobre la manera como tiene que articular sentido a un enunciado. En cambio, los conectores argumentativos actúan como marcadores interactivos. Relacionan lo que tienen a continuación con lo que lleva delante. No sólo comunican algo: interaccionan. Pueden ser de los siguientes tipos:

- Introductores de argumentos: *porque, ya que, dado que, en primer lugar, claramente, efectivamente, incluso*.
- Introductores de conclusiones: *por lo tanto, pues, entonces*.

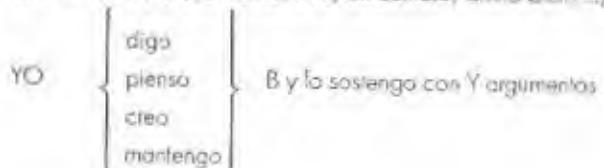
Por otra parte, Plantin (1998) establece que el enunciado de la conclusión en la argumentación puede tener forma constativa o inductiva, el primero se tratará de un enunciado en presente de indicativo, por ejemplo: *claramente, las cosas son así*. El segundo de un imperativo, *Hagamos pues esto!*

También, se puede utilizar el futuro de indicativo cuando se expresa un hecho aleatorio por naturaleza, pero cuya realización se presenta como muy probable. Por ejemplo: *no habrá framboosas este año*.

Para Calsamiglia y Tusón (1999), el texto argumentativo supondría la siguiente estructura, manifestada de forma total o parcialmente explícita:



SIN EMBARGO (contrariamente/en cambio/ahora bien/...)



Siguiendo a Cuenca (1995), se establece que los recursos lingüísticos y discursivos característicos de la argumentación son: la deixis personal, la antítesis, la modalización oracional y, muy especialmente, los conectores que articulan el discurso polémico. En cuanto a estos últimos plantean que se pueden organizar en:

- contrastivos (contradiccionarios): utilizados para poner de manifiesto relaciones de

oposición (*pero, en cambio, sin embargo, ahora bien*), sustitución (*sino, en lugar/vez de, por el contrario, antes bien, contrariamente*), restricción (*excepto si, a no ser que*) y contraste o concesión (*de todos modos, sea como sea, en cualquier caso, a pesar de, no obstante, con todo, aun así, después de todo, así y todo, con todo y con eso*). Con estos conectores el texto cambia de orientación en sentido contrario al segmento inmediatamente anterior, bien sea de forma total o parcial, la línea argumentativa sufre un culebre que indica que se abandona la primera orientación para tomar otra.

- Causales y consecutivos, utilizados para marcar las relaciones de causa (*a causa de ello, por eso, porque, pues, puesto que, ya que, dado que, por el hecho de que, en virtud de, gracias a, consecuencia de, así que, pues, luego, por eso, de modo que, de ello resulta que, así que, de donde se sigue, así pues, por lo tanto, de suerte que, por consiguiente, en consecuencia, en efecto, entonces*). Un tipo específico serían los conclusivos que introducen la conclusión (*en conclusión, en resumen, en supta, en resumidas cuentas, en total*).
- Distributivos: *por un lado, por otro; por una parte, por otra; éstos, aquellos*.
- Además de estos tres bloques, las autoras mencionan que existe también un uso habitual en argumentación de condicionales (*si, con tal de que, cuando, en el caso de que, según, a menos que, siempre que, mientras, a no ser que, siempre y cuando, sólo que, de tal modo que*), generalizadores (*sin ejemplos*), y exemplificadores (*por ejemplo, a saber, así, en concreto, pongamos por caso, sin ir más lejos*).

Por otra parte, las autoras, siguiendo esta vez a Cabaliero y Larreuri (1996), presentan los tipos de conectores argumentativos más habituales en el discurso filosófico:

- Causa: indican que los enunciados que lo siguen explican o dan razón de los enunciados antecedentes (*porque, pues, puesto que, dado que, ya que, por el hecho de que, en virtud de*).
- Certeza: indican que los enunciados que los siguen son enunciados ya probados por el autor (*tesis válidas*) o enunciados aceptados por una comunidad (*es evidente que, es indudable que, nadie puede ignorar que, es incuestionable que, de hecho, en realidad, está claro que*).
- Condición: siempre que en un texto aparece un condicional, éste va seguido de una consecuencia. Puede suceder que la palabra que introduce un enunciado como consecuencia de otro no vaya precedida de ningún conector. En estos se ha de hacer explícito, en el análisis del razonamiento, un entonces lo cualquier otro conector de consecuencia) aunque en el texto no esté (*si, con tal que, cuando, en el caso de que, según, a menos que, siempre que, mientras, a no ser que*).
- Consecuencia: indican que los enunciados que los siguen son efecto de los razonamientos antecedentes o de una condición fluego, entonces, *por eso, de manera que, de donde se sigue, así pues, así que, por lo tanto, de suerte que, por consiguiente, de ello resulta que, en efecto*).
- Oposición: Estos conectores señalan que los enunciados que vienen a continuación contienen alguna diferencia respecto de los que le preceden. La diferencia puede no ser más que un matiz o, por el contrario, puede ser algo totalmente opuesto a lo afirmado anteriormente (*pero, aunque, contrariamente, en cambio, no obstante, ahora bien, por el contrario, sin embargo, mientras que*).

Finalmente, las autoras, citan los elementos que se utilizan para introducir la propia opinión, como *entiendo, pienso, creo, desde mi punto de vista, a mi modo de ver, etc.* Asimismo, la polémica puede estar reforzada por elementos léxicos como *falacia, argumento, desacuerdo, demostración*.

Otro autor, que se ha convertido en clásico en los estudios de la argumentación, es lo Cascio (1998) quien al proponer una gramática de la argumentación sigue las ideas de Toulmin (1958), incorporando además las ideas de la gramática generativa para la conformación de las estructuras que explican la argumentación. Entre los rasgos lingüísticos propios de la argumentación podemos destacar lo que él denomina *indicadores de fuerza*, exponentes lingüísticos que sirven para conectar los enunciados y para indicar su papel dentro del discurso. Hacen las veces de conectivos textuales y pueden clasificarse según la función que realizan. Algunos marcan la tesis, otros los datos o los argumentos, otros la regla general y otros la reserva, la fuente, o categorías mayores como la argumentación misma. Así tenemos los indicadores de fuerza que:

- a) Introducen el macroargumento: *ahora me explico, el razonamiento es éste, ahora se demuestra por qué.*
 - b) Introducen un argumento o un dato: *[justificadores] puesto que, porque, de hecho, en efecto, dado que, ya que, ya que es cierto que, también porque, considerando que, partiendo del hecho que, y la prueba es que, y eso es porque, luego, uso del gerundio.*
 - c) Introducen una tesis o conclusión: *[conclusivos] por consiguiente, así pues, por tanto, he aquí que, por eso, se sigue que, por lo cual puede sostenerse que, por ello, si, entonces.*
 - d) Introducen la regla general: *[generalizadores] a partir de, dado que, y eso porque, dice que.*
 - e) Introducen la modalidad o el calificador: *[modales] quizás, probablemente, es probable que, necesariamente, poder + infinitivo, deber de + infinitivo, futuro (elemento morfológico con funciones modal y no de tiempo verbal).*
 - f) Introducen la fuente, la autoridad: *[garantes] como dice, según*
 - g) Introducen una reserva: *[relativizadores] o no ser que, salvo que, a menos que, excepto que, si/sí no, aunque.*
 - h) Introducen un refuerzo para la justificación presentada: *(refuerzos) sin contar con, si se tiene en cuenta el hecho de que, observemos que, no obstante, a pesar de que, si bien, aunque.*
 - i) Introducen una contracisión: *[alternantes] sin embargo, no obstante que, a pesar de que.*
- Entre estudios más contemporáneos del lexico argumentativo encontramos a Álvarez (2001) quien a propósito de los conectores pragmáticos y los organizadores textuales, plantea la conveniencia de recordar que, desde el punto de vista lingüístico, el estudio de la argumentación se centra principalmente en las relaciones de:
- a) Causa/consecuencia: *entonces, luego, así pues, así que, por (lo) tanto, por consiguiente, en consecuencia, por ello/eso, con ello, porque, puesto que, ya que, por el hecho de que, en virtud de, de suerte que, de ahí que.*
 - b) Oposición/contraste (contra-argumentativos): *pero, en cambio, sin embargo, no obstante, mientras tanto, ahora bien, contrariamente, en contra de, contra lo que pueda parecer, por el contrario, de todos modos, con todo, (ni) aun así*

- c) Introductor de premisas: *si...[entonces], dado que, considerando, teniendo en cuenta, ante.*
- d) Introductor de macroargumento: *ahora me explico, el razonamiento es éste, ahora se demuestra por qué.*
- e) Indicador de argumento: *porque, puesto que, pero, no obstante, con todo, sin embargo, de hecho, en efecto, es cierto que, y la prueba es que, y eso es porque, luego.*
- f) Marcador de conclusión: *en consecuencia, por consiguiente, por (lo) tanto, por eso (esto, lo cual), pues bien, (si)...entonces, de ahí que, así pues, se sigue que, por lo cual puede sostenerse que.*

Por último, Álvarez (2001) plantea que para reforzar la argumentación, es frecuente recurrir a enunciados con oraciones concesivas en los que el hablante menciona dos elementos de información que contrastan fuertemente entre ellos. A ello contribuye el uso de conectores concesivos del tipo: *aunque, a pesar de que, y eso que, si bien (es cierto que), aun + gerundio, con + infinitivo, por más que, así y todo, y más que...*; también se utilizan locuciones con matiz concesivo, tales como: *sin duda tiene vd. razón, aunque sea indispensable, debemos, etc.*

Para finalizar, expondremos la clasificación propuesta por Montolio (2001) para los conectores contraargumentativos, ésta se organiza en cuatro subgrupos:

- a) Por un lado, expresiones conectivas como *aunque, a pesar de (que), pese a (que) y si bien, de las que "aunque" es el conectador prototípico y que coinciden en estar integradas entonativa y sintácticamente en la oración.*
- b) Por otro, conectores como *pero, sin embargo, no obstante, ahora bien, con todo, aun así o a pesar de todo, grupo del que "pero" es el representante paradigmático.* Nótese que a excepción de la conjunción "pero", todos los demás elementos incluidos en este subgrupo presentan carácter parentético.
- c) Un tercer grupo está formado por conectores integrados en la oración tales como *mientras que, en tanto que, y si no que, y también, por los correspondientes parentéticos en cambio, por el contrario y antes bien.* Todos ellos comparten un significado básico de "corregir" en el segundo miembro algún aspecto de lo formulado en el primero.
- d) El cuarto y último subgrupo está compuesto por los conectores de *todas formas, de todas maneras y de todos modos.* Se caracterizan por el hecho de que, a pesar de no ser propiamente contraargumentativos, aparecen en numerosas ocasiones con un valor similar. Todos ellos coinciden en minimizar o cancelar la relevancia del segmento informativo anterior y anularlo para la prosecución del discurso.

Hasta aquí dejaremos nuestra revisión bibliográfica, ya que parece haber información suficiente para establecer, a partir de los rasgos lingüísticos aportados por los autores estudiados, una organización de los rasgos lingüísticos más relevantes de la argumentación. Para organizar esta información se establecerán categorías lingüísticas que permitan agrupar los ejemplos dados por los autores. Estas categorías son: A. Estructura argumentativa, B. Tipo de oraciones, C. Tipo de verbos, D. Tiempo, E. Modalidad, F. Modalizadores, G. Conectores, H. Marcas y locuciones de orden, I. Adverbios y locuciones adverbiales, J. Pronombres, K. Artículo determinado, L. Demostrativo, M. Léxico particular, N. Figuras retóricas, N. Expresiones y O. Marcas gráficas. En síntesis, podemos determinar 16 categorías

y 51 rasgos lingüísticos específicos que son característicos de la argumentación. A continuación, en la Tabla 1, se presentan las 16 categorías con sus respectivos rasgos lingüísticos y la exemplificación correspondiente.

Podemos determinar que la implicancia de haber podido determinar un conjunto de rasgos lingüísticos propios de la argumentación abre la posibilidad de reconocer con un alto grado de certeza, en un texto cualquiera, la presencia de argumentación y el grado en que ésta se presenta.

Tabla 1. Rasgos lingüísticos de la Argumentación

A. Estructura argumentativa (ver Colomiglio y Tuñón, 1999)	1. Hay quien $\left\{ \begin{array}{l} \text{dice} \\ \text{piensa} \\ \text{opina} \end{array} \right\}$ A y sostiene su $\left\{ \begin{array}{l} \text{afirmación} \\ \text{opinión} \\ \text{idea} \end{array} \right\}$ con X argumentos de $\left\{ \begin{array}{l} \text{causalidad} \\ \text{autoridad} \\ \text{certeza} \\ \text{experiencia} \end{array} \right\}$	SIN EMBARGO [contrariamente/en cambio/chora bien/]
	YO $\left\{ \begin{array}{l} \text{digo} \\ \text{pienso} \\ \text{creo} \\ \text{mantengo} \end{array} \right\}$ B y lo sostengo con Y argumentos	
	2. Adversativas	"No gastó el dinero en reparaciones de su casa sino que lo derrochó todo en una fiesta"
	3. Copulativas	"La economía de nuestros países está en quiebra y las devaluaciones de la moneda nos amenazan continuamente"
	4. Causales	"No le presenten atención a ese hombre, pues sólo está haciendo proselitismo a favor de su partido"
	5. Consecutivas	"Lo propuesto de la empresa le provocó desconfianza, por lo tanto , a última hora no firmó el contrato"
	6. Finales	"Los miembros del jurado buscarán nuevos testigos del accidente con el fin de que no negue lo sucedido"
	7. De opinión	Decir, pensar, creer, mantener, entender, dudar + [que]
	8. De causa/consecuencia	Causar, hacer, originar, generar, activar, ocasionar, producir, motivar, suscitar, determinar
	9. De operaciones argumentativas	Resultar, deducir, concluir, suponer, conjutar, inferir
B. Tipo de oraciones (ver Adam, 1992) * los ejemplos fueron extraídos de Munguía, Munguía y Roche (2000)	10. "Dicendi"	Decir, suscribir, afirmar, declarar, estimar, considerar, describir, expresarse, arbitrar, implicar, alegar, admitir, asentir, claudir, asegurar
	11. Representativos	Afirmar, suponer, aseverar, enunciar
	12. Directivos	Ordenar, prohibir, solicitar, rogar, invitar, desafiar, retar, preguntar
	13. Comisarios	Prometer, jurar, ofrecer
	14. Declarativos	Definir, explicar, precisar, entenderse como

D. Tiempo (ver Bassols y Torreal, 1998 y Planell, 1998)	15. Presente	
	16. Pretérito indefinido, pretérito imperfecto	
	17. Futuro	
	18. Asertiva	
	19. Interrogativa	
	20. Inyucción	
	21. Atenuadores (ver Toumin, 1958 y Lo Casco, 1998)	poder, parecer que, deber en condicional y en subjuntivo, es probable que, tal vez, a lo mejor, quizás, quizá
	22. Enfatizadores	deber, probar, confirmar, tener que, sin duda, por supuesto, siempre, nunca, evidente, de hecho, es cierto que, jamás, muy, en verdad, en realidad
	23. Verbos Modales	Poder+infinitivo, deber+infinitivo, querer+infinitivo, necesitar+infinitivo
	24. Adverbios en mente	Probablemente, posiblemente, indudablemente, evidentemente, decididamente, seguramente, ciertamente, definitivamente, obviamente.
	25. Negación	No, nunca, jamás, tampoco, ni, nadie, nada, ninguno, ninguno
	26. Introducen el macroargumento	ahora me explico, el razonamiento es éste, ahora se demuestra por qué
	27. Introducen un argumento o un dato: (justificadores)	puesto que, porque, de hecho, en efecto, dado que, ya que es cierto que, también porque, considerando que, partiendo del hecho que, y la prueba es que, y eso es porque, luego, uso del gerundio, dado que, en primer lugar, ciertamente, efectivamente, incluso
	28. Introducen una tesis o conclusión: (conclusivos)	por consiguiente, así pues, por lo tanto, he aquí que, por eso, se sigue que, por lo cual puede sostenerse que, por ello, si, entonces, pues, entonces, en conclusión, en resumen, en suma, en resumidas cuentas, en total, es evidente que, es indudable que, nadie puede ignorar que, es incuestionable que, de hecho, en realidad, está claro que, en consecuencia, por consiguiente, por esto, de ahí que, así pues
	29. Introducen la regla general o premisa: (generalizadores)	a partir de, dado que, y eso porque, dice que, considerando, teniendo en cuenta, ante
	30. Introducen la modalidad o el calificador: (modales)	quizás, probablemente, es probable que, necesariamente, poder+infinitivo, deber de +infinitivo, futuro (elemento morfológico con funciones modal y no de tiempo verbal)
	31. Introducen la fuente, la autoridad: (garantes)	como dice, según
	32. Introducen una reserva: (relativizadores)	a no ser que, salvo que, a menos que, excepto que, si/si no, aunque
	33. Introducen un refuerzo para la justificación presentada: (refuerzos)	sin contar con, al se tiene en cuenta el hecho de que, observemos que, no obstante, a pesar de que, si bien, aunque

Introducen una contraposición (argumentativos o alternantes) (ver Coloméglia y Tusón, 1999)	34. Oposición	Però, en cambio, sin embargo, ahora bien, o en su lugar de que, sino que
	35. Sustitución	Sino, en lugar/vez de, por el contrario, antes bien,
	36. Restricción	Excepto si, a no ser que, en tanto que, sólo en el caso de
Para estos colectores ver Coloméglia y Tusón (1999)	37. Contraste o concesión	Aunque, contrariamente, en contra de, contra lo que pueda parecer, por el contrario, de todos modos, sea como sea, en cualquier caso, a pesar de [que][todo], no obstante, con todo, aun así, después de todo, así y todo, con todo, con eso, y eso que, si bien [es cierto que], aun[queriendo], con[tinuando], por más que, pese a [que], ahora bien
	38. Causales	a causa de ello, por eso, porque, pues, puesto que, ya que, dado que, por el hecho de que, en virtud de, gracias a, por ello, con ello
	39. Consecutivos	de ahí que, pues, luego, por eso, de modo que, de ello resulta que, así que, de donde se sigue, así pues, por [lo] tanto de acuerdo que, por consiguiente, en consecuencia, en efecto, entonces, por eso, de manera que
	40. Distributivos	por un lado, por otro, por una parte, por otra, éstos, aquéllos,
	41. Condicionales	si, con tal de que, cuando, en el caso de que, según, a menos que, siempre que, mientras, a no ser que, siempre y cuando, sólo que, de tal modo que
	42. Ejemplificadores	por ejemplo, a saber, así, en concreto, pongamos por caso, sin ir más lejos
I. 43. Marcas y locuciones de orden (ver Bassols y Torrent, 1998)		Primero, segundo, tercero, en primer lugar, por un lado, por otra parte, finalmente, para acabar
I. 44. Adverbios y locuciones adverbiales (ver Bassols y Torrent, 1998)		Consecuentemente, en consecuencia, por consiguiente
I. 45. Pronombres (ver Perelman y Obréch-TYTECA, 1994 y Bassols y Torrent, 1998)		Se [impersonal], tercera persona singular y plural, se (reflexivo), cierto/a/s, uno/a/s
K. 46. Artículo determinado (ver Bassols y Torrent, 1998)		El, la, los, las
I. 47. Demostrativo (ver Perelman y Obréch-TYTECA, 1994) *los ejemplos son de Munguía, Munguía y Rocha (2000)		Este, ésta, esto, ése, ésa, eso, aquél, aquéllo, aquello, éstos, éstas, ésos, ésas, aquéllos, aquéllos
M. 48. Léxico particular (ver Perelman y Obréch-TYTECA, 1994 y Coloméglia y Tusón, 1999)		Indicio de, distinción, de familiaridad o de sencillez, Falacia, argumento, desacuerdo, demostración
N. 49. Figuras retóricas (ver Perelman y Obréch-TYTECA, 1994 y Coloméglia y Tusón, 1999)		Paráfrasis, antinomio
N. 50. Expresiones (ver Perelman y Obréch-TYTECA, 1994 y Coloméglia y Tusón, 1999)		Por el amor de, en consideración a, a causa de, desde mi punto de vista, a mi modo de ver
O. 51. Marcas gráficas (ver Bassols y Torrent, 1998)		Comillas, guiones y rayas

2. LA INVESTIGACIÓN

Como se desprende del apartado precedente, el foco de preocupación de este trabajo se centra en la identificación de los rasgos argumentativos presentes en textos. Para llevar a cabo la investigación debemos recordar que, como plantea Adam (1992), la mayoría de los textos poseen un conformación textual heterogénea y que, por lo tanto, presentarán secuencias diversas, es decir, la estructura de un texto puede estar construida por varias secuencias textuales diferentes. Sin embargo, siempre es posible distinguir un tipo de secuencia que destaque y que defina la estructura dominante del texto. Como también planteábamos anteriormente, en una argumentación tienen cabida secuencias introductorias, descriptivas, narrativas o de otros tipos, pero globalmente la secuencia que predomina es la argumentativa. Un texto determinado, pues, se puede considerar como argumentativo si se aproxima a este prototípo.

Ahora bien, la identificación de la secuencia argumentativa en un texto ejemplar no parece presentar mayores dificultades, sin embargo, ¿qué sucede si a partir de esta idea de dominancia de secuencias, se propone realizar un análisis que identifique argumentación en una gran cantidad de textos? La tarea no es tan fácil, puesto que en primer lugar se debe contar con herramientas que permitan interrogar una gran cantidad de textos, aspecto que está cubierto por la lingüística de corpus computacional. Y en segundo lugar, se debe contar con la posibilidad de disponer de un software identificador de secuencias textuales, lo cual parece ser muy difícil dado el carácter más bien difusa de la organización interna de las secuencias. Determinar dónde se inicia y dónde termina una secuencia cualquiera no parece ser todavía posible de formalizar en un programa de computadora.

Una posible solución puede ser inferida a partir de los mismos estudios de Adam (1992). Ésta estaría fundada en los rasgos lingüísticos que configuran una secuencia determinada. Así, podemos pensar que una secuencia argumentativa dominante o subordinada en un texto dado aportará al texto una serie co-ocurrente de rasgos lingüísticos propios de la argumentación. Y por ello sería posible identificar la argumentatividad del texto, ya no por la identificación de las secuencias dominantes, sino que por la acumulación e interrelación de los rasgos lingüísticos característicos. Esta propuesta supone entonces que la presencia y correlación de rasgos lingüísticos argumentativos sería un muy buen indicador de la argumentatividad presente en el texto.

La identificación de estos rasgos lingüísticos de la argumentación en los textos y la comparación de ellos con muchos otros es una tarea algo más fácil de realizar, dado que hoy en día existen no sólo las herramientas computacionales que permiten interrogar grandes corpus de textos a partir de sus rasgos morfológicos y sintácticos, por ejemplo BwanaNet para el español, sino que también el marco conceptual que permite describir y esbozar explicaciones del comportamiento distintivo de los textos a partir de la lingüística de corpus asistida por computación (Biber, 1988).

La característica principal de la lingüística de corpus computacional es el uso que hace de los medios informáticos para analizar y describir grandes cantidades de datos lingüísticos reales, esto es, datos del lenguaje natural (Biber, Conrad & Reppen, 1998).

En el análisis lingüístico son muchos los rasgos que se tienen que considerar, ya que es necesario comprender las formas en que estos actúan juntos, co-ocurren en los textos. Un

enfoque que responde a este problema es el análisis multidimensional, propuesto por Bibber (1988), según el cual, los patrones de co-ocurrencia basados en las distribuciones reales de los rasgos lingüísticos en un corpus de textos son identificados cuantitativamente.

Otro aspecto a considerar es que muy rara vez un registro (en términos de una variedad de lenguaje) puede ser identificado por un rasgo distintivo que se encuentra sólo en ese registro. Los registros comparten muchos rasgos lingüísticos y ellos más bien se distinguen entre sí por el uso relativo de esos rasgos. Así, las diferencias sistemáticas en el relativo uso de un conjunto básico de rasgos provee una base para distinguir entre estos registros.

Desde esta propuesta metodológica es posible desplegar la idea de que a partir de los rasgos de la argumentatividad, obtenidos a través del análisis bibliográfico realizado previamente, podemos caracterizar los textos de un corpus en términos de su argumentatividad.

En consecuencia, ¿es posible identificar presencia de argumentatividad en distintos tipos de textos a partir de la frecuencia de aparición de los rasgos característicos de la argumentación en ellos? Esta es, en definitiva, la interrogante que guiará nuestra investigación.

2.1. Metodología

Para llevar a cabo la investigación se utilizará entonces una metodología cuantitativa, a partir de la cual se determinará la frecuencia que presentan los rasgos de la argumentatividad en los corpora textuales del Proyecto FONDECYT 1020786 de la Pontificia Universidad Católica de Valparaíso.

Se llevará a cabo, en definitiva, un análisis comparativo de los corpora a partir de la frecuencia normalizada de los rasgos (ver Bibber, 1988: 75-78) como un intento de mostrar en primera instancia la ocurrencia y las relaciones de estos ocurrencias de los rasgos argumentativos presentes en los textos del corpus.

2.1.1. Corpora

Como ya hemos mencionado, utilizamos aquí el Corpora PUCV-2003 del Proyecto FONDECYT 1020786, confeccionado por el equipo de lingüística de la Pontificia Universidad Católica de Valparaíso y cuyo objetivo principal es caracterizar los textos utilizados por los estudiantes de educación técnico-científica de Valparaíso, Chile. Para cumplir con este objetivo se recolectó una muestra total de los textos utilizados por los alumnos de tres liceos técnico-científicos de la ciudad de Valparaíso, clasificándolos en 3 áreas: Marítima, Comercial e Industrial y en 12 clases textuales: *Manual, Guía Didáctica, Formulario, Reglamento, Glosa Legal, Instructivo, Leyes, Artículo Técnico, Tabla, Descripción Técnica, Diagrama y Glosario*. En total recolectaron y digitalizaron 74 textos con un total de 626.790 palabras. Este corpus es denominado Corpus Técnico-Científico (CTC).

Además, para poder llevar a cabo la descripción por comparación incluyeron un Corpus de Literatura Latinoamericana (CLL) conformado por una muestra de 12 textos literarios, entre cuentos y novelas leídos por los alumnos en la asignatura de Lenguaje y Comunicación, con un total de 459.860 palabras. Y una muestra de 4 transcripciones de entrevistas orales realizadas a los alumnos de los liceos técnicos de la región, con un total de 380.094 que constituyó el Corpus de Entrevistas Orales (CEO).

En resumen, nuestro investigación se lleva a cabo en 90 textos con un total 1.466.654 palabras. Una mayor descripción de estos corpora pueden ser encontradas en Parodi & Giomajo (2003), Parodi (2004), Parodi & Venegas (2004), Marinovich & Cademartori (2004).

El primer paso para llevar a cabo las comparaciones es intentar homogeneizar los corpora, ya que si observamos el porcentaje de distribución del total de palabras de CTC, CLL y CEO, podemos especificar que el CTC comprende el 43% (626.790) del total, el CLL el 31% (459.860) del total y el CEO el 26% (380.094) del total de las palabras.

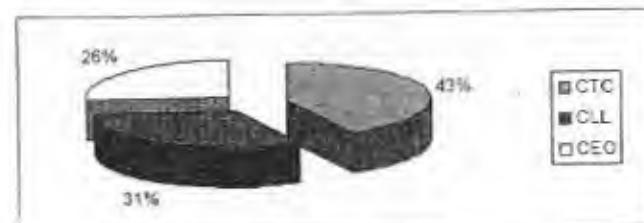


Gráfico 1. Distribución porcentual de palabras en los corpora del proyecto FONDECYT 1020786 PUCV

Es precisamente por esta diferencia en las cantidades totales de palabras que al comparar se debe tener el cuidado de normalizar las frecuencias de rasgos que aparecen en la muestra, de forma que los datos sean representativos y comparables entre sí.

2.1.2. Categorías de análisis

Como ya adelantamos la selección de los rasgos característicos de la argumentación realizados en la primera etapa alcanzaba a 51 rasgos en total. Sin embargo, debido a las posibilidades técnicas de búsqueda otorgadas por la interfaz del programa BwanaNet nos vimos obligados a reducir este número a 40 rasgos. A pesar de lo anterior, pensamos que es un número suficiente para dar cuenta de la argumentatividad presente en un texto cualquiera.

A modo de información los rasgos que han quedado fuera de este estudio son:

1. La estructura gramatical, 2. Las oraciones coordinadas copulativas, 3. Las subordinadas concesivas y causales se sintetizan en las de razón o causa, 4. La modalidad interrogativa, 5. Los conectores que introducen macroargumento, 6. Los artículos indeterminados, 7. Adverbios y locuciones adverbiales, 8. El léxico particular, 9. Las figuras retóricas, 10. Expresiones particulares y 11. Las marcas gráficas.

En definitiva los rasgos considerados en esta investigación y sus correspondientes criterios lingüísticos de búsqueda en los 3 corpora son los siguientes:

A. Tipo de oraciones

- Rasgo de Coordinación:
 - 1. Adversativas (Cody): Introducidas por: sino, mas, pero
- Rasgo de Subordinación:

2. Subordinadas Causales (Scous): Introducidas por: porque, ya que, debido a, por lo tanto, por ende, de este modo, así, puesto que
 3. Subordinadas Finales (Sfin): Introducidas por: con el fin de que y para +infinitivo

B. Tipo de Verbos:

4. De opinión (Vopin): decir, pensar, creer, mantener, entender, dudar + [que]
 5. De razón o causa (Vcau/cori): causar, hacer, originar, generar, activar, ocasionar, producir, motivar, suscitar, determinar
 6. De operaciones argumentativas (VopArg): resultar, deducir, concluir, suponer, conjutar, inferir
 7. "Dicend." (Vdic): decir, suscribir, afirmar, declarar, estimar, considerar, describir; expresarse, arbitrar, implicar, alegar, admitir, asentir, aludir, asegurar
 8. Representativos (Vrepr): afirmar, suponer, aseverar, enunciar
 9. Directivas (Vdir): ordenar, prohibir, solicitar, rogar, invitar, desafiar, relar, preguntar
 10. Comisarios (Vcomi): prometer, jurar, ofrecer
 11. Declarativos (Vdecl): definir, explicar, precisar

C. Tiempo

12. Presente (Pres)
 13. Pretérito indefinido, pretérito imperfecto (Pret ind/impl)
 14. Futuro y Futuro perifrásitico (Fut)

D. Modalidad

15. Indicativo (ModInd)
 16. Imperativo (ModImp)

E. Modalizadores

17. Atenuadores (Aten): poder, parecer que, crear, deber en condicional y en subjuntivo, probablemente, es probable que, posiblemente, seguramente, tal vez, a lo mejor, quizás, quizá
 18. Enfatizadores (Erf): deber, probar, confirmar, tener que, sin duda, por supuesto, claramente, indudablemente, siempre, nunca, evidente, evidentemente, de hecho, es cierto que, decididamente, definitivamente, jamás, muy, obviamente, en verdad, en realidad
 19. Verbos Modaños (VerMod): posibilidad (poder), necesidad, (es necesario que), obligación (es obligación que, tener que, deber que), volición (querer)
 20. Adverbios en mente (Adv-m): todos
 21. Negación (Neg): adverbio de negación (no), adverbios de tiempo (nunca, jamás, tampoco), conjunción "ni", pronombres (nadie, ninguno, ninguna), nada

F. Conectores

22. Introducen un argumento o un dato: (justificadores) (Introarg): puesto que, porque, en efecto, dada que, ya que, considerando que, luego, en primer lugar, claramente, efectivamente, incluso
 23. Introducen una tesis o conclusión: (conclusivos) (Introtes): por consiguiente, así pues, por lo tanto, por eso, por ello, pues, entonces, en conclusión, en resumen, en suma, en total, es evidente que, es indudable que, es incuestionable que, de hecho, en realidad, está claro que, en consecuencia, por esto

24. Introducen la regla general o premisa: (generalizadores) (Introprem), o partir de, dado que, y eso porque, dice que, considerando, teniendo en cuenta, ante
 25. Introducen la modalidad o el calificador: (modaños) (Intromod): quizás, probablemente, es probable que, necesariamente, poder+infinitivo, deber de +infinitiva
 26. Introducen la fuente, la autoridad: (garantes) (Intofuén): como dice, como plantea, como argumenta, como explica, según
 27. Introducen una reserva: (relativizadores) (Intores): o no ser que, salvo que, a menos que, excepto que
 28. Introducen un refuerzo para la justificación presentada: (refuerzos) (Introref): sin contar con, si se tiene en cuenta el hecho de que, observemos que, no obstante, a pesar de que, si bien, aunque

G. Conectores de Contraopinión: (contraargumentativos o alternantes)

29. Oposición (Opos): pero, en cambio, sin embargo, ahora bien, a pesar de que, sino que
 30. Sustitución (Sust): sino, en lugar/vez de, por el contrario, antes bien
 31. Restricción (Rest): excepto si, o no ser que, en tanto que, sólo en el caso de
 32. Contraste o concesión (Contral): aunque, en contra de, de todos modos, a pesar de, no obstante, si bien, ahora bien

H. Otros Conectores

33. Causales (Caus): a causa de ello, por eso, porque, pues, puesto que, dado que, por el hecho de que, en virtud de, gracias a, por ello, con ello
 34. Consecutivos (Cons): de ahí que, pues, luego, por eso, de modo que, de ello resulta que, así que, de donde se sigue, así pues, por (lo) tanto, de suerte que, por consiguiente, en consecuencia, en efecto, entonces, por eso, de manera que
 35. Distributivos (Dist): por un lado, por otro; por una parte, por otra
 36. Condicionales (Codic): si, con tal de que, cuando, según, a menos que, siempre que, mientras, o no ser que, siempre y cuando, sólo que, de tal modo que
 37. Ejemplificadores (Ejempl): por ejemplo, a saber, así, en concreto, pongamos por caso, sin ir más lejos

I. Categoría Mixta

38. Marcas y locuciones de orden (MldeOrd): primero, segundo, tercero, en primer lugar, por un lado, por otra parte, finalmente, para acabar
 39. Pronombres (Pron): tercera persona singular y plural
 40. Demostrativa (Den): éste, ésta, esto, ése, esa, aquél, aquella, aquello, éstos, ésos, ésas, aquéllos, aquéllas

2.1.3. Análisis e interpretación de los resultados¹

En este apartado se llevará a cabo el análisis del comportamiento estadístico de los rasgos, por medio de frecuencias porcentuales, en los tres corpus en estudio.

Para una mejor visualización de los datos se presentarán agrupados según las categorías a la cual se asignaron en el apartado anterior.

¹ Datos obtenidos a través de BwachNet durante el mes de julio de 2003

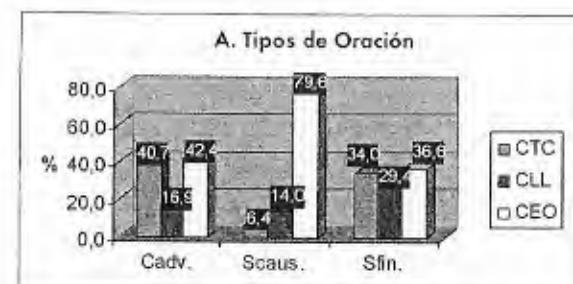


Gráfico 2. Tipos de Oración

Los tres primeros rasgos corresponden a los incluidos en la categoría Tipos de Oración y son: 1. Oración coordinada adversativa, 2. Oración subordinada causal y 3. Oración subordinada final.

Podemos observar en el Gráfico 2 que en el rasgo coordinada adversativa tiene un porcentaje de ocurrencia muy similar en el CTC (40,7%) y en el CEO (42,4), en tanto que baja considerablemente para el CLL (16,9%).

En cuanto al segundo rasgo podemos establecer una clara dominancia porcentual de CEO (79,6%) por sobre CTC y CLL, que en conjunto tienen un porcentaje del 20,4%. Esto indicaría que la subordinada causal permite diferenciar claramente los corpus y que el CEO tendría una presencia de mayor causalidad que en los otros corpus.

En términos de la subordinada final podemos afirmar que tampoco existen grandes diferencias en términos porcentuales, ya que en los tres corpus hay una presencia de alrededor de un 30%, dominando muy levemente el CEO, seguido de CTC y, finalmente, el CLL.

Los datos de esta categoría permiten establecer en primer lugar, que el rasgo que mejor diferencia los tres corpus es la subordinada causal. En segundo lugar es posible determinar que según los rasgos de esta categoría el CEO es el corpus que presenta mayor argumentatividad. En cuanto al CTC es posible determinar que presenta una frecuencia relativamente alta en los rasgos 1 y 3, en tanto que en el rasgo de subordinada causal la frecuencia es mínima. El CLL no presenta una frecuencia considerable de los rasgos 1 y 2, en tanto que en el rasgo 3 la frecuencia es relativamente alta, aunque menor que los otros dos corpus.

Los rasgos incluidos en esta categoría son: 4. Verbos de opinión (Vop), 5. Verbos de causa/consecuencia (Vcau/con), 6. Verbos de operaciones argumentativas (VopArg), 7. Verbos "Dicend", 8. Verbos representativos (Vrepr.), 9. Verbos directivos, 10. Verbos comisivos (Vcomi) y 11. Verbos declarativos (Vdec).

Podemos establecer que para el rasgo 4. Verbos de opinión hay una clara predominancia porcentual del CEO, el cual obtiene más del 50% de la frecuencia total. El CLL se ubica en segundo lugar con una frecuencia correspondiente al 37,5 y el CTC con muy escasa frecuencia porcentual de verbos de opinión.

En relación al rasgo 5. Verbos de causa/consecuencia, el CEO también domina porcentualmente con un 52,4% del total. En términos de este rasgo la frecuencia porcentual

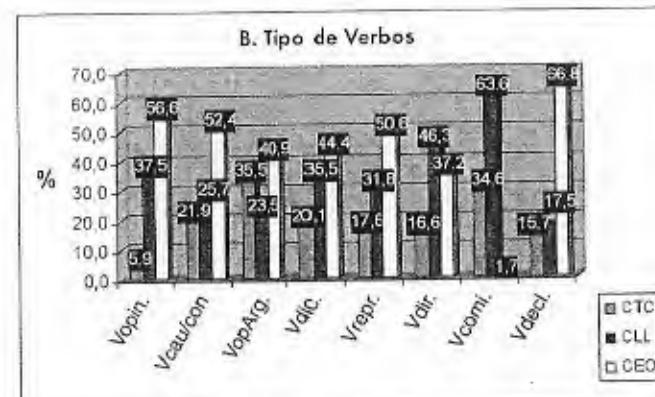


Gráfico 3. Tipo de Verbos

se reparte muy similarmente entre CLL y CTC, sin embargo, aún se presenta cierta dominancia de CLL sobre CTC.

En cuanto a los Verbos de operaciones argumentativas (rasgo 6), podemos establecer que se mantiene la dominancia de CEO por sobre CLL y CTC, sin embargo, se presenta una baja porcentual importante (cerca al 10%) con respecto a los rasgos anteriores. En este caso el CEO aumenta la frecuencia del rasgo a 35,5% siendo superior a CLL en 12%.

Para el rasgo 7. Verbos "Dicend" la mayor frecuencia aún la mantiene CEO con un 44,4%. Se presenta en este rasgo un aumento de la frecuencia porcentual en CLL, siendo esta vez superior a CTC en un 15% aproximadamente.

El rasgo 8. Verbos representacionales vuelve a presentarse con mayor porcentaje en CEO que en los otros corpus, manteniéndose también la diferencia entre el CLL (31,8%) y CTC (17,6%).

Para el rasgo 9. Verbos directivos la relación varía radicalmente, puesto que CLL presenta una frecuencia porcentual mayor que los otros dos corpus, seguida esta vez por CEO (37,2%) y luego por CTC (16,6%).

En relación al rasgo 10. Verbos comisivos, destaca fuertemente la frecuencia porcentual de CLL con un 63,6%, seguido por CTC, el cual presenta en este rasgo el mayor porcentaje de la categoría con respecto a los rasgos anteriores (34,6%). Finalmente el CEO en este rasgo presenta una frecuencia insignificante (1,7%).

El último rasgo de esta categoría es 11. Verbos declarativos. En este rasgo CEO destaca nuevamente, esta vez con una frecuencia porcentual muy alta (56,8%), en tanto que CLL y CTC están muy parejos con 15,7% y 17,5%, respectivamente.

Para concluir con la interpretación de esta categoría podemos decir que CEO es el corpus que presenta mayor argumentatividad, desde el punto de vista de los verbos identificados, siendo el rasgo más característico los verbos declarativos [definir, explicar, expresar], presenta además cierta homogeneidad en términos porcentuales en todos los rasgos [entre 40,9% y 56,6%], excepto en la frecuencia de los verbos comisivos, rasgo en el cual casi no tiene frecuencia.

El CLL presenta un rasgo con una frecuencia porcentual deslucible, los verbos comisarios (prometer, jurar, ofrecer), en tanto que en los otros rasgos no pasa del 46,3%. El rasgo con frecuencia menor son los verbos declarativos 17,5%.

Para el CTC podemos decir que no presenta en general una frecuencia media, presentando los rasgos Verbos de operaciones Argumentativas y Verbos comisarios con más alto porcentaje, alrededor de un 35%. El rasgo que se presenta con menor frecuencia es Verbos de opinión.

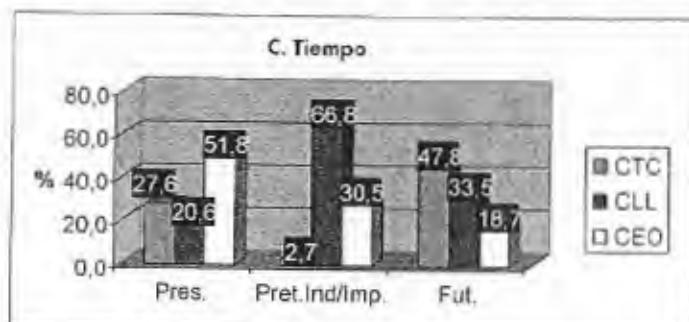


Gráfico 4. Tiempo

En cuanto a la categoría tiempo, los rasgos son: 12. presente, 13. pretérito indefinido e imperfecto y 14. futuro. Para el presente observamos que CEO predomina con cerca del 52%, en tanto que CTC presenta un porcentaje de un 27,6% y CLL un porcentaje de 20,6%. Los pretéritos indefinido e imperfecto tienen una muy alta ocurrencia en CLL (66,8%), seguido con un 30,5% en CEO y con una muy frecuencia bajísima en CTC. Para el 14. futuro la frecuencia porcentual más alta está en CTC con un 47,8%, en tanto que CLL presenta un 33,5% y CEO un 18,7%.

Esta categoría presenta una muy buena diferenciación en cuanto a la relación rasgo/corpus, sin embargo, no diferencia en términos argumentativos. De esta manera se puede plantear que el CEO se caracteriza por estar escrito en tiempo presente, el CLL un pretérito y el CTC en futuro, aunque con menor rango de diferenciación con respecto a los otros corpora.

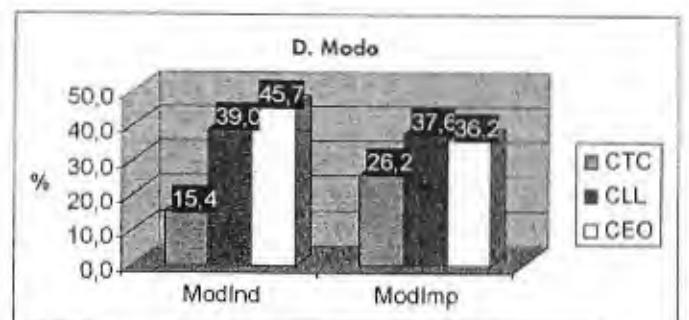


Gráfico 5. Modo

En la categoría modo encontramos los rasgos 15. modo indicativo y 16. modo imperativo. Para el primer rasgo distinguimos que el CEO presenta mayor ocurrencia, 45,7%, que el CLL (39%) y que el CTC (15,4%).

En el rasgo imperativo nos percatamos que no existe gran diferencia entre CLL y CEO, ambos con porcentajes cercanos al 37%, en tanto CTC manifiesta una ocurrencia algo mayor que en el rasgo anterior, pero aún muy baja (26,2%). Por lo tanto, podemos plantear que el modo imperativo no diferencia mayormente entre los corpus CLL y CEO y que el CTC presenta un porcentaje muy bajo. Sin embargo, en el indicativo se presenta una clara diferencia porcentual a favor de CEO, más pareciera ser que el uso de este modo por sí mismo no indicaría en ningún caso mayor o menor argumentatividad.

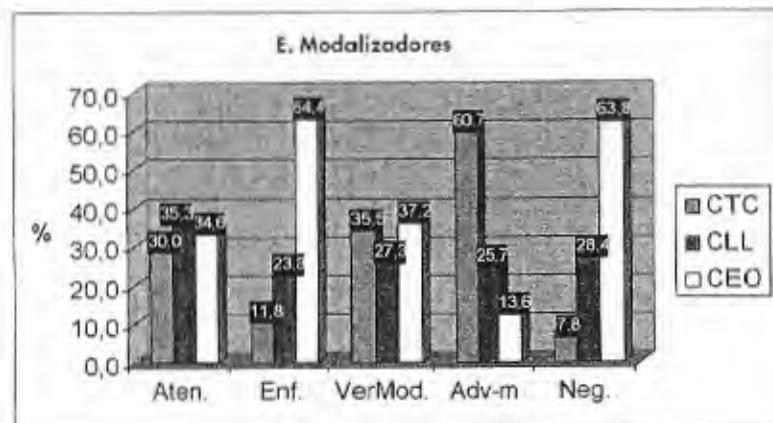


Gráfico 6. Modalizadores

En esta categoría encontramos los rasgos 17. atenuadores, 18. enfatizadores, 19. verbos modales, 20. adverbios de modo terminados en mente y 21. negación.

Para el rasgo 17 podemos plantear no existe gran diferenciación entre los corpus puesto que todos tienen una frecuencia más bien repartida de entre 30 y 35%, dominando levemente el CLL. El rasgo 18 por otra parte, manifiesta una alta frecuencia en CEO (64,4%), seguido por CLL y CTC con porcentajes más bien bajos, 23,8 y 11,8 respectivamente. Los verbos modales tampoco diferencian claramente los corpus, siendo el CEO que tiene el porcentaje mayor (37,2%) seguido por CLL con 35,5% y por CTC con un 27,3%. En el caso de los adverbios de modo terminados en mente, observamos la clara dominancia porcentual de CTC sobre los otros corpus, presentando un porcentaje de 60,7%. CLL y CEO presentan porcentaje bajos para este rasgo, 25,7% y 13,6 respectivamente. Finalmente, para el rasgo negación se invierte nuevamente la ocurrencia y CEO vuelve a presentar, como en los enfatizadores, un porcentaje muy alto (63,8%), en tanto que CLL presenta un porcentaje de un 28,4 y CTC un porcentaje de apenas un 7,8 por ciento. Esta categoría permite establecer que el CTC presenta en general una mayor argumentatividad que los otros dos corpus, sobre todo en términos de la relación existente entre enfatizadores y lo

negación. Por otra parte, los datos muestran que el CLL mantiene una ocurrencia muy similar para todos los rasgos de esta categoría, entre 23% y 28% aproximadamente, aunque presenta un porcentaje algo mayor en los atenuadores (35,3%).

En cuanto a CTC podemos establecer que se destaca por la utilización de los adverbios de modo terminados en mente (60,7%) y comparte con un porcentaje de alrededor del 36% el uso de los verbos modales con CEO. En general presenta una argumentatividad baja.

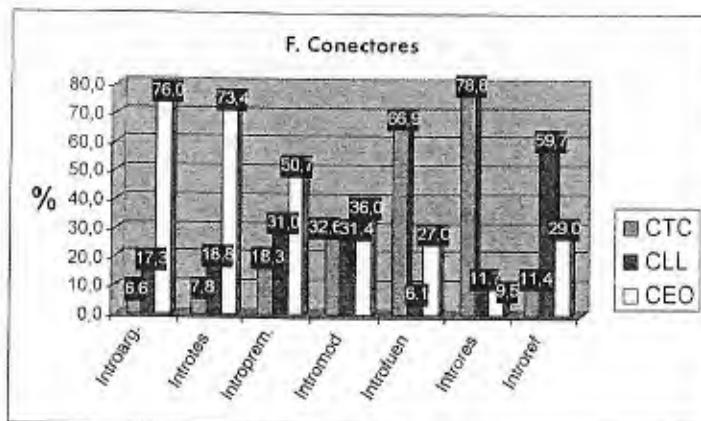


Gráfico 7. Conectores

En la categoría conectores encontramos los siguientes rasgos: 22. conectores que introducen argumento, 23. conectores que introducen tesis, 24. conectores que introducen premisa, 25. conectores que introducen modalización, 26. conectores que introducen fuente, 27. conectores que introducen reserva y 28. conectores que introducen refutación.

Así, para 22 podemos plantear que CEO presenta una muy alta ocurrencia del rasgo (76%), en tanto que para CLL presenta una ocurrencia baja (17,3%) y para CTC una muy baja (6,6%). En cuanto al segundo rasgo de esta categoría la situación es muy similar a la del rasgo anterior. Hay una muy alta ocurrencia para CEO, una baja para CLL y una muy baja para CTC. Para el rasgo 24, podemos decir que si bien sigue dominando porcentualmente CEO, lo hace esta vez con un porcentaje bastante menor (50,7%). CLL presenta una ocurrencia bastante mayor que los anteriores rasgos (31%), en tanto que CTC permanece igual. El rasgo 25 presenta una ocurrencia muy similar en los tres corpora. Así los porcentajes varían entre 36% para CEO, 32,6% para CLL y 31,4% para CTC. En cuanto al rasgo conector introductor de fuente hayamos que CTC presenta la mayor ocurrencia (66,9%), seguido por CEO con un bajo porcentaje (27%) y con un muy bajo porcentaje CLL (6,1%). El rasgo conector introductor de reserva se presenta con una muy alta frecuencia en CTC (78,8%), con una ocurrencia muy baja se presenta en CLL (11,7%) y una mucho menor en CEO (9,5%). Para el último rasgo de esta categoría podemos observar que es el único en el cual predomina el CLL con un 59,7%, seguido por CEO con un porcentaje de 29% y por CTC con un muy bajo porcentaje de 11,4%.

En síntesis, la categoría observada, que en sí es muy importante en la argumentación, presenta una frecuencia mayor para CEO, destacándose los rasgos 22 y 23 por su gran porcentaje de ocurrencia. CTC destaca en esta categoría por el gran porcentaje que tiene en los rasgos 26 y 27, aunque en el resto mantiene porcentajes muy bajos. CLL sólo un rasgo en el cual se destaca (rasgo 28). Por lo tanto, podemos decir que CEO presenta la mayor argumentatividad, según esta categoría.

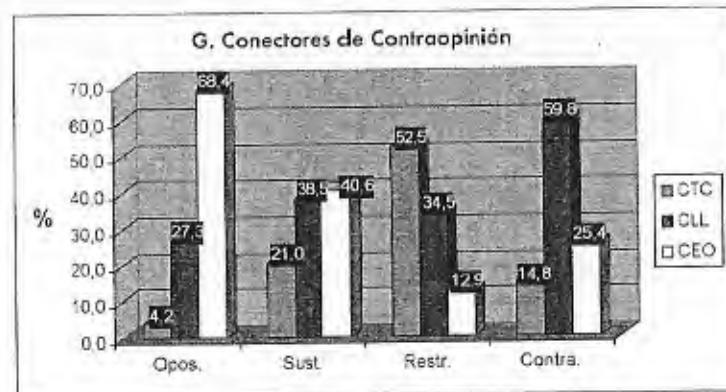


Gráfico 8. Conectores de Contraopinión

En la categoría Conectores de Contraopinión se incluyen los siguientes rasgos: 29. de oposición, 30. de sustitución, 31. de restricción y 32. de contraste.

En el primer rasgo de esta categoría observamos una clara predominancia de CEO con un 68,4%, al cual le sigue el CLL con el bajo porcentaje de 27,3% y continúa con un porcentaje mínimo el CTC (4,2%). El rasgo de sustitución presenta porcentajes muy similares para CEO (38,5%) y CLL (40,6%), en tanto con un porcentaje más bajo aparece CTC (21%). El siguiente rasgo (31) presenta un frecuencia mayor en CTC (52,5%), algo menor en CLL (34,5%) y muy baja en CEO (12,9%). Por último, el rasgo 32, presenta una alta frecuencia en CLL (59,8%), una frecuencia baja en CEO (25,4%) y una muy baja en CTC (14,8%).

Los resultados de esta categoría también son relevantes para determinar la argumentación, puesto que indican un momento clave en la argumentación. Sin embargo, no parece haber una dominancia a través de los rasgos, ya que cada corpus se presenta con mayor porcentaje que los otros en al menos un rasgo. Aún así el CEO presenta algo más de argumentación al tener un porcentaje muy alto en 29 y uno relativamente alto en 30. CLL se comporta de manera más o menos sistemática, manifestando frecuencias medianamente altas y una alta en 32.

Por último, CTC sólo presenta un rasgo con frecuencia medianamente alta, en 31, pero en el resto de los rasgos sus porcentajes son muy bajos.

En la categoría Otros conectores, podemos distinguir los siguientes rasgos: 33. causales, 34. concesivos, 35. distributivos, 36. condicionales y 37. exemplificadores. Para el primer

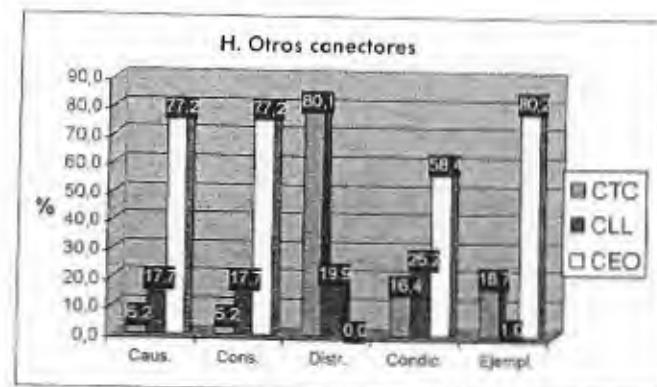


Gráfico 9. Otros conectores

rasgo de esta categoría podemos decir que el CEO predomina ampliamente con 77,2%, en tanto que CLL tiene un bajo porcentaje de frecuencia, seguido por CTC con porcentaje mínimo de 17,7%. En cuanto al rasgo 34, concesivos, CEO predomina de la misma manera que en el anterior y CLL y CTC presentan los mismo bajos porcentajes que en el rasgo anterior. Los conectores distributivos (35) ocurren con mucha mayor frecuencia en CTC (80,1%), en tanto que con un porcentaje de frecuencia de 19,9% aparecen en CLL y no aparecen en CEO. Los condicionales presentan una frecuencia mucha mayor en CEO (58,4%), seguida de CLL con 25,2% y por CTC con apenas 16,4%. Por último, los ejemplificadores aparecen mayoritariamente en CEO (80,2%), con un porcentaje bajo aparecen en CTC (18,7%) y casi no aparecen en CLL.

En resumen, se puede estimar que la mayor argumentatividad según estos conectores está dada por CEO. En tanto que en CLL y CTC casi no aparecen, con excepción del rasgo 35 en CTC que es muy recurrente.

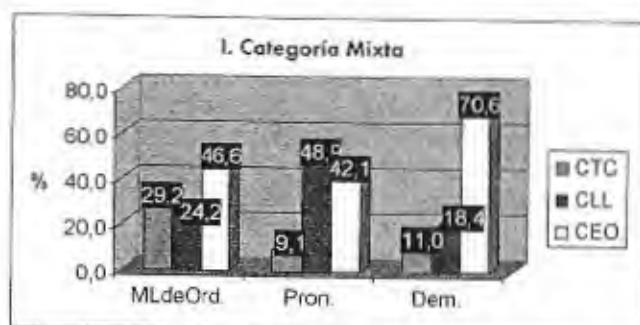


Gráfico 10. Categoría Mixta

Esta última categoría incluye elementos mencionados en la teoría pero que no sido posible de relacionar claramente con otros, de esta manera se configura una categoría que contiene los siguientes rasgos: 38. Marcas y locuciones de orden, 39. pronombres y 40. demostrativos. Observamos que en el primer rasgo de esta categoría, la mayor recurrencia se presenta en CEO (46,6%), seguido por CTC (29,2%) y luego por CLL (24,2%). En cuanto al rasgo 39, es posible afirmar que la mayor recurrencia está en CLL (48,9%), seguido muy de cerca por CEO (42,1%), en tanto que CTC tiene un porcentaje bajísimo.

Por último, para el rasgo que incluye a los demostrativos podemos observar que predomina fuertemente CEO (70,6%), en tanto que CLL y CTC presentan bajos porcentajes de recurrencia, 18,4% y 11% respectivamente. Podemos afirmar que hay predominancia argumentativa de CEO en esta categoría con porcentajes que van de 42% a 70,6%. El CLL se mantiene parejo, aun cuando tiene mayor presencia de pronombres de lecera plural y singular. El CTC presenta un grado muy bajo de ocurrencia en estos rasgos.

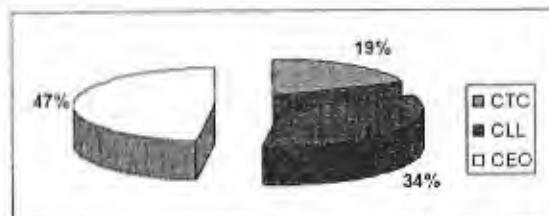


Gráfico 11. Porcentaje Total de los Rasgos por Corpus

Podemos observar que en promedio el mayor porcentaje de ocurrencia de los rasgos estudiados aparecen en CEO (47%), seguido por CLL (34%) y luego por CTC (19%). Estos datos nos permiten generalizar que el corpus CEO presenta una argumentatividad mayor que en los otros corpus. En tanto que la argumentatividad presente en CLL tiene una regular frecuencia y la de CTC es muy baja.

Ahora bien, con los datos obtenidos podemos demostrar que la argumentatividad se da mayormente en CEO, sin embargo, es interesante tener un visión global del comportamiento de los rasgos para determinar cuáles son en definitiva los rasgos que están configurando la argumentatividad. Para ello realizaremos una síntesis considerando sólo aquellas categorías y rasgos que son relevantes y que permiten diferenciar un corpus del otro. Veamos el siguiente gráfico:

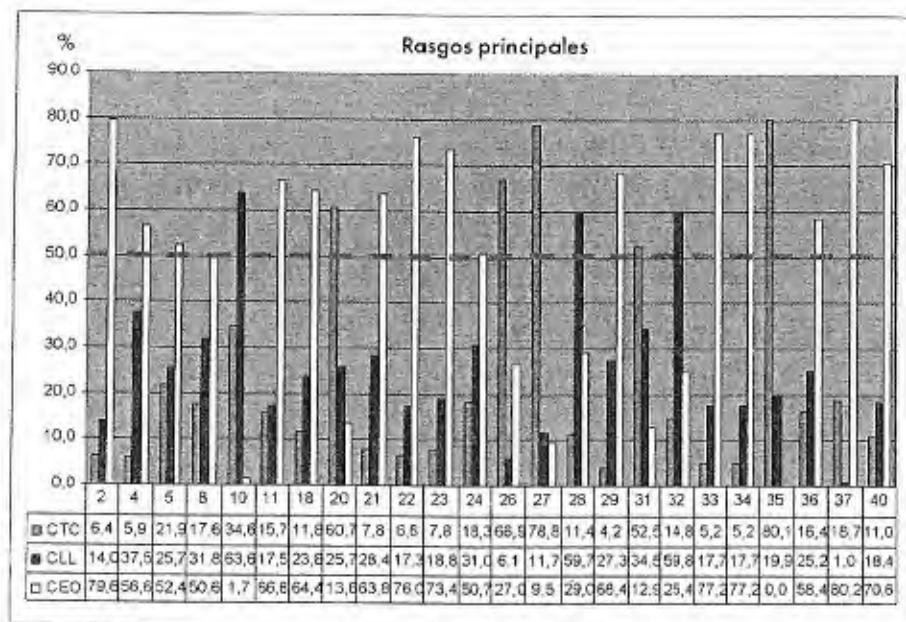


Gráfico 12. Rasgos principales por corpus

Si consideramos los puntajes sobre 50% para cada corpus, podremos reconocer cuáles son los rasgos que están intercruzando y, por lo tanto, que están conformando haces de rasgos que determinan con una mayor predominancia el grado de argumentatividad.

Así, los 16 rasgos que explican la argumentatividad en CEO se presentan en la Tabla 2.

Tabla 2. Rasgos de argumentatividad en CEO

Rasgo	Porcentaje	Ejemplo
1. Conectores de exemplificación	80,2%	"es como si yo dijera, por ejemplo, por qué los pájaros vuelan" (Chile0087.txt).
2. Oraciones subordinadas causales	79,6%	"...tenía una memoria bien especial, casi fotográfica, por decir, ya que, muchas cosas no se me olvidan con facilidad..." (Chile0089.txt).
3. Conectores causales	77,2%	"pongo mucha atención en clases, así que eso me ayuda y por eso me va bien ya" (Chile0090.txt).
4. Conectores consecutivos	77,2%	"sí, sí, sí sé, pero a ver, cuéntame pues también sé leer" (Chile0088.txt).
5. Conectores introductores de argumentos	76%	"se me quedó al tiro, la memoricé también porque eran más que nada conceptos" (Chile0089.txt).

6. Conectores introductores de tesis	73,4%	"cuando yo me pego la fuerza que yo con la cual yo me pego, o, entonces cuando tengo el casco puesto yo no sufro tanto daño" (Chile0087.txt).
7. Demostativos	70,6%	"le cuestan pronunciar las que llevan acento ya, éas no más" (Chile0090.txt).
8. Conectores de oposición	68,4%	"cuando yo sé que alguna cosa no la he comprendido, pero, sin embargo, me la puedo haber memorizado, yo sé que me va a ir mal en la prueba" (Chile0090.txt).
9. Verbos declarativos	66,8%	"explicaba cosas simples de una forma demasiado complicada" (Chile0087.txt).
10. Enfatizadores	66,4%	"durante la lectura de repente no entiendo palabras y tengo que estar preguntando qué es esto" (Chile0087.txt).
11. Negaciones	63,8%	"es que casi nunca leo los títulos al principio, nunca pescas" (Chile0087.txt).
12. Conectores condicionales	58,4%	"de repente los leo casi todos los días, según el diario" (Chile0089.txt).
13. Verbos de opinión	56,6%	"...me confunden, puede que alguien esté equivocado y yo pienso que yo estoy equivocado y contesta mal en la prueba" (Chile0090.txt).
14. Verbos de causa/consecuencia	52,4%	"los errores no me causan problema, simplemente, me los solto no más, para mí no quieren decir nada en el texto" (Chile0087.txt).
15. Conectores introductores de premisa	50,7%	"porque yo igual voy sacando la idea principal y a partir de la idea principal, yo me explico el resto" (Chile0088.txt).
16. Verbos representativos	50,6%	"- esto está incoherente ¿por qué? - porque se supone que la gente no sale a trabajar" (Chile0087.txt).

Para CTC sólo 5 rasgos obtienen un porcentaje mayor al 50%:

Tabla 3. Rasgos de argumentatividad en CTC

Rasgo	Porcentaje	Ejemplo
1. Conectores distributivos	80,1%	"En países desarrollados se han establecido estatutos legales para quienes operan el transporte multimodal, generando, por una parte , la normativa legal que necesitan para operar como tales y por otra , proporcionando a los usuarios un ente responsable legal y comercialmente" (Chile0042.txt)
2. Conectores introductores de causa	78,7%	"La exportación de mercancías no estará afecta al pago de derechos, a menos que ley las grave expresamente" (Chile0080.txt)
3. Conectores introductores de fuente	66,9%	"Nuestro código de comercio hace esta exigencia, según lo establece el artículo, que dice: 'Al abrir su giro todo comerciante hará en el libro de balances una enunciación estimativa de todos sus bienes, tanto muebles como inmuebles, y de todos sus créditos activos y pasivos'" (Chile0001.txt)
4. Adverbios terminados en mente	60,7%	"Hay que tener aquí presente que después del recocido deben enfriarse las piezas muy lentamente . Se cubren por esta razón frecuentemente con cono caliente, con lo cual se evita la aparición de nuevas tensiones" (Chile0027.txt)
5. Conectores de restricción	52,5%	"Nunca se use una lata o manera de palanca o alzaprima, pues su cinta es blanda y se dobla fácilmente, en tanto que el cuerpo es duro y muy quebradizo, de modo que basta una leve flexión para que se parte en dos" (Chile0031.txt)

Para CLL tenemos solamente 3 rasgos y con porcentajes en general más bajos:

Tabla 4. Rasgos de argumentatividad en CLL

Rasgo	Porcentaje	Ejemplo
1. Verbos comisivos	63,6%	"Lo más correcto sería que yo le ofreciese un noviazgo serio, muy serio , quizás demasiado serio" (Chile0082.txt)
2. Conectores introductores de refuerzo	59,7%	"No se sentía urgencia alguna y hasta las más primordiales necesidades, como comer, por ejemplo, o dormir, parecían olvidarse, anegarse por lo menos, sin contar con que en la plaza o en el ruelle se podía dormir, sentado, claro está" (Chile0085.txt)
3. Conectores de contraste	59,8%	"En contra de lo que temía, Isaías me recibió muy bien" (Chile0086.txt)

CONCLUSIONES

En este trabajo hemos realizado una extensa revisión de los rasgos que diversos autores han mencionado como rasgos caracterizadores de la argumentación. Esta revisión nos ha permitido generar una clasificación de 51 rasgos lingüísticos organizados en 16 categorías, a saber: A. Estructura argumentativa, B. Tipo de oraciones, C. Tipo de verbos, D. Tiempo, E. Modalidad, F. Modalizadores, G. Conectores, H. Marcas y locuciones de orden, I. Adverbios y locuciones adverbiales, J. Pronombres, K. Artículo determinado, L. Demostrativo, M. Léxico particular, N. Figuras retóricas, Ñ. Expresiones y O. Marcas gráficas.

Como conclusión de este primer momento de la investigación es posible establecer que los autores señalados como clásicos de la teoría de la argumentación se preocuparon de generar modelos de explicación de los procesos argumentativos, uno dando énfasis al público al cual se dirigen los argumentos y otro centrado en el estudio del argumento mismo. En este sentido son acercamientos holísticos al tema de la argumentación que no revelan aspectos tan específicos, como la identificación de rasgos lingüísticos particulares que nos puedan revelar la presencia o ausencia de argumentación en los textos.

Por otra parte, autores más bien centrados en la identificación de la composición estructural de los textos de carácter argumentativo han centrado su atención en los niveles retórico-estructurales que permiten configurar y reconocer un texto argumentativo. Sin embargo, es poca la atención que prestan a los rasgos lingüísticos propios de textos organizados argumentativamente, a pesar de que Adam (1992) esboza algunos rasgos que permitirían reconocer las secuencias argumentativas en la heterogeneidad textual.

En este sentido hemos descubierto que a partir de mediados de los años noventa la preocupación de los autores por la identificación de rasgos propios de los textos argumentativos ha ido en crecimiento continuo. Y es gracias a esto que hemos podido concretar nuestra primera fase de investigación, comparando y seleccionando los rasgos que parecen ser identificatorios de la argumentación en los textos.

En cuanto al apartado relacionado con la investigación, es posible determinar que si bien las herramientas utilizadas han favorecido mucha la investigación, estas tienen ciertas limitaciones, que se manifiestan explícitamente en la imposibilidad de identificar ciertos rasgos y que por ello no se han podido considerar en la investigación.

Por otra parte, los datos analizados nos arrojan ciertas luces respecto de la presencia de una alta frecuencia de un grupo de rasgos lingüísticos, que permiten establecer que el *Corpus de Entrevista Oral* presenta una alta probabilidad de encontrar en ellos trazas argumentativas importantes. Por otra parte, para el *Corpus Técnico-Científico*, que solo presenta cinco rasgos con porcentaje mayor a 50%, no es posible identificar un nivel de argumentatividad importante. Por último, el *Corpus Literario* sólo presenta tres rasgos posibles de considerar argumentativas por se, por lo mismo se hace muy difícil justificar la presencia de algún nivel de argumentatividad reconocible a partir de los rasgos lingüísticos identificados.

Se hace necesario precisar aquí que entre los rasgos que aparecen con mayor frecuencia resultan ser conectores de diversos tipos, por ejemplo: conectores de exemplificación, conectores causales, conectores consecutivos, conectores introductores de argumentos, conectores introductores de tesis, conectores de oposición, conectores condicionales, conectores introductores de premisa, conectores distributivos, conectores introductores de reserva, conectores de restricción, conectores introductores de fuente, conectores introductores de refuerzo y conectores de contraste. Esto es absolutamente compatible con la apreciación de los distintos autores en cuanto a que el rasgo lingüístico más común de la argumentación son los conectores (ver *La Cascia*, 1998; *Calsamiglia* y *Tusón*, 1999). Por otra parte advertimos una amplia presencia de distintos tipos de verbos, por ejemplo: verbos comisivos, verbos declarativos, verbos de opinión, verbos de causa/consecuencia y verbos representativos, lo que está acorde a lo expuesto por *Bassols* y *Torrent* (1998) y *Calsamiglia* y *Tusón* (1999). Por último, se advierte la presencia de elementos modalizadores como los adverbios terminados en -mente, los enfatizadores y las negaciones (ver *Toulmin*, 1958; *La Cascia*, 1998).

En definitivo sintetizaremos nuestros hallazgos diciendo que la mayor argumentación en este estudio comparativo se encuentra en los textos contenidos en el *Corpus Oral*, y que tanto los textos contenidos en el *corpus CTC* y como los incluidos en el *CII* presentan un reducido número de haces de rasgos de argumentación, lo cual hace pensar que en estos textos la argumentación es muy baja.

Finalmente, es posible observar que el uso de este tipo de metodología sirve para identificar la presencia de ciertas características propias de los textos. Sin embargo, aún se debe avanzar en los análisis, por una parte hacia una estadística más fina que permita reconocer agrupaciones significativas de rasgos, y por otra parte, hacia una identificación más concreta de la configuración textual de estos datos de manera de poder identificar las secuencias particulares que conforman el texto. Así, se podría llevar a cabo no sólo identificación de secuencias argumentativas, sino de secuencias de cualquier índole, presentes en el texto y a partir de esto categorizar los textos de una forma automatizada, sirviendo esto último como una herramienta que ayude y apoye más cuantitativamente las conclusiones de los investigadores en el análisis textual.

Capítulo 6

Especificidad, especialización y variabilidad verbal: Una aproximación computacional en estadística léxica

OMAR SABAJ M.

INTRODUCCIÓN

A pesar de que existe común acuerdo en que los verbos cumplen, como categoría lingüística (i.e. gramatical, morfológica, semántica y pragmática-discursiva), un rol central en las lenguas humanas (Wiemer-Hastings, Graesser & Wiemer-Hastings, 1998) su estudio ha quedado relegado casi exclusivamente a estudios de corte gramatical. Por otro lado, además, sabemos que los verbos, en muchas otras áreas del estudio lingüístico (semántica, terminología, etc.) han suscitado un menor interés que los sustantivos (Llorente, 2002; Markman, 1999).

A partir del uso de herramientas computacionales, en este capítulo se describe el comportamiento de los verbos, entendidos como unidades lexicográficas (lemas) con una ocurrencia establecida en un corpus determinado. El *corpus* analizado en esta investigación (*Corpus PUCV-2003*) está constituido por tres registros principales: a) un *corpus* de textos técnico-científicos (CTC), correspondiente a 74 textos recolectados en tres liceos técnico-profesionales de la ciudad de Valparaíso; b) un *corpus* de entrevistas orales (CEO), obtenido a partir de 150 entrevistas semiestructuradas a alumnos de cuarto medio; y c) un *corpus* de literatura latinoamericana (CII), compuesto por 12 textos, correspondientes a las obras literarias que principalmente se dan a leer en la enseñanza media en distintos liceos técnicos de la ciudad Valparaíso. El total de palabras del *Corpus PUCV-2003* supera el millón.

En la primera parte del trabajo, describimos de forma general el vínculo entre lingüística y computadores, especificando algunos principios metodológicos de la lingüística de corpus, disciplina en la cual se enmarca este estudio. Reseñaremos luego, algunas investigaciones que se han hecho en esta misma línea, determinando las diferencias o similitudes con el presente estudio. Luego, en el acápite de los antecedentes teóricos, definiremos conceptualmente las categorías que se estudiaron, a saber, los verbos especializados, la especificidad y variabilidad verbales. En una segunda parte, describimos los pasos metodológicos que se siguieron en este trabajo.

Los resultados obtenidos muestran las siguientes tendencias generales. En cuanto a la especialización de los verbos, existen muy pocos verbos que se pueden considerar especializados. Asimismo, se observa claramente una mayor comunalidad entre los verbos del CTC y del CII, mientras que los verbos del CTC contrastados con los del CEO son los que presentan el índice de comunalidad más bajo. Por último, se puede apreciar que la mayor variabilidad verbal se da en el CII y la menor variabilidad en el CEO, sin embargo, en este último se puede observar que hay un mayor porcentaje de verbos por cantidad total de palabras. De estos resultados se puede concluir que en las entrevistas orales se ocupan más verbos [que en los otros registros], pero que esos verbos tienden a ser los mismos. Finalmente, en la discusión y en las conclusiones, mostramos la interpretación y el alcance de estos resultados, respectivamente.

1. ANTECEDENTES TEÓRICOS

1.1 Lingüística y computadores

Aunque para muchos la lingüística de corpus es una subdisciplina de la lingüística general y es, por lo tanto, independiente del uso de medios informáticos, creemos que una concepción moderna del término "lingüística de corpus" implica necesariamente el uso de dichos medios, razón por la cual se establece un estrecho vínculo entre esta disciplina y una disciplina más general denominada "lingüística computacional" (Stubbs, 1996; Moreno, 1998). Así como ha sucedido en la mayoría de las ciencias, también en la lingüística el desarrollo de nuevas tecnologías ha permitido enormes avances para la adquisición de conocimientos nuevos en distintos campos científicos. La codificación del genoma humano en medicina o el descubrimiento de nuevos planetas en astronomía, son solo los ejemplos más típicos de tareas científicas en cuya realización ha intervenido, de forma certeza, el uso de computadores. La utilización de mecanismos informáticos en la lingüística ha derivado en diversas disciplinas que reciben distintas denominaciones tanto según sus objetivos como de la función que cumple el uso de computadores en el desarrollo de ella. A continuación, se presentan las principales denominaciones que han recibido estas disciplinas con una breve descripción de algunos de sus principales objetivos:

- Lingüística Computacional:** disciplina general que estudia los modelos computacionales de la estructura y función del lenguaje, su uso y su adquisición (Moreno, 1998; Joshi, 1999).
- Psicolingüística computacional:** estudio de los procesos de comprensión y generación del lenguaje a través de modelos y herramientas computacionales. Dentro de esta cam-

po se desarrollan también modelos formales de diversos tipos de gramáticas (Norris, 1999).

- Procesamiento del lenguaje natural:** subcampo de la Inteligencia Artificial cuyo objetivo es el desarrollo y uso de herramientas computacionales para el procesamiento del lenguaje natural (Jurafsky & Martin, 2000).
- Reconocimiento del habla:** subcampo del procesamiento del lenguaje natural que se centra en el estudio de distintas aplicaciones tecnológicas asociadas a la lengua oral (Jurafsky & Martin, 2000).
- Lingüística de corpus:** es el estudio del lenguaje a través de medios informáticos para analizar, describir y explicar fenómenos de distintos niveles lingüísticos en grandes cantidades de datos auténticos, esto es, partes del lenguaje natural estudiados a partir de modelos matemáticos probabilísticos (Biber, Conrad & Reppen, 1998; Stubbs, 1996, 2000; Svartvik, 1992; Kennedy, 1998; Caravedo, 1999).

Revisaremos a continuación, en mayor profundidad, algunos aspectos esenciales para comprender esta última línea de investigación.

1.2 Lingüística de corpus

Aunque el concepto de "corpus", entendido como un conjunto de datos lingüísticos de la lengua natural, existe ya hace aproximadamente unos 50 años en el ámbito de la sociolingüística (Gumperz, 1981; Labov, 1966) y de la etnografía del habla (Hymes, 1972, 1974); es sólo a partir del fin de los años sesenta, con la conformación del Brown Corpus (Francis & Kucera, 1964, 1979, 1982) y del Lancaster-Oslo/Batgen Corpus of British English (Johansson, Leech & Goodluck, 1978) que se funda la lingüística de corpus tal como la conocemos en la actualidad, ya que fue entonces la primera vez que se recolectaron grandes cantidades de datos lingüísticos que serían almacenados y procesados a través de medios informáticos.

A pesar del gran desarrollo que ha tenido la lingüística de corpus desde entonces, no se puede hacer una reseña de esta disciplina sin dar cuenta de las grandes dificultades que ha sufrido desde su aparición. Debido al auge y predominancia de la gramática generativa, desde los años 50 hasta los 80 y al correspondiente establecimiento de un nuevo objeto de estudio (la competencia lingüística), se dejó prácticamente de lado cualquier estudio del uso del lenguaje (McEnery & Wilson, 1996; Leech, 1991). De esta forma, después de una larga tradición en fonética, en la que los datos físicos ocupaban un lugar privilegiado, el lingüista de sillón, como lo caricaturizó Fillmore (1992), pasó a ser el protagonista de la escena, usando datos ideales (Stubbs, 1996, 2000), es decir, oraciones inventadas y aisladas, que eran analizadas según la intuición del lingüista como hablante nativo de una lengua dada. Según Chafe (1992), este desinterés por el uso lingüístico se debe al carácter modular de la teoría chomskiana, esto es, a la suposición de que el sistema lingüístico opera en forma independiente del sistema cognitivo:

"One consequence of the modular view is that its adherents are not particularly interested in observing the everyday use of language, since they believe that whatever is most interesting about language exists independently of its use" (Chafe, 1992: 81).

Más allá de apoyar o no la visión modularista de la lingüística, lo evidente es que sus principios iban y siguen yendo en contra de algunos de los postulados básicos de los estudios de corpus. Comentamos, a continuación, dos de estos postulados.

- El tipo de datos.** A diferencia de los estudios generativos, en lingüística de corpus los datos no son ideales (Stubbs, 1996, 2000), sino que corresponden a fragmentos lingüísticos auténticos, sean estos fragmentos de habla o textos escritos. En este sentido, se pasa del estudio de la competencia al estudio del uso lingüístico manifestado en distintos tipos de formato. Desde un punto de vista epistémico, los datos prevalecen sobre la teoría (*data driven*).
- El tipo de enfoque.** En lingüística de corpus se estudia el lenguaje como un fenómeno estadístico particularmente estocástico (Moreno, 1998). La lengua, en este sentido, es un conjunto de información condicionada por la teoría de la probabilidad. Al contrario de esta tendencia, en gramática generativa, se propician modelos simbólicos formales del lenguaje, utilizando datos ideales para demostrar cómo operan ciertas reglas o principios. El lenguaje es un mecanismo competencial de carácter biológico inherente al ser humano (Moreno, 1998).

1.3 Estudios acerca de los verbos en lingüística de corpus: similitudes y diferencias con esta investigación

En esta sección, abordaremos de forma sintética las orientaciones generales que han seguido la mayoría de los estudios verbales en lingüística de corpus, estableciendo de este modo las diferencias entre la presente investigación y aquellos trabajos.

Un primer punto interesante de destacar es que existen pocos estudios, en lingüística de corpus (sobre todo en español), que se concentren solamente en los verbos. Por su parte, los estudios que sí se focalizan en los verbos presentan, casi siempre, categorías exclusivamente sintácticas.

Un trabajo que, aunque no se concentra exclusivamente en los verbos, sino en un gran conjunto de rasgos lingüísticos es la descripción de la variación de registros de Biber (1988). A diferencia de las pesquisas de Biber (1988, 1993) y de Biber y otros (1998), la presente investigación se centra exclusivamente en los verbos.

Otro aspecto a tener en cuenta respecto a las indagaciones sobre los verbos que se realizan en lingüística de corpus es que la mayoría de estos estudios solo consideran las primeras frecuencias de los verbos que aparecen en un corpus. En este trabajo, por el contrario, se estimarán el total de las frecuencias de los lemas verbales. De este modo, podremos calcular cuáles son los verbos comunes y los verbos específicos de un subcorpus, dicho de otro modo, en esta investigación se toman en cuenta todo el rango de las frecuencias verbales del corpus que se analizará. En este sentido, este trabajo se encuentra muy cercano a una investigación pionera en lingüística de corpus (Johansson, 1981). Este autor desarrolló un sistema matemático para estimar los rasgos más característicos de un registro o un género (*distinctiveness coefficient*). En este estudio, se presentará una medida similar, aplicada esta vez no a un conjunto de rasgos sino solamente a los verbos. De esta forma, podremos establecer no solo cuáles son los verbos que mejor caracterizan a un grupo de

textos sino también cuáles son los verbos que son comunes para un conjunto de textos dados.

1.4 La lingüística estadística asistida por computador

Aunque el origen de la 'lingüística estadística' puede encontrarse de forma directa en los estudios de estadística léxica (del francés) de los años 40 y 50 (véase Lebart & Salem, 1994), desde un enfoque actual, el término puede circunscribirse a las investigaciones de lo que se conoce con el nombre de 'lingüística de corpus asistida por computador'. Antes de la aparición de esta disciplina, los enfoques estadísticos en español presentan una larga y arraigada tradición en fonética y fonología (Navarro, 1946, Quiles, 1999). Después de los estudios en fonética y fonología, las investigaciones estadísticas en nuestra lengua se desarrollaron en torno al tema de la disponibilidad léxica (Valencia & Echeverría, 1999). Como ya se señaló, sin embargo, la lingüística estadística, actualmente, debe circunscribirse a los estudios en lingüística corpus, englobando bajo ese rótulo dos grandes líneas de investigación: una de corte más computacional y otra, con propósitos más lingüísticos (Moreno, 1998). Allí bien, ambas orientaciones han utilizado medios estadísticos pero con objetivos distintos. La lingüística computacional es una disciplina general que estudia los modelos computacionales de la estructura y función del lenguaje, su uso y su adquisición (Moreno, 1998; Joshi, 1999). El rol de la estadística, en este caso, se centra en la modelación matemática del lenguaje, generalmente, con el objetivo de crear herramientas de análisis automático de textos, usando modelos estocásticos, cuyos resultados son automáticamente comprendidos por el sistema y continuamente retroalimentados y monitoreados por el investigador. La lingüística de corpus, por su parte, estudia el lenguaje a través de medios informáticos para analizar, describir y explicar fenómenos de distintos niveles lingüísticos en grandes cantidades de datos reales, esto es, partes del lenguaje natural estudiadas a partir de modelos probabilísticos, vectoriales o multidimensionales (Biber, et al. 1998; Stubbs, 1996, 2000; Svarnåk, 1992; Kennedy, 1998; Calavedo, 1999). La estadística, en este caso, es un medio para el análisis lingüístico y no para la modelación matemática del lenguaje.

Para configurar una relación pertinente entre estadística y lingüística se deben considerar algunos de los puntos sintéticos y sobresalientemente señalados por García (2004a, 2004b, 2004c). Además de un recorrido histórico de los estudios y procedimientos estadísticos, este autor nos ofrece una visión crítica sobre las utilidades y los riesgos de la aplicación de los enfoques estadísticos a las ciencias sociales, especialmente, a la lingüística. Desde este punto de vista, el autor enfatiza en que el objetivo de un lingüista es obtener conclusiones lingüísticas sobre un fenómeno, independientemente, de los medios (deductivos o inductivos) utilizados para llegar a dichas conclusiones. La integración de la lingüística con la estadística, entonces, requiere de tres procesos (García, 2004a):

- Atribuir números a hechos y manifestaciones lingüísticas.
- Someter las cifras a técnicas matemáticas seleccionadas con la intención de medir, comprobar, aceptar o rechazar fenómenos o suposiciones reflejadas y ponderadas en los índices e indicadores obtenidos.
- Interpretar lingüísticamente los resultados.

En la ejecución del punto (a) debe ponerse atención a la elección de una estadística adecuada. Esta etapa es, según García (2004a, 2004c), la más riesgosa. El investigador tiene que determinar qué tipo de variables va a ocupar (cuantitativas o categóricas) y establecer qué tipo de estadística es la adecuada para dichas variables, según la pregunta que se busca responder. Si el lingüista pasa por alto los puntos anteriores, los resultados de su investigación son simplemente aberrantes. El punto (b) es el estudio netamente estadístico. Los resultados se obtienen por medio de cálculos matemáticos que guardan una relación con los fenómenos lingüísticos. Desde el punto de vista matemático, dichos resultados son inquestionables. Finalmente, el punto (c) tiene un carácter estéticamente lingüístico, se basa en los resultados matemáticos, solamente en tanto son útiles para dar una explicación lingüística del fenómeno.

2. LOS VERBOS ESPECIALIZADOS

En este apartado revisamos una aproximación al estudio de los verbos en cuanto portadores de conocimiento especializado según Lorente (2002). La importancia del planteamiento de esta autora radica en que tanto teórica como metodológicamente su estudio está muy cercano a la presente investigación ya que, al igual que en esta pesquisa, se relaciona una tipología verbal a un corpus especializado. A pesar de que dicho estudio se enfoca más en la teoría de la terminología y que metodológicamente no sigue exactamente los mismos pasos del presente estudio, consideramos necesario tener en cuenta los aspectos señalados por Lorente (2002).

Revisamos, en primer lugar, algunas cuestiones generales respecto al rol de los verbos en la teoría sobre terminología. En segundo lugar, tratamos sintéticamente la clasificación verbal propuesta por Lorente (2002). Al final del apartado, presentamos un esquema de la estructura de los verbos especializados y algunas relaciones que se pueden establecer entre esta clasificación y las otras tipologías verbales descritas en este apartado.

Un primer punto interesante que señala Lorente (2002) es la constatación que presenta al comienzo de su trabajo. Según la autora, en los estudios terminológicos la presencia de unidades léxicas verbales es muy reducida. Esta comprobación puede ampliarse a distintas teorías donde el estudio verbal siempre ocupa un lugar secundario o simplemente no ocupa lugar alguno respecto al estudio de otras unidades lingüísticas. Como conclusión, Lorente (2002) sostiene que una descripción adecuada de los textos especializados debe, necesariamente, incorporar una representación de categorías léxicas verbales. Finalmente, la autora señala en forma exhaustiva las posibles causas de este desinterés por la categoría verbal en los diversos estudios terminológicos.

2.1 Los verbos discursivos

En el nivel más general de la clasificación, se encuentran un tipo de verbos que Lorente (2002) denomina verbos discursivos. Estos verbos están directamente ligados a la función textual que ejercen. Según la autora, los verbos discursivos no están ligados a un conocimiento especializado sino más bien a la competencia pragmática. Dichos verbos están, por lo tanto, relacionados a los géneros discursivos y a la tipología textual utilizados en los

comunicaciones de los especialistas (e.g. *presentar, decir, argumentar, defender, discutir*). En general, se dice que los verbos discursivos corresponden con tres tipos de unidades verbales:

- A los actos de habla del tipo 'opinar, comunicar, transmitir, decir'.
- A verbos que expresan la finalidad del discurso como 'presentar, convencer, instruir, demostrar'.
- A verbos que dan cuenta de la estructura del discurso como 'describir, narrar, dar instrucciones, argumentar'.

Además de estos tres tipos de verbos, Lorente (2002) incluye dentro de la categoría de verbos discursivos a los denominados verbos psicológicos del tipo 'pensar, opinar, temer, dudar'.

2.2 Los verbos conectores

Los llamados verbos conectores corresponden básicamente a los verbos copulativos y pseudocopulativos de la gramática tradicional. Según Lorente (2002), estos verbos expresan básicamente relaciones de equivalencia, igualdad, similitud, dependencia o atribuyen cualidades o valores. Aunque se sostiene que estos verbos tampoco tienen por se un valor especializado, se argumenta que son básicos para la expresión de cualquier tipo de conocimiento, es decir, las afirmaciones sobre la igualdad o la identificación de referentes no pueden generarse sin el conocimiento de una materia. En este sentido, se explica que, a pesar de que los verbos conectores no son verbos especializados, forman parte de lo que Lorente (2002) denomina Unidades de Conocimiento Especializado (UCE). Otras verbos que según la autora se incluirían dentro de la categoría de los verbos conectores, son los verbos de las operaciones metalingüísticas o los verbos que aparecen en las definiciones de conceptos. Algunos ejemplos típicos de los verbos conectores son: 'ser, parecer, equivaler, corresponder, identificar, diferenciarse'.

2.3 Los verbos fraseológicos

Los verbos fraseológicos se refieren a un tipo especial de verbos que adquieren un valor especializado solo cuando aparecen combinados con otras unidades en una construcción sintagmática y en un contexto especializado. De acuerdo a Lorente (2002), estos verbos son característicos de los textos de especialidad. Son utilizados en estos textos para expresar acciones, procesos y estados pero su significado no se distingue de un uso en contextos de lengua no especializada. Este tipo de verbos pueden adquirir un valor especializado cuando forman parte de unidades sintagmáticas, preferentemente, en construcciones de verbo en infinitivo más un complemento directo. Algunos ejemplos de estas construcciones son: 'generar energía', 'instalar penitencia', 'producir recursos', 'disminuir costos', etc. Dentro de los verbos fraseológicos también se integran los llamados verbos de sopo-*le*, ya que, con una menor carga semántica y mayores restricciones gramaticales, forman parte de combinaciones léxicas bastante fijadas en el uso y muy cercanas al comportamiento léxico (e.g. 'dar tratamiento', 'hacer causa').

La presencia de estos verbos en descripciones y aplicaciones del discurso especializa-

do estará limitado, según Lorente (2002), a la inclusión de unidades fraseológicas enteras o bien a la introducción de los lemas verbales en aquellos casos en que el funcionamiento gramatical o el significado difieren de los habituales en contextos no especializados.

A pesar de la importancia de los verbos fraseológicos, estos no serán considerados como unidades de análisis en el presente estudio ya que para determinar su carácter especializado se debe analizar sintácticamente el contexto de uso de estas unidades, análisis que, como ya dijimos anteriormente, no constituye una parte de la metodología de la presente investigación.

2.4 Los verbos-término o verbos especializados

Como parte final de esta clasificación aparecen un tipo de verbos que la misma autora denomina verbos-término. Estos son la clase más básica de verbos de esta clasificación y corresponden básicamente a verbos cuyo contenido semántico está estrechamente ligado a un área específica de conocimiento. Según define Lorente (2002), esta clase de verbos estaría formada por aquellas unidades cuyos lexemas y significados están vinculados exclusivamente, o de manera referida, a un ámbito de especialidad. Verbos, como *capitalizar* (arquitectura), *eutrofizar* (ecología), *acetificar* (química), son un buen ejemplo, y se caracterizan por estar en correlación con unidades léxicas de categoría nominal o adjetiva (*capitalizar*, *capitalizado*, *eutrofizar*, *eutrofización*). Entre los verbos-término son muy frecuentes las unidades formadas por derivación y muchos presentan variantes sintácticas formadas por el verbo *hacer* más un derivado nominal (Lorente, 2002).

Ahora bien, en la Figura 1 se puede observar la estructura inclusiva de los distintos verbos de conocimiento específico según los hemos descrito en este apartado.

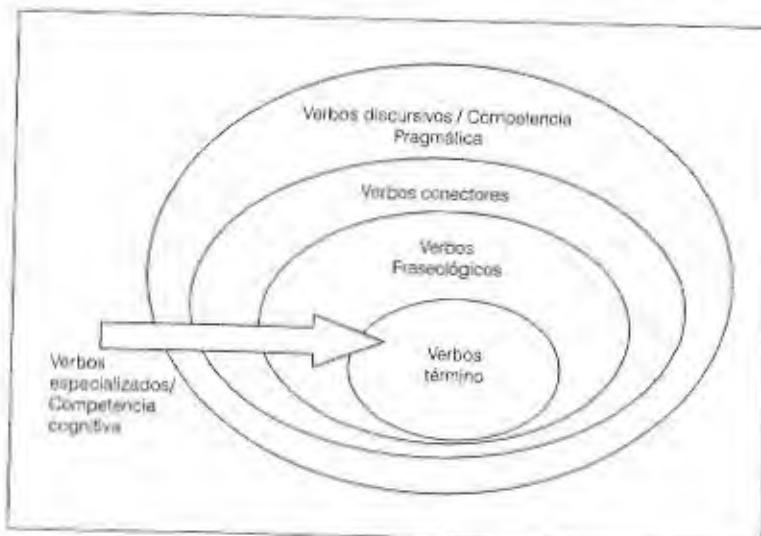


Figura 1. Estructura de los verbos de especialidad. Adaptado de Lorente (2002).

Según la Figura 1, los verbos-término, corresponden a aquellos verbos que portan un conocimiento especializado y se encuentran, por lo tanto, en el nivel más específico de la clasificación. Tanto los verbos discursivos como los verbos conectores y los verbos fraseológicos solo adquieren un valor especializado en algunos contextos determinados y, en este sentido, tienen un valor más difuso en cuanto unidades de conocimiento específico.

3. ESPECIFICIDAD Y COMUNALIDAD VERBAL

Si el grado de especialización es un aspecto más bien lexicográfico del verbo (la relación que guarda el significado de un verbo con un área específica del conocimiento), en el caso de la especificidad, tal como la entendemos en este estudio, es una relación de carácter estadístico que se establece entre un lema verbal (unidad lexicográfica) y su aparición en un corpus dado. Así entendida, la especificidad siempre debe referirse a un corpus determinado (es decir, a una muestra lingüística) y no a la lengua en general o al léxico mental. En este sentido, este trabajo se encuentra muy cercano a una investigación pionera en lingüística de corpus (Johansson, 1981). Como ya señalamos, este autor desarrolló un sistema matemático para hacer una estimación de los rasgos más característicos de un registro o un género (*distinctiveness coefficient*). En la presente investigación se utilizará un coeficiente similar, aplicado en este caso, exclusivamente a los lemas verbales. La estimación del grado de especificidad verbal se realizará en este estudio según el siguiente supuesto: el grado de especificidad verbal de un registro o subcorpus depende de la cantidad de verbos exclusivos de ese registro contrastada con los otros subcorpus o registros que componen el corpus general, o dicho de otra forma, el grado de especificidad verbal de un corpus aumenta en la medida en que sus verbos no aparecen en los otros corpora que sirven de contraste. Opuesta a la idea de especificidad está la idea de communalidad verbal. En términos estadísticos un corpus presenta un alto grado de communalidad verbal, en cuanto sus verbos aparecen en más de uno de los registros o subcorpora que componen el corpus general. Aunque sabemos que, tanto el grado de especificidad como de communalidad dependen, en gran medida, de la cantidad de subcorpora que integren el corpus estudiado y, considerando además, que en el presente estudio solo se incorporan tres registros principales; creemos que interpretando los resultados obtenidos en su límite exacto, estos igualmente pueden contribuir a un conocimiento más acabado de los aspectos verbales del corpus analizado.

4. VARIABILIDAD VERBAL

La noción de variabilidad, en general, tiene una larga tradición en distintos estudios lingüísticos, principalmente, de carácter léxico. Al igual que en el caso anterior (especificidad/comunalidad), la variabilidad es una medida que depende en un alto grado de las características del corpus que se está investigando. Una forma típica de medir variabilidad se refiere a la tasa que se obtiene de la división de los tipos (*types*) por los casos (*tokens*), lo que tradicionalmente se aplicaba usando como unidad de análisis la palabra (Chafe & Danielewicz, 1987). Los tipos, en ese caso, corresponden al total de palabras diferentes en un corpus y los casos al número total de palabras. El grado de variabilidad es un coeficiente que tiene un rango que va desde 0 a 1. Si el resultado del coeficiente tiende a 1, el

corpus analizado es más variable. Por el contrario, si el resultado tiende a 0, decimos que se trata de un corpus poco variable. Gracias a herramientas computacionales de etiquetaje morfológico es posible aplicar el coeficiente tipos/casos a una categoría gramatical específica, en este caso, a los verbos. La variabilidad verbal, así entendida, se refiere al coeficiente entre los tipos de lemas verbales y la cantidad total de las formas verbales del corpus. La variabilidad verbal, sin embargo, no es una medida independiente del corpus de análisis, por lo cual, para obtener un índice más exacto es necesario complementar la estimación de la variabilidad con el cálculo del porcentaje verbal del corpus. De esta forma para obtener un índice adecuado de variabilidad deben aplicarse los siguientes dos fórmulas:

- a) Variabilidad verbal
- $$\frac{\text{Tipos de lemas verbales}}{\text{Total de casos verbales}}$$
- b) Porcentaje verbal del corpus
- $$\frac{\text{Total de casos verbales} * 100}{\text{Número total de palabras}}$$

5. LA INVESTIGACIÓN

El estudio del que da cuenta este capítulo constituye parte de un proyecto mayor cuyos resultados conforman en gran medida el contenido de este libro, razón por la cual, omitiremos los pasos metodológicos de descripción y procedimiento de recolección del corpus puesto que ya han sido descritos en los capítulos [2, 3 y 4]. Nos limitaremos, entonces, a describir el procedimiento de recolección y procesamiento de los datos, a establecer qué tipo de estadística usamos y a declarar las preguntas que buscamos responder.

5.1 Preguntas de investigación

Dado nuestro marco teórico y considerando las facilidades de las herramientas tecnológicas, las preguntas que esperamos responder en esta investigación son las siguientes:

1. ¿Cuáles y cuántos son los verbos especializados del Corpus PUCV-2003?
2. ¿Cuán específicos o cuán comunes son, respecto a sus verbos, los distintos subcorpora que conforman el corpus general?
3. ¿Cuán variables son, respecto a sus verbos, los subcorpora que constituyen el Corpus PUCV-2003?

5.2 Tipo de estudio

La investigación es de tipo descriptivo no experimental, no paramétrico y se enmarca dentro de los procedimientos metodológicos de la lingüística de corpus y de la estadística léxica. Es importante señalar que a pesar del énfasis en lo estadístico, el presente estudio no es puramente de índole cuantitativa sino que también implica un análisis propio de los enfoques cualitativos.

5.3 Las variables

Las variables que investigaremos para responder las preguntas que nos hemos propuesto son las siguientes:

- A → Grado de especialización de un verbo
- B → Índice de especificidad y de comunidad verbal de los subcorpora
- C → Índice de variabilidad verbal de los subcorpora

Puesto que no se trata de un estudio correlacional, no existe un estatus causal entre las variables de estudio, es decir, no se pueden distinguir variables dependientes de independientes. A pesar de lo anterior, se pueden sostener algunas relaciones o cruces entre las variables arriba mencionadas que revisaremos en la discusión.

5.4 Procedimiento de recolección, codificación y análisis estadístico de los datos

La recolección de los datos para esta investigación comenzó con la obtención de listas de tipos de lemas verbales (junto a su frecuencia de casos) tanto los distintos subcorpora como del corpus general. Para realizar esta tarea se utilizaron las herramientas de consulta (Ver Interfaz El Grial, www.elgrial.cl), desarrolladas en el Laboratorio Computacional de Ciencias del Lenguaje del Programa de Postgrado en Lingüística de nuestra Universidad. La codificación de los datos se realizó según la variable que se estuviera midiendo. En el primer caso, el grado de especialización de los verbos, no se trata de una variable cuantificable sino que suponen un análisis cualitativo (variable categórica) que indirectamente puede representarse como datos numéricos. Para realizar esta tarea se adoptó un enfoque netamente cualitativo que consistió en buscar en un diccionario (www.rae.es) caso a caso todos los tipos de lemas que aparecieron en el corpus. De esta forma se determinó que un verbo era especializado según las marcas (explícitas e implícitas) de especialidad que aparecen en su definición en el diccionario y apoyándose en el conocimiento del investigador como hablante nativo, se agruparon los verbos según las especialidades que emergieron del análisis. Indirectamente, se traspasaron estos resultados a una representación numérica y se determinó cuánto (porcentualmente) del total de los verbos del corpus correspondía a verbos especializados. En el segundo caso, el índice de especificidad y comunidad de los subcorpora se trata de una variable cuantificable que se calcula según la cantidad de verbos que aparecen exclusivamente en un registro (el índice de especificidad) y la cantidad de verbos que son compartidos entre dos o más subcorpora (el índice de comunidad). Estos índices implican el uso de una estadística de proporciones representadas en porcentaje.

Lo último variable en estudio es la variabilidad verbal de los registros del corpus general. Como ya mostramos en el marco teórico, también se trata de una variable cuantificable que se puede medir según las fórmulas a) y b), introducidas en el marco teórico. La primera fórmula arroja una estadística por medio de un rango que varía de 0 a 1, números que representan una baja y una alta variabilidad respectivamente. La segunda fórmula supone una estadística de proporciones que se manifiesta en un porcentaje dado, que corresponde a la cantidad de formas verbales en el total de palabras.

6. RESULTADOS

6.1 Los verbos especializados en el Corpus PUCV-2003

Después de obtener la lista total de los tipos de lemas verbales (3.558 tipos correspondientes a 254.794 casos de formas) y de consultar uno a uno el significado de esos tipos en un diccionario se obtuvieron los resultados que se muestran en la Tabla 1:

Tabla 1. Los verbos especializados en el Corpus PUCV-2003

	Verbo (lema)	Frecuencia		Verbo (lema)	Frecuencia
1	Abarrocar	1	24	Granillar	1
2	Abarrocar	1	25	Infringir	6
3	Abogar	8	26	Insumir	1
4	Abolir	18	27	Jarrar	2
5	Absolver	4	28	Lastrar	2
6	Acerar	72	29	Mandillar	2
7	Achelonear	6	30	Mineralizar	1
8	Acodar	2	31	Ministrar	6
9	Adicionar	19	32	Minutar	2
10	Aducir	4	33	Niquelar	4
11	Alicatar	10	34	Nitratar	7
12	Amolar	1	35	Nitruar	5
13	basamentar	1	36	Oxidar	36
14	Biselar	8	37	Oxigenar	10
15	Bilar	1	38	Refundir	3
16	Bromar	3	39	Roblar	1
17	Bunkar	4	40	Sinterizar	2
18	Calafatear	1	41	Soldar	206
19	Carenar	2	42	Treñar	1
20	Cintar	16	43	Troneter	1
21	Desguazar	1	44	Variar	2
22	Estibar	58	45	Zarpar	17
23	Gorrir	3	46	Zunchar	2

La Tabla 1 muestra el total de los tipos de lemas verbales de especialización (46) en el Corpus PUCV-2003. En la primera columna, aparece el número de tipo verbal. En la segunda columna, se muestra el lema y, en la última columna, aparece la frecuencia bruta de los casos asociados a ese lema. Estos resultados muestran que los verbos especializados representan apenas un 1.29 % del total de tipos de verbos del corpus. Las especialidades que emergieron del análisis del significado de estos verbos se presentan en la Tabla 2.

Tabla 2. Verbos por especialidad

MARÍTIMA	METALURGIA	CONSTRUCCIÓN	LEGAL	ECONOMÍA
Abarrocar	Acerar	Abuscar	Abogar	Aduanar
Amar	Bullar	Abullonar	Absolver	Insumir
Bilar	Granillar	Acodar	Abolir	
Calafatear	Mandillar	Alicatar	Aducir	
Carenar	Mineralizar	Basamentar	Infringir	
Desguazar	Niquelar	Biselar	Ministrar	
Estibar	Nitratar	Bromar	Ministar	
Gorrir	Nitruar	Cintar		
Iastrar	Oxidar	Jarrar		
Troneter	Oxigenar			
Variar	Refundir			
Zarpar	Roblar			
	Sinterizar			
	Soldar			
	Treñar			
	Zunchar			

A partir de la aplicación del Anexo 1, en el que aparecen las entradas léxicas de estos verbos en el diccionario (www.roe.es), se agruparon los verbos según las áreas o disciplinas que aparecen en la primera fila de la Tabla 2. Los contextos de utilización de los verbos de la Tabla 2 estarían estrechamente ligados a las prácticas, usos y comunidades discursivas de esas áreas o disciplinas. A partir de los datos de la Tabla 2, podemos deducir que es el área metalúrgico lo que presenta un mayor número de tipos verbales especializados. Cabe señalar que, aunque no fue objeto de esta investigación existe una relación especial entre las áreas de especialidad de la Tabla 2 y el subcorpus de textos técnico científicos. Tal como se describe en los capítulos 2, 3 y 4 este corpus se subdivide en áreas de especialidad (marítima - industrial metalmeccánica - comercial), que inicialmente no fueron consideradas en este estudio. Sin embargo, esas divisiones resultan interesantes de comparar con los resultados de la Tabla 2. Al respecto, se puede observar que todas las áreas del corpus CTC (marítima - industrial metalmeccánica - comercial) aparecen más o menos directamente en las especialidades que emergen de un análisis del significado de esos verbos. Pero también, aparecen otras especialidades que no se consideran en las distinciones de dicho corpus, como son, el área legal o el ámbito de la construcción. Así también, se puede observar que los registros al interior del corpus técnico científico no implican, en algunos casos, un gran número de verbos especializados asociados léxicamente a un área del conocimiento, como es el caso del área comercial, que a pesar de estar relacionada con el área de economía de la Tabla 2, presenta apenas 2 verbos especializados.

El mayor número de apariciones de este tipo de verbos en un área específica se aso-

ciencia a una mayor importancia de los verbos como transmisores del conocimiento especializado de una disciplina.

6.2 El índice de especificidad y comunalidad verbal en los registros del Corpus PUCV-2003

Por medio del análisis computacional y estadístico del total de los casos de verbos del Corpus PUCV-2003, se extrae el número de los verbos exclusivos de cada corpus. Estos datos se presentan en la Tabla 3.

Tabla 3. Índice de especificidad verbal

Subcorpus	Número de verbos exclusivos	Porcentaje
CTC	581	26.35 %
CLL	1119	39.24 %
CEO	66	6.21 %

En la segunda columna de la Tabla 3 aparece el total de los verbos exclusivos de cada corpus. El porcentaje que se presenta en la tercera columna representa la proporción de los verbos exclusivos en el total de verbos que aparecen en cada corpus. Según este índice, el CLL es el corpus con un mayor porcentaje de verbos que no aparecen en los otros registros. En un nivel intermedio de ese índice se ubica corpus técnico científico con un 26.35% de verbos específicos. Por último, apenas un 6.21% de los verbos del corpus de entrevistas orales aparece exclusivamente en ese registro. En este sentido el grado de especificidad en el Corpus PUCV-2003 presenta, de forma ascendente, el orden siguiente CLL>CTC>CEO.

A partir de la lista total de los verbos del corpus general, analizamos el grado de comunalidad verbal existente entre los distintos subcorpora (y todas sus agrupaciones posibles) del corpus general. En la Tabla 4, se presentan estos resultados.

Tabla 4. Índice de comunalidad verbal entre registros

Subcorpora	Número de verbos compartidos	Porcentaje
CTC/CLL	798	22.42 %
CLL/CEO	170	4.77 %
CTC/CEO	61	1.71 %
CTC/CEO/CLL	771	21.66 %

En la primera columna de la Tabla 4, aparecen las distintas agrupaciones posibles del Corpus PUCV-2003. En la segunda columna, se entrega el número de verbos compartidos en cada agrupación. Por último, en la tercera columna, se muestra el índice de comunalidad verbal que se calcula estimando el porcentaje de verbos compartidos por el total de tipos de verbos (3.558). Los resultados muestran que los registros más comunes des-

de el punto de vista verbal son el Corpus Técnico Científico y el Corpus de Literatura Latinoamericana. Por el contrario, la agrupación del corpus técnico científico con el corpus de entrevistas orales presenta una comunalidad verbal mínima. La comunalidad entre los verbos del corpus de literatura latinoamericana y el corpus de entrevistas orales, aunque es mayor que la anterior, tampoco es muy elevada (4.77%). El grado de comunalidad verbal de los tres registros del Corpus PUCV-2003 es de un 21.66%. Es interesante notar que esa cifra se acerca bastante a la comparación entre el CTC y el CLL. De esto se puede deducir que la cantidad de verbos comunes que aporta el CEO al cálculo total (CTC/CLL/CEO) es casi insignificante.

6.3 Variabilidad verbal de los registros del Corpus PUCV-2003

A través del análisis estadístico del total de los verbos del Corpus PUCV, obtuvimos los resultados que aparecen en la Tabla 5.

Tabla 5. Variabilidad verbal del Corpus PUCV-2003

	PUCV		CTC		CLL		CEO	
	Tp	Tk	Tp	Tk	Tp	Tk	Tp	Tk
V	3.558	254.794	2.205	84.155	2.852	85.434	1.062	85.205
Tp/Tk	0.139		0.262		0.333		0.124	
%V / N	17 %		13 %		19 %		23 %	

V

= Verbos

Tp = Tipos (cantidad de verbos distintos)

Tk = Casos (cantidad total de verbos)

Tp/Tk = Tasa de variabilidad verbal

%V / N = Porcentaje de verbos en el número total de palabras

En la primera columna de la Tabla 5, se puede observar el total de tipos verbales del Corpus PUCV-2003, correspondientes a 3.558 tipos de verbos, que representan el universo de las formas verbales del corpus (254.794). A partir de las observaciones en cada registro, se estimó la tasa de variabilidad que aparece en la cuarta fila de la (Tp/Tk) de la Tabla 5. Este índice debe complementarse con la información porcentual que se muestra en la última fila (%V / N) de la tabla, a saber, el porcentaje de verbos en el total de palabras.

En primer lugar, apreciamos que existe una diferencia en la cantidad de tipos verbales en las distintas subdivisiones del corpus. Es interesante destacar que es el CLL donde hay una mayor cantidad de tipos verbales (2.852) y también un mayor índice de variabilidad (0.333). Por el contrario, el CEO es el corpus donde hay una menor cantidad de tipos verbales (1.062) y también una menor tasa de variabilidad (0.124). Pasando por alto la tasa de variabilidad, puesto que es una medida que está demasiado influenciada por el número total de palabras, comentamos a continuación la relación entre esta tasa de variabilidad y el porcentaje de verbos que hay en el corpus. En general, el rango porcentual de los verbos en el corpus va desde el 13% en el CTC al 23% en el CEO. A partir de estos

simples datos se puede concluir que el CEO, a pesar de tener la tasa de variabilidad más baja (0.124), cuenta con el mayor porcentaje de verbos en relación a la cantidad total de palabras. Esto significa que en ese corpus (CEO) se utilizan más verbos por palabras que en los otros corpora pero que esos verbos tienden a ser los mismos, es decir, no hay mayor variabilidad verbal. Un último aspecto interesante respecto a la Tabla 5 es el siguiente. En el subcorpus CTC, el porcentaje de verbos por número total de palabras es comparativamente el más bajo. Esto permite inferir que en ese registro los verbos tienen menor importancia, comparativamente hablando, que otras categorías gramaticales principales (como los sustantivos o los adjetivos). El caso contrario se puede decir del CLL y el CEO que, a la luz de los datos, son corpora con una mayor tendencia verbal, aunque no necesariamente, como ya dijimos en el caso del CEO, con una gran variabilidad verbal.

7. DISCUSIÓN

Los verbos especializados del Corpus PUCV-2003 no representan un gran número de tipos, pero a partir de su análisis léxico emergen cinco áreas de especialidad. Tres de esas cinco áreas aparecen representadas más o menos directamente en el corpus (marítima, industrial metalmeccánica y económica). En este sentido, es interesante constatar que una de las áreas que no aparece representada (construcción) en las áreas de especialidad del corpus y que si emerge del análisis léxico de estos verbos, ocupa el tercer lugar respecto a la cantidad de tipos de verbos especializados. Por otro lado, aunque el área de economía si tiene una relación directa con una de las divisiones del corpus CTC, a saber, el área comercio, no presenta un gran número de verbos de este tipo. Estos resultados nos permiten reflexionar en distintas direcciones. Una explicación para el bajo número de estos verbos en el Corpus PUCV-2003 puede ser que los verbos, en general, juegan un papel menos importante que los sustantivos en la transmisión de conocimientos especializados. Esto sería corroborado por las aprehensiones de Lorente (2002) respecto del menor interés que suscitan los verbos respecto a otras categorías gramaticales (específicamente los sustantivos) en los estudios de terminología.

La emergencia de áreas del conocimiento que no aparecen directamente representadas en el corpus, pero que sí se pueden deducir del análisis léxico de dichos verbos, implica algunas consecuencias para el estudio de unidades terminológicas a través de corpus, puesto que las divisiones por área de especialidad en un conjunto de datos lingüísticos no garantiza, en primer lugar, un gran número de apariciones, ni, en segundo lugar, que aparezcan solo unidades relacionadas a dichas áreas. La razón de esto es más o menos obvia. Por otras investigaciones (ver Capítulo 1), sabemos que el grado de especialización de un texto más que una categoría fija se constituye como un *continuum*. Esto implica que una especialidad no es un conjunto hermético de rasgos lingüísticos fijos sino que mantiene relaciones conceptuales o lingüísticas con otras áreas del conocimiento o con otros tipos textuales de la lengua general. Así por ejemplo, aunque el área legal no aparece en las subdivisiones temáticas del corpus CTC, si aparece en el análisis léxico de los verbos del corpus porque está asociada conceptualmente (aunque no directamente) con el área comercial del mismo.

Tal como aparece en los resultados, el grado de especificidad, de comunalidad y

variabilidad del Corpus PUCV-2003 se puede resumir en los siguientes puntos:

- El corpus de literatura latinoamericana (CLL) es el más específico y, a la vez, el más variable.
- El corpus de entrevistas orales (CEO) resulta ser el menos específico y el menos variable aunque, tal como vimos, es el que mayor porcentaje de verbos por cantidad total de palabras presenta.
- El corpus de textos técnicos científicos se sitúa en un nivel intermedio (entre los extremos CLL y CEO), tanto en el índice de especificidad como en el de variabilidad y es, claramente, el corpus donde el porcentaje de verbos por cantidad total de palabras es el menor.
- El mayor grado de comunalidad se da entre los subcorpora CTC y CLL, mientras que el contraste entre CTC y el CEO es el que arroja el nivel más bajo en este índice.

Estos hallazgos nos permiten deducir que los índices estudiados dependen fuertemente de las funciones (macrofuncionales) que cada tipo de corpus analizado presenta. Al momento de interpretar estos resultados, entonces, es necesario establecer relaciones entre las características funcionales de estos corpora y los índices estadísticos que hemos estimado. El corpus CTC, en este sentido, cuya función principal es la transmisión directa de conocimientos técnicos otorga una menor importancia a los verbos en comparación a otras categorías gramaticales, específicamente, los sustantivos, los cuales son más pertinentes para actualizar la función referencial propia de ese corpus. Los textos de literatura latinoamericana, por su parte, cuya objetivo es narrar acciones, eventos o estados presenta, por lo tanto, no solo un mayor porcentaje de verbos por cantidad total de palabras, sino también, debido a las funciones estéticas de estos textos, un uso más variable de unidades verbales. Por último, las restricciones contextuales del corpus de entrevistas orales redundan en el menor índice de variabilidad verbal que aparece en dicho corpus, pero también, debido a que la función de la oralidad es principalmente interactiva, este corpus presenta el mayor porcentaje de verbos por cantidad total de palabras.

CONCLUSIONES

En esta sección, se presenta un resumen del presente estudio junto con una reflexión de los alcances, las limitaciones y las proyecciones de los resultados obtenidos en el mismo.

Aunque siempre el tema de los verbos ostenta un capítulo muy completo (y complejo) en cualquier tipo de enfoque gramatical, su estudio es escaso y secundario en muchas otras disciplinas, que en conjunto conforman las ciencias del lenguaje. A partir de las nociones de especialización, especificidad (comunalidad) y variabilidad, en este trabajo se ha investigado los verbos en un corpus lingüístico informatizado desde un enfoque estadístico y léxico.

La especialización de un verbo depende de la asociación léxica de su significado con los conocimientos de un área o disciplina que, en términos de Swales (1990), denominamos comunidad discursiva. Según el marco de referencia de Lorente (2002), el estudio de los verbos, al menos, en el ámbito de la investigación en terminología, juega un papel mínimo. Este enfoque (Lorente, 2002) proporciona una definición de los verbos que permite visualizar la relación entre su significado y los conocimientos propios de un área o especificidad técnica o científica.

Los nociones de especificidad y comunidad verbal, a diferencia de la noción anterior, se definen a partir de una relación estadística, a saber, la ocurrencia de un lema verbal en un corpus dado. La especificidad y communalidad, tal como se entienden en este estudio, son un índice porcentual de los verbos exclusivos (en un corpus) y comunes (entre dos o más corpora). La estimación aplicada es una adaptación exclusiva a los verbos del cálculo del conjunto de unidades léxicas más características (léase exclusivos) de un corpus realizado por Johansson (1981).

La variabilidad léxica, entendida en términos tradicionales, corresponde al coeficiente entre las unidades (generalmente, palabras) distintas (tipos) y las unidades totales (casos) de un corpus (Chafe & Danielewicz, 1987). De nuevo, la variabilidad, tal como se entiende y se aplica en este estudio, es una adaptación de esa estimación tradicional. Primero, se aplica solamente a las unidades léxicas verbales y, segundo, se complementa con el cálculo del porcentaje de verbos por el total de palabras, lo que permite superar el alto grado de dependencia que en cálculo de variabilidad a secas (i.e. los tipos divididos por los casos) presenta respecto al número total de palabras.

Dentro de las principales limitaciones de este estudio, se puede mencionar el carácter no paramétrico de los resultados. En este sentido, sabemos que de las variables analizadas podrían variar significativamente su comportamiento si se ampliara la muestra y la heterogeneidad del corpus investigado, a saber, la especialización y los índices de especificidad y communalidad verbal. Si ampliáramos la muestra, aparecerían otros verbos especializados y emergirían otras áreas de conocimientos asociados a ellos. Así también (y más radicalmente), los índices de especificidad y communalidad serían completamente distintos si incluyéramos en nuestro análisis otros registros o subcorpora. La estimación de la variabilidad, en cambio, resulta ser una variable más paramétrica o preyectable a un universo mayor ya que es muy posible que dicho cálculo siga la misma tendencia en otros corpora similares en función a los analizados en este trabajo, datos que se pueden comprobar en investigaciones futuras.

A pesar de las limitaciones recién expuestas, esta investigación presenta grandes fortalezas y proyecciones en diferentes ámbitos, que, a nuestro juicio, significan un avance tanto en las investigaciones verbales como en la caracterización de distintos registros. Una de las principales fortalezas del trabajo es la gran cantidad de lemas verbales analizados (3.558). Esto permite obtener una amplia descripción del comportamiento de esta unidad en un conjunto de datos lingüísticos determinados. Por otra parte, creemos que el enfoque metodológico adoptado constituye un modo (en cierta medida) innovador para el estudio de los verbos que supera la exclusividad de la sintaxis en los estudios verbales, accediendo así, a responder preguntas que habitualmente no se formulan. Las proyecciones de este trabajo aparecen vinculadas a distintas áreas lingüísticas. Primero, constituye un antecedente para los investigadores en terminología o el estudio de la transmisión lingüística de conocimientos especializados. Segundo, los resultados obtenidos pueden ser utilizados para complementar otras investigaciones descriptivas sobre otras categorías lingüísticas en corpus por medio de métodos estadísticos. En tercer y último lugar, la identificación de verbos especializados en este corpus, puede ser aplicada para medir, por medio de pruebas psicolingüísticas, el dominio conceptual que un conjunto de sujetos tiene sobre un área especializada.

ANEXO 1

El anexo 1 muestra el total de las entradas del diccionario (www.rae.es) de los verbos especializados del Corpus PUCV-2003.

1	Abalorar (peros, de balorar) tr. Situar [un buque] de costado casi en contacto con otro buque o con un muelle, etc. SIN. Barlear.
2	Abiselar tr. Hacer biselar [en un cristal, una moldura, etc.]. SIN. biselar.
3	Abogar (lat. advocate, llamar cerca de sí) intr. Defender en juicio. 2 fig. Interceder, hablar en favor de alguien. ~ por, a contra, alguien o alguna cosa. CONJUG. como llegar.
4	Abolir (lat. abatere) tr. defect. Derogar, dejar sin vigencia una ley, precepto, costumbre, etc.
5	Absolver (lat. aere) tr. Dar por libre de algún cargo. 2 Remitir [a un penitente de sus pecados]. 3 ant. Resolver [una dificultad]. 4 der. Dar por libre [al reo]. CONJUG. como mover; pp. ineg., absuelto. SIN. 1, 2 y 4 v. Perdonar.
6	Acerar (de acero) tr. Dar [a un hierro] las propiedades del acero, especialmente convertir en acero [el corte o los puntas de las armas o herramientas]. ~ un sable. 2 Dar [a un líquido] propiedades medicinales con tintura de acero o apagando en el acero hecho ásca. 3 Dar los grabadores un baño de acero [a las planchas de cobre]. 4 tr-ppnl. fig. Fortalecer, vigorizar; ésta penalidad aceró su ánimo; acerarse en la lucha.
7	Achaflanar tr. Dar [a una esquina] forma de chaflán. También chaflanar.
8	Acodar (der., del lat. cubilu, codo) tr-ppnl. Apoyar uno el codo sobre alguna parte; acodó el brazo; se acodó. 2 tr. Enterrar [el vástago de una planta] en forma de codo y sin separarlo del tronco para que eche raíces. 3 Poner códales [en la superficie de una piedra o de un madero] para ver si está plana. 4 Acodillar (doblar). 5 Disponeer en ángulo piezas de maquinaria. 6 org. Acodillar. 7 veter. Clavar mal [los clavos] al hierar, desviándolos sobre las partes sensibles. SIN. 2 Cachar, ensamblar, si se trata de vides.
9	Aduanar tr. Registrar en la aduana [los géneros o mercaderías], y pagar los derechos que adeuden.
10	Aducir (lat. adducere) tr. Presentar, alegar [pruebas, razones]. 2 Añadir (agregar) como conductor. incor. aducista.

11	Alicatar (gr. alikat, espejuelo) tr. Azulejar. 2 Cortar los azulejos para darles la forma conveniente.
12	Arrufar (hol. roef) tr. mar. Dar arrufadura [al buque] en su construcción. 2 intr. Hacer arrufadura el buque arrufo por la popa. 3 Grñir los perros enseñando los dientes. 4 pml. Embriácerse, irritarse.
13	Basamentar tr. arq. Poner los basamentos de una edificación.
14	Biselar tr. Hacer biselos [en un cristal, una moldura, etc.] SiN. Abiselar.
15	Bitar tr. Amarar y asegurar [la cadena del ancla] a las bitas, abajo.
16	Bromar tr. Roer la brama [la madera].
17	Berilar tr. Grabar [figuras o adornos] en los metales con el buril. ~ en cobre.
18	Calafatear (gr. mod. kalaphatein) tr. Cerrar [los junctos de las maderas de las naves] con estopa y brea para que no entre el agua. 2 p. ext. Cerrar [otras juncturas].
19	Carenar (lat. carinare) tr. Reparar el casco [de una nave]. ~ de flame un buque, repararlo completamente. 2 Dar forma aerodinámica [a la carrocería de un vehículo].
20	Cintar tr. arq. Poner [cintas o fajas imitadas], como adorno, en las construcciones.
21	Desguazar tr. Desbastar con el hacha [un madero]. 2 Deshacer [un buque] total o parcialmente. 3 Desmontar o deshacer cualquier estructura; esp. vehículos, maquinarias, etc. CONJUG. como realizar.
22	Estivar (lat. stipare) tr. Apretar [materiales o cosas sueltas] para que ocupen el menor espacio posible. ~ la lana al ensacarla. 2 Distribuir convenientemente [todos los pesos] del buque. 3 p. ext. Cargar y descargar mercancías de los buques en cada puerto. 4 Amér. Distribuir y edificar [mercancías] en un local.
23	Garrar intr. Cejar un buque arrastrando el ancla por no haber ésta hecho presa.
24	Granellar tr. Reducir [un metal] o granular.
25	Infringir (lat. -ere) tr. Quebrantar [la ley, un convenio, etc.] CONJUG. como dirigir.

26	Insumir (lat. insu mere, girar) tr. Entre economistas, invertir [dinero y otros bienes] en una producción industrial, agrícola, comercial, etc.
27	Jarrar tr. fam. Jaharrar. Jaharrar [oraci. del ár. yeyyar, encolar, der. de yr, coll] tr. Revocar [una pared].
28	Lastrar tr. mar. Poner lastre [a la embarcación]. 2 Afilar [una cosa] cargándola de peso.
29	Mandrilar tr. Perforar un metal con un mandril II. 2 Ensanchar y pulir los agujeros de las piezas de metal con el mandril.
30	Mineralizar tr. Comunicar una sustancia [a otra], en el seno de la tierra, las condiciones de mineral: el azufre mineraliza el hierro. 2 pml. Cargarse el agua de sustancias minerales. CONJUG. como realizar.
31	Ministrar (lat. -are) tr.-intr. Servir [un oficio, a ministerio]; ministro en la Audiencia, ministro la justicia - 2 tr. Dar, suministrar [una cosa].
32	Minutar (de minuto) tr. Hacer la minuta [de un contrato, escritura, etc.].
33	Niquelar tr. Cubrir con un baño de níquel [otro metal].
34	Nitratar tr. Abonar [la tierra] con nitratos.
35	Nitrurar tr. Endurecer superficialmente los metales ferrosos mediante acción del nitrógeno caliente.
36	Oxidar (de óxido) tr.-pml. Combinar [una sustancia] con oxígeno. 2 Quitar hidrógeno [a un compuesto] por la acción del oxígeno. CONTR. 1 Reducir.
37	Oxigenar (de oxígeno) tr. Oxidar. 2 pml. fig. Airearse, respirar al aire libre.
38	Refundir (lat. -ere) tr. Volver a fundir o liquidar [los metales]. 2 fig. Dar nueva forma o disposición [a una obra de ingenio, comedia, díctirso, etc.]. 3 fig. Comprenderse o incluir. ~ una poesía en un teatro, refundirse dos obras. 4 intr. fig. Redundar: el trabajo refundió a su favor. 5 pml. Amér. Extraviarse, perderse. 6 Guat., Méx. Guardar algo con mucho ahínco.
39	Roblar (v. robar) tr. Roblar. 2 Doblar o remachar [una pieza de hierro] para que esté firme.

40	Sinterizar (del al. sinter, escoria, ceniza) tr. Metal. Producir piezas de gran resistencia y dureza calentando, sin llegar a la temperatura de fusión, conglomerados de polvo, grante, metálicos, a los que se ha modelado por presión. CONJUG. como realizar.
41	Soldar (v. soldar) tr. Pegar sólidamente [dos cosas] a partes de una misma cosa. 2 esp. Unir entre sí [dos partes o piezas de metal] por medio de una soldadura. 3 fig. Enmendar [un desacierto] con acciones o palabras. CONJUG. como contar.
42	Trefilar tr. Transformar en hilo o alambre [un metal] pasándolo por la hilera.
43	Troneras tr. Altroneras. Itronera [de tronar] 1: Abertura para disparar con acierto y seguridad los cañones; la ~ de un buque; la ~ de un parapeto. 2 Ventana pequeña y angosta.
44	Varar (de varo; en b. lat. varare) intr. Encollar la embarcación. 2 fig. Quedar detenido un negocio. 3 Amér. Quedarse detenido en un vehículo por avería. 4 tr. Sacar a la playa [una embarcación]. 5 desus. Echar [un barco] al agua.
45	Zarpar (tr. ant. zarpare, de orig. incierto) tr.-intr. Levar anclas; hacerse a la mar; la escuadra zarpó del puerto. 2 intr. Partir a salir embarcado.
46	Zunchar tr. Reforzar [una cosa] con zunchos.

Capítulo 7

Foco Narrativo y Foco Informativo: Dos dimensiones para una descripción de los manuales en la formación técnico-profesional

JUANITA MARINKOVICH
YANINA CADEMÁRTORI

INTRODUCCIÓN

El lenguaje escrito juega un papel fundamental en la construcción, producción y divulgación del conocimiento técnico. A pesar de su importancia, en Chile se registra un número reducido de investigaciones acerca de los textos que se leen o se utilizan en el ámbito de la enseñanza técnico-profesional.

El presente capítulo pretende ser un aporte en dicha dirección. El estudio presentado aquí es parte de una investigación mayor, en la que, siguiendo la metodología de la lingüística de corpus, se aplicó un análisis estadístico multivariado (Biber, 1986, 1988) sobre 65 rasgos lingüísticos (ver capítulos 2 y 3) en tres corpora correspondientes a tres registros: Técnico-Científico (CTC), Literario (CIL) y Entrevista Oral (CEO), pertenecientes al Corpus PUCV-2003. Producto de este análisis cuantitativo, se configuraron 5 dimensiones, denominadas Foco Contextual e Interactivo, Foco Narrativo, Foco Compromiso, Foco Modalizador y Foco Informativo (ver Capítulo 3). Estas dimensiones deben ser entendidas como interpretaciones funcionales de patrones de co-ocurrencia de rasgos lingüísticos. Posteriormente, se determinó estadísticamente un punto para cada registro (CTC, CIL y CEO) en relación con cada dimensión. Así, sobre la base del punto que cada uno de ellos alcanzó con respecto a cada dimensión, se compararon los tres corpora, caracterizándolos en un perfil de dimensiones. En el caso de las dimensiones Foco Narrativo e Informativo, los tres registros se comportan de un modo muy diferenciado: los textos del registro técnico tienen puntos relativamente altos en la dimensión Foco Narrativo y bajos en la Foco Informativo.

livos altos en la dimensión Foco Informativo; en cambio, los del registro literario y de la entrevista oral, puntuaron relativamente alto en la dimensión Foco Narrativa.

Para los propósitos de nuestro trabajo, se ha seleccionado el manual al interior del registro técnico-científico, la clase textual con mayor representatividad en ese registro (ver Capítulo 4). El manual destaca como un exponente del discurso divulgativo en el ámbito didáctico y, como tal, está construido para responder a múltiples exigencias comunicativas. En ese sentido, junto a la informatividad, expresada a través del modo de organización textual expositivo/informativo, es posible suponer la presencia en los manuales de algún grado de narratividad, expresada a través de rasgos de la dimensión correspondiente.

Dicho en términos de las dimensiones, suponemos que, a pesar de la baja relación detectada por el análisis cuantitativo entre la dimensión Foco Narrativo y el registro técnico-científico, en los manuales dicha dimensión también ocupa un lugar.

El objetivo de esta investigación es, entonces, complementar el análisis cuantitativo realizado en el marco del proyecto Fondecyt 1020/786 (ver Capítulo 4), mediante un análisis de tipo cualitativo, e indagar el modo en que se manifiestan en 24 manuales técnicos la Dimensión Foco Narrativo y la Dimensión Foco Informativo.

Especificamente, el estudio se llevará a cabo mediante el análisis de fragmentos textuales, seleccionados debido a que en ellos se manifiesta la co-ocurrencia de los rasgos lingüísticos que en cada dimensión se han revelado como altamente significativos.

1. MARCO DE REFERENCIA

El manual técnico puede ser definido como un macrogénero, analizable desde cuatro ejes: las relaciones lógico-semánticas, la estructura esquemática, las realizaciones lingüísticas y el propósito comunicativo. Este tipo de manual privilegia ciertas relaciones lógico-semánticas, tales como propósito-resultado, causa-condición, razón y contradicción. En lo que respecta a la estructura esquemática, ésta debe estar al servicio de establecer contacto con el lector y orientarlo, induciendo a la acción, anticipando y resolviendo problemas. Los rasgos lingüísticos, por su parte, permiten realizar patrones textuales de evaluación, definición, descripción, especificación, recomendación e instrucción, etc. (Lasseri, 2003).

Por otra parte, Menéndez (1999) plantea que los libros de texto o manuales de estudio son representaciones de cuáles son y cómo se enseñan las diversas materias en el ámbito escolar. Estos adquieren, en función de su importancia dentro del proceso educativo, un valor de legitimación social de los conocimientos que transmiten.

Los manuales técnicos usados en el ámbito instruccional parecen responder a un propósito general de lo que se denomina la accesibilidad, que exige de los autores un estilo claro y sencillo con el fin de permitir el acceso al saber técnico-científico a un público no especializado (Killingsworth & Gilbertson, 1992). En este sentido, la exigencia de que los manuales cumplan con este requisito explica el uso de un estilo más personal que impersonal, más concreto que abstracto (Lasseri, 2003). Es dable suponer, entonces, una explícita participación de las personas del discurso y de la deixis personal. Además, al concretarse la experiencia en un espacio y tiempo determinados, aparecen indicadores de tiempo y lugar, lo que ayuda a la contextualización de los eventos involucrados (Adam & Lorda, 1999).

En otras palabras, la accesibilidad de los manuales podría asociarse con la función divulgativa de los mismos para lo cual se deberá utilizar una serie de recursos que contribuyan a dar cuenta, de un modo didáctico, del conocimiento especializado a un lector no informado (Lasseri, 2003).

Uno de estos recursos es la narratividad. Es posible relacionar la narratividad como recurso presente en los manuales, con la modalidad narrativa de Bruner (1988), uno de los modos de explicación de la realidad, que la sitúa en un tiempo y en un espacio concreto. El orden narrativo hace comprensible los elementos individuales integrándolos al todo que éstos configuran. El proceso de ordenación opera uniendo diversos sucesos a lo largo de una dimensión temporal e identificando el efecto que un elemento tiene sobre otro (Polkinghorne, 1988).

Por su parte, Adam y Lorda (1999) constatan que la narración, si bien está dotada de organización interna propia, mantiene una relación de dependencia/independencia con otras unidades o secuencias del conjunto más amplio del que forma parte el texto. Así, pues, la dimensión narrativa surge en los manuales debido a la necesidad de apoyarse en la experiencia del mundo y en las representaciones del sujeto, que, como Jeannieret (1994) señala, permiten dar existencia a la realidad que describe, produciendo un universo imaginario cargado de la ilusión de realidad.

Además de la accesibilidad, vista, en este caso, desde la narratividad, los manuales con propósito didáctico también deben cumplir con la exigencia de la informatividad, propia de los discursos del ámbito académico. Dada su orientación referencial, los autores de los manuales técnicos integran y compactan gran cantidad de información, al tiempo que buscan la precisión y la objetividad en la entrega de la misma. En sus observaciones prima la modalidad descriptiva que marca la necesidad y certeza de los juicios expresados (Hyland, 1998). Asimismo, la impresión de objetividad se logra conformando una atmósfera impersonal, evitando la mención explícita del agente y omitiendo las referencias temporales inmediatas.

En este sentido, la informatividad puede relacionarse con la otra modalidad de funcionamiento cognitivo de Bruner (1988), la lógico-científico, de indudable presencia en los manuales. En esta modalidad, la posibilidad de explicarse el mundo proviene de la capacidad de abstraer los eventos de los contextos particulares y de descubrir relaciones que se establecen en diversos casos, conformando así una categoría al margen del contexto espacial o temporal. Por esta razón, el manual técnico privilegia la función comunicativa referencial y favorece un modo expositivo de estructuración del discurso que facilita la presentación de la información, con una organización detallada y homogénea en la que abundan los ejemplos, tablas y recursos no verbales.

En síntesis, si bien en los manuales predomina la informatividad, ésta coexiste, en momentos funcionales diferentes y en muy variada proporción, con la narratividad. Esta característica contribuye a que el manual cumpla con variadas exigencias comunicativas, de claridad y precisión, por una parte, y de concreción y accesibilidad, por otra.

2. EL ESTUDIO

2.1. Las dimensiones

Para los fines del estudio, la narratividad y la informatividad se han operacionalizado a través de lo que se ha llamado dimensión "Foco Narrativo" y dimensión "Foco Informativo". La dimensión, como ya se expresó, es una interpretación funcional de un patrón sistemático de co-ocurrencia de determinados rasgos lingüísticos de los que se inferen funciones subyacentes compartidas, es decir, se asume que un grupo de rasgos lingüísticos co-ocurren frecuentemente en ciertos textos porque estos rasgos son usados de modo recurrente para cumplir un conjunto de funciones comunicativas (Biber, 1988; Parodi, 2004).

En la dimensión Foco Narrativo co-ocurren, entre otros, los pronombres personales, el pretérito indefinido e imperfecto, los adverbios de lugar, tiempo y negación y los verbos de volición. La presencia de estos rasgos permite asociar esta dimensión con la expresión de una sucesión cronológica de eventos, principalmente en el pasado, con mención explícita de las circunstancias temporales y de lugar (Weinrich, 1974).

Por su parte, la dimensión Foco Informativo está conformada por la co-ocurrencia de verbos modales de obligación, el modo subjuntivo, nominalizaciones, participios en función adjetiva y frases preposicionales. Estos rasgos dan cuenta de la necesidad y certeza de los juicios expresados, correspondiendo fundamentalmente a una modalidad deontica (Calsamiglia & Tusón, 1999). Permiten, además, la estructuración de discursos de mayor complejidad sintáctica y la integración y compactación de información altamente abstracta, propia de un discurso académico (Longacre, 1983; Chafe, 1985; Ficilio, 1999). De ahí que esta dimensión caracterice a textos de alta informatividad, privilegiando la función referencial y prestando limitada atención a las relaciones interpersonales y la temporalidad de los eventos.

2.2. El corpus

El corpus está compuesto por 24 manuales técnicos que constituyen 463.468 palabras. Esta cifra representa el 74% del total de palabras y el 32,4% de los textos del CTC, lo que confirma la representatividad de esta clase textual en relación con todo el corpus. Como se puede apreciar en la Tabla 1, el área marítima del CTC está representada por 4 manuales con un total de 84.324 palabras (18,1% del total de palabras de los manuales); el área industrial está representada por 15 manuales con un total de 241.953 palabras (52,2% del total de palabras de los manuales); el área comercial está representada por 5 manuales con un total de 137.191 palabras (29,6% del total de palabras de los manuales).

Tabla 1. Cantidad de palabras en los manuales del CTC por área de especialidad.

	Número de palabras	Número de textos
Corpus Técnico-Científico (CTC)	626.790	74
Manuales del CTC	463.468	24
Área Marítima	84.324	4
Área Industrial	241.953	15
Área Comercial	137.191	5

3. METODOLOGÍA

Se procedió a revisar los manuales con el objeto de seleccionar fragmentos donde se presentaran los rasgos de las dos dimensiones estudiadas. Para los fragmentos narrativos, el criterio de selección consistió en que el texto estuviera orientado hacia la presentación de eventos concatenados, acaecidos en el pasado, y con la presencia de actores de esos eventos. Para los fragmentos expositivo-informativos, el criterio se basó en que el texto estuviera orientado hacia la presentación de una información acerca de un referente, sin precisiones temporales y sin participación explícita de los agentes.

En los fragmentos seleccionados, se constató la co-ocurrencia de los rasgos lingüísticos característicos de la dimensión Foco Narrativo. Correspondientemente, el mismo procedimiento se siguió para los rasgos de la dimensión Foco Informativo.

4. ANÁLISIS DEL CORPUS

4.1. La narratividad en los manuales

En 15 de los 24 manuales es posible distinguir la presencia de fragmentos narrativos. En los 9 restantes no se verificó presencia alguna de este tipo de fragmento. Un análisis más fino permite determinar que hay diferencias entre las áreas. De 4 manuales del área marítima, 3 (75%) contienen fragmentos narrativos; en cambio, de 5 del área comercial, 2 (60%) de los manuales presentan secuencias narrativas; la misma proporción se da en el área industrial, donde, de 15 manuales, 9 contienen este tipo de secuencia.

Sobre la base de los resultados del área marítima es posible conjutar que los manuales de dicha área privilegian una orientación más didáctica en la presentación de sus contenidos, supuesto que habría que verificar con un estudio más detallado.

El análisis ha permitido establecer, además, que las secuencias de foco narrativo en los manuales cumplen una importante función: aparecen sistemáticamente en partes determinadas de la superestructura del texto, como lo es el prólogo, la introducción o la presentación de los antecedentes acerca del tema que se desarrolla, complementando así la fuerza expositiva e informativa del texto.

Los siguientes fragmentos corresponden a secuencias cuya foco es narrativo. En estos fragmentos se identifican los rasgos lingüísticos de narratividad y se interpretan funcionalmente.

1) "Metro: la unidad de longitud arbitraria, base del Sistema Métrico Decimal, fue creada con el fin de unificar las diferentes unidades de longitud empleadas en los diferentes países durante el siglo XVIII y anteriores, cuya equivalencia entre sí constituye una gran dificultad. En un congreso internacional celebrado en Francia en 1799 se estableció que se tomaría como base del Sistema Métrico una unidad derivada de las dimensiones de nuestro planeta lo que impuso la necesidad de la medición de la Tierra, problema que ha sido tanteado por Aristóteles hacia el año 340 a.C. y por otros filósofos griegos y médicos árabes. De esos antiguos ensayos cabe mencionar el debido a Erastóstenes, quien fixó la equivalencia de un grado de meridiano en 094'4 estadios, valor muy aproximado a despecho de los rudimentarios elementos que se disponía" (Manual laboratorio de especialidad industrial: 3).

El ejemplo (1) da cuenta de la presencia de ciertos rasgos lingüísticos de la dimensión "Foco Narrativo", tales como los tiempos verbales en el pasado, tanto en pretérito indefinido (*estableció, impuso, fijó*) como en pretérito imperfecto (*constituyó, disponía*), circunstancias temporales [*año 340 a.C., siglo XVIII, en 1799*] y de lugar [*en diferentes países, en un congreso internacional, en Francia*], conjunción aditiva (*y*), como también la alusión a personas (*Aristóteles, Eratóstenes*) que participaron en los eventos que se presentan. Todos estos rasgos apuntan a personalizar el contenido y a traer a la memoria del lector eventos pasados que dicen relación con la creación de la unidad de medida denominada "metro". Cabe señalar que este trozo se sitúa en las primeras páginas del manual, específicamente, corresponde a los principios generales de metrología.

(2) "los armadores eran sociedades de responsabilidad limitada donde los socios tomaban tantas acciones como le permitían sus recursos, pero con el advenimiento de la Revolución Industrial y la especialización consecuente, al capital necesario para los viajes creció de tal modo que dificultó la posibilidad de que un comerciante fuese al mismo tiempo propietario de la nave, o sea, armador, separándose estas dos figuras jurídicas hacia fines del siglo XVIII" (Manual del curso taller: sistema operativo y documental del comercio exterior de Chile: 13).

El ejemplo (2) ilustra claramente la secuencia narrativa, ya que aparecen verbos en pretérito indefinido (*creció, dificultó*) y en pretéritos imperfectos (*eran, tomaban, permitían*), con sus agentes respectivos (*los armadores, los socios y el capital*).

Completan la secuencia dos estructuras que dan cuenta de la ubicación temporal de los eventos señalados (*Revolución Industrial, Hacia fines del siglo XVIII*). Este fragmento entrega un dato histórico importante para comprender las relaciones comerciales de Chile con el exterior a través del transporte marítimo.

(3) "Todos sabemos también que, en este caso, el cuerpo sigue moviéndose algún tiempo, pero finalmente se detiene. ¿Por qué? Aristóteles creía que era porque cesaba la fuerza que lo movía, mientras Galileo Galilei, científico italiano del siglo XVII, quien realizó los primeros experimentos científicos destinados a conocer el comportamiento de los cuerpos que se mueven, por el contrario, con instrumental muy simple, llegó a las siguientes conclusiones: si no se ejerce fuerza alguna sobre un cuerpo en reposo, éste permanece en reposo indefinidamente y si no se ejerce fuerza alguna sobre un cuerpo en movimiento, éste seguirá moviéndose continuamente" (Manual de mecánica: 1).

El texto (3) presenta el principio de la mecánica clásica, la *inercia*, explicándolo mediante mecanismos narrativos, tales como la sucesión de eventos en tiempos verbales en pasado (pretérito indefinido: *realizó, llegó*; y pretérito imperfecto: *creía, era, cesaba, movía*) y las personas relacionadas con estos eventos (*Aristóteles y Galileo Galilei*). Cabe mencionar también la presencia de la conjunción adversativa *pero* y de circunstancias de tiempo (*siglo XVII*), recursos que apoyan la secuencia narrativa. Este fragmento ha sido extraído de la primera página del manual, en donde se establecen a manera de introducción los principios de la mecánica clásica de Newton.

(4) "Con el objeto de tender a la unificación y conseguir un sistema de pesos y medidas, cuyo empleo se extendiese todo lo posible, se ideó el sistema métrico

decimal, que respondió perfectamente a cuanto de él se esperaba" (Manual de metrología: 1).

En el ejemplo (4), se observa nuevamente un intento por narrativizar el surgimiento de un sistema en el ámbito científico, acudiendo al empleo de las formas verbales del pretérito indefinido (*ideó, respondió*) y del pretérito imperfecto del indicativo (*esperaba*). El trozo corresponde nuevamente a la primera página del manual, poniendo el acento en el carácter introductorio del mismo.

(5) "En el decenio de 1930, el mundo atravesaba por un período de depresión económica intensa, por lo que muchos gobiernos trataron de protegerse con la creación de obstáculos al comercio de diversas formas, como: aranceles elevados, contingentes de importación, control de cambios, etc" (Manual de comercio exterior: 67).

El fragmento (5) presenta una secuencia narrativa típica que sitúa un evento en un momento determinado, apoyándose para ello en los tiempos verbales en el pasado (pretérito indefinido: *trataron* y pretérito imperfecto: *atravesaba*) y en una circunstancia temporal (*en el decenio de 1930*). Además, se personaliza la acción con un agente (*muchos gobiernos*). Este texto se ubica en el Capítulo IV del manual en su sección "Origen e Historia".

(6) "Este genial avance fue debido al empuje y visión de Malcolm P. McLean, en aquel momento un transportista de carretera norteamericano, que decidió resolver los problemas que le planteaban las regulaciones dimensionales y de peso de los diversos Estados, concibiendo una cadena de transporte integrado, inicialmente terrestre-marítima, mediante el uso de un estabilón clave normalizado, el contenedor, que podía ser transportado de unos a otros vehículos de transporte con facilidad con mínimos daños a la mercancía" (Manual estructura y operaciones de transporte aéreo, terrestre e intermodal: 112).

En el ejemplo (6), se pretende dar cuenta del origen de un elemento clave en el transporte intermodal, cual es, el contenedor. Aparece en una sección introductoria del texto como una manera de orientar al lector acerca de las circunstancias que llevaron a su creador a la implantación de este artefacto en el ámbito del transporte. Para ello, la secuencia narrativa precisa al agente mediante el nombre propio (*Malcolm P. McLean*) y el pronombre personal (*l/e*), como también la acción llevada a cabo en el pasado por dicha agente a través del pretérito indefinido (*fue, decidió*), que señala los sucesos, y, por tanto, el dinamismo de las acciones. Se complementa la temporalidad en el pasado con el pretérito imperfecto (*plantearon, podía*), que describe situaciones y circunstancias en las que las acciones se desarrollan (Koyacci, 1993).

4.2. La informatividad en los manuales

La dimensión Foco Informativo tiene una significativa presencia en los manuales arribados. Su aparición es sistemática en todos los partes de la superestructura de estos textos. Este resultado confirma el objetivo primordial de los manuales: informar y dar cuenta de un conocimiento especializado que será posteriormente utilizado en el ámbito profesional.

Los fragmentos expositivo-informativos ocurren preferentemente a continuación de títulos

y subtítulos, a modo de explicación y presentación de carácter más general del tema o del concepto que se va a tratar. En los manuales que presentan normas legales o de uso, los párrafos expositivo-informativos son menos frecuentes y aparecen especialmente en las definiciones.

Los siguientes párrafos se caracterizan por poseer una alta carga informativa, asociada a una entrega regulada de la información. Carecen de agentes que realicen acciones, las formas verbales conjugadas son escasas y el único tiempo usado es el presente que tiene un carácter atemporal. Los ejemplos permiten apreciar la relación entre la co-ocurrencia de determinados rasgos lingüísticos y el cumplimiento de propósitos comunicativos específicos.

[1] "Si se admite una repartición uniforme de las fuerzas de cohesión sobre toda la sección transversal, a cada mm² de la superficie cargada le corresponde una fuerza perfectamente definida que se puede determinar numéricamente mediante ensayos parecidos, llamados ensayos de rotura" (Alrededor del trabajo de los metales: 59).

El ejemplo [1] sucede a un subtítulo y corresponde a una definición. Presenta nominalizaciones (repartición, cohesión, sección) participios en función adjetiva (cargada, definida, parecidos, llamados) y frases preposicionales (repartición de, cohesión sobre, mm² de, ensayos de). Estos tres rasgos tienen la función común de compactar la información. Adicionalmente, las frases preposicionales precisan el significado de los términos usados. El tiempo verbal es un presente atemporal. De este modo, todo el párrafo se orienta hacia el referente que presenta.

[2] "Lo anterior se hace fundamental en una economía que tiene problemas de financiamiento. El capital extranjero es importante que llegue a un país, especialmente si se trata de efectuar obras encaminadas a la capitalización. La verdad es que la mayoría de los países en vías de desarrollo y los subdesarrollados necesitan de la ayuda externa. Esta debe ser bien empleada, es decir, deberá ser destinada a obras productivas para todo el país y no para determinados grupos" (Manual de comercio exterior: 126).

El ejemplo [2] trata de la estructura del comercio exterior chileno y aun cuando hay cierta evaluación acerca de lo que debe ser, el tono es expositivo. De ahí la presencia de nominalizaciones (financiamiento, capitalización), participios en función adjetiva (encaminadas, empleada, destinados), frases preposicionales (problemas de, mayoría de, vías de, obras para), modo subjuntivo (llegue) y verbos modales de obligación (debe, deberá). Todos estos rasgos se orientan a entregar información condensada y precisa.

[3] "Un cargador deberá completar, firmar y entregar este documento al transportista siempre que se trate de un transporte de mercaderías peligrosas. Este documento proporciona información detallada, incluyendo información referente a la clasificación de las mercaderías peligrosas de acuerdo a las normas relativas al transporte de las mismas" (Estructura y operaciones de transporte aéreo, terrestre e intermodal: 143).

El propósito del párrafo anterior es normativo, y para mayor efectividad, se expone con

precisión la manera en que deber realizarse los procedimientos descritos. Las nominalizaciones (información, clasificación) objetivizan procesos, los participios en función adjetiva (detallada, relativas, referente) contribuyen a la descripción breve y precisa; las frases preposicionales (clasificación de, transporte de) compactan la información. Por su parte, el modo subjuntivo (trate) y verbos modales de obligación (deberá) corresponden a la intención normativa.

[4] "Aduana: Servicio público encargado de vigilar y fiscalizar el paso de mercancías por las costas, fronteras y aeropuertos de la República, de intervenir en el tráfico internacional para los efectos de la recaudación de los impuestos a la importación, exportación y otros que determinen las leyes, y de generar las estadísticas de ese tráfico por las fronteras, sin perjuicio de las demás funciones que le encienden las leyes" (Manual de aduanas: 1).

El ejemplo [4] corresponde a una definición dentro de un tratado de normas legales. Sobresalen las frases preposicionales (paso de, mercancías por, tráfico internacional para, efectos de, recaudación de, los impuestos a, estadísticas de, tráfico por, perjuicio de) que permiten integrar información de modo sucinto, pero claro. Abundan las nominalizaciones (recaudación, importación, exportación, funciones) que expresan nociones de contenido abstracto y, a la vez, compactan oraciones completas. Asimismo, el modo verbal usado es el subjuntivo (determinar, encienden) que otorga a la acción verbal el significado de posibilidad eventual subordinando ideas.

[5] "El aire comprimido se transmite y distribuye a varias áreas de uso por tubería interconectada diseñada por líneas, tubos colectores, alimentadores, líneas de caño y líneas múltiples. El sistema de distribución permite que el aire corra libremente del receptor a los varios puntos de uso. Un sistema de distribución bien diseñado permite que los compresores suministren a los cilindros y otros componentes la cantidad de aire propia para la operación deseada" (Neumática: 40).

El fragmento está situado junto a una figura y bajo un subtítulo al que explica. Co-ocurren varios rasgos frecuentes en el discurso técnico y científico por su capacidad de compactar la información, por tanto, se registra una notable frecuencia de frases preposicionales (áreas de, designada por, líneas de, sistema de, receptor a, puntos de, sistema de, cantidad del y de participios en función adjetiva (comprimido, interconectada, designada, diseñado, deseada). Otra rasgo presente es la nominalización (distribución en dos ocasiones, operación), que como ya se ha señalado, presenta la terminología de la especialidad. Llama la atención la escasez de formas verbales, usándose sólo un presente atemporal, lo que reafirma el alto grado de densidad informativa del párrafo.

[6] "Las juntas realizadas entre adherentes no porosos (tales como metal o plástico), con adhesivos activados con solvente fallan debido a la deficiente evaporación del solvente. El solvente alrededor de los rebordes de la unión se estabilizan, formando un sistema de sellado que impide la evaporación posterior del solvente. Esto reduce la superficie efectiva de su unión, disminuyendo su fuerza" (Tecnología mecánica, procesos y materiales: 164).

El ejemplo [6] es un texto que acompaña una ilustración. El propósito expositivo infor-

mativo es evidente. Su ubicación en el texto exige el máximo de densidad informativa y de claridad. Por ello, coexisten los participios (*realizadas, activadas, selladas*) con frases preposicionales que desarrollan la información de modo compacto (*adhesivos [] con, evaporación de, solvente [] de, rebordes de, sistema de, evaporación [] del, superficie [] de*). Tal como sucede en los ejemplos anteriores, el tiempo es un presente atemporal.

CONCLUSIONES

Los rasgos lingüísticos en co-ocurrencia proveen pistas seguras y sólidas para determinar las funciones que los textos cumplen. Sin embargo, estos rasgos y estas funciones necesitan encontrar un constructo que los subsuma, como lo es la noción de dimensión, que justamente permite interpretar cualitativamente los patrones de co-ocurrencia en términos de funciones situacionales, sociales y cognitivas compartidas por los rasgos lingüísticos. Esto es, rasgos lingüísticos co-ocurren en los textos porque reflejan funciones compartidas. Ejemplos claros son los que encontramos en nuestra análisis. Tiempos verbales del pasado y agentes que asumen las acciones en ellos tiempos están relacionados con la narratividad; por el contrario, las nominalizaciones, frecuentemente modificadas por participios en función adjetiva o frases preposicionales, llevan la marca de la informatividad. De lo primero podemos inferir que la temporalidad en el pasado, con presencia limitada de los agentes, es el rasgo-fuerza de la narratividad en los manuales y, de lo segundo, que la nominalización, las frases preposicionales y los adjetivos de participio contribuyen a textualizar la informatividad. Este hallazgo coincide con estudios que siguen la línea de la lingüística de corpus, como aquellos realizados especialmente por Biber (1988) y Biber, Conrad y Reppen (1998). Es evidente que la lingüística de corpus con su enfoque multidimensional, aplicado en estudios sobre la lengua inglesa y otras lenguas, constituye un aporte en la búsqueda de una síntesis entre metodologías de análisis cuantitativas y cualitativas, desvirtuando así lo que la crítica señala de esta lingüística al considerarla como un planteamiento puramente cuantitativo y meramente descriptivo.

En el caso de los manuales técnicos con propósitos didácticos, nuestro estudio ha podido comprobar que las dimensiones Foco Narrativo y Foco Informativo permiten explicar aspectos aparentemente contradictorios de esta clase de textos, pero que corresponden a la realidad compleja y multidimensional de los mismos.

Las secuencias cuyo foco se centra en la narratividad y en la informatividad, en diferente proporción, ocupan un lugar determinado en la estructura de los textos analizados. Las secuencias cuya foco se centra en la narratividad tienden a ubicarse en las secciones introductorias o en las instancias en que se da cuenta del origen o de la evolución histórica del fenómeno o proceso técnico en cuestión. Por su parte, las secuencias cuyo foco se centra en la informatividad son de gran frecuencia en los manuales. Preferentemente, se insertan en el cuerpo del texto, en donde se desarrolla en toda su extensión dicho fenómeno o proceso, aunque es posible también encontrarlas en las otras partes de la superestructura.

Una implicancia pedagógica de esta constatación es la necesidad de articular lo lingüístico con lo funcional comunicativo, revalorando la importancia del estudio de la gramática como base para la mejor comprensión y producción del discurso en la comunicación cotidiana y académica. Tal como lo expresa DiTullio (1997: 14):

"... se trata de que a través de la gramática se comprenda mejor el significado de una construcción, se reconozca su inserción en el sistema de la lengua y se seleccione y defienda un análisis entre las varias propuestas alternativas".

Capítulo 8

Las estrategias de reformulación: El paso desde un texto-fuente a un texto de divulgación didáctica

JUANITA MARINKOVICH

INTRODUCCIÓN

El presente estudio parte de la premisa de que la mayoría de los textos que se utilizan en el ámbito educativo son textos que han debido adecuarse en sus contenidos para así ser más accesibles a los estudiantes. En otras palabras, estos textos que podríamos llamar de divulgación didáctica se derivan normalmente de textos originales o fuente que han sido sometidos a reformulaciones. En nuestro caso, los textos corresponden al material textual utilizado en la especialidad Operación Portuaria de la Formación Diferenciada, Sector Marítimo, del cuarto año de la Enseñanza Media Técnico-Profesional de un liceo de la ciudad de Valparaíso (Chile). Específicamente, los textos fuente son dos textos legales o, más bien, dos leyes que se han transformado en dos textos expositivos, respectivamente, que pretenden facilitar la comprensión de las mismas, convirtiéndose así en textos de divulgación. La metodología para determinar qué tipo de estrategia de reformulación se ha aplicado para convertir las leyes en textos expositivos consiste en comparar ambos textos, la ley y su reformulación, determinando si la diferencia entre ellos se debe a una estrategia de expansión, de reducción o de variación del contenido. En suma, este estudio persigue determinar hasta qué punto un texto con fines didácticos –un texto expositivo que contiene la explicación acerca de una ley– es reformulación de un texto primario o fuente, en este caso, un texto legal.

1. MARCO DE REFERENCIA

1.1 Los textos especializados

Clasificar los textos en especializados o no especializados es una cuestión que preocupa a los estudiosos del texto. Esta preocupación ha sido compartida por un numeroso grupo de lingüistas de la corriente europea de la lingüística textual y por otros de orientación funcionalista o enunciativa. Ciapuscio (1994, 2003), en Argentina, establece una exhaustiva valoración de las tipologías de las décadas de los setenta y de los ochenta y parte de los noventa. La misma autora señala que una mayor atención se ha puesto en los lenguajes especiales, entendidos como comunicación entre especialistas (Hoffmann, 1985, 1987; Sager & Dungworth, 1980; Schröder, 1991).

Una propuesta interesante es la de Gläser (1993), quien plantea un modelo multinivel para una tipología de textos especializados, estableciendo tres niveles que reflejan la dimensión situacional y lingüística de estos textos. En un primer nivel, se hace la distinción entre comunicación entre expertos y comunicación de experto a no experto. A partir de esta distinción, en un segundo nivel, entre expertos, se derivan los géneros interpersonales, informativos, directivos y didácticos y, de expertos a no expertos, los géneros didácticos, los géneros divulgativos y los géneros normativos, que contienen diferentes propósitos y exigen diferentes estrategias textuales. El tercer nivel contempla, entre expertos, en el género informativo, textos primarios, derivados, pretextos y cuastextos.

Por su parte, Goldman y Bisanz (2002) señalan que los textos especializados o científicos no constituyen un conjunto monológico, sobre todo en lo que concierne a las funciones que cumplen y a las formas que reflejan dichas funciones. En este sentido, los autores plantean como hilo conductor de una clasificación de estos textos el concepto de género, entendido, siguiendo a Swales (1990: 58), como:

“una clase de eventos comunicativos con propósitos y metas compartidas. Los géneros están situados dentro de comunidades discursivas y son creados y etiquetados por los miembros de dichas comunidades” (*la traducción es nuestra*).

Así, la comunidad de científicos se comunica mediante el género científico que tiene un carácter formativo (por ejemplo, los apuntes, las ponencias) o integrativo (por ejemplo, el artículo científico publicado en una revista de prestigio). Esta comunidad genera una literatura científica primaria.

Las comunidades asociadas a los medios masivos de comunicación y la industria editorial (periodistas, escritores especializados) se concentran en torno al género de divulgación. Aquí se hace una distinción entre lo que es de conocimiento público a través de los medios masivos con la intención de informar (comunicados de prensa, anuncios publicitarios, mensajes de servicio público) y el aprendizaje informal (artículos de primera plana, panfletos, libros de interés especial, biografías, etc.). Esta comunidad crea una literatura secundaria, ya que la información es extraída de una fuente primaria.

Por otra parte, las comunidades que tienen como propósito preparar a los individuos para acceder a la comunidad científica tanto para ocupar un lugar en la sociedad como para llegar a ser ciudadanos científicamente letreados, conforman el género didáctico. Este

género se realiza preferentemente mediante el libro de texto o manual, concebido en el contexto de los estándares curriculares fijados por los gobiernos y los debates políticos acerca de las reformas educativas imperantes. Esto no significa que otros textos no surjan de este entorno, como es el caso de los cuadernillos de ejercicios, manuales de instrucciones, documentos técnicos, etc. Los encargados de dar forma a los textos de este género pueden ser los científicos y los educadores, generando una literatura secundaria que sirve los intereses de los estudiantes.

1.2 Los textos de divulgación didáctica

Los antecedentes aportados hasta aquí permiten situar tanto los textos primarios o fuente como aquéllos reformulados con propósitos didácticos. Estos últimos son textos secundarios, originados en textos previos (Mortureux, 1985).

De acuerdo con el criterio funcional de clasificación de los géneros especializados tanto de Gläser (1993) como de Goldman y Bisanz (2002), los textos primarios mayoritariamente corresponden al género científico propiamente tal, la literatura primaria generada por expertos para expertos. Los textos reformulados, por su parte, corresponden al género de divulgación, entendido como aquél que intenta dar cuenta de la información científica primaria, ya sea a un público en general o a un público constituido por estudiantes y profesores. Si lo que orienta es el conocimiento público de la información científica, podemos denominar a este proceso como de divulgación científica propiamente tal y si, por el contrario, interesa transmitir la información científica previamente descubierta por el experto a estudiantes, nos encontramos en el ámbito de la divulgación didáctica.

Parte de los textos especializados en la educación media técnico-profesional corresponden al género de divulgación didáctica y, por tanto, deberían constituir estructuras retóricas dinámicas que puedan ser manipuladas de acuerdo a sus condiciones de uso. La naturaleza genérica de estos textos los hace ser solidarios con la metodología de la disciplina que sirven y se caracterizan por su gran heterogeneidad. Además, constituyen un apoyo a los módulos o bloques unitarios de aprendizaje que integran tanto el saber como el saber hacer (la tecnología y la práctica de taller o laboratorio). Estos textos, por tanto, deben ser comprendidos cabalmente por los estudiantes, es decir, analizados, interpretados y sintetizados en términos de la información que conllevan.

La divulgación didáctica está íntimamente ligada al proceso de reformulación y a las estrategias que se aplican para llevar a cabo dicho proceso. La reformulación es originalmente un proceso discursivo intralingüístico que permite el avance de la información en un texto mediante repeticiones, reconsideraciones de lo dicho anteriormente y retrocesos en los que se retoma total o parcialmente lo que se expresó.

Para Ciapuscio (2001), la reformulación es el procedimiento básico de producción textual que caracteriza los textos de divulgación en general y, por tanto, los textos de divulgación didáctica. La reformulación, definida como una simplificación textual tanto a nivel global como local, opera mediante tres tipos de estrategias: *expansión*, *reducción* y *varación*. En el orden global, las estrategias resuelven problemas de modalidad discursiva, función textual y de nivel de especialización. En el orden local, estas mismas estrategias solucionan principalmente obstáculos terminológicos, pero también de naturaleza gramatical.

A modo de mayor precisión, tomaremos las definiciones de Ciopuscio (2001) para explicar en qué consisten estos estrategias e ilustraremos su funcionamiento. En primer lugar, la *expansión* corresponde a la adición de contenido de naturaleza informativa o emotiva que no está presente en el texto fuente, como, por ejemplo, en el intento del divulgador por explicar el término "un compuesto", quien acude a la comparación, ausente en la fuente, "como si fuera glucosa".

La *reducción*, por el contrario, significa supresión de información no necesaria para el texto de divulgación y puede aparecer también como *condensación*. Un ejemplo es la supresión de léxico especializado, como "mecanismos moleculares" o de la expresión "se adaptan genéticamente a un compuesto". Esta reducción va la mayoría de las veces asociada a la *variación*, concebida como la paráfrasis o cambio en el contenido léxico o gramatical del texto, como también en el formato o presentación formal de la información y la modalidad enunciativa de la misma.

Un ejemplo de esta última estrategia en un nivel global, es la transformación de un texto fuente oral en un texto divulgativo escrito. El texto escrito puede considerarse una reformulación paráfrásica del primero, en tanto existe entre ambos una coincidencia en el tema. Sin embargo, los mismos textos pueden variar en cuanto a la función textual, uno puede intentar explicar un método a especialistas y, el otro, informar al gran público acerca de una investigación y persuadirlo de su importancia. Como ejemplo en un nivel local, se presenta el cambio de la expresión de un científico "lo interesante de es/estar esto acá" por la expresión más emotiva del divulgador "lo más fascinante de todo esto", lo que muestra la emergencia del rasgo afectivo de la divulgación.

Por otra parte, estas estrategias pueden subsumirse bajo lo que Jeanneret (1994) denomina la divulgación como *recontextualización*, cuya planteamiento ha sido desarrollado y expresado en un modelo por Cassany, López y Martí (2000).

En la Figura 1 se representa el proceso de transformación del conocimiento científico en uno de tipo divulgativo. Las fuentes de este último son el discurso científico y el general, pero lo que va a ser transformado es un conocimiento científico que se entenderá como una red intrincada de conceptos especializados (Neopponen, 1997). En este sentido, la *recontextualización* es un concepto más abarcador, pero más afinado, a la vez, que la *reformulación*, ya que limita la red conceptual del discurso científico y los recursos expresivos de éste, buscando redes y contextos más generales (*re-conceptualizar*), usando rasgos discursivos más apropiados a la divulgación (*re-textualizar*) y elaborando formas acordes con el género divulgativo (*denominar*).

En otras palabras, una tarea cognitiva principal que debe acometer quien divulga consiste en reelaborar la red conceptual del conocimiento científico, realizando dos tareas: la *reducción* o limitación de las conexiones entre los conceptos y la inclusión de vínculos entre conceptos especializados y no especializados. Cassany, López y Martí (2000) ilustran este proceso con el concepto de "transgénico", que en el ámbito científico está asociado con otros conceptos especializados, tales como "genoma", "DNA recombinante", "composición cromosómica", etc., en cambio, en el discurso divulgativo, el mismo concepto se relacionará con "dieta", "salud", "industria agroquímica", "comercialización", etc.

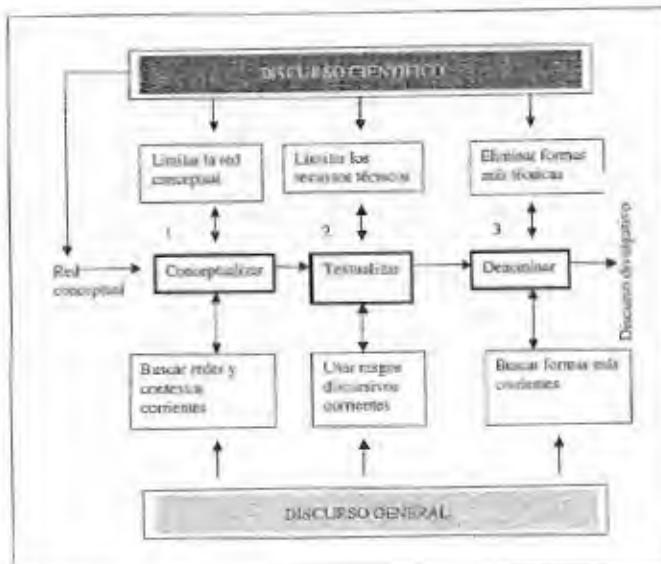


Figura 1. Del discurso científico al divulgativo (Cassany et al. 2000)

Una vez reelaborada la red conceptual, procede decidir acerca de su textualización, recurriendo a una serie de mecanismos expresivos que resultan de la negociación entre los distintos géneros que están intercambiando el científico, el general y el divulgativo. Por ejemplo, si la divulgación se canaliza a través de la prensa escrita, en géneros informativos, los recursos expresivos están condicionados por ese contexto. Finalmente, otra exigencia divulgativa la constituye la *denominación de conceptos*, esto es, la elección de las formas lingüísticas para referirse a cada concepto que debe ser transmitido, lo que se relaciona con la terminología específica de cada disciplina y el esfuerzo por hacerla accesible al leño.

En suma, Cassany (2001) señala que la *recontextualización* supone decisiones en distintos niveles. En primer lugar, en el nivel de género, pasamos de los artículos científicos, protocolos, etc., a los artículos periodísticos, manuales educativos, etc. En segundo lugar, los objetivos varían desde incrementar el conocimiento, formalizarlo o convencer o informar, entretejer o instruir. En tercer lugar, desde el punto de vista del conocimiento, un conocimiento científico se transforma en un conocimiento general o divulgativo. Por último, los interlocutores en la comunidad científica son los científicos y en el ámbito de la divulgación, el periodista o el divulgador para un público amplio.

Chevallard (1991), por su parte, utiliza el término *transposición didáctica*, un proceso complejo de transformación por el cual el conocimiento erudito se constituye en conocimiento u objeto susceptible de ser enseñado y éste a su vez es objeto de enseñanza o conocimiento enseñado. Esto significa que nos enfrentamos a dos instancias textuales: una, la del texto original o primario, y otra, la del texto intervenido, secundario o de divulgación. La versión secundaria es un texto reformulado puesto que el núcleo del contenido textual ha sido formulado en otro ámbito discursivo y a través de otras clases textuales, propias de ese

ámbito. Para Foucault (1970), esta versión podría acercarse a lo que el autor denomina el *comentario*, es decir, el destase entre el primer y segundo texto, con comentarios solidarios, aunque el segundo tiene el cometido de decir en forma definitiva lo que solo está articulado en el primero.

En este sentido, es también pertinente incorporar lo que Bernstein (1993) señala a propósito del discurso pedagógico al considerarlo como un discurso recontextualizado, ya que se ha seleccionado y descolocado el contexto primario del discurso para recolocarlo en un contexto secundario de la reproducción del discurso.

2. EL ESTUDIO

Tal como se enunciara previamente, el objetivo que persigue este estudio es identificar el tipo de estrategia de reformulación que se utiliza en dos textos expositivos de divulgación didáctica que pretienden informar a los estudiantes de la enseñanza media técnico-profesional acerca de dos leyes relacionadas con el sistema de exportaciones en Chile.

El corpus está constituido, por una parte, por textos legales que corresponden a dos textos primarios o fuente, a saber, la Ley N° 18.480 que establece el sistema simplificado de reintegro a exportadores (publicada en el Diario Oficial de Chile en 1985) y la LEY N° 18.708, que norma el sistema de reintegro de derechos y demás gravámenes aduaneros que indica en favor de los exportadores (publicada en 1988). Por otra parte, está conformado por textos expositivos, textos reformulados, que hemos denominado *glosa legal*, documentación que respalda el Módulo Documentación en la Operación Portuaria, obligatorio de la Especialidad Operación Portuaria de la Formación Diferenciada, Sector Marítimo, del cuarto año de la enseñanza media técnico-profesional en Chile.

La ley 18.480 contiene 9 artículos, de los cuales cuatro, el 2, 3, 5 bis y 6, están subdivididos por letras y con un listado de mercancías. Los artículos 4 y 5 aparecen con bis. La ley 18.708 contiene 9 artículos y dos transitorios. Solo el artículo 2 se presenta con subdivisiones marcadas con letras.

Las leyes tienen una estructura normativa, puesto que su función primordial es la de regular diversas situaciones de la vida individual y social. Como obra humana, la ley no puede escapar totalmente a la subjetividad, pero es posible reducirla gracias a la utilización de un lenguaje especializado adecuado. Cuando nos referimos a especializado, cabe la mención a lo que Cabré, Doménech, Morel y Rodríguez (2000:18) entienden por texto especializado, es decir, aquellos textos que transmiten:

“...un conocimiento que se ha codificado en relación con un esquema exacto pre establecido. Por lo tanto, el factor que determina el carácter general o especializado no es el contenido por sí mismo, sino la manera como se conceptualiza dicho contenido.”

El lenguaje legal, es decir, el lenguaje que se emplea en la redacción de normas y de leyes tiene dos tipos de destinatarios. Por un lado, los ciudadanos comunes que, al igual que los receptores de la documentación administrativa, exigen a los textos judiciales una garantía de inteligibilidad; y, por otro, los profesionales del derecho que además de destinatarios son usuarios de dichos textos.

Son muchos los estudios publicados sobre técnica legislativa y redacción de leyes (Duarte & Martínez, 1995). Entre los aspectos analizados, podemos encontrar elementos relativos a la estructura organizativa del texto (distribución del contenido de las disposiciones finales de la ley -transitorios, adicionales, finales-, recurso a definiciones al inicio de la parte dispositiva, empleo de títulos para cada artículo) y otros elementos de orden estilístico y lingüístico. Entre los factores lingüísticos, cabe señalar, en español, el uso del futuro, basándose en la posición de quien redacta o aprueba la norma y pensando en sus efectos futuros, aunque también aparece el presente y la perifrasis de obligación (“debe presentar”). Por otra parte, los artículos se numerarán en cardinales arábigo; de haber uno solo, este se indicará como ‘artículo único’.

Por su parte, se entiende por glosa legal o “la aclaración o comentario a una ley. Su función primordial es referencial. Su estructura predominante es la expositiva” (Parodi & Gramajo, 2003: 218). El carácter referencial y expositivo de la glosa legal la hace, por un lado, dar cuenta de hechos, cosas o ideas de la realidad o del contexto en que se inserta y, por otro, informar o exponer acerca de un tema en forma objetiva para así facilitar la comprensión.

2.1 Metodología

La metodología de análisis consiste en enfrentar los textos legales con su correspondiente glosa legal, identificando las diferencias y similitudes entre ellos en cuanto a la información contenida en cada texto. Una vez determinadas dichas diferencias y similitudes, se procede a tipificarlas según la estrategia de reformulación utilizada en cada caso. Cabe destacar que entre ambos textos debe existir una coincidencia en la temática o núcleo conceptual.

Las categorías de análisis aplicados corresponden a las estrategias de reformulación según la clasificación de Crapuccio (2001), expuesta anteriormente, a saber, expansión, reducción y variación. Se complementa esta clasificación con la estrategia de repetición.

A manera de síntesis, la expansión se refiere a la adición de información en el texto de divulgación didáctica que no aparece en el texto original o fuente. En el caso de la reducción, ésta se define como la sucesión de la información del texto original en el texto divulgativo. La reducción puede también derivar hacia la condensación o resumen de la información del original en el texto secundario. La variación, por su parte, corresponde al cambio léxico-gramatical o conceptual, como también de formato o de función, entre el texto fuente y el texto meta. Finalmente, la repetición consiste en mantener la información en forma literal del original en el texto de divulgación.

La unidad de medida para identificar cada estrategia será el segmento de texto que la contenga, salvo en el caso de la variación en el formato y en la función textual, que son aspectos textuales más globales y que solo se detectan por su presencia o ausencia.

3. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

Para el análisis de las estrategias de reformulación detectadas en las glosas legales, se presentan, en primer lugar, los resultados obtenidos a nivel local. En la Tabla 1, a continuación, podemos apreciar los porcentajes de aquellas estrategias que han podido ser cuantificadas.

Tabla 1. Número y porcentaje de estrategias de reformulación utilizadas en las glosas legales

Estrategias de reformulación	Glosa legal 1	Glosa legal 2
Expansión	0 (0%)	0 (0%)
Reducción	51 (63,7%)	31 (68,8%)
Variación	10 (12,5%)	5 (11,1%)
Repetición	19 (23,7%)	9 (20%)
Totales	80 (100%)	45 (100%)

Un análisis de las estrategias utilizadas permite determinar, según la Tabla 1 ya desplegada, que en ambas glosas legales predomina la reducción de información, aspecto que el autor del texto reformulado considera importante para los estudiantes que se inician en el conocimiento de los documentos legales relacionados, en este caso particular, con la operación portuaria. Sigue en una proporción menor la repetición literal de la información y la variación. Por último, la expansión no se presenta en ningún caso.

En cuanto a la estrategia de reducción, cabe señalar que en el caso de la glosa legal 1, se elimina información de la Ley 18.480, que establece un sistema simplificado de reintegro a exportadores, en lo que dice relación con el organismo que la publica, el valor de los productos exportados, los insumos nacionalizados que dan derecho al exportador al reintegro, las personas que determinan el valor neto de los insumos, el beneficio del reintegro, las mercancías excluidas y el lugar de presentación de la solicitud de reintegro. En resumen, de los nueve artículos originales se suprime completamente el 3, el 4, el 4 bis, el 8 y el 9. Los artículos 1, 2, 5, 5 bis, 6 y 7 contienen supresiones parciales y sus contenidos aparecen condensados en forma de subtitulos, manteniendo en estos casos el contenido de la ley original. Este último data de cuenta de la estrategia de repetición del texto primario.

La glosa legal 2, por su parte, reformulación de la Ley 18.708, que establece el sistema de reintegro de derechos y demás gravámenes aduaneros en favor del exportador, presenta la supresión de los datos de publicación del proyecto de ley, como también de una serie de definiciones especializadas, las personas responsables de determinar el reintegro, cómo se determinará el reintegro y qué sucede con las personas que venden a las Zonas Francas mercancías nacionales. En síntesis, en la glosa legal 2 se prescinde de los artículos 2, 3, 5, 8, 9 y los dos artículos transitorios. Los artículos 1, 4, 6 y 7 se mantienen casi literalmente en el texto de divulgación didáctica.

En lo que respecta a la estrategia de variación en las dos glosas legales, ésta no implica una reformulación profunda, ya que opera solo en un nivel superficial, como, por ejemplo, los desplazamientos en las formas verbales del futuro a formas en el presente, formas imperativas a formas en el presente, oraciones simples a subordinadas, pasiva con ser a pasiva con se y cambios en los conectores. En lo que respecta al léxico propiamente tal, no se observa un proceso de reconceptualización, puesto que los términos especializados se mantienen tal como aparecen en los textos originales, salvo en el caso de la palabra "reintegro" que se explica en la glosa legal 2 a través del sinónimo "devolución", que se coloca entre paréntesis.

A modo de ilustración, las tablas 2 y 3, respectivamente, muestran ejemplos de la estrategia de variación en las dos glosas legales en relación con las leyes fuente. En ocasiones, la

variación está también asociada a la condensación del contenido del texto fuente, tal es el caso del ejemplo 4 y 7 en la Tabla 2, que constituyen a su vez subtitulos de la glosa legal 1.

Tabla 2. Ejemplos de la aplicación de la estrategia de variación en la glosa legal 1

Ley 1 (texto fuente)	Glosa legal 1 (texto reformulado)
Establécese	ley que establece
El reintegro será	Establece que el reintegro será
Podrán acceder	Pueden acceder
Las tasas de reintegro aplicables a las mercancías definidas en los incisos precedentes, así como los tramos de exportaciones afectas a dichas tasas, serán las siguientes:	Tasas de Reintegro y Tramos de Exportaciones afectas a dichas Tasas (Subtítulo)
Se considerarán	Se considera
Del mismo modo	Además
No podrán accederse al sistema de reintegro simplificado	Restricciones para Acogerse al Sistema de Reintegro Simplificado (Subtítulo) No pueden accederse al Sistema de Reintegro Simplificado:
Tampoco tendrán	No tendrán
los exportaciones de mercancías que se clasifican en la Subposición 41.01.00.00, 41.01.02.00 y 41.01.03.00 del Arancel Aduanero.	Las exportaciones de mercancías que se clasifican en las subposiciones 74.01.05.00, 41.01.01.00, 41.01.02.00 y 41.01.03.00 del Arancel Aduanero.
Todo aquél que perciba	El que perciba

Tabla 3. Ejemplos de la aplicación de la estrategia de variación en la glosa legal 2

Ley 2 (texto fuente)	Glosa legal 2 (texto reformulado)
Para todos los efectos legales	Para este beneficio
Los servicios prestados al exterior	Se considerará exportación los servicios prestados al exterior
serán consideradas exportación	
En este caso	Si embargo
Servicio exportado	Servicio prestado
Establecida por esta ley	Señalado en esta ley

El hecho de que en la glosa legal 2 aparezca una menor cantidad de estrategias de reformulación se debe a la menor extensión de dicho texto. Como una manera de explicar este comportamiento, en la Tabla 4 podemos observar el número de palabras de los textos fuente como también el de los textos reformulados. Asimismo, se incorpora a modo de resumen el número de palabras que han sido suprimidas en los textos reformulados, como también el número de palabras que corresponden a las reformulaciones y a la información que se repite desde las leyes fuente a las glosas legales.

Tabla 4. Comportamiento de las leyes y las glosas legales en cuanto al número de palabras que contienen

Textos	Total de palabras	Palabras repetidas	Palabras reformuladas
Ley 1	2824 (100%)		
Ley 2	1424 (100%)		
Glosa legal 1	727 (25,7%)	689 (24,3%)	68 (2,4%)
Glosa legal 2	315 (22,1%)	258 (18,1%)	15 (1,05%)

En la Tabla 4, podemos apreciar, en primer lugar, que el número de palabras entre la ley 1 y la glosa legal correspondiente se disminuye fuertemente, quedando esta última reducida en un 25,7% de las palabras de la primera. La glosa legal 2 se comporta en forma similar a la glosa legal 1 con respecto al texto fuente, disminuyendo el número de palabras en un 22,1% del texto fuente. Asimismo, es significativo constatar en estas otras el porcentaje de palabras repetidas desde el original al texto reformulado en ambos casos (24,3% y 18,1%, respectivamente), como también la cantidad mínima de palabras reformuladas propiamente tales (2,4% y 1,05%, respectivamente). En el anexo 1, se presentan los textos originales con las modificaciones marcadas para así visualizar los cambios operados.

Ahora bien, un análisis cuantitativo de los resultados a nivel global permite establecer que la estrategia de reformulación que destaca es la variación y reducción, ya que las glosas legales rompen con la estructura típica de una ley en términos de artículos y letras y el comportarse como textos expositivos hacen uso de subtitulos y viñetas. Además, condensan y redistribuyen la información original. Esta transformación que afecta tanto la estructura del texto como su función es lo que en términos de Cassany, López y Martí (2000) se denominó textualización, en donde el contexto comunicativo en que se enmarca el texto divulgativo impone sus limitaciones de género. En los formatos siguientes se observa cómo los textos normativos de ambas leyes se transforman en textos expositivos con una organización diferente a la de los textos originales.

GLOSAS LEGAL 1 (texto reformulado 1)

«Definición
[ley que establece un sistema simplificado de reintegro de gravámenes...]

Merancías de Origen Nacional
Se considera mercancías de origen nacional...

Merancías que pueden acceder al Beneficio

Podrán acceder al beneficio que regulan las normas establecidas en la legislación...

Tasas de Reintegro y Tramos de Exportaciones afectos a dichas Tasas

- 10% para aquellas mercancías que, durante 1990, se hubieren exportado...
- 5% para que las mercancías que, durante 1990, se hubieren exportado...
- 3% para aquellas mercancías que, durante 1990, se hubieren exportado...

Restricciones para Acogerse al Sistema de Reintegro Simplificado

- No pueden acogerse al Sistema de Reintegro Simplificado...
- Las exportaciones de mercancías que, luego de su introducción en el país, dentro de los últimos doce meses...
- Las exportaciones que...

Plazo para Solicitar Beneficio

El plazo para solicitar el beneficio ante el Servicio de Tesorerías es de 120 días.

Sanciones

El que perciba indebidamente el reintegro señalado en esta ley, proporcionando...

Ley 18.480 (texto fuente 1)

«Art. 1. Establecerse un sistema simplificado de reintegro de gravámenes...»

Art. 2. Podrán acceder al reintegro establecido en el artículo 1º, todos los mercancías exportadas integralmente en el país con insumos nacionales...

Art. 5 bis. No podrán acogerse al sistema de reintegro simplificado...

Art. 6. El plazo para solicitar el reintegro...

Art. 7. Todo aquél que perciba indebidamente el reintegro señalado en esta ley...

"Art. 1. Los exportadores podrán obtener el reintegro de los derechos y demás gravámenes aduaneros pagados, respecto de materias primas, artículos o media elaboración, partes o piezas importadas por el propio importador o portadores...

Para todos los efectos legales, los servicios prestados al exterior serán considerados exportación...

Art. 4. No podrá solicitarse reintegro con cargo a declaraciones de importación de más de dieciocho meses... A su vez, el reintegro deberá imponerse, a lo más dentro del plazo de nueve meses...

Art. 6. Los exportadores que se acojan al beneficio establecido por esta ley no podrán, por un mismo producto, imponer el sistema simplificado...

Art. 7. El que fraudulentamente perciba el reintegro señalado en esta ley será sancionado..."

"Alcances:

Los exportadores podrán obtener el reintegro [devolución] de los derechos y demás gravámenes aduaneros pagados, respecto de materias primas, artículos a media elaboración, partes o piezas importadas por el propio importador o portadores...

Para este beneficio, se considerarán exportación los servicios prestados al exterior...

Plazo Para Solicitar el Beneficio

No podrá solicitarse reintegro con cargo a declaraciones de importación de más de dieciocho meses. A su vez, el reintegro deberá imponerse, a lo más dentro del plazo de nueve meses...

Montos a Solicitar de Beneficio

No podrá solicitarse reintegro por un monto inferior a (100) cien dólares...

Restricciones

Los exportadores que se acojan al beneficio de esta ley no podrán por un mismo producto, imponer el sistema simplificado de reintegro de gravámenes...

Sanciones

El que fraudulentemente perciba el reintegro señalado en esta ley será sancionado...

Si comparamos los resultados obtenidos con reformulaciones en otros ámbitos del conocimiento, como aquella que da cuenta del paso desde las entrevistas orales con un científico y su correspondiente reformulación en una noticia de divulgación escrita (Crapuscio, 2001) o la de la divulgación de la enfermedad de las vacas locas en España (Cassany & Martí, 2000), las diferencias en la riqueza de recursos es significativa sobre todo en el plano léxico, ya que en ambos trabajos el problema u obstáculo terminológico requiere acudir a procedimientos paráfrácticos que en el caso que nos ocupa no parecen ser tan importante de aplicar.

Planteamos que las leyes analizadas, si bien abordan un tema específico, no implican una reelaboración profunda para ser comprendido por los usuarios. Por ejemplo, términos que aparecen relacionados con la temática de las 'exportaciones', que es común a ambas leyes, tales como 'mercancías de origen nacional' o 'tasas de reintegro', exigen solo destacarlos mediante algún procedimiento formal en el texto reformulado, suprimiendo toda aquella información que no aporta precisión al término en cuestión. Tal es el caso particular de la supresión en la glosa legal 2 de las definiciones de una serie de términos que se mencionan al inicio del texto ("materia prima", "artículos a media elaboración", "partes" o "piezas"), que en la Ley 18.708 están contenidas en el artículo 2. Esta decisión podría explicarse por lo que Hurtado (2001) plantea cuando se refiere a que existen lenguajes especializados de diversos tipos y con diferentes grados de especialización: algunos presentan un alto grado de especialización, como lo es la Física, la Biología, etc., otros ocupan un terreno más intermedio, como el de la Barca y el Derecho y otros se sitúan en la frontera de la lengua común, como los Deportes. Los diferentes grados de especialización tendrían, por tanto, su correlato en distintos niveles de divulgación o reformulación, según las áreas del conocimiento. Estos niveles o grados de reformulación afectarían también la divulgación didáctica.

CONCLUSIONES

Los procesos de reformulación del discurso especializado en uno divulgativo general constituyen una base para el estudio del discurso divulgativo didáctico, permitiendo determinar las diferencias y similitudes entre ambos procesos. No obstante, es necesario profundizar en qué medida los distintos niveles de especialización del discurso primario afectan los niveles de reformulación del discurso secundario y en qué medida estos niveles de reformulación están influidos por factores extratextuales.

La respuesta a estas interrogantes podría consolidar una clase textual denominada "reformulación didáctica" que, como dice Crapuscio (2001), representaría una "solución preformada" a problemas recurrentes en la producción de textos; y, agregamos, en los procesos de comprensión de discursos especializados en la enseñanza.

Estamos seguros que la reformulación es un proceso complejo y que está fuertemente ligado a subprocesos que, en el caso de la divulgación didáctica, no parecen estar lo suficientemente asumidos o, nos atrevemos a decir, sistematizados. No cabe duda que el problema es de naturaleza cultural y discursiva. La divulgación didáctica, por una parte, al igual que todo tipo de divulgación, está sometida a las características que le imponen las distintas disciplinas, los propósitos, los autores y los contextos. Por otra parte, la divulgación,

como cualquier forma discursiva, es el resultado de la negociación entre los participantes de la misma, es decir, aquellos que detentan el discurso especializado y aquellos que tienen que transformarlo en discurso divulgativo.

Conseguir que, en nuestro caso, los estudiantes accedan al discurso especializado de las distintas disciplinas de la formación técnico-profesional requiere de los divulgadores un conocimiento de los procesos de reelaboración del texto fuente que se centra principalmente en las redes conceptuales especializadas y a partir de ellas buscar redes y contextos corrientes que permitan una adecuada textualización y denominación de las mismas. La experiencia lograda en el presente estudio a partir del enfrentamiento de las leyes con las glosas legales permite adelantar que existe un conocimiento intuitivo por parte del divulgador de que el texto original debe sufrir ciertas reducciones o reformulaciones para poder ser trabajado por el estudiante, aunque esta tarea no siempre se alcanza con los logros deseados. Prueba de ello es una reformulación que se limita a la reducción de información, con escasa variación o nula expansión de la misma, transformando el texto en un producto opaco que no logra cumplir en toda su extensión con el propósito divulgativo esperado.

Finalmente, la reformulación está intimamente ligada a una serie de intervenciones a las que se ven expuestos los discursos de divulgación didáctica. Una primera intervención está presente en el proceso de conversión de un discurso fuente en un discurso destinado a lectores no especializados en las temáticas abordadas. Una segunda intervención se opera en los discursos ya generados en la primera intervención, aplicando para ello criterios pedagógicos, tales como secuencia de aprendizaje, contenidos de los programas de estudio vigente y accesibilidad del texto. Por último, una tercera intervención está relacionada con aquella que se realiza en el aula, es decir, aquella del docente a cargo de la enseñanza de los distintos contenidos disciplinares.

ANEXO 1

Leyenda

Cursiva	Información suministrada en el texto reformulado
Negrita	Información que permanece literal
Subrayado	Información reformulada

TEXTO 1: LEY N° 18.480

Ministerio de Hacienda

LEY N° 18.480

ESTABLECE SISTEMA SIMPLIFICADO DE REINTEGRO A EXPORTADORES

(Publicada en el DO de 19.12.85)

La Junta de Gobierno de la República de Chile ha dado su aprobación al siguiente Proyecto de ley:

Art. 1. Establecése un sistema simplificado de reintegro de gravámenes que inciden en el costo de los insumos de las exportaciones menores no tradicionales.

El reintegro será de 10 %, de 5% o de 3% de los correspondientes productos exportados de origen nacional, de acuerdo con las normas que más adelante se indican. Para este efecto se entenderá como valor de los productos exportados el valor FOB de los respectiva mercancía, excluidas las comisiones y cualquier otro gasto deducible en el resultado final de la operación de exportación, en dólares de Estados Unidos de América, al tipo de cambio establecido en el artículo 22 de las Ordenanza de Aduana.

Los insumos nacionales o nacionalizados que se destinan a la fabricación, elaboración o producción de bienes exportados, aunque estos últimos se encuentren excluidos o se excluyan del beneficio de reintegro, si se consuman en la producción de éstos, darán derecho al exportador, respecto de dichos insumos, al reintegro a que se refiere el inciso anterior sobre el valor neto facturado, una vez cumplida la exportación. Los Directores Regionales o Administradores de Aduanas determinarán, a petición del exportador, el valor neto de los insumos incorporados o consumidos directamente en la producción del bien exportado, de acuerdo con los requisitos, modalidades, procedimientos y sistemas de control que lleve el Director Nacional de Aduanas. El reintegro al respectivo insumo será incompatible con cualquier sistema devolutivo de derechos y gravámenes aduaneros, como asimismo, no tendrán acceso a este beneficio aquellas mercancías que no pueden acogerse al régimen establecido en la ley.

NP 18.708. Los insumos nacionalizados a que hace referencia este artículo deberán cumplir con las exigencias establecidas en el inciso primero del artículo 5 para aquellas mercancías que no elaboradas son íntegramente en el país con insumos nacionales.

El beneficio del reintegro regirá respecto de todas las mercancías susceptibles de acogerse a esta ley hasta la fecha de vigencia de la lista que las excluya en conformidad a los artículos 2, 3 y 4.

los exportadores podrán renunciar, en todo o parte, al beneficio que establece esta ley, debiendo dejar expresa constancia de dicha renuncia en la Declaración de Exportación.

Art. 2. Podrán acceder al reintegro establecido en el artículo 1º todas las mercancías exportadas de origen nacional, clasificadas en el Arancel Aduanero en la fecha de aceptación a trámite de la declaración de exportación emitida por el Servicio Nacional de Aduanas y que, al 31 de diciembre de 1990, estuvieren afectas a este beneficio.

También podrán acceder a esta ley aquellas mercancías exportadas de origen nacional, clasificadas en el Arancel Aduanero en la fecha de aceptación a trámite de la declaración de exportación emitida por el Servicio Nacional de Aduanas y que, al 31 de diciembre de 1990, se encontraren excluidas del reintegro, siempre y cuando el monto exportado por partida arancelaria, según su clasificación en la fecha de aceptación a trámite, haya sido, en 1990, igual o menor al valor FOB de US\$ 5.000.000, moneda de los Estados Unidos de Norteamérica, según lo certifique el Servicio Nacional de Aduanas.

Las tasas de reintegro aplicables a las mercancías definidas en los incisos precedentes, así como los tramos de exportaciones afectas a dichas tasas, serán los siguientes:

a) Diez por ciento para aquellas mercancías que, durante 1990, se hubieren exportado, por partida arancelaria, por montos iguales o inferiores a US\$ 10.000.000, valor FOB en moneda de los Estados Unidos de Norteamérica, según lo certifique el Servicio Nacional de Aduanas.

b) Cinco por ciento para aquellas mercancías que, durante 1990, se hubieren exportado, por partida arancelaria, por montos superiores a US\$ 10.000.000, pero iguales o inferiores a US\$ 15.000.000, valores FOB, en moneda de los Estados Unidos de Norteamérica, según lo certifique el Servicio Nacional de Aduanas.

c) Tres por ciento, para aquellas mercancías que, durante 1990, se hubieren exportado, por partida arancelaria, por montos superiores a US\$ 15.000.000, pero iguales o inferiores a US\$ 18.000.000, valores FOB, en moneda de los Estados Unidos de Norteamérica, según lo certifique el Servicio Nacional de Aduanas.

Los montos de exportaciones señalados en las letras a), b) y c) se readjutarán anualmente de acuerdo con el índice de precios promedio relevante para el comercio exterior de Chile, según lo certifique el Banco Central de Chile, tomando como base el año 1990, y servirán para fijar la lista anual de exclusiones que dispone el artículo 3º.

Art. 3. Anualmente antes del 31 de marzo, mediante decreto expedido por el Ministerio de Economía, Fomento y Reconstrucción, y suscrito, además, por el Ministerio de Hacienda, se fijará una lista de las mercancías excluidas, clasificadas según las posiciones arancelarias vigentes en la fecha de confección de la misma, que estará constituida por:

a) Aquellas mercancías que, de acuerdo con el artículo 2º inciso segundo, no accedieren a los beneficios de esta ley.

b) Aquellas mercancías que, en conformidad con el artículo 2º, quedaren marginadas de las tasas de 10%, 5% y 3% de reintegro, por haber superado, en el año calendario anterior los límites de US\$ 10.000.000, US\$ 15.000.000 y US\$ 18.000.000, en moneda de los Estados Unidos de Norteamérica, debidamente readjustados según la norma del artículo 2º, inciso final.

Asimismo, a la referida lista se podrán incorporar:

1) Aquellas mercancías que correspondieren a proyectos de inversión que hayan sido diseñados para producir exportaciones que excedan los US\$ 10.000.000, en moneda de los Estados Unidos de Norteamérica, debidamente readjustados de acuerdo con lo establecido en el inciso final del artículo 2º.

dos Unidos de Norteamérica, debidamente readjustados de acuerdo con lo establecido en el inciso final del artículo 2º.

2) Aquellos materiales primos o insumos que constituyeron el componente principal de productos exportados no acogidos al sistema establecido en esta ley. Para que opere lo anterior, deberá existir una solicitud fundada, presentada ante el Ministerio de Economía, Fomento y Reconstrucción, en la que se demuestre que el valor de la materia prima o de los insumos para los cuales se solicita la exclusión de este beneficio, constituye un componente del valor FOB del producto final exportado no inferior al 10%. Además, para que proceda el retiro de la materia prima o del insumo del beneficio que otorga esta ley, en el o en los productos finales exportados deberá haberse utilizado, individual o colectivamente, al menos el 20% de las ventas internas de esta materia prima o insumo en el mercado nacional, durante los últimos dos años calendario anteriores a la fecha de la solicitud.

3) Aquellas mercancías cuya posición arancelaria, vigente en el momento de confección de la lista, no alcanzare en el promedio de los tres últimos años calendario, un incremento en los montos exportados, debidamente readjustados, conforme al inciso final del artículo 2º, igual o superior a 1,5 veces el crecimiento promedio del Producto Geográfico Bruto en el mismo período.

Todas las declaraciones de exportación correspondientes a mercancías incorporadas en la lista de exclusiones, tendrán derecho a percibir el reintegro cuando hayan sido aceptadas a trámite por el Servicio Nacional de Aduanas con anterioridad a la publicación del decreto que las excluya.

Art. 4. No obstante lo dispuesto en el artículo 3º, mediante decreto supremo fundado, expedido por intermedio del Ministerio de Economía, Fomento y Reconstrucción, en la forma indicado en el inciso primero del artículo 2º, se podrá ampliar, en cualquier tiempo, las listas de exclusión del beneficio de esta ley, respecto de una determinada mercancía, cuando se acrediten las causales establecidas en los incisos segundo, número 1 y 2 del artículo 3 para incorporar nuevas exclusiones.

Art. 4, bis Para el efecto de la confección de las listas de exclusión a que se refieren el artículo 3 de esta ley, la Dirección Nacional de Aduanas informará anualmente al Ministerio de Economía acerca de los montos de aquellos insumos exportados, con indicación de su respectiva posición arancelaria, por los cuales se haya reconocido el derecho al beneficio del reintegro en conformidad a lo dispuesto en el inciso tercero del artículo 1. Para determinar la exclusión de los insumos aludidos, del beneficio del reintegro, el Ministerio de Economía, Fomento y Reconstrucción sumará los montos respectivos informados por la Dirección Nacional de Aduanas.

Art. 5. Se considerarán mercancías de origen nacional, para los efectos de esta ley, las elaboradas íntegramente en el país con insumos nacionales, así como aquellas en cuya elaboración se utilicen materias primas, artículos a media elaboración, o partes o piezas importadas cuyo valor CIF no supere el 50% del valor FOB del producto a exportarse y que en su elaboración hayan experimentado una transformación que les confiera una nueva individualidad, clasificada en el Arancel Aduanero en una posición diferente a la del o de los componentes importados.

Del mismo modo se considerarán de origen nacional los productos de los reinos mineral, vegetal y animal extraídos, cosechados o recolectados, nacidos y criados, incluyendo los de la caza y la pesca capturados dentro del territorio nacional.

En las mercancías cuya calificación de origen sea de difícil determinación, el Servicio de Teso-

terías podrá definir previamente aquellas a las que se les exigirá un certificado de origen, emitido por el organismo competente que, para estos efectos, se establecerá por decreto del Ministerio de Economía, Fomento y Reconstrucción suscrito, además, por el Ministerio de Hacienda.

Art. 5. bis No podrán accogese al sistema de reintegro simplificado:

a) las exportaciones de mercancías que tengan incorporados insumos extranjeros que hubieren sido ingresados al país mediante el uso de mecanismos aduaneros suspensivos o devolutivos de aranceles o de franquicias aduaneras especiales.

No obstante, esta limitación no afectará a las mercancías que tengan incorporados insumos importados que se encuentren negociados en un régimen arancelario preferencial en el marco del Tratado de Montevideo, de 1980, o en otro Tratado debidamente ratificado que contempla regímenes arancelarios preferenciales;

b) las exportaciones regidas por la ley N° 18.483 o las acogidas a sus beneficios.

c) los exportadores que, individualmente, en el curso de los últimos doce meses hubieren embarcado una mercancía afecta al reintegro del 10% respecto de la posición arancelaria correspondiente, en la parte que excede el valor FOB de US\$ 10.000.000, en moneda de los Estados Unidos de Norteamérica, debidamente reajustado. Sobre dicho exceso, accederán automáticamente al beneficio del 5% que confiere el artículo 2º de esta ley.

No tendrán derecho a reintegro las mercancías exportadas por un mismo beneficiario respecto de la posición arancelaria correspondiente, en la parte que excede el valor FOB de US\$ 15.000.000, en moneda de los Estados Unidos de Norteamérica, debidamente reajustado, en el mismo período de doce meses a que se refiere el inciso anterior. Sobre dicho exceso, accederán automáticamente al beneficio del 3% que confiere el artículo 2º de esta ley.

Tampoco tendrán derecho a reintegro aquellos exportadores que, individualmente, en el mismo período señalado en los incisos anteriores, hubieren embarcado una mercancía afecta a los beneficios de esta ley, en la parte que excede el valor FOB de US\$ 18.000.000, en moneda de los Estados Unidos de Norteamérica, debidamente reajustado.

Las exclusiones referidas en los incisos anteriores se aplicarán indistintamente a las mercancías que se exporten como producto final o como insumos incorporados a un producto final.

d) Las exportaciones de mercancías que se clasifican en la subposición 74.01.05.00 del arancel Aduanero, y

e) Los productos nacionalizados que se exporten sin cumplir las condiciones señaladas en esta ley.

f) Las exportaciones de mercancías que se clasifican en la Subposición 41.01.01.00, 41.01.02.00 y 41.01.03.00 del Arancel Aduanero.

g) Aquellas mercancías exportadas, cuya materia prima o insumo principal esté excluido del sistema de reintegro establecido en esta ley y represente a lo menos el 85% del valor FOB del producto final exportado.

Esta exclusión será aplicable también cuando el producto exportado esté elaborado con más de una materia prima excluida del reintegro y que en conjunto alcancen a lo menos, dicho valor.

Art. 6. Los exportadores deberán presentar su solicitud de reintegro al Servicio de Tesorerías, acompañada de:

a) Una constancia suficiente, a juicio de Servicio de Tesorerías, del valor obtenido por las mercancías exportadas. El referido Servicio podrá objecar dicho valor en caso de que, de conformidad con los antecedentes proporcionados por el Servicio Nacional de Aduanas, resultare ser superior al que la respectiva mercancía tiene corrientemente en el mercado internacional;

b) Una declaración jurada en que manifiesten que la mercancía exportada es de origen nacional en los términos definidos en la presente ley, y que no se encuentra incluida en la lista frada en conformidad al artículo 3 si en las situaciones a que se refiere el artículo 5 bis.

El reintegro se efectuará mediante cheque girado por el Servicio de Tesorerías, a la orden del exportador, y se entregará a éste dentro de los cinco días hábiles siguientes a la recepción de la solicitud a que se refiere el inciso anterior.

El plazo para solicitar el reintegro ante el Servicio de Tesorerías será de 120 días, contado desde el vencimiento del plazo de retorno de la exportación por la cual se pida el beneficio.

Si el Servicio de Tesorerías no da curso a una solicitud de reintegro de gravámenes a que se refiere el artículo 1, los interesados podrán solicitar reconsideración al Tesorero General de la República, quien se pronunciará en definitiva acerca de la aceptación o rechazo de la misma previo informe de una Comisión Técnica, integrada por un representante del Ministerio de Economía, Fomento y Reconstrucción, quien la presidirá; un representante del Ministerio de Hacienda; uno del Servicio Nacional de Aduanas y uno del Banco Central de Chile. Esta Comisión se pronunciará dentro de un plazo de 60 días contado desde la petición de informe. El interesado deberá presentar la reconsideración dentro del plazo de 30 días, contado desde que el Servicio de Tesorerías le comunique su negativa o dar curso a la solicitud de reintegro. Presentada la reconsideración, el Tesorero General de la República deberá, dentro del término de 10 días hábiles, solicitar el informe de la Comisión Técnica y pronunciarse en definitiva dentro del plazo de 20 días hábiles, contado desde la fecha de recepción del informe de la Comisión.

En su comiendo la Comisión deberá ponderar los antecedentes recibidos y los planteamientos que los interesados hagan valer.

Los integrantes de la Comisión Técnica serán nombrados, a proposición de la entidad que representan, mediante resolución del Ministro de Economía, Fomento y Reconstrucción, la que deberá publicarse en el Diario Oficial.

Art. 7. Jodó aquél que perciba indebidamente el reintegro señalado en esta ley, proporcionando antecedentes material o ideológicamente falsos será sancionado con la pena de reclusión menor en sus grados medio a máximo y multa equivalente al triple de las sumas percibidas.

Sin perjuicio de lo anterior, determinado administrativamente por el Servicio Nacional de Aduanas que el reintegro se percibió indebidamente, este Servicio deberá emitir un cargo para el cobro del reintegro indebidamente percibido. Esta suma se restituirá reajustada en el mismo porcentaje de variación que haya experimentado el Índice de Precios al Consumidor en el período comprendido entre el mes anterior a aquél en que se hizo efectivo el cobro del reintegro y el mes anterior al de la restitución. Estos cargos tendrán mérito ejecutivo y su cobro se sujetará a las normas de procedimiento establecidas en el Código Tributario. El Servicio Nacional de Aduanas podrá formular estos cargos dentro del plazo de tres años, contado desde el pago del reintegro. Los cargos formulados serán reclamables según lo dispuesto en el artículo 132 de la Ordenanza de Aduanas. Para interponer la reclamación, no será preciso restituir previamente el reintegro.

Art. 8. El gasto que demande el pago de los reintegros que establece esta ley, se cargará a un ítem excedible que anualmente se consultará en la Ley de Presupuestos del Sector Público.

Art. 9. Créase en el ítem 50-01-03-25-31 del Presupuesto vigente, la asignación 008: «Reintegro simplificado de gravámenes a exportadores» Moneda Nacional, Miles de \$ 1.000.000. JOSE T. MERINO CASTRO, Almirante, Comandante en jefe de la Armada, Miembro de la Junta de Gobierno. FERNANDO MATTHEI AUBEL, General del Aire, Comandante en jefe de la Fuerza Aérea, Miembro de la Junta de Gobierno. RODOLFO STANGE OELCKERS, General Director de Carabineros, Miembro de la Junta de Gobierno. CESAR RAÚL BENAVIDES ESCOBAR, Teniente General de Ejército, Miembro de la Junta de Gobierno.

Por cuanto he tenido a bien aprobar lo precedente ley, la sanciono y la firmo en señal de promulgación. Lévese a efecto como Ley de la República.

Registrese en la Contraloría General de la República, públíquese en el Diario Oficial e insértese en la Recopilación Oficial de dicha Contraloría.

Santiago, 17 de diciembre de 1985. AUGUSTO PINOCHET UGARTE, General de Ejército, Presidente de la República. Hernán Bache Boc, Ministro de Hacienda. Jorge Valenzuela Durán, Coronel de Ejército, Ministro de Economía, Fomento y Reconstrucción Subrogante.

Lo que transcribe a Ud. para su conocimiento. Saluda atentamente a Ud. Manuel Brito Viales, Subsecretario de Hacienda Subrogante.

TEXTO 2: LEY N° 18.708

Reintegro de gravámenes a exportadores

LEY N° 18.708

Reintegro de gravámenes aduaneros Identificación Norma: (LEY-18708 Fecha Publicación: 13.05.1988- Fecha Promulgación: 11.05.1988 Organismo: MINISTERIO DE HACIENDA Ultima Modificación: LEY 19036 Fecha Ultima Modificación: 25.01.1991 Estado: ACTUALIZADO Actualizaciones: Última Versión ESTABLECE SISTEMA DE REINTEGRO DE DERECHOS Y DEMAS GRAVAMENES ADUANEROS QUE INDICA EN FAVOR DE LOS EXPORTADORES

La Junta de Gobierno de la República de Chile ha dado su aprobación al siguiente Proyecto de ley:

Art. 1. Los exportadores podrán obtener el reintegro de los derechos y demás gravámenes aduaneros pagados respecto de materias primas, artículos a media elaboración, partes o piezas importadas por el propio exportador o por terceros, cuando tales insumos hubieran sido incorporados o consumidos en la producción del bien exportado. *Este beneficio no comprende los sobretasas y derechos compensatorios que se establecieren de conformidad con el artículo 10 de la ley N° 18.525.*

Para todos los efectos legales, los servicios prestados al exterior serán considerados exportación, debiendo el exportador dar cumplimiento a las exigencias y formalidades establecidas para las exportaciones. En este caso, el reintegro no podrá exceder del porcentaje correspondiente a la tasa general de los derechos de aduana del valor del servicio exportado, correspondiendo su determinación al Servicio Nacional de Aduanas, de acuerdo con las normas siguientes, en lo que fuerá procedente.

Art. 2. Para los efectos de la presente ley se entenderá por:

a) **Materia Prima:** Toda sustancia, elemento o materia necesarios para obtener un producto, incluidos aquellos que se consumen o intervienen directamente en el proceso de producción o manufacturación, o sirven para conservar el producto de exportación. Se considerarán como materia prima las etiquetas, envases y otros artículos necesarios para la conservación y transporte del producto exportado. Sin embargo, no se considerarán como materia prima los combustibles o cualquier otra fuente energética cuando su función principal sea la de generar calor o energía para la obtención del producto exportado. Tampoco se considerarán materia prima los repuestos y artículos de recambio que se consuman o empleen en la obtención de estos mercancías.

b) **Artículos a Media Elaboración:** Aquellos elementos que requieren de procesos posteriores para adquirir la forma final en que serán incorporados al producto exportado.

c) **Pieza:** Aquella unidad previamente manufacturada, cuya ulterior división física produzca su inutilización para la finalidad a que estuvo destinada.

d) **Parte:** El conjunto o combinación de piezas, unidas por cualquier procedimiento de sujeción, destinado a constituir una unidad superior.

e) **Desperdicios sin Carácter Comercial:** Aquellos restos o residuos no aprovechables que resulten del proceso de producción o manufacturación, los que para estos efectos se considerarán incorporados o consumidos en el bien exportado.

Art. 3. Correspondrá a los Directores Regionales o Administradores de Aduanas determinar el reintegro a que se refiere el artículo 1º, conforme a la proporción de los insumos incorporados o consumidos directamente en la producción del bien exportado, de acuerdo a los requisitos, modalidades, procedimientos y sistemas de control que fije el Director Nacional de Aduanas, pudiendo exigir, cuando las circunstancias lo aconsejen y con cargo al solicitante, estudios, análisis o dictámenes de técnicos o consultores externos, previamente calificados por dicha Servicio.

En caso de la primera solicitud, el Servicio Nacional de Aduanas deberá pronunciarse respecto a la procedencia y monto del reintegro dentro del plazo de 15 días hábiles. Tratándose de las siguientes solicitudes respecto de un mismo producto y exportador deberá pronunciarse dentro del término de 8 días hábiles, contados, ambos plazos, desde la fecha de recepción de la solicitud respectiva.

El rechazo de una solicitud deberá ser fundado. Los efectos por solicitudes rechazadas, tendrán derecho a solicitar reconsideración ante el Director Nacional de Aduanas basándose en nuevos antecedentes.

En todo caso, en cada oportunidad no podrá solicitarse reintegro por un monto inferior a 100 dólares de los Estados Unidos de América, debiendo los reintegros inferiores a dicha suma agruparse para entregar o superar ese monto.

Art. 4. No podrá solicitarse reintegro con cargo a declaraciones de importación de más de dieciocho meses, contados desde la fecha de la exportación. A su vez, el reintegro deberá imponerse, a lo más dentro del plazo de nueve meses, contado desde la fecha anterior. No obstante ambos plazos podrán ser prorrogados en casos calificados por el Director Nacional de Aduanas.

Art. 5. El reintegro se determinará mediante certificado expedido por el Servicio de Aduanas, expresado en dólares de los Estados Unidos de América. El Servicio de Tesorerías, dentro del

plazo de cinco días hábiles, contado desde la fecha de recepción del certificado, procederá a liquidar y pagar el reintegro al tipo de cambio establecido en el artículo 122 de la Ordenanza de Aduanas, vigente a la fecha de emisión del referido certificado. El pago se hará mediante cheque nominal a favor del solicitante.

Art. 6. Los exportadores que se acojan al beneficio establecido por esta ley no podrán, por un mismo producto, imponer el sistema simplificado de reintegro de gravámenes, a que se refiere la ley No. 18.480. Los que infringieren esta disposición, serán sancionados con las penas señaladas en el artículo 7º de esta ley.

Art. 7. El que fraudulentamente perciba el reintegro señalado en esta ley será sancionado con las penas del artículo 470, No. 8, del Código Penal. Sin perjuicio de lo anterior, el infractor deberá restituir la suma indebidamente percibida, readjustada en el mismo porcentaje de variación que haya experimentado el Índice de Precios al Consumidor en el período comprendido entre el mes anterior a aquel en que se hizo efectivo el cobro del reintegro y el mes anterior al de la restitución.

Art. 8. Las personas que vendan a las Zonas Francas mercancías nacionales en las que se hayan incorporado o consumido directamente insumos nacionalizados a que se refiere el artículo 1º, por un valor CIF no inferior al 10% del valor de venta, podrán obtener el reintegro de los derechos y demás gravámenes aduaneros pagados en la importación de tales insumos en conformidad a las normas precedentes.

El ingreso de estas mercancías a Zona Franca, se veña a Zona Franca de Extensión, así como su posterior salida de dicha zona, se sujetará a las normas establecidas en el artículo 10 bis del Decreto de Hacienda No. 341, de 1977.

El reintegro de dichas mercancías al resto del país, se sujetará a las normas que rigen para las mercancías importadas, quedando afectos al pago de derechos aduaneros solamente aquellas que habiéndose consumido o incorporado en las mercancías vendidas a Zona Franca, hayan sido beneficiadas con el reintegro establecido en esta ley.

Art. 9. Derógase el decreto con fuerza de ley No. 409, de 1970.

Art. 1 Transitorio. No obstante lo dispuesto en el artículo 9º, las solicitudes presentadas con anterioridad a la vigencia de la presente ley y en conformidad al decreto con fuerza de ley No. 409, de 1970, podrán seguir sujetas a las normas establecidas en dicho decreto.

Asimismo, concédese un plazo, hasta el 12 de mayo de 1990, para solicitar acogerse a las normas del decreto con fuerza de ley No. 409, de 1970, del Ministerio de Hacienda, respecto de los derechos, impuestos y demás gravámenes aduaneros pagados, incluidos los derechos compensatorios y sobre tasas del artículo 10 de la ley No. 18.525, correspondientes a insumos utilizados en la elaboración de productos exportados entre el 12 de mayo de 1987 y el 12 de mayo de 1988.

Art. 2 Transitorio. Sin perjuicio de lo dispuesto en el artículo 3º, hasta el 31 de diciembre de 1992, corresponderá al Director Nacional de Aduanas determinar el reintegro a que se refiere este texto legal.

JOSE T. MERINO CASTRO, Almirante, Comandante en Jefe de la Armada, Miembro de la Junta de Gobierno. RODOLFO STANGE OELCKERS General Director de Carabineros, Miembro de la Junta de Gobierno. HUMBERTO GORDON RUBIO, Teniente General de Ejército, Miembro de la Junta de Gobierno. RAMON VEGA HIDALGO, General de Aviación, Comandante en Jefe de la Fuerza Aérea y Miembro de la Junta de Gobierno Subrogante.

Por cuento he tenido a bien aprobar la precedente ley, la sanciono y la firmo en señal de promulgación. Llévese a efecto como ley de la República. Regístrese en la Contraloría General de la República, publíquese en el Diario Oficial e insértese en la recopilación Oficial de dicha Contraloría.

Santiago, 11 de mayo de 1988. AUGUSTO PINOCHET UGARTE, Capitán General, Presidente de la República. Hernán Büchi Buc, Ministro de Hacienda. Manuel Concha Martínez, Brigadier General, Ministro de Economía, Fomento y Reconstrucción. Lo que transcriba a Usted para su conocimiento. Saluda atentamente a Usted. Dante Santoni Compliano, Teniente Coronel, Subsecretario de Hacienda.

Capítulo 9

BUCÓLICO: Aplicación computacional para el análisis de textos. Hacia un análisis de rasgos de la informatividad

GIOVANNI PARODI

RENÉ VENEGAS

INTRODUCCIÓN

No hay duda de que la irrupción de las actuales tecnologías de la información ha afectado el mundo contemporáneo y ha determinado un cambio progresivo de enormes proporciones en prácticamente todos los ámbitos: la economía, la educación, la cultura, el entretenimiento, la ciencia y la política. Hoy en día es difícil imaginar alguna actividad humana que no haya sido impactada por las técnicas digitales. Es un hecho que la computación, sus diversas proyecciones multimediales y las tecnologías de la comunicación han obligado a re-estructurar el modo en que interactuamos en la vida cotidiana y en la laboral y, más importante aún, están impactando fuertemente la manera en que pensamos. Hay quienes incluso –arriesgadamente– afirman que estamos en presencia de un nuevo ser humano. En este capítulo estamos lejos de tan osada afirmación, sobre todo, en lo que a cuestiones ontológicas y epistemológicas se refiere.

Como es de esperar, las ciencias del lenguaje no han estado alejadas de estos desarrollos. Aunque lentamente, se han producido avances tecnológicos de magnitud. En esta línea, por ejemplo, cabe consignar que los traductores automáticos (no importando aquí su calidad), los diccionarios electrónicos monolingües y multilingües, diversos atlas lingüísticos computarizados, variados software de tipo educativo para el desarrollo de competencias lingüísticas específicas, ya han hecho su aparición y muchos se encuentran disponibles gratuitamente en línea a través de Internet.

Las diversas (inter)disciplinas más clásicas de la lingüística tales como la psicolingüística, la neurolingüística, la dialectología, la sociolingüística y la lexicografía –entre otras– se proyectan renovadamente gracias a los instrumentos digitales que han venido en su complemento. A la vez, se ha dado origen a otras inter o transdisciplinas, por ejemplo, la psicolingüística computacional y, otras no necesariamente con base inicial en lingüística, pero si fuertemente arraigadas en la tecnología computacional: procesamiento del lenguaje natural, recuperación, extracción y explotación de datos, tecnologías del habla, etc.

Al mismo tiempo, en la mayoría de los países hispanohablantes se ha venido tomando conciencia de la necesidad de realizar reformas educacionales centradas en formar un ciudadano capaz de construir sus propios aprendizajes, más preocupado por adquirir nuevas y mejores competencias y menos por acumular conocimientos específicos. Las reformas en marcha apuntan a la contextualización del conocimiento y a la colaboración en el aprendizaje.

Estas dos fuerzas, la del cambio tecnológico y la de la reforma educacional, nos han llevado a reflexionar sobre las prácticas lingüísticas en nuestros países hispanoparlantes, y muy en particular, sobre el rol de las TIC's (tecnologías de la información y de la comunicación) en la investigación en el ámbito de la lingüística y sus inter y transdisciplinas.

Desde este marco, en este capítulo exponemos los componentes esenciales de un programa computacional desarrollado en el marco del proyecto de investigación que estima este libro (ver Capítulos 1, 2 y 3), así como también daremos cuenta de un estudio en el marco de la lingüística de corpus con apoyo del programa que aquí presentamos. En la primera parte del trabajo, entregamos una breve revisión del ámbito de la lingüística y los computadores. En lo que sigue, abordamos la descripción y ejemplificación del programa computacional BUCÓLICO [Buscador de Concordancias lingüísticas en Corpus]. En la última parte, con el doble objetivo tanto investigativo descriptivo como ejemplificador del uso de la herramienta computacional, indagamos algunos rasgos de la informatividad en el Corpus PUCV-2003 y sus tres subcorpus (corpus técnico-científico -CTC-, corpus de entrevistas orales -CEO- y corpus de literatura escrita latinoamericana, CLL). El objetivo específico de este análisis es describir comparativamente en los tres subcorpus la ocurrencia de estructuras lingüísticas tradicionalmente identificadas como portadoras de una alta carga informativa, a saber, sustantivos (comunes y propios), nominalizaciones, frases preposicionales como complemento del nombre y participios en función adjetiva.

1. MARCO DE REFERENCIA

1.1 Lingüística y desarrollos tecnológicos

Se debe reconocer que a pesar de grandes esfuerzos e importantes inversiones, en lo que respecta a tecnologización de la investigación en Chile –en el ámbito de la lingüística– se está lejos de los estándares de los centros internacionales de avanzada. A pesar de ello, se realizan importantes proyectos con apoyo gubernamental (FONDECYT, FONDEF, MECESUP, entre otros) cuyo objetivo transversal es impulsar y elevar el nivel del desarrollo académico en términos generales y de la investigación y de la transferencia tecnológica, de modo más específico. No obstante lo anterior, es innegable que la tecnologización de la investigación en lingüística no se encuentra en el nivel deseado y que se debe impulsar

un cambio progresivo y sostenido que permita posicionarse a nivel internacional. Si no se logra esta actitud, se corre el riesgo de quedar postergados de manera exponencial en el breve tiempo.

En el ámbito aplicado y educacional, no es una novedad que los niveles de comprensión y producción escrita que nuestros alumnos revelan hoy en día están por debajo de lo esperado. Existe abundante bibliografía que da cuenta de los escasos logros (Arnoux, Noguera, & Silvestri, 2002; Peronard, Gómez, Parodi, & Núñez, 1998; Parodi, 2003a). Como se sabe, los programas computacionales (ya sea en CDROM o a través de Internet) pueden ser una alternativa metodológica emergente para resolver algunas de las dificultades lingüísticas de nuestros alumnos y, al mismo tiempo, convertirse en una herramienta que posibilite cambios metodológicos en las prácticas didácticas (Ferreira, Campos, & Ruggeri, 1998; Echeverría, 2002; Echeverría & Ramos, 2002; Véliz, 2002; Ferreiro-Cabrera & Atkinson-Abudity, 2002; Parodi, 2002a; Parodi, Núñez, & Gramajo, 2003; Peronard, Parodi & Marinkovich, 2004; Chapelle, 2001; Graesser, VanLehn, Rose, Jordan, & Hartel, 2001; Cobot, 2000; Warschauer & Kern, 2000).

Por otro lado, en el área de investigación más pura, los programas computacionales necesarios para llevar a cabo estudios en lo que se ha denominado "lingüística de corpus" y "lingüística computacional" (Moreno, 1998), se han desarrollado, mayoritariamente, para el español en algunos centros académicos de Europa y, en algunos casos, en países no hispanoparlantes. En efecto, la mayor disponibilidad de ellos es en lenguas diferentes al castellano.

Los avances desde las ciencias computacionales y desde áreas conexas como son las tecnologías del habla, el procesamiento del lenguaje natural, la recuperación y el tratamiento de información, entre otros, exigen que la investigación lingüística contemporánea abandone el trabajo con textos o corpus ejemplares de reducido número de palabras, a no ser para investigaciones de índole exclusivamente cualitativa (Manning & Schütze, 1999; Jurafsky & Martin, 2000; Jackson & Moulinier, 2002; Bed, 2003). En este nuevo contexto, es vital contar con grandes muestras de textos (de mayor poder explicativo) que requieren ser digitalizadas y procesadas con programas computacionales de alto poder descriptivo. Para estos efectos, es imprescindible contar con etiquetadores morfológicos (taggers) y sintácticos (parsers), con lematizadores y desambiguadores lingüísticos y estocásticos. También se debe construir bases de datos e interfaces que permitan la interrogación de los corpus digitalizados y marcados según las medidas estandarizadas internacionalmente (Códigos SGML: Standard Generalized Markup Language). Todo ello asociado a paquetes estadísticos que permitan los análisis pertinentes.

1.2 Rasgos distintivos de la informatividad (RDI)

Una de las cuestiones relevantes del estudio del discurso especializado es la indagación de los rasgos lingüísticos que cumplen funciones específicas y que permiten distinguir este tipo de discurso de otros. En este marco, nos interesa indagar el carácter diferenciador y distintivo de un conjunto de rasgos como representativos de la función informativa de los textos especializados. La mayoría de estos rasgos emergen del estudio multirrasgos y multifunciones que determinó un Foco Informativo (Dimensión 5) y que se reveló como la

principal dimensión caracterizadora del corpus especializado (CTC) en comparación con los corpus de tipo literario (CIL) y oral (CEO); esta dimensión resultó relevante para distinguir oralidad de escritura (al respecto, ver análisis detallado en Capítulo 4).

El conjunto de rasgos seleccionados como caracterizadores de la informatividad son cuatro: sustantivos (comunes y propios), nominalizaciones, frases preposicionales como complemento del nombre y participios en función adjetiva. La elección de estos cuatro elementos gramaticales como "rasgos distintivos de la informatividad" (RDI) obedece a razones esencialmente funcionales, es decir, fueron elegidos pues clásicamente se caracterizan por ser representantes de funciones asociadas a la carga informativa de un texto. Diversos investigadores destacan que la presencia en un texto de estos cuatro rasgos permite identificar cuestiones como la integración y la compactación de información altamente abstracta, propia de un discurso de tipo referencial (Chafe, 1982, 1985; Janda, 1985; Bibel, 1986, 1988; Ciapuscio, 1992; Burdach, 2000; Picallo, 1999; Albentosa & Moya, 2000; Gotti, 2003). También se identifica protótipicamente a la menorreferencialidad, la carencia o escasez de emoción, la precisión y la concisión como características clásicas de los textos de alta especialización, todas ellas representadas a través de los cuatro tipos de rasgos distintivos de la informatividad (RDI).

En términos funcionales específicos y revisando uno a uno los cuatro rasgos en cuestión, destacamos que los sustantivos han sido clásicamente los portadores principales del significado referencial en un texto y, por ende, una alta frecuencia de ellos en un texto es un claro indicador de alta densidad de información (Bibel, 1988). Por su parte, las nominalizaciones han sido caracterizadas como procesos de formación de nombres derivados de bases verbales, adjetivas o nominales mediante la adjunción de sufijos no apreciativos (Lang, 1997). Ellas permiten integrar información en pocas palabras (Chafe, 1982, 1985; Ciapuscio, 1992; Gotti, 2003) y reducir oraciones completas en series más compactas y eficientes de frases nominales (Janda, 1985); además, estas construcciones de reordenación gramatical tienen la función de transportar información altamente abstracta (Bibel, 1986). Las nominalizaciones, como procesos de recategorización o derivación, unidas a un alto grado de organización estructural pre/postmodificadora conllevan una alta densidad léxica, que también se manifiesta, entre otros, a través de la concisión y ausencia de participantes humanos. Complementariamente, Albentosa y Moya (2000) plantean que las nominalizaciones contribuyen a la abstacción del discurso en la medida que la desliga del contexto inmediato y de los participantes, así como también, contribuyen a la compactación del discurso, ya que integran información, en particular junto a preposiciones. Por último, para Charcudeau (2004), las nominalizaciones presentan –en algunos casos– una función de "distanciamiento" con fines de neutralización de la subjetividad del sujeto hablante (como en los textos administrativos o científicos).

De manera similar, las frases preposicionales, en este caso como complemento de nombre, sirven para integrar altas cantidades de información en un texto de modo conciso y compactado. Ellas también constituyen un dispositivo importante que logra condensar altas cantidades de información (Bibel, 1988). Chafe (1982, 1985) especifica que las preposiciones funcionan como mecanismos que integran información en unidades de sentido y, al mismo tiempo, apoyan la expansión de la cantidad de información contenida dentro de una unidad de sentido.

Los estudios que abordan la descripción del participio, lo sitúan preferentemente en el discurso escrito más que en el oral y mayoritariamente se consigna que ellos cumplen una función de integración de información a través de elaboraciones estructurales muy compactas (Bibel, 1988; Ciapuscio, 1992). Por su parte, Janda (1985) establece que se utilizan en la toma de apuntes porque son más compactos e integrados y, por ello, sirven para la producción de un discurso altamente informativo cuando el tiempo es limitado.

En síntesis, los rasgos lingüísticos que hemos agrupado aquí están fundamentalmente orientados a cumplir una función prototípica respecto de la información en los textos. Esto es, los rasgos distintivos de la información (RDI) permiten reconocer en los textos una alta concentración de la información en unidades y estructuras lingüísticas más pequeñas, que presentan los datos de modo más conciso y precisamente posible, revelando una alta carga informativa fundamentalmente de carácter referencial (Bibel, 1988; Halliday & Martin, 1993; Lang, 1997; Burdach, 2000; Albentosa & Moya, 2000; Gotti, 2003).

2. BUCÓLICO: UNA PROPUESTA EN EL MARCO DE LA LINGÜÍSTICA DE CORPUS

El programa computacional BUCÓLICO (Buscador de Concordancias Lingüísticas en Corpus) tiene como propósito principal administrar los datos obtenidos a partir del estudio descriptivo del Corpus PUCV-2003. El procedimiento de creación de este programa incluye el modelamiento de una base de datos constituida por las frecuencias normalizadas de rasgos lingüísticos del Corpus PUCV-2003, obtenidas a través de la interfaz de interrogación computacional en Internet denominada El Grial (www.elgrial.cl). Esta base de datos, BUCÓLICO, fue diseñada utilizando el programa Microsoft Access, en tanto que el lenguaje utilizado para la programación fue Visual Basic.

El programa permite llevar a cabo diversas tareas de consulta acerca del corpus. Es posible obtener información respecto de la constitución de cada corpus ingresado, hecho que posibilita conocer las áreas, las clases textuales y las siglas utilizadas en la investigación. Además, como se muestra en la Figura 1, el usuario tiene acceso a cada uno de los textos utilizados, así como a los sesenta y cinco rasgos lingüísticos, agrupados en 16 categorías (ver Capítulo 4).

El programa también fue diseñado para llevar a cabo análisis cuantitativos de los textos incluidos en el Corpus PUCV-2003; de esta manera, es posible conocer la frecuencia de una palabra objetivo en un texto dado o en todo el corpus. Además, junto con la frecuencia, es posible conocer el contexto oracional (concordancia), es decir, se pueden conocer las palabras a la derecha y a la izquierda que acompañan a la palabra que está siendo buscada (ver Figura 2).

A modo de ejemplo, en la Figura 2, se presenta la búsqueda de la palabra "puerto" en los cuatro manuales del Área Marítima del corpus Técnico-Científico. Podemos observar en la ventana *Buscar* la identificación de la clase textual, el área y el corpus en el cual se realiza la búsqueda de la palabra. La palabra objetivo puede ser buscada de forma sensitiva, es decir, tal como ha sido escrita o, en forma no sensitiva, esto es, se incluyen en la búsqueda las distintas maneras en las cuales la palabra ha sido escrita (considerando mayúsculas y/o minúsculas). Además, se puede determinar cuántas palabras se requieren visualizar en el contexto oracional. Al seleccionar *Procesar*, se despliega la ventana *Resul-*

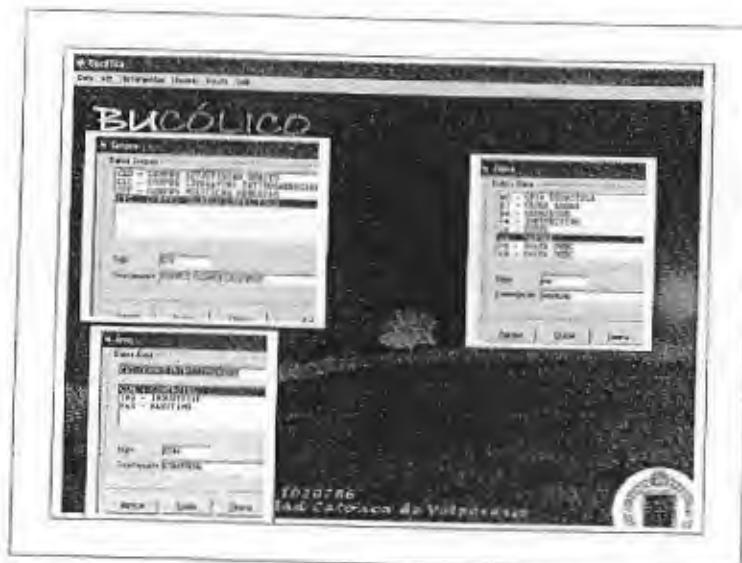


Figura 1. Información respecto del *corpus* en las áreas y las clases textuales

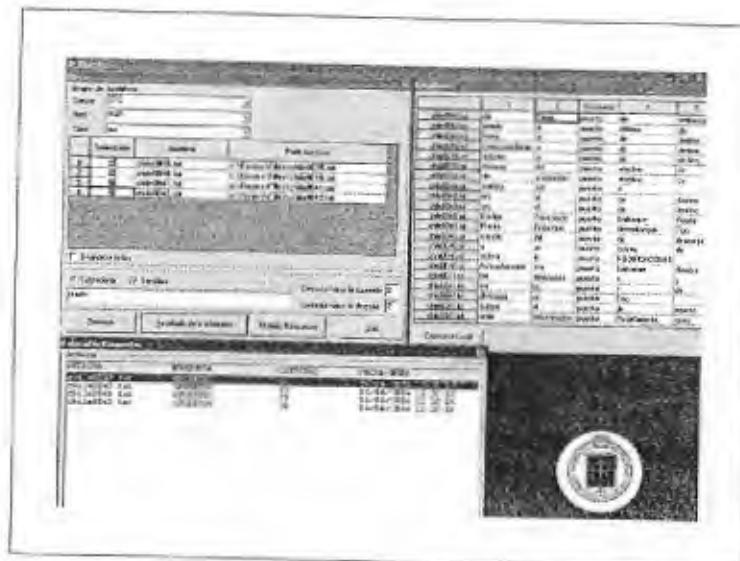


Figura 2. Herramienta de búsqueda de palabras y de conteo de frecuencia

tadas en la cual se destaca la palabra objetivo y se presentan las palabras en cada texto de la clase textual seleccionada. Por último, al seleccionar *Resultado de la Búsqueda* se despliega la ventana *Historial de Búsquedas*, en la cual se presenta el archivo en el cual fue buscada la palabra, la palabra objetivo, la frecuencia de aparición de la palabra en cada texto y la fecha y hora de búsqueda. En esta ventana se va registrando la última búsqueda realizada, en tanto que en la ventana *Historial de Búsquedas* (no desplegada en el ejemplo) se muestran todas las búsquedas de palabras realizadas en los textos.

Otra tarea posible de llevar a cabo sobre textos no marcados es la caracterización estadística de los textos, esto es, se puede obtener información correspondiente al número de palabras, al largo promedio de palabras, al número de oraciones y al largo promedio de oraciones de cada uno de los textos a analizar, datos que son fundamentales para aplicaciones posteriores, por ejemplo, en fórmulas de lecturabilidad (ver Figura 3).

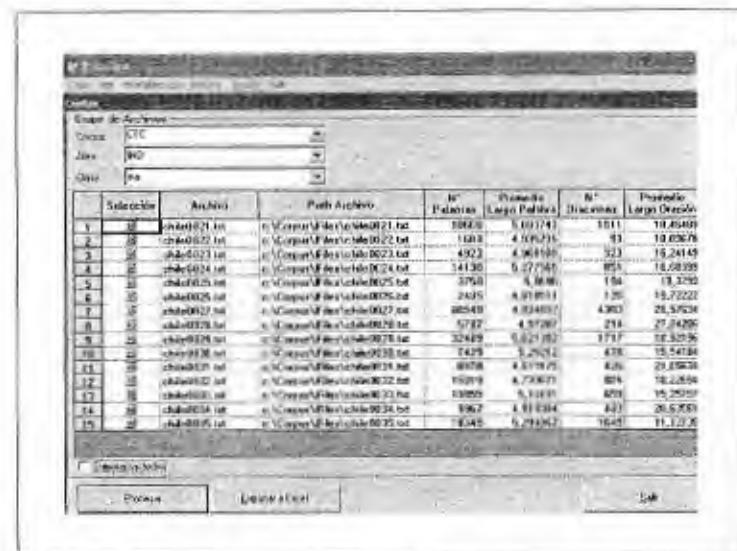


Figura 3. Herramienta Contar

Otro análisis posible de realizar es uno de tipo probabilístico: *N*-Gramas. A través de esto se calcula un conjunto finito de variables aleatorias encadenadas (secuencia de palabras) que tienen una probabilidad conjunta. Es un modelo que intenta predecir la ocurrencia de una palabra, dada una secuencia de palabras que la preceden.

Por último, en relación a los textos no marcados, se puede llevar a cabo análisis de frecuencias asociadas a palabras con y sin repetición, en uno o más archivos de texto seleccionados, esto es conocido como análisis *Type/Token*.

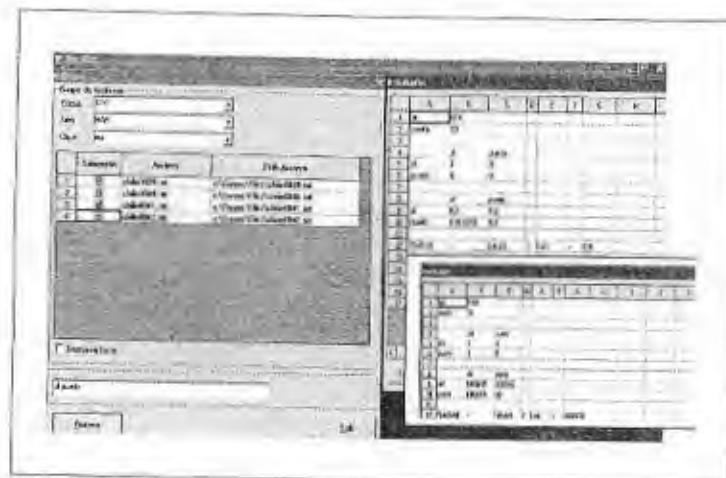


Figura 4. Herramienta N-Gramas.

En la Figura 4, se presenta un breve ejemplo del uso de la herramienta N-Gramas, la cual determina la probabilidad de aparición de una secuencia de palabras. En este caso, se buscó, en los Manuales Técnicos (MT) del Área Marítima, la secuencia "el puerto" y "del puerto", obteniéndose un probabilidad mayor para la primera secuencia (0,0133) que para la segunda (0,0046).

Por otra parte, el programa almacena una matriz de frecuencias, obtenida a partir del análisis de frecuencia realizado de los rasgos lingüísticos en cada uno de los textos constituyentes del corpus de textos marcados morfológicamente. Los resultados fueron almacenados en el programa en forma de una base de datos, a partir de la cual se pueden comparar y graficar las frecuencias de aparición de rasgos lingüísticos entre los distintos corpus y/o textos, tal como lo muestra la Figura 5.

Con esta herramienta es posible comparar la frecuencia de una selección de rasgos en dos grupos de textos distintos. En la Figura 5, se presenta una comparación de las frecuencias de los rasgos sustantivos, nominalizaciones, frases preposicionales y participios en función adjetiva en los MT del Área Comercial (CTC-COM-maj) versus el Área Marítima (CTC-MAR-maj). Como se observa en el gráfico resultante, los manuales del Área Marítima presentan mayor frecuencia de los rasgos seleccionados que los MT de Área Comercial.

Es relevante comentar que todos los resultados obtenidos a partir de estas herramientas son exportables al programa Microsoft Excel para posteriores análisis y graficación.

En síntesis, BUCÓLICO es un programa que permite administrar la información obtenida a partir de textos etiquetados morfológicamente, en forma de matriz de frecuencias, y procesar textos en lenguaje natural.

Cabe señalar la importancia que tiene la integración multidisciplinaria en el desarrollo de aplicaciones tecnológicas de este tipo (lingüistas, ingenieros computacionales, estadísticos, diseñadores gráficos, entre otros), y que permitió desarrollar una herramienta compe-

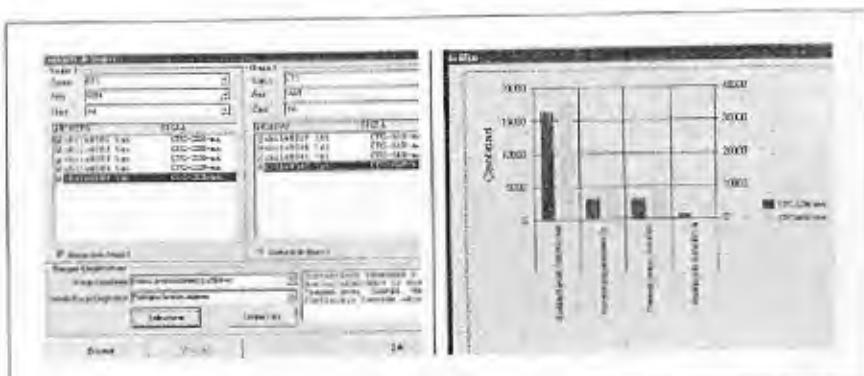


Figura 5. Herramienta contraste de grupos.

rativamente rudimentaria con los avances tecnológicos detectados fuera de Latinoamérica, pero que se constituye en un primer avance en el diseño de instrumentos computacionales elaborados en el marco del programa de investigación en desarrollo en la Pontificia Universidad Católica de Valparaíso, Chile.

3. LA INVESTIGACIÓN PUCV

En los apartados precedentes a este capítulo se ha aportado antecedentes detallados tanto del marco referencial como de los procedimientos metodológicos seguidos en el proyecto global de investigación. Por ello, en lo que sigue, nos concentraremos en tres tipos de resultados empíricos: 1) una descripción global del Corpus PUCV-2003, 2) una breve descripción de los rasgos lingüísticos a indagar en el corpus, y 3) una comparación de frecuencias de los rasgos distintivos de la información-sustantivos, nominalizaciones, frases preposicionales como complemento de nombre y participios en función adjetiva en los corpus, las áreas del corpus CTC y en la clase textual Manual Técnico del corpus CTC.

3.1 Descripción del Corpus PUCV-2003

Tal como se ha indicado en los capítulos precedentes de este libro, la conformación general del Corpus PUCV-2003 se compone de 90 textos que equivalen a 1 466 744 palabras. Este corpus general está dividido en tres grandes registros o subcorpus (Corpus Técnico Científico -CTC-, Corpus de Literatura Latinoamericana Escrita -CLL- y Corpus de Entrevistas Orales -CEO-). Inicialmente, se recolectó el CTC y, posteriormente, con el objetivo de llevar a cabo procedimientos comparativos entre diversos registros que brinden una profunda y certera descripción del CTC y, cumpliendo procedimientos de rigurosidad en el marco de la lingüística de corpus, se recolectaron otros dos corpus, a saber, el CLL y el CEO. La siguiente tabla muestra su distribución en número de textos y palabras.

Tabla 1. Constitución global del Corpus PUCV-2003

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

3.1.1 Corpus de textos técnico-científicos (CTC)

Tal como se muestra en la Tabla 1, el Corpus Técnico-Científico (CTC) está integrado por setenta y cuatro textos con un total de 626.790 palabras, recolectado en establecimientos secundarios técnico-profesionales de la ciudad de Valparaíso, Chile, en tres diferentes orientaciones profesionales. Estas tres diferentes áreas del conocimiento técnico especializado dicen relación con la formación de tres diferentes profesionales técnicos, a saber, sector marítimo (Especialidad Operación Portuaria), sector metal/mecánico (Especialidad Mecánica Industrial), y sector de administración y comercio (Especialidad Contabilidad). El desglose de esta información se entrega en la Tabla 2.

Tabla 2. Constitución del CTC

Área Técnica CTC	Número de textos	Número de palabras
Marítima (Operación Portuaria)	36 (49%)	155.150 (25%)
Industrial (Mecánica)	18 (24%)	246.374 (39%)
Administración y Comercio (Contabilidad)	20 (27%)	225.256 (36%)
Totales	74 (100%)	626.790 (100%)

Estas cifras revelan una cierta heterogeneidad respecto a la configuración del corpus de acuerdo a cada ámbito de especialización y también muestran que no existe una relación directa entre área técnica y porcentaje de textos y palabras.

Los tipos de textos que conforman el corpus CTC se presentan en la Tabla 3.

Tabla 3. Las clases textuales en el CTC

	Nº Textos	Nº Palabras
1 Artículo Técnico	1	9.346
2 Descripción Técnica	8	25.170
3 Diagrama	2	115
4 Formulario	3	2.287
5 Guía Didáctica	14	20.063
6 Glosa Legal	2	4.142
7 Glosario	4	9.747
8 Instructivo	9	17.864
9 Leyes	3	67.905
10 Manual Técnico	24	463.468
11 Reglamento	2	4.485
12 Tablas	2	2.197
Totales	74	626.790

Para una mejor visualización de la distribución porcentual de la ocurrencia de palabras en estos tipos textuales dentro del CTC, los datos se entregan en el siguiente gráfico.

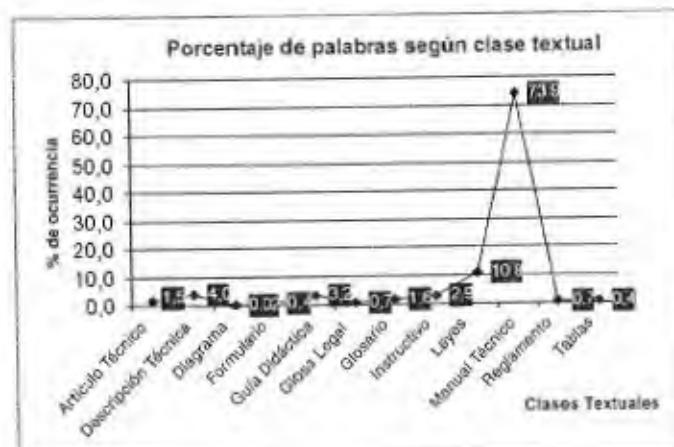


Gráfico 1. Porcentaje de distribución de palabras entre las clases textuales del Corpus CTC

Como se observó, los Manuales Técnicos concentran cerca del 74% del total de palabras del corpus, haciendo de esta clase textual la más representativa del corpus CTC. Debido a esto último, se optó por utilizar esta clase textual por área de especialización (Marítima, Comercial e Industrial) como submuestra de las clases textuales en la investigación que se presentará más adelante.

Según lo expuesto anteriormente, a continuación, se entrega una definición operacional de Manual Técnico, así como la distribución de los textos y cantidad de palabras de los manuales por área (ver Tabla 4).

Manual Técnico: Es un tratado de carácter didáctico enmarcado dentro de un área profesional técnico-científica. Es rico en ejemplos, tablas y recursos multimodales lo que facilita su comprensión y el acceso a información especializada. Su función primordial es la referencial, pudiendo tener secundariamente una función apelativa. Su estructura textual predominante es expositiva-normativa.⁴ (Parodi & Gramajo, 2003: 218)

Tabla 4. Distribución de textos y palabras en los Manuales Técnicos del CTC

Manuales por área	Número de textos	Número de palabras
Marítima (Operación Portuaria)	4 (16,7%)	84.484 (18,2%)
Industrial (Mecánica)	15 (62,5%)	242.244 (52,2%)
Administración y Comercio (Contabilidad)	5 (20,8%)	137.225 (29,5%)
Totales	24 (100%)	463.953 (100%)

3.2 Rasgos lingüísticos

Respecto a los rasgos lingüísticos indagados, se elaboró un conjunto inicial de diecisésis categorías representativas de características gramaticales y funcionales del español. Estos rasgos lingüísticos fueron rastreados a partir de bibliografía relevante en el tema y atendiendo a las posibilidades comunicativa-funcionales de estos. A partir de estas diecisésis categorías iniciales, se procedió a construir una matriz más específica de los rasgos caracterizadores del español según cada uno de estos grandes lineamientos preliminares; de este modo, se llegó a un total de sesenta y cinco rasgos lingüísticos de importancia gramatical y funcional. Para un mayor detalle acerca de los 65 rasgos, ver Capítulo 4.

4. RESULTADOS

4.1 Estudio de los rasgos distintivos de la informatividad (RDI)

Tal como se dijo en la introducción de este capítulo, abordamos ahora el estudio de frecuencia en base a cuatro rasgos lingüísticos, a saber: sustantivos (comunes y propios), nominalizaciones, frases preposicionales como complemento del nombre y participios en función adjetiva. Esta indagación busca describir las ocurrencias de estos rasgos en los corpus, en las áreas del corpus CTC y en la clase Manual Técnico de cada área del corpus CTC. De manera más específica, nos interesa explorar la relevancia y posible carácter diferenciador de un conjunto de rasgos prototípicos de la informatividad en los textos del corpus y en las áreas técnico-científicas en comparación.

Recordemos que el estudio de frecuencias, que a continuación presentamos, también cumple el propósito de exemplificar la utilidad del Programa BUCÓLICO, a través del cual se ha obtenido la información para describir la distribución porcentual de los RDI en tres

niveles de análisis. Un primer nivel de análisis es el que corresponde a la comparación de los RDI entre los tres subcorpus del Corpus PUCV-2003; un segundo nivel de análisis toma en cuenta la comparación de los RDI entre las tres áreas técnicas del corpus CTC; y, un tercer nivel de análisis corresponde a la comparación de los RDI entre los Manuales Técnicos del CTC. Para hacer posibles las comparaciones de frecuencias entre los diversos niveles se han normalizado las frecuencias de ocurrencias multiplicando cada frecuencia por 1000 y luego dividiendo el resultado por el número de palabras de cada texto ($FrN = FrR \times 1000 / Tpal$). Así, por ejemplo, si queremos estimar la frecuencia normalizada del rasgo sustantivo en los tres corpus:

Tabla 5. Ejemplo de normalización

Corpus	FrR (sustantivo)	Base de normalización	Tpal	FrN
CTC	143508	X 1000	626 790	17730
CLL	85052	X 1000	459 860	2255
CEO	44219	X 1000	380 094	466,72

Si observamos los datos no normalizados (FrR) de la Tabla 6, podemos establecer que, si los comparamos sin considerar el número total de palabras, la diferencia existente entre el CTC y el CLL es apenas de 1,7 veces, en tanto que la diferencia entre el CTC y el CEO es de 3,2 veces. En tanto que, si normalizamos las frecuencias (FrN), la diferencia entre el CTC y el CLL es de 7,9 veces y entre el CTC y el CEO es de 38 veces. Esto nos demuestra que las diferencias son mucho más significativas de lo que era posible observar sin considerar el total de palabras de cada corpus, siendo los resultados normalizados mucho más significativos y confiables.

4.2 Nivel 1: RDI por corpus

En este primer nivel de análisis, describiremos las ocurrencias de los RDI en los tres subcorpus de PUCV-2003. De esta manera, en el Gráfico 2, se presenta la distribución porcentual de los RDI de manera agrupada.

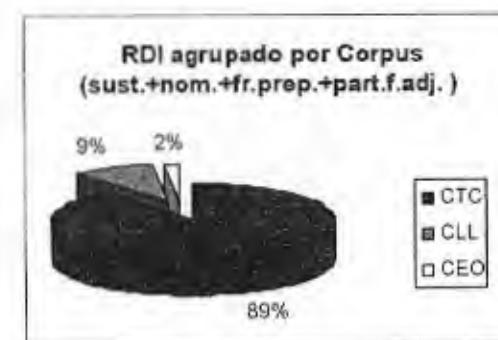


Gráfico 2. Porcentaje de los RDI en el Corpus PUCV-2003

Según se observa en el Gráfico 2, la mayor presencia de RDI se encuentra en el subcorpus CTC, con un 89% de ocurrencia. Ello revela el carácter distintivo de los rasgos en estudio al interior del discurso especializado escrito (CTC), en oposición a lo no especializado (CLL) y a la oralidad (CEO). La alta ocurrencia de los cuatro rasgos indagados en el discurso técnico y científico escrito da cuenta del carácter distintivo de la compactación e integración de la información en los textos especializados. Por su parte, tal como se muestra con claridad en el Gráfico 2, los textos de índole narrativo-literario y de entrevistas orales no destacan por una ocurrencia significativa de estos cuatro rasgos lingüísticos, lo que evidentemente los posiciona en el eje opuesto del continuum informatividad/no-informatividad; ello, en términos comparativos respecto del CTC. Parodi (ver Capítulo 4) ha demostrado, por medio de un estudio estadístico de análisis factorial, que estos dos grupos de textos (CLL y CEO) se caracterizan principalmente por lo que denomina un "Foco Contextual e Interactivo" y un "Foco Narrativo". La focalización en estos funciones está expresada por medio de un grupo de rasgos lingüísticos identificadores, entre otros, adverbios de tiempo y lugar, tiempo presente y pasado, pronombres demostrativos, pronombres y desinencias verbales de primera y segunda persona singular. En este sentido, los datos aportados en el Gráfico 2, a través de un nuevo estudio de frecuencias (ahora comparativa) más detallado, confirman las distinciones ya detectadas y nos aporta nuevos datos más precisos.

A continuación, en el siguiente gráfico, se entrega un desglose pormenorizado de las ocurrencias normalizadas de cada uno de los cuatro rasgos, en cada uno de los subcorpus en estudio.

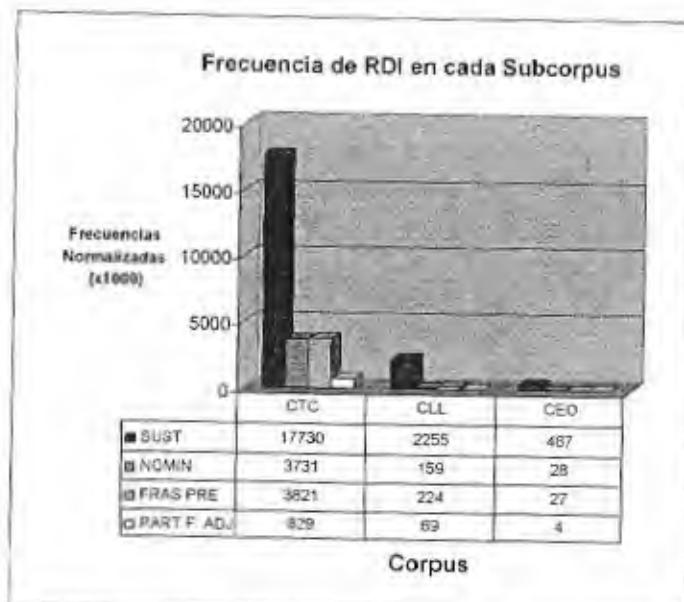


Gráfico 3. Los RDI en los tres subcorpus del FUCV-2003

En el Gráfico 3 se detecta que, en los tres subcorpus, el sustantivo es la categoría gramatical que se presenta como el rasgo distintivo de la información más frecuente. Sin embargo, y en consonancia con el Gráfico 2, podemos determinar que el sustantivo es 7,8 veces más frecuente en el CTC que en el CLL y 37,9 veces más frecuente que en el CEO. Por otra parte, también se observa que la frase preposicional secunda en frecuencia al sustantivo en el CTC; ello con una frecuencia normalizada de 3731 ocurrencias, teniendo una frecuencia 4,8 veces menor a la del sustantivo en el CTC. Además, la frase preposicional en el CTC es 17 veces más frecuente que en el CLL y 141 veces más que en el CEO. La nominalización, con una frecuencia muy similar a la frase preposicional en el CTC, se ubica en tercera posición, siendo 24 veces más frecuente que en el CLL y 134 veces más frecuente que en el CEO. Por último, el participio en función adjetiva, es el RDI que se presenta con menor frecuencia en el CTC, con una ocurrencia de 829 veces, siendo 21,3 veces menos frecuente que el sustantivo. Sin embargo, este rasgo es 12 veces más común en el CTC que en el CLL y es 200 veces más frecuente que en el CEO.

En síntesis, es posible establecer fehacientemente que los RDI se presentan con mucha mayor frecuencia en el corpus CTC, reafirmando con esto la función distintiva de estos rasgos en relación a la concentración y densidad de la información, planteada ya por diversos investigadores. En este sentido, podemos aseverar que los textos incluidos en el corpus CTC presentan una alta densidad informativa y una mayor compactación e integración de la información que en los textos de tipo literarios y en los de entrevistas orales.

4.3 Nivel 2: RDI por área de especialización

En lo que viene, se indagan los cuatro RDI en las tres áreas de especialización técnico-científica del subcorpus CTC. El siguiente gráfico entrega los citados porcentajes.



Gráfico 4. Los RDI en conjunto en las tres áreas del CTC

En el Gráfico 4 se aprecia que la mayor frecuencia porcentual de RDI se concentra en el Área Marítima del CTC con 46%, seguida en porcentaje por el Área Comercial del CTC y finalmente por el Área Industrial del CTC. Es interesante comprobar, al menos en este

punto, que los textos de las tres áreas técnico-científicas no presentan una gran homogeneidad en cuanto a la ocurrencia de estos rasgos.

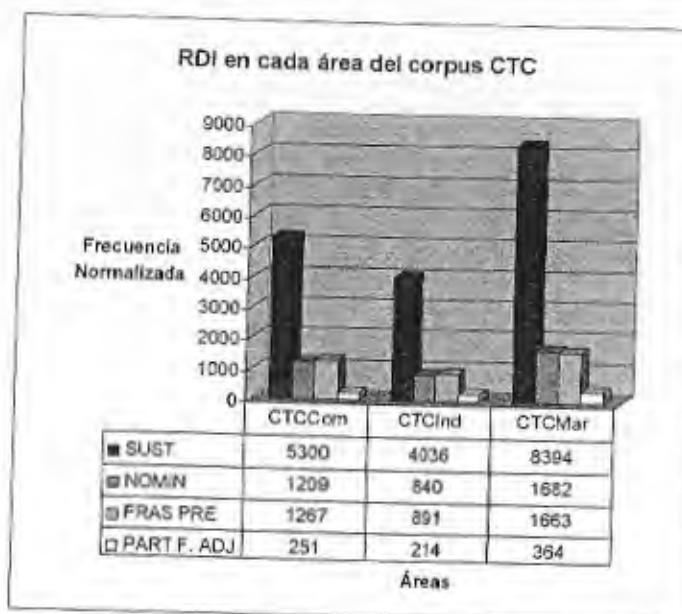


Gráfico 5. Los RDI en las tres áreas técnicas del CTC

En este gráfico se constata que el rasgo sustantivo presenta la mayor frecuencia entre los cuatro RDI en estudio al interior de cada una de las tres áreas, siendo nuevamente el Área Marítima la que aparece con una mayor frecuencia normalizada de sustantivos (8 394). Esto muestra que este rasgo es 1,6 veces más frecuente en el Área Marítima que en el Área Comercial, la cual presenta una frecuencia normalizada de 5.300 y que es 2,1 veces más frecuente que en el Área Industrial, con una frecuencia de 4036. En cuanto al rasgo nominalización es posible determinar que predomina en el Área Marítima, presentando una frecuencia de 1.682 ocurrencias, esto es, 1,4 veces más que en el Área Comercial y el doble más frecuente que en el Área Industrial. El rasgo frase preposicional, levemente menos frecuente que la nominalización, también predomina en el Área Marítima por sobre las otras áreas, siendo apenas 1,3 veces más frecuente que en el Área Comercial y 1,9 veces más que el Área Industrial. Por último, el rasgo participio en función adjetiva, que se presenta como el rasgo menos frecuente en todos los áreas, predomina en el Área Marítima con 364 ocurrencias, siendo 1,5 veces más frecuente que en el Área Comercial y 1,7 veces más frecuente que en el Área Industrial.

Cabe señalar que en todas las áreas los rasgos presentan una distribución de frecuencias muy similar, esto es, predomina el sustantivo, seguido ya sea por la nominalización o la frase preposicional, normalmente con diferencias muy bajas y finalmente, seguido por el participio en función adjetiva.

En vista de estas cifras, es posible establecer que entre las áreas la distribución de los rasgos es muy similar, existiendo una alta densidad informativa, así como compactación e integración de la información en estos textos. Sin embargo, cabe señalar que el Área Marítima presenta una mayor frecuencia de estas características, confirmándose lo representado en el Gráfico 4. Resulta interesante comprobar que son los sustantivos en primera instancia los que se constituyen en el rasgo lingüístico caracterizador de la informatividad en estos textos; no obstante ello, se debe recordar que una determinada función como la informatividad se compone de un conjunto de rasgos que co-ocurren sistemáticamente y por ello la co-ocurrencia significativa de estos cuatro elementos gramaticales da fuerza a la indagación específica.

4.4 Nivel 3: RDI en los Manuales Técnicos del CTC

A continuación, se indagan los cuatro RDI en los veinte y cuatro Manuales Técnicos del CTC que, en términos de palabras, dan cuenta del 74% del Corpus PUCV-2003.



Gráfico 6. Los RDI en los Manuales Técnicos del CTC

En el Gráfico 6 se presenta la distribución de los RDI en los Manuales Técnicos (MT), agrupados por cada área del CTC. Podemos observar que los MT del Área Industrial son los que reúnen el mayor porcentaje de frecuencias, 58% del total. Por otra parte, los MT del Área Comercial presentan la segunda mayor frecuencia porcentual, con un 26%, seguido los MT del Área Industrial con un 16% del total.

En el siguiente gráfico y Gráfico 7, se entrega un análisis más pormenorizado de los datos y cifras de los cuatro RDI en el MT por cada ámbito de especialización.

En este gráfico se detecta que el sustantivo es el rasgo que predomina en todos los manuales, sin embargo, es mucha más frecuente en los manuales del Área Industrial con una frecuencia de 3.382 ocurrencias, siendo el doble más frecuente que en los manuales del Área Comercial y 3,9 veces más frecuente que en los manuales del Área Marítima. Por otra parte, la frase preposicional es el segundo más frecuente de los rasgos en las áreas industrial y comercial, en tanto que en la marítima se ubica bajo la nominalización con uno

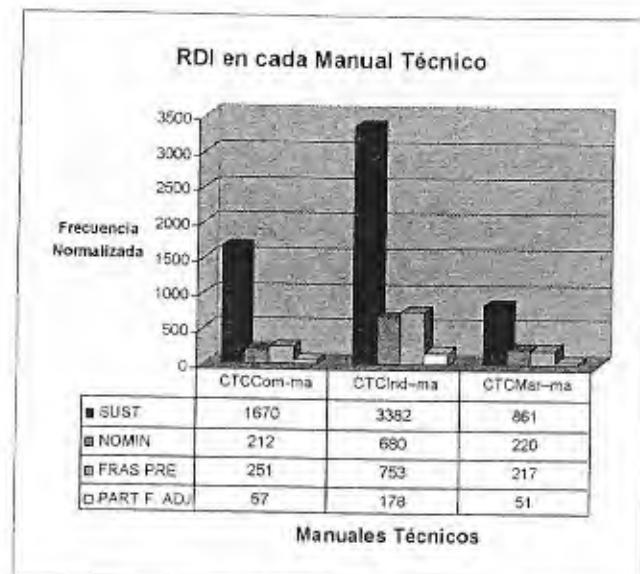


Gráfico 7. Los RDIs en los Manuales Técnicos por área de especialización

diferencia mínima entre ambos rasgos. En este sentido, la frase preposicional es 3 veces más frecuente en el Área Industrial que en el comercial y 3,4 veces más frecuente que en la marítima. La nominalización se presenta como el tercer rasgo más frecuente en las Áreas Industrial y Comercial y el segundo más frecuente en el Área Marítima. La presencia de nominalización en el Área Industrial es 3,2 veces más frecuente que en el Área Comercial y 3 veces más frecuente que en el Área Marítima. Finalmente, el rasgo participio en función adjetiva se presenta como el rasgo menos frecuente en todos los manuales, siendo 3,1 veces más frecuente en los manuales del Área Industrial que en el Área Comercial y 3,4 veces más frecuente que en el Área Marítima.

En suma, podemos establecer que si se observa la presencia de los RDIs en los Manuales Técnicos de las tres áreas del CTC, la mayor frecuencia de estos rasgos se presenta en los manuales del Área Industrial y que el rasgo sustantivo es el más frecuente en todos los áreas, seguido por la frase preposicional y la nominalización en proporciones muy similares. Esto nos permite pensar que los manuales del Área Industrial tienen una mayor densidad informativa y que compactan e integran información con mayor frecuencia que los manuales de las otras áreas. Al respecto, Marinkovich y Cademártori (ver Capítulo 7) por medio de un estudio de tipo cualitativo muestran que algunos de estos manuales técnicos contienen secuencias de tipo narrativo en su carácter de textos académicos de divulgación didáctica. Es factible, desde esta óptica, sugerir que la posible diferencia porcentual de los RDIs tenga alguna relación con lo anteriormente comentado, esto es, que los manuales de las áreas comercial y marítima pudieran presentar un estilo más didáctico ocultando a cierto tipo de estrategias de reformulación divulgativas.

4.5 Las nominalizaciones y los sufijos derivacionales nominales en el Manual Técnico

Dado que el estudio ha permitido distinguir entre sustantivos propiamente tales y las denominadas nominalizaciones que se constituyen en sustantivos de tipo derivacional a partir de otra categoría gramatical, resulta interesante explorar las frecuencias de ocurrencias de las nominalizaciones y sus diversos mecanismos de sufijación derivacional. Para ello, inicialmente, focalizamos el análisis de las nominalizaciones por área de especialización en la clase textual Manual Técnico (MT), ya que –como se ha dicho anteriormente– esta clase textual representa el 74% de las palabras del Corpus CTC.

La cuantificación de las formas nominalizadas en cada una de las tres áreas de especialización arrojó los siguientes resultados:



Gráfico 8. Ocurrencia de las nominalizaciones en los manuales

De acuerdo a las cifras entregadas en este gráfico, se puede apreciar la relativamente homogénea distribución porcentual de la ocurrencia de las nominalizaciones. Se comprueba así que su ocurrencia es sumamente pareja en las tres áreas de especialización (comercial, industrial y marítima), solo con un leve aumento de 7 puntos porcentuales el Área Marítima. Esto indica que en cuanto a este recurso lingüístico esta clase textual se presenta con bastante homogeneidad, siendo su ocurrencia una estructuración gramatical importante como modo de presentación e integración de la información que pretende llegar a una audiencia no experta en las temáticas tratadas.

Las nominalizaciones constituyen un tipo de las llamadas metáforas gramaticales, término acuñado por Halliday (1993) desde la lingüística sistemática funcional, mediante el cual se realiza un doble proceso de compactación de información y de recategorización de unidades lingüísticas. Esta transformación es un desplazamiento desde una estructura lingüística a otra en que la semanticidad suele permanecer intacta o escasamente alterada, pero el modo lingüístico de su expresión ha variado, muchas veces comprimiéndose cierta información y ejecutándose un proceso de reducción de piezas lingüísticas. En el siguiente gráfico, se entregan cifras de la ocurrencia de nominalizaciones formadas por sufijos derivativos deverbiales (es decir, nominalizaciones que compactan procesos) y deadjetivales (aquellas que se construyen a partir de un adjetivo). En particular, hemos seleccionado el conjunto de sufijos derivativos no apreciativos constituido por el sufijo derivativo deadjetival (-dad) y los sufijos derivativos deverbiales (-ción, -sión, -miento).

Con el fin de profundizar en estos resultados, a continuación, se exploran y comparan las ocurrencias de estos sufijos en los manuales técnicos estudiados. El Gráfico 9 muestra porcentualmente esta ocurrencia en los MT.

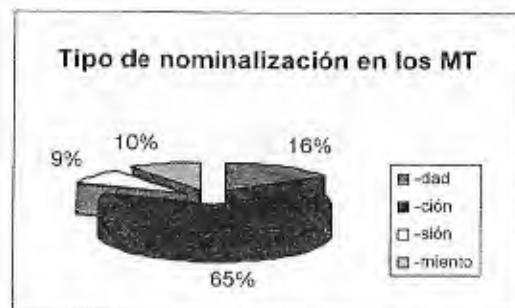


Gráfico 9. Ocurrencia de los tipos de nominalizaciones en los MT

Según se aprecia en este gráfico, la mayor ocurrencia del tipo de derivación nominal se presenta con el sufijo deverbal *-ción* (65%), en tanto con porcentajes bastante bajos se observan las ocurrencias de *-dad*, *-miento* y *-sión* (16%, 10% y 9%, respectivamente). Si recordamos que según el DRAE (1992), el sufijo *-ción* se utiliza para formar sustantivos verbales, que expresan acción y efecto y que además de su significado abstracto, *-ción* y sus variantes pueden denotar objeto, lugar, etc., es posible sostener que la nominalización terminada en *-ción* es un tipo de agrupación nominal cuya función principal es "empaquetar" procesos verbales a través de los cuales se expresan acciones (con agentes, procesos y meta) y efectos (de la acción realizada por alguien sobre algo y su(s) correspondientes consecuencias). De esta manera, el uso prioritario de este tipo de nominalización indica el ocultamiento de los agentes y pone en relieve el proceso en sí mismo.

Como se sabe, las nominalizaciones son un recurso que, entre otros, sirve para establecer la cadena coreferencial, ya que a través de ellas se puede retomar y resumir lo expresado en otra parte del texto; así, este tipo de metáfora gramatical o nominalización es un medio eficiente para dar garantía del establecimiento de cohesión textual por parte de un lector experto. Una alta ocurrencia de nominalizaciones, tal como la detectada en los Manuales Técnicos del Corpus PUCV-2003 es prueba de un tipo de discurso que destaca por complejos encadenamientos de razonamientos típicos de textos científicos (Halliday, 1993; Martín, 1993) y que apunta a un alto grado de abstracción, hechos que –seguramente– implican una serie de dificultades para su comprensión por parte de lectores semileglos y no poseedores de estrategias de lectura eficientes. Desde este punto de vista, si bien los manuales técnicos son un tipo de texto que debe apoyar al semilego en su acercamiento al conocimiento especializado, de modo casi iniciático, su organización compleja –según los datos aquí entregados– exigiría del lector habilidades de comprensión tendientes a revelar inferencialmente los agentes, las causas y las consecuencias compactadas por las nominalizaciones.

CONCLUSIONES

En primer lugar cabe señalar que la herramienta computacional BUCÓLICO mostró ser un recurso tecnológico poderoso para la interrogación y análisis de corpus no marcados morfológicamente. Parte de los datos aquí estudiados no han requerido un etiquetaje lingüístico, sino que se han trabajado como formas gramaticales en su ocurrencia normalizada a partir de los textos del Corpus PUCV-2003. Por una parte, esto muestra la utilidad y posibilidad de agregar nuevos corpus al programa BUCÓLICO sin la necesidad de contar con tecnología que previamente implique su marcaje automático. Por otra, se revelan las proyecciones de investigaciones, apoyados en este tipo de recursos digitales, que trabajen sobre extensos corpus de textos lingüísticos dando así origen a resultados mucho más robustos y confiables. También se debe destacar que esta herramienta, relativamente rudimentaria, construida sin altos costos monetarios ni conocimientos de alta especialización, se constituye en una prueba concreta de que en nuestros ámbitos académicos es factible efectuar desarrollos entre lingüística e informática de manera contundente.

Ahora bien, en cuanto al estudio de los cuatro rasgos de distinción informativa (RDI), una vez descritas las frecuencias porcentuales y normalizadas ($\times 1000$) de los RDI en cada uno de los subcorpus (CTC, CLL, CEO), en cada una de las áreas del CTC (Comercial, Marítima e Industrial) y en los manuales de cada área del CTC, es posible concluir que los RDI estudiados son mucho más frecuentes en el CTC que en los otros dos subcorpus del PUCV-2003, siendo el sustantivo el rasgo que predomina en el CTC, así como en los otros corpus. En un segundo orden de frecuencias aparece la nominalización y la frase preposicional, resultando mucho mayor la presencia de estos rasgos en el CTC que en el CLL y en el CEO. Estas amplias diferencias en la ocurrencia en el CTC nos permite establecer que este corpus se distingue por la alta densidad informativa, así como por la compactación e integración de la información, en congruencia con otros tipos de resultados como los aportados por Parodi (2004).

Más específicamente, respecto a las áreas técnico-científicos del CTC, es posible precisar que la mayor frecuencia de RDI ocurre en el Área Marítima, replicándose la distribución de los rasgos vista en los corpus a través de cada una de las áreas. Esto permite pensar que los textos incluidos en esta área presentan una función referencial muy marcada, asociada a una alta densidad informativa.

En relación al tercer nivel análisis podemos establecer que comparativamente los Manuales Técnicos del Área Industrial presentan una frecuencia de RDI correspondiente a más del doble de las frecuencias de los Manuales Técnicos de las áreas comercial y marítima. Se reconoce nuevamente el patrón de distribución de los cuatro rasgos en donde el sustantivo ocupa la primera posición en todos los manuales, destacando considerablemente la muy alta frecuencia de estos en el Área Industrial, así también aparecen la frase preposicional y la nominalización, esta vez algo más diferenciados entre sí en favor de la frase preposicional. Y, finalmente, aparece el participio en función adjetiva ocupando el último lugar en la distribución de frecuencias, tal como ha aparecido en todos los niveles estudiados. Estos datos confirmán la definición de Manual Técnico presentada por Parodi y Gramajo (2003), ya que estos rasgos lingüísticos nos permiten, desde un análisis más cuantitativo, identificar la función referencial y la orientación informativa predominante en textos de este tipo.

El análisis más fino llevado a cabo sobre las nominalizaciones y los sufijos deverbiales y adjetivales, constitutivos de las mismas, permite concluir que estas organizaciones lingüísticas que recategorizan elementos gramaticales y comparten –muchas veces- altas cantidades de información son un recurso importante en los Manuales Técnicos de los tres ámbitos de especialización en indagación. El sufijo deverbal –*-ción* se posiciona como el mayoritariamente empleado para efectuar este proceso. Sin lugar a dudas, estas construcciones que aglutinan información se transforman en unidades de importante abstracción y se vuelven cruciales para el adecuado procesamiento del discurso y su correspondiente comprensión lingüística.

Por último, los datos aportados en esta investigación constituyen antecedentes vitales para pensar en la necesidad de una didáctica del discurso especializado escrito. Si bien es cierto en los últimos años se han hecho esfuerzos ingentes para apoyar la educación lingüística en la educación escolar básica y secundaria, es evidente que también se hace urgente alfabetizar a las comunidades técnico-científicas. De modo particular, se debe proyectar una didáctica especializada en la comprensión y producción del discurso escrito de los textos que efectivamente circulan en los diferentes medios especializados. Como bien sabemos, los textos escritos vehiculan gran parte del saber comunitario particular, por ello se constituyen en un medio central para la paulatina transformación de los actuales legos en los futuros expertos del conocimiento especializado. Leer y escribir, por lo tanto, son herramientas fundamentales en el proceso de incorporación a comunidades discursivas especializadas. Esta es una de las sendas en espera a que la investigación lingüística le brinde luces; si la lingüística se vale de la alianza con la informática, evidentemente la hará mejor y en plazos más breves.

Postfacio

Así como un prólogo debe despertar el entusiasmo del posible lector, la retórica del postfacio impone continuar la reflexión estimulada por la lectura de la obra. A esto nos invita también Giovanni Parodi en el cierre del libro: *"los datos aportados en esta investigación constituyen antecedentes vitales para pensar en la necesidad de una didáctica del discurso especializado escrito"*, y alrededor de esta propuesta nos moveremos en la breve intervención que sigue. Asimismo, la posición *"post"* nos libera de rigor expositivo y de la enunciación ordenada de méritos y nos permite recorrer con cierta espontaneidad algunos caminos abiertos por la propuesta.

Quiero destacar la importancia de abordar la problemática de la lectura y la escritura en los establecimientos técnico-profesionales. Habitualmente se la deja de lado por considerar que un buen desempeño en este campo no es fundamental en la futura tarea profesional o que la comprensión acabada de los textos escritos, en general, no incide mayormente en el éxito de estos estudios ya que *"lo necesario"* es la adquisición de conocimientos asociados con –y resultado de– la *"práctica"* en las disciplinas específicas. Esta obra cuestiona ese lugar común y en ese sentido es necesario avanzar.

En investigaciones recientes que nuestro equipo de la Universidad de Buenos Aires ha realizado en un colegio industrial de Buenos Aires a alumnos del último año, que obtienen en las asignaturas propias de la especialidad resultados superiores a los habituales en otras escuelas del mismo tipo, se evidenciaron dificultades serias de integración de fuentes informativas y de reconocimiento de voces

diferentes a la del enunciador básico en pruebas de comprensión y una escasa capacidad de reformulación escrita. Debemos adorar que trabajamos con textos del amplio campo de las Ciencias Sociales y las Humanidades, no con los de especialidad. Los planteos del texto que nos precede nos impulsan –mérito no menor de su lectura– a contrastar los datos que habíamos obtenido con otros que provengan de pruebas a partir de textos técnico-científicos. Sin embargo, aunque los resultados pudieran ser superiores, queda en pie un problema político-pedagógico en el que tenemos que reflexionar. El formar alumnos para un buen desempeño en las carreras profesionales no debe implicar dejar de lado una formación que les permita participar como ciudadanos, para lo cual se requiere el desarrollo de una capacidad crítica que los ayude en la interpretación de los discursos sociales que invaden sus vidas cotidianas. Debemos pensar, además, que la comunidad discursiva a la que deberá ingresar laboralmente no es la única que pueda integrar el egresado en su vida adulta ya que, por un lado, los individuos participan de ámbitos sociales o institucionales variados y, por otra, las aceleradas transformaciones que sufre la sociedad contemporánea van a exigir múltiples adaptaciones a nuevas situaciones, lo que una educación de excelencia no puede desdeñar. Asimismo, para que pueda ascender profesionalmente va a necesitar un dominio lector y escritor propio de instancias de gestión. Sabemos, en ese sentido, que esas adquisiciones, si sus bases no se asientan en la escuela media, se logran luego parcial y dificultosamente, incluso si se siguen carreras universitarias, en las cuales, por otra parte, raramente se contempla un espacio para la lectura y la escritura. Todo esto nos lleva a insistir en que la escuela –si desea construir una sociedad más democrática– no debe solo preparar para un buen dominio de los discursos especializados sino también de otros tipos de discursos y diseñar para ello espacios en los cuales se ofenda a un entrenamiento sostenido y reflexivo en lectura y escritura. Se podrá recurrir, sin duda, al *corpus* literario, que es particularmente apropiado para explorar sistemáticamente las múltiples posibilidades del lenguaje, y evaluar luego la incidencia en el campo técnico-científico de los aprendizajes realizados. Esto no implica, por cierto, dejar de lado un trabajo particular con los textos propios del área de especialización involucrada, al que nos convoca energicamente la excelente investigación realizada, y remitimos al capítulo inicial que hace un lúcido diagnóstico y enuncia propuestas didácticas ineludibles para un cambio efectivo. Tanto las actividades con los discursos especializados como las que se puedan realizar con los provenientes de otras prácticas sociales sostendrán en el campo educativo la dimensión humanista implicada en el programa cuyos resultados comentamos y su valoración de la diversidad de "mundos" en los cuales los sujetos participan.

En la línea de lo que he señalado, algunos resultados son particularmente significativos y nos permiten, también, seguir pensando en la didáctica del texto escrito. El estudio sobre los verbos, por ejemplo, suministra datos y llega a conclusiones de notable interés para nuestro recorrido. Por un lado, muestra la menor variabilidad del texto oral en relación con el texto escrito, lo que nos advierte sobre la importancia del trabajo con la escritura para la ampliación del léxico verbal que podrá ser utilizado luego en géneros orales más formales que la entrevista, propios de los últimos años de la escuela media y de tareas socialmente más complejas. Esto se refuerza por el alto índice de comunalidad entre los textos escritos, lo que impulsa a implementar estrategias pedagógicas diversas y cuyas entradas textuales puedan ser variadas ya que tenderán todas a logros próximos. Por otro lado, la escasa

presencia de términos verbales de especialización, cuyo aprendizaje se alcanza gracias a las prácticas específicas que los requieren y en las que circulan, nos lleva a insistir en la apelación a otros géneros y campos culturales y en la puesta en contacto asidua y sistemática con materiales escritos. Entre estos, los literarios, que en el *corpus* analizado muestran el mayor porcentaje de verbos que no aparecen en los otros registros, son, como dijimos, particularmente apropiados.

El relevamiento de los Índices de argumentatividad en los corpora considerados, si bien confirma algunas apreciaciones derivadas de investigaciones cualitativas, tiene un valor indudable que, en realidad, atraviesa toda la obra, el de articular rigurosamente fenómenos gramaticales con una perspectiva discursiva. No dudamos de que los datos obtenidos van a enriquecer la pedagogía de la lengua en el nivel medio facilitando el reconocimiento de rasgos y su apropiación para el desarrollo más eficaz de trabajos escritos. Este capítulo muestra el *continuum* que va, en los textos razonados, de los expositivo-explicativos a los netamente argumentativos y cómo los géneros que circulan por los espacios escolares de aprendizaje son, en general, o explicativos o narrativos. Si bien las prácticas orales muestran el mayor índice de argumentatividad, debemos recordar que esto no implica que los alumnos estén en condiciones de reconocer la orientación argumentativa de textos razonados complejos o de producir textos argumentativos de cierta extensión como los propios de algunos géneros escolares de los últimos años o de los estudios superiores. En este campo también es importante insistir en el entrenamiento claramente pautado que permite el desarrollo de habilidades en comprensión y producción de textos argumentativos aunque el *corpus* técnico-científico presente "un número reducido de haces de rasgos de argumentación".

En la misma línea de reflexión se inscribe el capítulo sobre la co-ocurrencia de rasgos lingüísticos que permiten definir las dimensiones. La lingüística textual ha considerado reiteradamente las secuencias discursivas y su articulación con posiciones dominantes o subordinadas pero este trabajo muestra cómo determinados rasgos lingüísticos –tiempos y modos verbales, tipos de sintagmas, etc.– sostienen, por ejemplo, la dimensión narrativa o informativa que conviven en los textos del *corpus* técnico-científico. Otra vez se destaca el vínculo entre lo gramatical y lo discursivo a partir del cual se pueden elaborar estrategias pedagógicas que permitan alcanzar el dominio del género "manual" atendiendo al entramado propio de las áreas del conocimiento involucrados que determinan, además, las zonas más sensibles a uno u otro foco. En el aula se podrán trabajar, así, los aspectos gramaticales enmarcados en la oración con los despliegues secuenciales.

Los géneros didácticos como el manual a menudo son reformulaciones de textos de expertos o de documentos "reales". El contraste entre texto fuente y texto meta permite reconocer las operaciones de reformulación realizadas, desde el borrado o la expansión hasta la sustitución o la comutación. La regularidad que el investigador pueda determinar en esas operaciones le permitirá formular hipótesis acerca de la representación del género y del destinatario que el "nuevo" autor tiene y cómo resuelve la tensión entre fidelidad al texto primero y necesidad de economía en la elaboración del texto segundo. La complejidad de la tarea de reformulación lleva a veces a la producción de un texto que plantea dificultades en los niveles cognitivo, retórico o comunicacional. El conocimiento de esta problemática es fundamental para evaluarlas y, a partir de ello, considerar tanto las marcas

facilitadores como los problemas de lectura que pueden implicar. Pero constituyen, también, estímulos para la reflexión en el marco de la formación docente ya que el maestro o el profesor se inscribe en una cadena de voces reformuladoras. A hacer consciente y, por lo tanto, controlable, esta práctica habitual nos impulsa el capítulo leído.

Notablemente, el capítulo realizado respecto de los rasgos de informatividad de los *corpus* considerados apoya la señalada. Al revelar la alta ocurrencia de nominalizaciones en los manuales técnico-científicos -lo que es propio, como señalan los autores, de textos científicos caracterizados por "complejos encadenamientos de razonamientos", e implica "un alto grado de abstracción"- muestran el desconocimiento que, en muchos casos, los autores tienen de las competencias del destinatario. Los datos obtenidos exponen cómo las reformulaciones didácticas que no tienen en cuenta fenómenos lingüísticos y su incidencia en la comprensión lectora pueden no alcanzar su objetivo y exigir habilidades de lectura que no poseen los que se inician. Estos resultados, producto de una aplicación computacional al análisis de textos, podrán servir de orientación a los autores de libros de texto o, si las transformaciones a las que nos invita la obra no se realizan, podrán guiar los modos de lectura colectivos realizados por el docente en el aula. Por otra parte, tienen un inesimable interés para el análisis de las tensiones a la que toda actividad reformuladora se ve sometida y, especialmente, la que debe desplegar el discurso divulgativo didáctico.

Para terminar, quiero resaltar la importancia, como esta obra lo demuestra a lo largo de todos los capítulos, de acudir a programas computacionales cuando se trabaja con materiales extensos. Es, indudablemente, un avance metodológico significativo que no solo permite obtener datos más confiables sino que también puede llegar a interrogar algunos supuestos teóricos. En el campo que nos interesa, que es el del análisis del discurso aplicado a la pedagogía de lenguas, permite delimitar con mayor claridad los géneros que circulan en el ámbito escolar y articular rasgos gramaticales y dominancias textuales. Se podrá retomar, a partir de ello, justificadamente, la enseñanza de la gramática oracional y el dominio del metalingüaje correspondiente, que constituyen no solo un estímulo para el desarrollo intelectual sino también un andamiaje esencial para la comprensión y producción de textos complejos.

ELVIRA NARVAIZ DE ARNOUX
Universidad de Buenos Aires
Argentina

Buenos Aires, noviembre de 2004

Referencias bibliográficas

- Adis, J. (1991). Intuition-based and observation-based grammars. En K. Aljmer & B. Altenberg (Eds.), *English corpus linguistics. Studies in honor of Jan Svartvik* (pp. 44-62). London: Longman.
- Adam, J. (1992). *les textes: types et prototypes. Récit, description, argumentation, explication et dialogue*. Paris: Nathan.
- Adam, J. & Lorda, C. (1999). *Lingüística de los textos narrativos*. Barcelona: Ariel.
- Aljmer, K. (2002). Modal adverbs of certainty and uncertainty in an English-Swedish perspective. En H. Hasselgard, S. Johansson, B. Behrens & C. Fabricius-Hansen (Eds.), *Information structure in a cross-linguistics perspective* (pp. 97-113). Amsterdam: Netherlands.
- Aljmer, K. & Altenberg, B. (Eds.) (1991). *English corpus linguistics. Studies in honor of Jan Svartvik*. London: Longman.
- Albertosa, J. & Moya, A. (2000). La reducción del grado de transividad de la oración en el discurso científico en lengua inglesa. *Revista Española de Lingüística*, 2: 445-468.
- Alarcos Llorach, E. (1999). *Gramática de la lengua española*. Madrid: Espasa Calpe.
- Alcón, J. & Bleoza, J. (1975). *Gramática española*. Barcelona: Ariel.
- Alcoba, S. (1999). La flexión verbal. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 4915-4992). Madrid: Espasa Calpe.

- Alvar, M. (2000). *Introducción a la lingüística española*. Barcelona: Ariel.
- Álvarez, T. (2001). *Textos expositivos-explicativos y argumentativos*. Barcelona: Octaedro.
- Amoux, B., Nogueira, S. & Silvestri, A. (2002). La construcción de representaciones enunciativas: el reconocimiento de voces en la comprensión de textos políticos. *Annual Review of Applied Linguistics*, 13, 141-172.
- Arianzen, C. (2001). Las estrategias discursivas en el relato de Julio Román Ribeyro [En línea]. Disponible en <http://folk.uio.no/jmaria/lund/2000/textos/1.pdf>
- Aristóteles (1964). *la retórica*. Madrid: Aguilar.
- Aristóteles (1978). *Metafísica*. Argentina: Planeta.
- Arroyo, J. (2000). La presencia de lo oral en la literatura: sobre la variable futuro verbal en una muestra del teatro español contemporáneo. En M. Muñoz, G. Fernández, A. Rodríguez & V. Benítez [Eds.], *IV Congreso de Lingüística General* (pp. 267-292). Cádiz: Universidad de Cádiz.
- Ávila, A. (2000). Hacia una caracterización gramatical del corpus de lengua hablada. En M. Muñoz, G. Fernández, A. Rodríguez y V. Benítez [Eds.], *IV Congreso de Lingüística General* (pp. 151-158). Cádiz: Universidad de Cádiz.
- Bajtin, M. (1982). *Estética de la creación verbal*. México: Siglo Veintiuno.
- Barrientos, R., Cluente, H., Lagos, D., Muñoz, S., & Tassara, G. (1996). Proyecto de Atlas Lingüístico y Etnográfico del Centro de Chile (ALECECH). *Nueva Revista del Pacífico*, 39-40, 24-29.
- Bassols, M., & Torrest, A. (1997). *Modelos textuales, teoría y práctica*. Barcelona: Euro Editorial.
- Becke, R. (2005). El metadiscursión interpersonal en artículos de investigación. *Signos* (38), 57 [en prensa].
- Bentivoglio, P. (1992). La estructura argumental preferida en el español moderno. En M. Vaquero & A. Morales [Eds.], *Homenaje a Humberto López Morales* (pp. 107-120). Madrid: Arco/Ibros.
- Bernalde, M. (2002). En torno al léxico general del sur de Chile. En G. Parodi (Ed.), *Lingüística e interdisciplinariedad: desafíos del nuevo milenio. Ensayos en honor a Marianne Peronard* (pp. 457-467). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- Bernstein, B. (1993). *La estructura del discurso pedagógico. Clases, códigos y control*. Madrid: Morata.
- Biber, D. (1986). Spoken and written textual dimensions in English: resolving the contradictory findings. *Language*, 62, 384-414.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

- Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, 19, 243-258.
- Biber, D. (1994). Using register-diversified corpora for general language studies. En S. Armstrong (Ed.), *Using large corpora* (pp. 180-201). Cambridge, Massachusetts: MIT Press.
- Biber, D. (2003). Variation among university spoken and written registers: A new multidimensional analysis. En P. Leisyna & Ch. Meyer [Eds.], *Corpus analysis: language structure and language use* (pp. 47-70). Amsterdam: Rodopi.
- Biber, D. & Finegan, E. (1986). An initial typology of English text types. En J. Aarts & W. Meijis [Eds.], *Computer linguistics: Recent developments in the use computer corpora in English language research* (pp. 19-46). Amsterdam: Rodopi.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Reppen, R., Clark, V., & Walter, J. (2001). Representing spoken language in university settings: The design and construction of the spoken component of the T2K-SWAL Corpus. En R. Simpson & J. Swales [Eds.], *Corpus linguistics in North America. Selections from the 1999 Symposium* (pp. 48-57). Ann Arbor: The University of Michigan Press.
- Bod, R. (2003). Introduction to elementary probabilistic theory and formal stochastic language theory. En R. Bod, J. Hay & S. Janreddy [Eds.], *Probabilistic linguistics* (pp. 11-37). London: MIT Press.
- Boschi, A. (2000). Verbos de comunicación y discurso. En de J. Bustos, P. Charaudeau, J. Girón, S. Iglesias & C. López [Eds.], *Lengua, discurso, texto. I Simposio Internacional de Análisis del Discurso VI* (pp. 253-262). Madrid: Visor.
- Bosque, I. (1990). *Las categorías gramaticales*. Madrid: Síntesis.
- Bosque, I. (1999). El nombre común. En I. Bosque. & V. Demonte [Coords.], *Gramática descriptiva de la lengua española* (pp. 3-76). Madrid: Espasa Calpe.
- Brinker, K. (1988). *Linguistische textanalyse*. Berlin: E. Schmid.
- Briscoe, T. & Carroll, J. (1994). Generalized probabilistic LR parsing of natural language corpora with unification-based grammars. En S. Armstrong [Ed.], *Using large corpora* (pp. 25-59). Cambridge: MIT Press.
- Bruner, J. (1988). *Realidad mental y mundos posibles*. Madrid: Gedisa.
- Bunt, H. & Nijholt, A. (2000). New parsing technologies. En H. Bunt & A. Nijholt [Eds.], *Advances in probabilistic and other parsing technologies* (pp. 1-12). Netherlands: Kluwer Academic Publishers.
- Burdach, A. (2000). El léxico científico y técnico: un recurso publicitario persuasivo. *Onomazón*, 5, 189-208.

- Caballero, F. & Larraín, M. (1996). El análisis de los textos filosóficos. *Textos de didáctica de la lengua y de la literatura*, 8: 17-26.
- Cabot, C. (2000). Computer applications to reading and writing abilities and the culture of the Spanish speaking world. *Eurocall Journal*, 184: 208.
- Cabré, M. (1993). *La terminología: teoría, metodología y aplicaciones*. Barcelona: Aníbal.
- Cabré, M. (1999). Hacia una teoría comunicativa de la terminología: aspectos metodológicos. *Revista Argentina de Lingüística*, 15: 24-38.
- 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 J. García & M. Fuentes (Eds.), *Texto, terminología y traducción* (pp. 122-187). Barcelona: Almar.
- Cabré, M., Doménech, M., Morel, J., & Rodríguez, C. (2001). Las características del conocimiento especializado y la relación con el conocimiento general. En T. Cabré & J. Felú (Eds.), *La terminología técnica y científica* (pp. 173-186). Barcelona: Instituto Universitario de Lingüística Aplicada, Universidad Pompeu Fabra.
- Cademártori, Y. (2003). La inscripción de las personas en textos de divulgación científica. *Revista Latinoamericana de Estudios del Discurso*, 3 (1): 9-28.
- Cademártori, Y., Gramajo, A. & Porodí, G. (2003). Texts classes in three specialized areas in technical secondary education. *Meeting of the Society for Text & Discourse, Madrid* [July]. Universidad Autónoma de Madrid: España.
- Calsamiglia, E. (2000). Decir la ciencia: Las prácticas divulgativas en el punto de mira. *Discurso y Sociedad*, 2 (2): 3-8.
- Calsamiglia, H. & Tusón, A. (1999). *Las cosas del decir, manual de análisis del discurso*. Barcelona: Ariel.
- Comacho, J. (1999). La coordinación. En I. Bosque & V. Demonte (Coords.), *Gramática Descriptiva de la Lengua Española* (pp. 2635-2694). Madrid: Espasa Calpe.
- Cameron, L. (2003). *Metaphor in educational discourse*. London: Continuum.
- Caravedo, R. (1990). *La competencia lingüística. Crítica de la génesis y del desarrollo de la teoría de Chomsky*. Madrid: Gredos.
- Caravedo, R. (1999). *Gramática española: enseñanza e investigación. Apuntes metodológicos: Lingüística del corpus*. Salamanca: Ediciones Universidad de Salamanca.
- Cassany, D. (2001). Análisis de la divulgación científica. Modelo teórico y estrategias divulgativas. *Conferencia dictada en el XIV Congreso de la Sociedad Chilena de Lingüística*. Universidad de Los Lagos, Osorno, Chile.
- Cassany, D. & Martí, J. (2000). Estrategias de divulgación de un concepto científico: el

- prón. En J. Bustos, P. Charauobau, J. Grón, S. Iglesias & C. López (Eds.), *Lengua, discurso, texto. I Simposio Internacional de Análisis del Discurso VI* (pp. 2665-2680). Madrid: Visor.
- Cassany, D., López, C. & Martí, J. (2000). La transformación divulgativa de redes conceptuales científicas. Hipótesis, modelo y estrategias. *Discurso y Sociedad*, 2 (2): 73-103.
- Castellano, A. (2000). Ambigüedad y variación del pronombre personal sujeto. En M. Muñoz, G. Fernández, A. Rodríguez & V. Berlizé (Eds.), *IV Congreso de Lingüística General* (pp. 521-531). Cádiz: Universidad de Cádiz.
- Cepeda, C. (1998). El movimiento antícadencial en el español de Valdivia: ejemplos. *Estudios Filológicos*, 33: 23-40.
- Cepeda, G. (2002). Enunciación, actitud y modalidad. *Estudios Filológicos*, 37: 7-28.
- Chafe, W. (1982). Integration and involvement in speaking, writing and oral literature. En D. Torrieri (Ed.), *Spoken and written language: exploring orality and literacy* (pp. 35-53). Norwood, N.J.: Ablex.
- Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. En D. Olson, N. Torrence & A. Hidiard, (Eds.), *Literature, language and learning: The nature and consequences of reading and writing* (pp. 105-123). Cambridge: Cambridge University Press.
- Chafe, W. (1992). The importance of corpus linguistics to understand the nature of language. En J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 79-97). Berlín: Mouton de Gruyter.
- Chafe, W. (1994). *Discourse, consciousness and time*. Chicago: The University of Chicago Press.
- Chafe, W. & Danilewicz, J. (1987). Properties of spoken and written language. En R. Horowitz & J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83-113). San Diego: Academic Press.
- Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.
- Charaudeau, P. (2004). La problemática de los géneros. De la situación a la construcción textual. *Signos*, 37 (56): 23-39.
- Charniak, E. (1996). *Statistical language learning*. Cambridge: MIT Press.
- Chevillard, Y. (1991). *La transposition didactique*. Paris: La Pensée Sauvage Editions.
- Chomsky, N. (1957). *Syntactic structure*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1969). Quine's empirical assumptions. En D. Davidson & Hinkka, J. (Eds.), *Words and Objections. Essays on the Work of W.V. Quine*. Dordrecht: Reidel.

- Christie, F. (1998). Science and apprenticeship. The pedagogic discourse. En J. Martín & Veel, R. (Eds.), *Reading science. Critical and functional perspectives on discourse of science* (pp. 152-180). London: Routledge.
- Christie, F. & Martin, J. (Eds.) (1997). *Genre and Institutions. Social processes in the workplace and school*. London: Continuum.
- Church, K. & Marcel, R. (1993). Introduction to the special issue on computational linguistics. Using large corpora. *Computational Linguistics*, 9 (1): 1-24.
- Clapuscio, G. (1992). Impersonalidad y desagentivación en la divulgación científica. *Lingüística española actual*, 2: 183-205.
- Clapuscio, G. (1994). *Tipos textuales*. Buenos Aires: Eudeba.
- Clapuscio, G. (2000). Hacia una tipología del discurso especializado. *Discurso y Sociedad*, 2 (2): 39-71.
- Clapuscio, G. (2001). Procesos y Recursos de Producción Textual en la Divulgación de Ciencia. En J. Brumme (Ed.), *La historia de los lenguajes ibero-románicos de especialidad. La divulgación de ciencia*. Barcelona: Vervuert/iberoamericana.
- Clapuscio, G. (2003). *Textos especializados y terminología*. Barcelona: Instituto Universitario de Lingüística Aplicada Universidad Pompeu Fabra.
- Clapuscio, G. (2004). El verbo poder en el discurso científico oral. *Revista de la Sociedad Argentina de Lingüística*, 1: 81-99.
- Clapuscio, G. & Kuguel, I. (2002). Hacia una tipología del discurso especializado. En J. García & M. Fuentes (Eds.), *Texto, terminología y traducción* (pp. 37-73). Salamanca: Almor.
- Clapuscio, G. & Oñate, I. (2002). Las conclusiones de los artículos de investigación desde una perspectiva contrastiva. *Revista del Instituto de Investigaciones Lingüísticas y Literarias Hispanoamericanas*, 15: 117-133.
- Conrad, S. & Biber, D. (2001). Multi-dimensional methodology and the dimensions of register variation in English. En S. Conrad & D. Biber (Eds.), *Variation in English. Multi-dimensional studies* (pp. 13-42). London: Longman.
- Contreras, C. (2000). Unidad temática y variedad textual: un tipico social en tres relatos cortos. *Estudios Filológicos*, 35: 23-39.
- Contreras, L. (1984). *Nuevo texto gramatical práctico, reestructurado, ampliado y actualizado conforme a las recientes normas actualmente en vigencia de la Real Academia Española*. Santiago: Instituto Geográfico Militar.
- Cornille, B. (2003). Subjetivización, predicciones de anclaje y modales del español. *Fora Hispánico*, 23: 41-57.
- Coseriu, E. (1977). *El hombre y su lenguaje. Estudios de teoría y metodología lingüística*. Madrid: Gredos.

- Citado de Val, M. (1962). *Psicología del idioma español*. Madrid: Editorial Aguilar.
- Clymire, A. (1989). *Talking with readers. Metadiscourse as rhetorical act*. New York: Peter Lang Publishers.
- Cubo de Severino, L. (2002). Evaluación de estrategias retóricas en la comprensión de manuales universitarios. *Revista del Instituto de Investigaciones Lingüísticas y Literarias Hispanoamericanas*, 15: 69-84.
- Cuenco, M. (1995). Mecanismos lingüísticos y discursivos de la argumentación. *Comunicación, lenguaje y Educación*, 22: 23-40.
- De Beaugrande, R. & Dressler, W. (1981). *Introducción a la lingüística del texto*. Barcelona: Ariel.
- De Kock, J. & Gómez, C. (2002). *Gramática española. Enseñanza e investigación. Apuntes metodológicos*. Salamanca: Ediciones Universidad Salamanca.
- De Miguel, E. (1999). El aspecto léxico. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 2977-3060). Madrid: Espasa Calpe.
- Delbecque, N. & Lamiray, B. (1999). La subordinación sustantiva: las subordinadas enunciativas en los complementos verbales. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 1965-2082). Madrid: Espasa Calpe.
- Demonte, V. (1997). *La subordinación sustantiva*. Madrid: Cátedra.
- Demonte, V. & Varela, S. (1997). Las infinitivas nominales eventivas del español. *Signo y Seña*, 7: 123-156.
- Di Tullio, A. (1997). *Manual de gramática del español. Desarrollos teóricos. Ejercicios. Soluciones*. Buenos Aires: Edícol.
- Duarte, C. & Martínez, A. (1995). *El lenguaje jurídico*. Buenos Aires: A-Z Editores.
- Echeverría, M. (2002). Programas computacionales para el español como lenguaje matemático. *Signos*, 35 (51-52): 163-193.
- Echeverría, M. & Ramos, M. (2002). Antex' 98: Un tutorial interactivo para el análisis de textos. En G. Parodi (Ed.), *Lingüística e interdisciplinariedad: Desafíos del nuevo Milenio. Ensayos en honor a Marianne Peronard* (pp. 375-385). Valparaíso: Ediciones Universitarias de Valparaíso de la Pontificia Universidad Católica de Valparaíso.
- Fernández, O. (1999). El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 1209-1274). Madrid: Espasa Calpe.
- Ferreira-Cabrera, A. & Atkinson-Abudity, J. (2002). A model for generating explanatory web-based natural language dialogue interactions for document filtering. *Journal of Research and Practice in Information Technology*, 43 (1): 2-19.

- Ferreira, A., Campos, D. & Ruggeri, E. (1998). VFRBUM: Una aplicación multimedia para la enseñanza del latín. *Estudios Clásicos*, 114, 121-134.
- Filmore, Ch. (1992). Corpus linguistics and computer-aided armchair linguistics. En J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 35-60). Berlin: Mouton de Gruyter.
- Fontanella, M. (1999). Sistemas pronominales de tratamiento usados en el mundo hispánico. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 1399-1426). Madrid: Espasa Calpe.
- Foucault, M. (1970). *El orden del discurso*. París: Collège de France.
- Francis, N. (1979). A tagged corpus: problems and prospects. En S. Greenbaum, G. Leech & J. Svartvik (Eds.), *Studies in English linguistics for Randolph Quirk* (pp. 192-209). London: Longman.
- Francis, N. & Kucera, H. (1964 / 1979 / 1981). *Manual of information to accompany a standard sample of present-day edited American English, for use with digital computers*. Providence: Department of Linguistics, Brown University.
- Fries, Ch. (1952). *The structure of English*. New York: Harcourt, Brace & Company.
- Galán, C. (1999). La subordinación causal y final. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 3597-3642). Madrid: Espasa Calpe.
- Gamero, S. (2001). *La traducción de textos técnicos*. España: Ariel.
- García, A. (2004a). Los procedimientos matemáticos en estudios e investigaciones lingüísticos: utilidad y riesgo. [En línea]. Disponible en: http://angarmegia.tresuvesdables.com/riesgos_y_beneficios.htm
- García, A. (2004b). Los métodos estadísticos y la investigación lingüística: breve recorrido histórico. [En línea]. Disponible en: <http://angarmegia.tresuvesdables.com/HistMetS.pdf>
- García, A. (2004c). Beneficios y servidumbres de los métodos estadísticos. [En línea]. Disponible en: «<http://angarmegia.tresuvesdables.com/BenServS.pdf>» <http://angarmegia.tresuvesdables.com/BenServS.pdf>
- Genette, G. (1972). *Figures III*. París: le Seul (Points).
- Gili Gaya, S. (1980). *Curso superior de sintaxis española*. Barcelona: Vox.
- Gläser, R. (1993). A multi-level model for a typology of LSP genres. *Fachsprache International Journal of LSP*, 15(1-2):18-26.
- Gläser, R. (1982). *The problem of style classification in LSP (ESP)*. Ponencia presentada en el 3º European Symposium on LSP. Copenhague.
- Goldman, S. & Bisanz, G. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. En J. Otero, J. León & Graesser, A. (Eds.), *The psychology of science text comprehension* (pp. 19-50). Mahwah, N.J.: Lawrence Erlbaum.

- Gómez Macker, L. (1991). Los matices del hombre y el lenguaje. *Signos*, 30(24): 34-49.
- Gómez Macker, L. (1998). Dimensión social de la comprensión verbal. En M. Peronard, L. Gómez, G. Parodi & F. Núñez (Eds.), *Comprensión de textos escritos de la teoría a la sala de clases* (pp. 34-58). Santiago de Chile: Andrés Bello.
- Gómez Macker, L. & Peronard, M. (1988). *El lenguaje humano. Léxico fundamental para la iniciación lingüística*. Valparaíso: Ediciones Universitarias de Valparaíso.
- Gómez, I. (1999). Los verbos auxiliares. Las perifrasis verbales de infinitivo. En I. Bosque & V. Demonte (Coords.), *Gramática Descriptiva de la Lengua Española* (pp. 3323-3390). Madrid: Espasa Calpe.
- González, J. (2000). Morfología Nominal. En M. Alvar (Dir.), *Introducción a la lingüística española* (pp. 177-190). Barcelona: Ariel.
- Gott, M. (2003). *Specialized discourse: Linguistic features and changing conventions*. Berna: Peter Lang.
- Graesser, A., M. Gernsbacher & Goldman, S. (2003). Introduction to the handbook of discourse processes. En A. Graesser, M. Gernsbacher & Goldman, S. (Eds.), *Handbook of discourse processes* (pp. 1-23). Mahwah, N.J.: Lawrence Erlbaum.
- Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Horte, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22: 39-51.
- Guipériz, J. (1981). Communicative competence. En D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 323-334). Washington: Georgetown University press.
- Gunnarsson, B., Linell, P. & Nordberg, B. (Eds.), (1997). *The construction of professional discourse*. London: Longman.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (2001). *Análisis multivariante*. Madrid: Prentice Hall.
- Halliday, M. (1975). *Learning how to mean: Explorations in the development of language*. London: Edward Arnold.
- Halliday, M. (1978). *Language as a social semiotics: The social interpretation of language and meaning*. London: Edward Arnold.
- Halliday, M. (1985). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. (1991). Corpus studies and probabilistic grammars. En K. Aljmer & B. Altenberg (Eds.), *English corpus linguistics. Studies in honor of Jan Svartvik* (pp. 31-43). London: Longman.
- Halliday, M. (1992). Language as a system and language as a instance: the corpus as a theoretical construct. En J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 61-77). New York: Mouton de Gruyter.

- Halliday, M. (1993). On language and physical science. En M. Halliday & J. Martin (Eds.), *Writing science. Literacy and discursive power* (pp. 54-68). Pittsburgh: University of Pittsburgh Press.
- Halliday, M. (1994). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. (1994). On language in relation to the evolution of human consciousness. *Nobel Symposium The relation between language and the mind* (8-12). Stockholm: August.
- Halliday, M. (1998). Things and relations. Regrammaticising experience as technical knowledge. En J. Martin & R. Veel (Eds.), *Reading science. Critical and functional perspectives on discourse of science* (pp. 185-235). London: Routledge.
- Halliday, M. & J. Martin. (1993). *Writing science. Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.
- Harvey, A. (2002). Representación e Imagen del quehacer científico en los Medios de Comunicación. En G. Parodi (Ed.), *Lingüística e Interdisciplinariedad: Desafíos del nuevo milenio. Ensayos en honor a Marianne Peronard* (pp. 335-353). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- Harvey, A. (2004). La evaluación en el discurso de informes escritos por estudiantes universitarios. Ponencia presentada en el III Encuentro Nacional de Análisis del Discurso 28 de septiembre al 1 de octubre: Valdivia.
- Hernández, C. (1996). *Gramática funcional del español*. Madrid: Gredos.
- Hernández, C. (2000a). Morfología del verbo. La auxiliaridad. En M. Alvar (Dir.), *Introducción a la lingüística española* (pp. 195-213). Barcelona: Ariel.
- Hernández, C. (2000b). Sintaxis: La subordinación. En M. Alvar (Dir.), *Introducción a la lingüística española* (pp. 391-407). Barcelona: Ariel.
- Hernanz, M. (1999). El infinitivo. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 2197-2356). Madrid: Espasa Calpe.
- Hoffmann, L. (1987). *Fachsprach Instrument und Objekt*. Leipzig: Enzyklopädie.
- Hoffmann, L. (1985). *Kommunikationsmittel Fachsprache*. Tübingen: Niemeyer.
- Horowitz, R. & Samuels, J. (1987). Comprehending oral and written language: Critical contrasts for literacy and schooling. En R. Horowitz, & J. Samuels (Eds.), *Comprehending oral and written language* (pp. 1-52). San Diego: Academic Press.
- Hurtado, A. (2001). *Traducción y traductología*. Madrid: Cátedra.
- Hyland, K. (1998). *Hedging in scientific research articles*. Philadelphia: John Benjamins.
- Hymes, D. (1972). Models of interaction of language and social life. En J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics* (pp. 35-71). New York: Holt, Reinhart & Winston.

- Hymes, D. (1974). *Foundations in sociolinguistics*. Pennsylvania: UP Press.
- Ivanic, R. (1998). *Writing and identity. The discursive construction of identity in academic writing*. Philadelphia: John Benjamins.
- Jackson, P. & Moulinier, I. (2002). *Natural language processing for online applications. Text, retrieval, extraction and categorization*. Amsterdam: John Benjamins.
- Jakobson, R. (1961). *Linguistics and poetics*. New York: Wiley.
- Janda, R. (1985). Note-taking as simplified register. *Discourse Processes*, 84, 437-454.
- Jeannelet, L. (1994). *Écrire la science. Formes et enjeux de la divulgation*. Paris: Presse Universitaire de France.
- Johansson, S. (1981). Word frequencies in different types of English texts. *ICAME NEWS*, 5, 1-13.
- Johansson, S. (1991). Times change, and so do corpora. En A. Aljmer & B. Altenberg (Eds.), *English corpus linguistics. Studies in honor of Jan Svartvik* (pp. 305-314). London: Longman.
- Johansson, S., Leech, G. & Goodluck, H. (1978). *Manual of information to accompany the Lancaster-Oslo/Bergen corpus of British English, for use with digital computers*. Oslo: University of Oslo.
- Joshi, A. (1999). Computational linguistics. En R. Wilson & F. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 162-164). Massachusetts: MIT Press.
- Jurafsky, D. (2003). Probabilistic modelling in psycholinguistics: Linguistics comprehension and production. En R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 38-95). London: MIT Press.
- Jurafsky, D. & Martin, J. (2000). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. New Jersey: Prentice Hall.
- Kaiser, D. (2002). La presencia del autor en los textos académicos: un estudio contrastivo de trabajos de estudiantes de Venezuela y Alemania. *Boletín de Lingüística* 17, 53-68.
- Kearney, G. (1998). *An introduction to corpus linguistics*. New York: Longman.
- Killingsworth, M. & Gilbertson, M. (1992). *Good style*. London: Chapman & Hall.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kittridge, R. (1982). Variation and homogeneity of sublanguages. En R. Kittridge & J. Lehrberger (Eds.), *Sublanguages. Studies of language in restricted semantic domains* (pp. 145-189). Berlin: Walter de Gruyter.
- Kovacci, O. (1993). La didáctica de la lengua materna. Experiencias en la Argentina. *Actas del I Congreso Internacional sobre la Enseñanza del Español*. Madrid: CEMIP.

- Kovac, O. (1999). El Adverbio. En I. Bosque, & V. Demonte, (Coords.), *Gramática descriptiva de la lengua española* (pp. 705-786). Madrid: Espasa Calpe.
- Kucher, S. (1985). The making of meaning: reading and writing as parallel processes. *Written Communication*, 2 (3), 317-356.
- Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago: The University of Chicago Press.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Lagos, D., Cifuentes, H., Tassara, G., Fuentes, K., Reyes, J. P., & Vélez, R. (1999). *Atlas lingüístico etnográfico del centro de Chile* (ALECECH). Centro meridional: Informe de proyectos Universidad de Playa Ancha, 18.
- Iakoff, G. (1972). A study in meaning criteria and the logic of fuzzy concepts. *Chicago Linguistics Society*, 8: 183-288.
- Lang, M. (1997). *Formación de palabras en español*. Madrid: Cátedra.
- Langacker, R. (1990). Subjecdification. *Cognitive Linguistics*, 1 (1): 5-38.
- Lassen, I. (2003). *Accessibility and acceptability in technical manuals*. Amsterdam: John Benjamins.
- Lebart, L. & Salem, A. (1994). *Statistique textuelle*. Paris: Dunod.
- Lebart, L., Salem, A., & Bécaud, M. (2000). *Análisis estadístico de textos*. Lleida: Editorial Milenio.
- Leech, G. (1991). The state of the art in corpus linguistics. En K. Ajmer & B. Altenberg (Eds.), *English Corpus Linguistics. Studies in honor of Jan Svartvik* (pp. 8-29). London: Longman.
- Leech, G. (2002). Sobre la importancia de los corpus de referencia. *Donostia*, 24-25: 1-3.
- Leech, G. (1992). Corpora and theories of linguistic performance. En J. Svartvik (Ed.), *Directions in corpus linguistics: proceeding of Nobel symposium* (pp. 105-122). Berlin: Mouton de Gruyter.
- Lledó, E. (1995). Usos lingüísticos y género. *Textos de didáctica de la lengua y la literatura*, 6: 29-34.
- Lo Caso, V. (1998). *Gramática de la argumentación*. Madrid: Alfonzo.
- Longacre, R. (1983). *The Grammar of discourse*. New York: Plenum Press.
- Lope Blanch, J. (1977). *Estudios sobre el español hablado en las principales ciudades de América*. UNAM.
- López, C. (2002). Aproximaciones al análisis de los discursos profesionales. *Signos*, 35 (51-52), 195-215.
- Lorente, M. (2002). Verbos y discurso especializado. [En línea]. Disponible en: <http://ales.rediris.es/ales16/lorente.htm>

- Louwense, M., McCarthy, P., McNamara, D., Graesser, A. (2004). Variation in language and cohesion across written and spoken registers. En K.D. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- Lu, H. (2003). *Oraciones complejas y modo subjetivo en español*. Taipei: Kuan Tong.
- Manning, C. & Schütze, H. (Eds.) (1999). *Foundations of statistical natural language processing*. Cambridge: MIT Press.
- Marcos Martín, F. (1975). *Aproximación a la gramática española*. Madrid: Editorial Círculo.
- Marinkovich, J. (2003). Rasgos lingüísticos asociados a los manuales técnico-didácticos. *Ponencia leída en el XV Congreso de la Sociedad Chilena de Lingüística (Octubre)*. Universidad Metropolitana de Ciencias de la Educación, Santiago de Chile.
- Marinkovich, J. & Cademártori, Y. (2004). Foco narrativo y foco informativo: Dos dimensiones para una descripción de los manuales en la formación técnico-profesional. *Signos*, 37 (55), 31-40.
- Markkanen, R. & Schröder, H. (2000). Hedging: A challenge for pragmatics and discourse analysis. [En línea]. Disponible en: <http://sw2.euv-frankfurt.o.de/Publikationen/Hedging/markkanne.html>
- Martínez, A. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum.
- Martín Zorráquin, M. & Portolés, J. (1999). Los marcadores del discurso. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 4051-4213). Madrid: Espasa Calpe.
- Martín, G. (1986). *Curso de redacción*. Madrid: Paraninfo.
- Martín, J. (1986). Intervening in the process of writing development. En C. Pointer & Martín, J. (Eds.), *Writing to mean: teaching genres across the curriculum* (pp. 11-43). Sydney: Applied Linguistics Association of Australia.
- Martín, J. (1992). *English text, system and structure*. Amsterdam: John Benjamins.
- Martín, J. (1997). Analyzing genre: functional parameters. En F. Christie & J. Martín (Eds.), *Genre and institutions. Social processes in the workplace and school* (pp. 3-39). London: Continuum.
- Martín, J. (1998). Discourse of science: Recontextualisation, genesis, intertextuality and hegemony. En J. Martín & Vélez, R. (Eds.), *Reading science. Critical and functional perspectives on discourse of science* (pp. 3-14). London: Routledge.
- Martín, J., Christie, F., & Rothery, J. (1987). Social processes in education. En F. Reid (Ed.), *The place of genre learning* (pp. 67-149). Victoria: Centre for Studies of Literary Education.
- Mathiassen, C. (1998). Constraining processes of consciousness: From the commonsense model to the uncommonsense model of cognitive science. En J. Martín & R. Vélez

- (Eds.), *Reading science. Critical and functional perspectives on discourse of science* (pp. 327-356). London: Routledge.
- Matus, A. (2002). Corrección académica: ideal panhispánico y norma culta. En G. Parodi (Ed.), *Lingüística e interdisciplinariedad: desafíos del nuevo milenio. Ensayos en honor a Marianne Peronard* (pp. 389-401). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Mendikoetxea, A. (1999a). Construcciones inacusativas y pasivas. En I. Basque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 1575-1630). Madrid: Espasa Calpe.
- Mendikoetxea, A. (1999b). Construcciones con *se*: medias, pasivas e impersonales. En I. Basque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 1631-1722). Madrid: Espasa Calpe.
- Menéndez, S. (1999). El discurso del libro de texto: una propuesta estratégico-pragmática. *Discurso y Sociedad*, 1 (2): 85-104.
- Minker, W., Waibel, A. & Mariani, J. (1999). *Stochastically-based semantic analysis*. Netherlands: Kluwer Academic Publishers.
- Molner, M. (1986). *Diccionario de uso del español*. Madrid: Gredos.
- Montalio, E. (2001). *Conectores de la lengua escrita*. Barcelona: Ariel.
- Montureux, M. (1985). Linguistique et vulgarisation scientifique. *Information sur les Sciences Sociales*, 24 (4): 825-845.
- Moreno, A. (1998). *Lingüística computacional: Introducción a los modelos simbólicos, estadísticos y biológicos*. Madrid: Síntesis.
- Moyano, E. (2000). *Comunicar Ciencia. El artículo científico y las comunicaciones a congresos*. Buenos Aires: Universidad Nacional de Lomas de Zamora.
- Munguía, I., Munguía, M. & Rocha G. (2000). *Gramática de la lengua española. Reglas y ejercicios*. México D.F.: Larousse.
- Narbona, A. (2000). Sintaxis coloquial. En M. Alvar (Dir.), *Introducción a la lingüística española* (pp. 463-476). Barcelona: Ariel.
- Navarro, T. (1946). *Estudios de fonología española*. New York: Las Américas Publishing Company.
- Norma, D. (1999). Computational psycholinguistics. En R. Wilson & F. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 168-169). Massachusetts: MIT Press.
- Nuopponen, A. (1997). A model for systematic terminological analysis. *Ponencia leída en el LSP Symposium*. Copenhagen.
- Oakes, M. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

- Olbertz, H. (1998). *Verbal periphrases in a functional grammar of spanish*. Berlin-New York: Mouton de Gruyter.
- Osorio, M. (2000). Categorías gramaticales de modalidad y aspecto en la lengua de señas colombiana. [En línea]. Disponible en <http://www.geocities.com/fenascolv/ponencia151.html>.
- Otal, J., Fontanet, I. & Codina, V. [Eds.] (1997). *Estudios de Lingüística Aplicada*. Castellón: UJI.
- Otero, J., León, J. & Graesser, A. (Eds.) (2002). *The psychology of science text comprehension*. Mahwah, NJ: Lawrence Erlbaum.
- Oyanedel, M. (2004). Secuencias descriptivas en informes escritos de estudiantes universitarios. Ponencia presentada en el III Encuentro Nacional de Análisis del Discurso 28 de septiembre al 1 de octubre: Valdivia.
- Oyanedel, M. & Samaniego, L. (1998). Notas para un nuevo perfil lingüístico de Santiago de Chile. *Boletín de Filología de la Universidad de Chile*, 37: 899-913.
- Palmer, F. (1974). *The English verb*. London: Longman.
- Parodi, G. (1999). Conexiones entre comprensión y producción de textos escritos: Una aproximación psicolingüística. En G. Parodi (Ed.), *Discurso, cognición y educación. Ensayos en honor a Luis Gómez Macker* (pp. 59-90). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- Parodi, G. (2002a). Comprensión lingüística: ¿Hacia dónde vamos desde donde estamos? En G. Parodi (Ed.), *Lingüística e interdisciplinariedad: Desafíos del nuevo Milenio. Ensayos en Honor a Marianne Peronard* (pp. 47-67). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- Parodi, G. (2002b). El análisis del discurso especializado: aproximación a los tipos de textos utilizados en la formación secundaria técnico-profesional. *Ponencia leída en el I Congreso Internacional de Análisis del Discurso: Lengua, Cultura, Valores* (Noviembre). Universidad de Navarra, Pamplona, España.
- Parodi, G. (2003a). *Relaciones entre lectura y escritura: una perspectiva cognitiva discursiva. Antecedentes teóricos y resultados empíricos*. Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- Parodi, G. (2003b). Identificación de rasgos lingüísticos relevantes en textos técnico-científicos del ámbito educativo profesional. *Ponencia leída en el V Coloquio de la Asociación Latinoamericana de Estudios del Discurso (ALED)* (Octubre). Benemérita Universidad Autónoma de Puebla, Puebla, México.
- Parodi, G. (2004). Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada en corpus computarizado. *Estudios Filológicos*, 39: 7-36.
- Parodi, G. (2005a). *Reading-writing connections: Discourse-oriented research. Reading and Writing: Interdisciplinary Journal* (en prensa).

- Parodi, G. (2005b). Lingüística de corpus y análisis multidimensional: exploración de la variación en el Corpus PUCV-2003. *Revista Española de Lingüística*, 35 (1) (en prensa).
- Parodi, G. & Gramajo, A. (2003). Los tipos textuales del Corpus PUCV-2003: una aproximación multiniveles. *Signos*, 36 (54): 207-223.
- Parodi, G., Núñez, P. & Gramajo, A. (2003). LECTES, Programa para mejorar las competencias de lectura y escritura a través de la Web. *Ponencia leída en el II Congreso Internacional Cátedra UNESCO Lectura y Escritura. Comprensión y producción de textos escritos: de la reflexión a la práctica en el aula* (Mayo). Pontificia Universidad Católica de Valparaíso, Chile.
- Parodi, G., Sabaj, O. & Venegas, R. (2005). *Lingüística de corpus*. Buenos Aires: Eudeba, (en prensa).
- Parodi, G. & Venegas, R. (2004). BUCÓLICO: aplicación computacional para el análisis de textos. Hacia un análisis de rasgos de la informatividad. *Lingüística y Literatura*, 15: 223-251.
- Perelman, Ch. & Olbrecht-Tyteca, L. (1958/1994). *Tratado de la argumentación. La nueva retórica*. Madrid: Gredos.
- Pérez-Rioja, J. (1971). *Gramática de la lengua española*. Madrid: Tecnos.
- Peronard, M. (1994). La evaluación de la comprensión de textos escritos: el problema del resumen. *Lenguas Modernas* 21: 91-93.
- Peronard, M. (1998). ¿Qué significa comprender un texto escrito? En M. Peronard, L. Gómez, G. Parodi & P. Núñez (Eds.), *Comprensión de textos escritos: de la teoría a la sala de clases* (pp. 55-78). Santiago de Chile: Editorial Andrés Bello.
- Peronard, M. & Gómez, L. (1985). Reflexiones acerca de la comprensión lingüística: hacia un modelo. *Revista de Lingüística Teórica y Aplicada*, 23: 19-32.
- Peronard, M., Gómez, L., Parodi, G. & Núñez, P. (1998). *Comprensión de textos escritos: de la teoría a la sala de clases*. Santiago de Chile: Editorial Andrés Bello.
- Peronard, M., Marinovich, J. & Parodi, G. (2004). Comprensión y producción del discurso escrito: Aproximaciones multidisciplinarias. Ponencia presentada en las Primeras Jornadas Internacionales sobre comprensión del texto y del discurso, 9-11 Diciembre: Madrid.
- Picollo, M. (1999). La estructura del sintagma nominal: la nominalización y otros sustantivos con complementos argumentales. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 363-394). Madrid: Espasa Calpe.
- Plantin, Ch. (1998). *La Argumentación*. Barcelona: Ariel.
- Platón (1985a). *Gorgias*. Madrid: Altamira.
- Platón (1985b). *Fedro*. Madrid: Altamira.

- Polkinghorne, D. (1988). *Narrative knowing and human sciences*. Albany: State University New York Press.
- Portolés, J. (1998). *Marcores del discurso*. Barcelona: Ariel.
- Quilis, A. (1999). *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- R.A.E. (1973). *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa Calpe.
- Rabanal, A., & Contreras, L. (1986). El habla culta de Santiago de Chile. Materiales para su estudio. *Boletín de Filología*, 1, (Anexo 2): 46-57.
- Reppen, R. & Simpson, R. (2002). Corpus linguistics. En N. Schnitt (Ed.), *An introduction to applied linguistics* (pp. 92-111). New York: Arnold.
- Reppen, R., Fitzmaurice, S. & Biber, D. (2002). *Using corpora to explore linguistic variation*. Amsterdam: John Benjamins.
- Rojo, G. (2002). Sobre la lingüística basada en el análisis de corpus. *Hizkunza-corpusak. Oraria eta geroa*, 1-17.
- Rojo, G. (2001). La explotación de la base de datos sintácticos del español actual (BDS). En J. Kock (Ed.), *Lingüística con corpus. Catorce aplicaciones sobre el español* (pp. 255-286). Salamanca: Universidad de Salamanca.
- Rose, D. (1997). Science, technology and technical literacies. En F. Christie & J. Martin (Eds.), *Genre and institutions. Social processes in the workplace and school* (pp. 40-72). London: Continuum.
- Sager, J. & Dungworth, D. (1980). *English special languages*. Wiesbaden: Verlag.
- Sánchez, C. (1999). La negación. En I. Bosque & V. Demonte (Coords.), *Gramática descriptiva de la lengua española* (pp. 2561-2634). Madrid: Espasa Calpe.
- Santana, O., Pérez, F., Carreras, J., Duque, Z., Hernández R. & Rodríguez, G. (1997). FLAVER: flexionador y lematizador automático de formas verbales que trata conjugación y pronombres encílicos. *Lingüística Española Actual*, 19 (2): 229-282.
- Schröder, H. (1991). Linguistic and text-theoretical research on languages for special purposes. A thematic and bibliographical guide. En H. Schröder (Ed.), *Subject-oriented texts. Languages for special purposes and text theory* (pp. 1-48). Berlin-New York: Walter de Gruyter.
- Schröder, H. (Ed.) (1991). *Subject-oriented texts. Languages for special purposes and text theory. Research in text theory*. Berlin: Walter de Gruyter.
- Shannon, C. (1949). The mathematical theory of communication. En C. Shannon & W. Weaver (Eds.), *The mathematical theory of communication* (pp. 2-34). Illinois: The University of Illinois Press.
- Silvestri, A. (1995). *Discurso instruccional*. Encyclopedie Semiológica. Buenos Aires: Instituto de Lingüística.

- Simpson, R., & Swales, J. (2001). Introduction to North American perspectives on corpus linguistics at the millennium. En R. Simpson & J. Swales (Eds.), *Corpus linguistics in North America. Selections from the 1999 Symposium* (pp. 1-14). Ann Arbor: The University of Michigan Press.
- Sinclair, J. (1982). Reflections on computer corpora in English language research. En S. Johansson (Ed.), *Computer corpora in English language research* (pp. 1-6). Bergen: Norwegian Computing Centre for the Humanities.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1996). The Empty lexicon. *International Journal of Corpus Linguistics*, 1, 99-199.
- Stubbs, M. (1996). *Text and corpus analysis. Computer-assisted studies of language and culture*. Malden: Blackwell.
- Stubbs, M. (2000). Using very large text collections to study semantics schemas: a research note. [En línea]. Disponible en: <http://www.uni-trier.de/uni/fb2/anglistik/Projekte/stubbs/largetext.htm>
- Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Svartvik, J. (1992). [Ed.]. *Directions in corpus linguistics: Proceeding of Nobel symposium*. Berlin: Mouton Gruyter.
- Svartvik, J. (1992). Corpus linguistics comes of age. En J. Svartvik (Ed.), *Directions in corpus linguistics: Proceeding of Nobel symposium* (pp. 7-13). Berlin: Mouton Gruyter.
- Swales, J. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tottie, G. (1983). *Much about "not" and "nothing": a study of the variation between analytic and synthetic negation in contemporary American English*. Lund: CWK Gleerup.
- Toulmin, S. (1958). *The Uses of Argument*. Londres: Cambridge.
- Valencia, A. & Echeverría, M. (1999). *Disponibilidad léxica en estudiantes chilenos*. Santiago: Ediciones UCH & UDEC.
- Valencia, A. (2002). Aspectos del habla femenina de Santiago de Chile. En G. Parodi (Ed.), *Lingüística e interdisciplinariedad: desafíos del nuevo milenio. Ensayos en honor a Marianne Peronard* (pp. 439-456). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- van Dijk, T. (1983). *La ciencia del texto: un enfoque interdisciplinario*. Barcelona: Paidós.
- van Dijk, T. (1987). *La Ciencia del Texto*. Barcelona: Paidós.
- Vásquez, G. (2001). *Guía didáctica del discurso académico escrito. ¿Cómo se escribe una monografía?* Madrid: Edinumen.

- Veel, R. (1997). Learning how to mean -scientifically speaking: apprenticeship into scientific discourse in the secondary school. En J. Christie & J. Martin (Eds.), *Genre and institutions. Social processes in the workplace and school* (pp. 161-194). London: Continuum.
- Veel, R. (1998). The greening of school science: Ecogenesis in secondary classrooms. En J. Martin & R. Veel (Eds.), *Reading science. Critical and functional perspectives on discourses of science* (pp. 114-151). London: Routledge.
- Véliz, M. (2002). Desarrollo de estrategias de lectura por medio del computador: evolución de una experiencia. En G. Parodi (Ed.), *Lingüística e interdisciplinariedad: Desafíos del nuevo milenio. Ensayos en honor a Marianne Peronard* (pp. 161-178). Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso.
- Venegas, R. (2003). Los rasgos lingüísticos de la argumentatividad en corpora textuales. Ponencia leída en el XV Congreso de la Sociedad Chilena de Lingüística (Octubre). Santiago de Chile: Universidad Metropolitana de Ciencias de la Educación.
- Villaseñor, L., Montes, M., Pérez, M. & Vaufreydaz, D. (2002). Comparación léxica de corpus para generación de modelos de lenguaje. *Proceeding del IBERAMIA Workshop on Multilingual Information Access and Natural Language Processing*, [Noviembre]. Sevilla: España.
- Wagner, C. (1998). El Atlas lingüístico y etnográfico de Chile por regiones (ALECh). *Estudios Filológicos*, 33, 119-129.
- Wagner, C. (1999). Uso a cuestas. *Atlas lingüístico y etnográfico de Chile (ALECh)*. *Estudios Filológicos*, 34, 193-200.
- Warschauer, M. & Kern, R. (2000). Network-based language teaching. Concepts and practice. New York: Cambridge University Press.
- Weber, E. & Bentivoglio, P. (1991). Verbs of cognition in spoken Spanish: A discourse profile. En S. Fleischman & L. Waugh (Eds.), *Discourse pragmatics and the verb: The Evidence from Romance* (pp. 114-198). London: Routledge.
- Weinrich, H. (1974). *Estructura y función de los tiempos en el lenguaje*. Madrid: Gredos.
- Wiemer-Hastings, P., Graesser, A. & Wiemer-Hastings, K. (1998). Inferring the meanings of verbs in context. En M. Gernsbacher & S. Derry, (Eds.), *Proceedings of the twelfth annual conference of cognitive science* (pp. 1142-1147). Wisconsin: Lawrence Erlbaum.
- Zarzalejos, J. (2001). En el titular está el periódico, [En línea]. Disponible en <http://cvc.cervantes.es/obref/congresos/valladolid/presentaciones/elespanolensociedad/11aprensoenespanol/zarzalejos.htm>