



Estudios Filológicos

ISSN: 0071-1713

efil@uach.cl

Universidad Austral de Chile

Chile

Parodi, Giovanni

Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada  
en corpus computarizado

Estudios Filológicos, núm. 39, septiembre, 2004, pp. 7-36

Universidad Austral de Chile

Valdivia, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=173413833001>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

[Inicio Web Revistas](#) [Web Biblioteca](#) [Contacto](#)

**Revistas Electrónicas UACH**

Sistema de Bibliotecas UACH

Artículos [Búsqueda artículos](#)

[Tabla de contenido](#) [Anterior](#) [Próximo](#) [Autor](#) [Materia](#) [Búsqueda](#) [Inicio](#) [Lista](#)



## Estudios filológicos

ISSN 0071-1713 *versión impresa*

- [Como citar este artículo](#)
- [Agregar a favoritos](#)
- [Enviar a e-mail](#)
- [Imprimir HTML](#)

Estud. filol. n.39 Valdivia sep. 2004

---

Estudios Filológicos, N° 39, septiembre 2004, pp. 7-36

### Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada en corpus computarizado<sup>\*</sup>

Specialized texts and discourse technical-professional communities: a computerized corpus-based approach

**Giovanni Parodi**

Pontificia Universidad Católica de Valparaíso. Programas de Postgrado en Lingüística. Avda. Brasil 2830, 9 piso, Chile. E-mail: [gparodi@ucv.cl](mailto:gparodi@ucv.cl)

\* Investigación parcialmente financiada por el Proyecto FONDECYT N° 1020786.

---

Cualquier intento por apoyar la incorporación como miembros activos de futuros profesionales técnicos a sus respectivas comunidades discursivas está determinado por un manejo eficiente del discurso especializado a través del cual el conocimiento de cada ámbito se genera, transmite y disemina. El objetivo de esta investigación es realizar una descripción de los tipos de textos técnico-científicos que leen alumnos de 3 y 4 año de E.M. de tres colegios técnico-profesionales de diferentes ámbitos de especialización en la ciudad de Valparaíso, Chile. Para ello, se entregan el marco global del proyecto, el sustento metodológico, y se da cuenta de doce tipos de textos determinados a partir de criterios funcionales, situacionales y textuales.

corpus: el manual técnico-científico. Los resultados muestran gran heterogeneidad en los tipos de textos utilizados en todas las áreas técnicas. El manual técnico-científico revela cierta similitud en cuanto al comportamiento de los rasgos lingüísticos.

**Palabras clave:** discurso especializado, lingüística de corpus, manual técnico.

Any attempt to support active membership participation of technical professionals in discourse communities is determined by the efficient competence of the specialized discourse through which the knowledge of each domain is generated, transmitted, and disseminated. The objective of this investigation is to describe the technical-scientific text types that 3rd and 4th high school students read, when attending three professional schools in three different areas of specialization in the city of Valparaíso, Chile. In this article, a general research theoretical framework and methodological procedures are presented; also, twelve text types identified on functional, situational and textual criteria are described. Based on computational tools, we conducted a deep linguistic description of sixty five grammatical features in the most common text type detected in the corpus: technical-scientific manuals. The results show interesting text type heterogeneity among the specialized areas. Manuals reveal a degree of similarity in respect to the occurrence of the linguistic features involved.

**Key words:** specialized discourse, corpus linguistics, technical manual.

---

## 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 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, transmite, disemina y adquiere el conocimiento de cada ámbito. El conocimiento acabado de los textos que circulan tanto en esos ambientes educativos profesionales como en las organizaciones e instituciones laborales es prioritario para un proyecto de educación técnico-científica secundaria.

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

La llave de la cultura y de la sabiduría, cuestiones mucho más allá que la mera información, no reside en la cantidad de esta última que seamos 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 exigida 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

este punto, donde la lingüística y sus diversas ramas juegan un rol central. Los lingüistas debemos hacernos cargo del conocimiento deficiente disponible en determinadas áreas técnico-profesionales y buscar los medios que permitan, en este caso particular, 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.

Me parece relevante en este contexto poner de relieve que en Chile no conocemos los tipos de textos que se emplean en el sistema educacional secundario técnico-profesional diferenciado. Tampoco contamos 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 ha de 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 diversas ópticas, 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. 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.

De manera más específica, en este artículo me concentraré en una descripción general del proyecto, en una caracterización inicial de los corpora recolectados y en una descripción preliminar a partir de un grupo de rasgos lingüísticos de 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 al 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, entregaré una revisión de los pasos metodológicos efectuados. Finalmente, presentaré los resultados cuantitativos acerca de los tipos textuales detectados en el corpus y efectuaré una descripción y análisis de las frecuencias de ocurrencia de sesenta y cinco rasgos lingüísticos en veinticuatro manuales técnicos en tres ámbitos de especialidad. Cerraré el trabajo con proyecciones para el área y el proyecto en cuestión.

## 1. MARCO DE REFERENCIA

DISCURSO ESPECIALIZADO. No cabe duda de que la investigación en torno al llamado discurso especializado, en particular aquella que se centra en sus diferentes modalidades de divulgación científico-tecnológica ([Gläser 1982](#); [Calsamiglia 2000](#); [Ciapuscio 1994, 2000, 2003](#); [Cassany, López y Martí 2000](#); [López 2002](#); [Gunnarsson, Linell y Nordberg 1997](#)) y en su uso al interior de la sala de clases en el nivel secundario profesional constituye un elemento nuclear y estratégico en la planificación lingüística y en el proceso de alfabetización progresiva de estas comunidades discursivas en formación ([Christie y Martin 1997](#); [Veel 1998](#); [Otero, León y Graesser 2002](#); [Biber, Reppen, Clark y Walter 2001](#); [Cameron 2003](#)).

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 lo que ocurre toda vez que requerimos tipologizar o clasificar 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

2002). Estoy cierto de que no se debe simplificar lo que no corresponde, pero un poco de orden para sobrevivir al caos parece sano.

En este contexto, 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 (Gläser 1982; Martin, Christie y Rothery 1987; Halliday y Martin 1993; Rose 1997; Christie 1998; Goldman y Bizans 2002; Cameron 2003; Cabré 2002; Ciapuscio 2003). Desde esta óptica, coincidimos 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, de las tradiciones de una comunidad discursiva específica (Brinker 1988; Schröder 1991).

Determinar dicotómicamente si un texto se clasifica como de especialidad o no-especialidad o de tipo general es por supuesto un problema teórico y descriptivo. La precisión de las categorías a partir de las cuales debería organizarse el sistema de oposiciones, rasgos discretos y valores se vuelve su gran debilidad. En este respecto, existe consenso hoy en día en una postura en favor de un *continuum* (Gläser 1982; Schröder 1991; Halliday 1993; Jeanneret 1994; Peronard 1998; Ciapuscio 1994, 2000; Cabré 2002; Goldman y Bisanz 2002), proveniente de lo difícil que resulta aceptar en la realidad la idea de límites estrictos cuando en verdad son más bien difusos (*fussy categories*) (Lakoff 1972). 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 límites; pero de cualquier modo el discurso especializado debería ser una categoría reconocible para cualquier hablante de español (Parodi y Gramajo 2003).

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 acudo a los tres aspectos del contexto situacional o tres variables del registro. Ellas como bien se sabe son campo, tenor y modo (Halliday 1985, 1998; Martin 1986, 1992, 1997; Halliday y Martin 1993). En la primera, se considera el espacio institucional o de acción social y se centra en el tópico 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 esta se presente de manera oral o escrita (modo).

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

Campo: tecnolectos, procesos descriptivos, taxonomías.

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):

Any variety of language, whether functional or dialectal, occupies an extended space, a region whose boundaries are fussy and within which there can be considerable variation. But it can be defined and recognized, by certain syndromes, patterns of co-occurrence among features at one or another linguistic level typically features of the expression in the case of a dialect, features of the content in the case of functional variety of 'register'. Such syndromes are what makes it possible to talk of 'language of science'.

Dentro de los discursos de especialidad, los síndromes de rasgos lingüísticos varían en cada

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 supuestos 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 y Gómez 1985; Gómez 1998; Parodi 2003a](#)).

LINGÜÍSTICA DE CORPUS: ¿PARADIGMA O METODOLOGÍA? Durante un prolongado período de tiempo, el enfoque generativo transformacional de [Noam Chomsky \(1957, 1965\)](#) influyó en la postergación de diversos programas de investigación enfocados al estudio de la variación lingüística y opacó los desarrollos alternativos en otros polos del mundo académico ([Francis 1979; Sinclair 1982, 1991; Aijmer y Altenberg 1991; McEnery y 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 desconocimiento o falta de reconocimiento al rol que este lingüista desempeñó en el escenario de quiebres 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 generan 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 (por ejemplo, SEU Corpus, Cobuild Corpus, LL Corpus, Brown Corpus); ellos constituyen el eje de avanzada de esta nueva reposición. 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.

[Sinclair \(1991: 1\)](#) ilustra claramente lo acontecido en lingüística debido a lo limitado del enfoque generativista:

Starved of adequate data, linguistics languished indeed it became almost totally introverted. It became fashionable to look inwards to the mind rather than outwards to society. Intuition was the key, and the similarity of language structure to various formal models was emphasized. The communicative role of language was hardly referred to.

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, generó un impacto que influyó negativamente en ciertos desarrollos en lingüística ([Leech 1991; Sinclair 1991; Stubbs 1996; McEnery y Wilson 1996; Kennedy 1998](#)). Ello debido, en parte, a que se sostendría que 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 ([Red. 2003; Jurafsky](#)



Como se sabe, la noción de probabilidad proviene de la teoría matemática de la comunicación, propuesta inicialmente por [Shannon \(1949\)](#), y se basa en los modelos de cadenas de Markov. Desde este marco, los enfoques probabilísticos o estocásticos para el estudio del lenguaje humano se basan en técnicas estadísticas a partir de corpus de lenguaje natural, los que se apoyan y se oponen, por ende, a modelos gobernados por reglas lógicas absolutas propuestas apriorísticamente ([Oakes 1998](#); [Jurafsky y Martin 2000](#)).

Ahora bien, los requerimientos de análisis semiautomáticos y exhaustivos de textos sobre la base de herramientas computacionales (tales como etiquetadores morfosintácticos de tipo *tagger* y *parser*) derivó en descripciones en términos probabilísticos y llevó al desarrollo de gramáticas independientes del contexto (*context-free-grammars*). En 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. La probabilidad de una secuencia de palabras se determina por la suma de las probabilidades individuales de todas las estructuras. 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 esas estructuras ([Halliday 1992](#); [Aarts 1991](#); [Stubbs 1996](#)). Una de las contribuciones clave de gramáticas y de etiquetadores probabilísticos es que permiten la desambiguación en el análisis lingüístico eligiendo la alternativa más plausible, según sea el caso. Tal elección se basa en las frecuencias relativas de ocurrencia de todas las combinaciones posibles en un corpus representativo ([Briscoe y Carroll 1994](#); [Minker, Waibel y Mariani 1999](#); [Jurafsky y Martin 2000](#); [Bunt y Nijholt 2000](#)).

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 venimos enfrentando un renacimiento del empirismo sin los resabios de la lingüística estructural de corte behaviorista ni de la psicología conductista imperante en los años cincuenta. La oposición entre métodos basados en el conocimiento ([Church y Mercer 1993](#)) y métodos empiricistas, así como la oposición entre una "lingüística del sillón" versus una "lingüística de corpus" ([Fillmore 1992](#)), son distinciones dicotómicas que debieran repensarse en busca de posturas más colaborativas e integradoras.

[Church y Mercer \(1993: 21-22\)](#) explican que existen tres desarrollos que han gatillado 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 y Biber 1998](#); [Chafe 1992](#); [Sinclair 1991](#); [Leech 1991](#), [Kennedy 1998](#); [McEnery y Wilson 1996](#); [Moreno 1998](#); [Graesser, Gernsbacher y 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 \(1996: 231\)](#):

what is typical in the language, and which can use findings about frequently recurring patterns to construct a theory of the relation between routine and creative language use.

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 \(1996\)](#), 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 diversas (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, indaga 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 uno u otro tipo de lingüística y, en definitiva, se inclina por denominaciones más genéricas que no provoquen disputas clásicas: introspección/experimentación ([Chafe 1992, 1994](#)).

En el marco de lo dicho hasta aquí, cabe la pregunta: ¿Cómo pueden los investigadores del discurso beneficiarse con la aplicación de métodos asistidos por computadores? Una respuesta es que los avances recientes en tecnologías computacionales han hecho accesible un análisis sistemático de grandes cantidades de textos de manera rápida y con creciente confiabilidad, es decir, que la tecnologización de la investigación se ha tornado un aliado de gran utilidad. Existe gran variedad de posibilidades para los lingüistas interesados en contrastar sus hipótesis acerca del lenguaje a través de evidencia empírica a gran escala, superando en parte los estudios de textos ejemplares. También es una buena alternativa para validar hallazgos iniciales a partir de pequeñas muestras de textos que luego se trabajan con corpus extensos. Un principio empírico no extraordinariamente novedoso, pero sí muy valioso que guía estas investigaciones dice que en cuanto mayor sea la cantidad de información disponible para el análisis, mayor poder tendrán las conclusiones a las que se arribe. 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 y Mercer 1993](#)). Al respecto, [Biber, Conrad y Reppen \(1998: 106\)](#) opinan que: "Most studies do not use quantitative methods to describe the extent to which different discourse structures are used, and relatively few of these studies aim to produce generalizable findings that hold across texts".

Ahora bien, aunque en el mundo hispánico se detectan interesantes avances en investigación con textos auténticos lo mismo que polos de desarrollo tecnológico, en ciertos casos con equipos multidisciplinarios ([Santana, Pérez, Carreras, Duque, Hernández y Rodríguez 1997](#); [Rojo 2001](#); [Villaseñor, Montes, Pérez y Vaufreydaz 2002](#); [Ferreira, Campos y Ruggeri 1998](#); [Ferreira-Cabrera y Atkinson-Abutridy 2002](#); [Lorente 2002](#); [Echeverría 2002](#); [Véliz 2002](#); [Núñez, Gramajo y Parodi 2003](#); [Parodi y Gramajo 2003](#)), es muy cierto que aún la tecnologización de la investigación es escasa y que tanto el desarrollo de programas computacionales como de tecnologías requeridas todavía no se encuentran accesibles de manera expedita. En mi opinión, la superación de esta barrera metodológica y tecnológica no puede esperar si queremos, efectivamente, producir investigación competitiva y de primer orden; la docencia de pregrado y de postgrado exigen que así sea para que entre otros 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 y Biber 1998](#); [Simpson y Swales 2001](#)). En nuestro caso, los costos implicados son de relevancia, y los apoyos institucionales y académicos de investigación son escasos, contemplando un gran porcentaje de infraestructura



Fundación ANDES, ECOS, ALFA, y otros).

En suma, en esta investigación adherimos a los principios metodológicos de la lingüística de corpus, cuyo marco de acción resulta muy útil para indagar fenómenos del lenguaje en uso y su 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 en el sentido de [Khun \(1970\)](#), 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 caminos por los que transitará. Al respecto, en un artículo en homenaje a [Jan Svartik, Johansson \(1991: 313\)](#) concluía así:

In spite of the great changes in the less than three decades since the first computer corpus, there is one way in which the role of the corpus in linguistic research has not changed. The corpus remains *one* of the linguist's tools, to be used together with introspection and elicitation techniques. Wise linguists, like experience craftsmen, sharpen their tools and recognize their appropriate uses. It is no coincidence that Jan Svartik has done distinguished work in the areas of both corpora and elicitation. His example is worth bearing in mind, whatever further changes may follow in the future.

## 2. PROYECTO EN MARCHA: MARCO GENERAL Y PASOS INICIALES

En un sentido muy amplio, dentro de este programa de investigación en desarrollo nos interesa el estudio del discurso de especialidad, en su modalidad de divulgación didáctica, como eje de la alfabetización de jóvenes estudiantes de educación secundaria técnico-profesional diferenciada y en su posterior relación con su vida laboral ([Gläser 1982](#); [Halliday y Martin 1993](#); [Martin 1997, 1998](#); [Rose 1997](#); [Veel 1998](#); [Cameron 2003](#); [Lassen 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 el profesor entrega a sus alumnos para que sean leídos y así accedan al conocimiento disciplinar. También es escasa o nula la información disponible acerca de la vinculación entre esta formación lingüística especializada y los materiales que circulan en el 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 estas tendrán relación en su modo de incorporarse a su actividad profesional y de compartir con sus pares. Del mismo modo, entre otros, nos preocupa el grado de conciencia que posean los profesores de especialidad en cuanto al valor que ellos asignen a los textos escritos, por ejemplo, como generadores de la progresiva inscripción en la respectiva comunidad discursiva o como meros objetos reproductores de saberes y no de competencias.

### 2.1. LA INVESTIGACIÓN: OBJETIVOS Y METODOLOGÍA

*Objetivos.* A continuación se entregan los objetivos generales del programa de investigación. Cabe puntualizar que en este artículo sólo algunos de ellos serán abordados de manera parcial:

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

Buscar criterios para clasificar los textos según una aproximación multiniveles: lingüística, funcional y comunicativa ([Parodi y Gramajo 2003](#)).

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 ([Parodi 2003b](#); [Parodi y Gramajo 2003](#); [Marinkovich 2003](#)).

Comparar el corpus técnico-científico (CTC) internamente con el fin de detectar diferencias y/o similitudes entre los textos que lo componen ([Parodi y Gramajo 2003](#); [Venegas 2003](#)).

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.

*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 metalmecánico (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 enfrentan 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 entregan listas de libros, artículos o apuntes existentes en la biblioteca escolar.

Se entregan apuntes o artículos fotocopiados en la clase.

Se entregan 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 entregan 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 de los *corpora* 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 *corpora* es factible de manera parcial.

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

1) La representación del universo de textos leídos por los alumnos de las tres

2) El tamaño de las muestras (en cuanto al número de textos y número de palabras) para asegurar la utilización de los programas estadísticos requeridos.

3) 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 cuyo apoyo desinteresado no habría sido posible tener acceso a los textos ni menos lograr los objetivos iniciales en breve plazo.

*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](#); [Ciapuscio 2000, 2003](#); [Ciapuscio y Kugel 2002](#)). La idea era construir una primera matriz de rasgos multiniveles (por ausencia o presencia) que nos orientara en la diferenciación y descripción de los tipos de textos recolectados. Buscábamos elaborar una tipología multiniveles 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 psicosociolingüística 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	Audiencia original	Autor explícito	Función comunicativa	Estructura textual	Tema	Multimedialidad	Requiere Escritura
Escolar Laboral	Lego Semilego Experto	Sí o No	Referencial Expresiva Apelativa Conativa Poética Metalingüística	Expositiva Argumentativa Normativa	Monotemático Pluritemático	Sí o No	Sí 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

recurrir a denominaciones clásicas para los tipos textuales que esperamos en la mayoría de los casos los alumnos y/o profesores del sistema institucional reconozcan y posiblemente utilicen intuitivamente. Ellas son las siguientes:

01. Artículo Técnico
02. Descripción Técnica
03. Diagrama
04. Formulario
05. Glosa Legal
06. Glosario
07. Guía Didáctica
08. Instructivo
09. Ley
10. Manual Técnico
11. Reglamento
12. Tabla

Por razones de espacio y foco, en este trabajo nos hemos limitado a una presentación esquemática de criterios y rasgos de clasificación de los textos del proyecto. Una descripción pormenorizada acerca de los criterios tipológicos empleados y definiciones de cada tipo se encuentran en [Parodi \(2002\)](#); [Cademártori, Gramajo y Parodi \(2003\)](#) y [Parodi y Gramajo \(2003\)](#).

*Metodología 3: Análisis multirrasgos (AMR) y Análisis Multidimensional (AMD).* Como marco de referencia metodológico para investigar la variabilidad de rasgos lingüísticos entre textos de los *corpora*, adscribimos en líneas generales al Análisis Multirrasgos (AMR) y Multidimensiones (AMD) propuesto por [Biber y Finegan \(1986\)](#) y [Biber \(1988\)](#). Este enfoque metodológico fue desarrollado originalmente para el estudio analítico de las variaciones entre registros, con el objetivo particular de: (1) identificar 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 ([Biber 1988, 1994](#); [Biber, Conrad y Reppen 1998](#); [Conrad y Biber 1998](#); [Reppen, Fitzmaurice y Biber 2002](#)). En nuestro caso particular, el enfoque presenta gran utilidad para describir y comparar los textos de las tres especialidades tanto entre tipos textuales preliminarmente determinados como entre los ámbitos mismos.

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 funciones 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-ocurre frecuentemente en ciertos textos, porque ellos son usados para expresar un conjunto de funciones comunicativas. Estos conjuntos de rasgos, determinados estadísticamente por medio de técnicas multivariadas ([Hair, Anderson, Tatham y Black 1999](#)), son lo que se denomina una dimensión; en otras palabras, una dimensión se constituye mediante la interpretación funcional del conjunto de los rasgos co-ocurrentes (factor), agrupados entre otros mediante análisis factorial.

Ahora bien, el AMR y el AMD se encuadran metodológicamente dentro de la lingüística de corpus ([Sinclair 1991](#); [Leech 1991](#); [Svartvik 1992](#)). Ello implica el diseño de corpus representativo, programas computacionales automáticos para etiquetar morfosintácticamente los textos y programas estadísticos para determinar las relaciones entre las variables. [Biber, Reppen, Clark y Walter \(2001\)](#) aportan cuatro ventajas importantes para optar por un enfoque desde 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 (semi)automático de los textos mediante el uso de

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. Posteriores investigaciones pueden utilizar los mismos *corpora*, u otros *corpora* pueden ser analizados con las mismas herramientas computacionales.

*Marcación estructural 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) de Barcelona, 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 en el IULA por parte del equipo chileno-catalán consistió en:

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

En virtud del convenio firmado con el IULA de la UPF, los documentos digitalizados del corpus PUCV-2003 se encuentran disponibles a través de una interficie computacional en Internet (Bwananet). Ello permite un acceso directo y expedito a la totalidad de los textos recopilados y a una diversidad de alternativas de interrogación con resultados cuantificados y ejemplificados en cada caso, limitadas exclusivamente por el tipo de marcación 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 se encuentra en el estadio final de prueba de 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) proveer alternativas de análisis no contempladas en la interficie del ILUA (Bwananet), tales como N-grams, 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 una herramienta comparativamente rudimentaria con los avances tecnológicos detectados fuera de Latinoamérica, pero se constituye en un primer avance en el diseño de instrumentos computacionales elaborados en el marco del programa en desarrollo ([Parodi y Venegas 2004](#)).

*Rasgos lingüísticos.* Respecto a los rasgos lingüísticos que se indagarán, se decidió elaborar un conjunto inicial de dieciséis categorías representativas de características gramaticales y funcionales del español. Ellos se rastrearon a partir de bibliografía relevante en el tema. El listado de estos dieciséis 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 adjetivos
G. Especificidad léxica	Ñ. Marcadores de coordinación
H. Formas estativas activas	O. Marcadores de negación

A partir de estas dieciséis 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 gramatical y funcional. Paralelamente, la matriz de rasgos debió ser cotejada con las posibilidades de marcación e interrogación que brindan los procedimientos tecnológicos del IULA de la Pompeu Fabra. 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**

<p><b>A. Marcadores de tiempo verbal</b></p> <ol style="list-style-type: none"> <li>1. Pretérito indefinido (indicativo)</li> <li>2. Pretérito imperfecto (indicativo)</li> <li>3. Pretérito perfecto (indicativo y subjuntivo)</li> <li>4. Presente (indicativo y subjuntivo)</li> <li>5. Futuro (indicativo y subjuntivo)</li> <li>6. Futuro perifrástico</li> </ol> <p><b>B. Marcadores de modo verbal</b></p> <ol style="list-style-type: none"> <li>7. Indicativo/imperativo</li> <li>8. Subjuntivo/imperativo</li> <li>9. Modo indicativo</li> <li>10. Modo subjuntivo</li> <li>11. Modo imperativo</li> </ol> <p><b>C. Desinencias verbales de persona</b></p> <ol style="list-style-type: none"> <li>12. Primera singular</li> <li>13. Segunda singular</li> <li>14. Tercera singular</li> <li>15. Primera plural</li> <li>16. Segunda plural</li> <li>17. Tercera plural</li> </ol> <p><b>D. Pronombres personales</b></p> <ol style="list-style-type: none"> <li>18. Primera persona singular</li> <li>19. Primera persona plural</li> <li>20. Segunda persona singular</li> <li>21. Segunda persona plural</li> <li>22. Tercera persona singular</li> <li>23. Tercera persona plural</li> <li>24. Demostrativo</li> </ol> <p><b>E. Formas nominales</b></p> <ol style="list-style-type: none"> <li>25. Nominalizaciones</li> <li>26. Sustantivos (comunes y propios)</li> </ol> <p><b>F. Formas Pasivas</b></p> <ol style="list-style-type: none"> <li>27. Pasivas con "se"</li> <li>28. Pasivas con "ser" sin agente</li> <li>29. Pasivas con "ser" con agente</li> <li>30. Pasivas con "estar"</li> </ol> <p><b>G. Especificidad léxica</b></p> <ol style="list-style-type: none"> <li>31. Relación clase/tipo (<i>type/token</i>) por forma</li> <li>32. Relación clase/tipo (<i>type/token</i>) por lema</li> </ol>	<p><b>H. Formas estativas activas</b></p> <ol style="list-style-type: none"> <li>33. "Ser"</li> <li>34. "Estar"</li> </ol> <p><b>I. Tipos verbales</b></p> <ol style="list-style-type: none"> <li>35. Públicos</li> <li>36. Privados</li> <li>37. Persuasivos</li> <li>38. Perceptivos</li> </ol> <p><b>J. Verbos modales</b></p> <ol style="list-style-type: none"> <li>39. Posibilidad</li> <li>40. Necesidad</li> <li>41. Obligación</li> <li>42. Volición</li> </ol> <p><b>K. Marcadores de modalidad</b></p> <ol style="list-style-type: none"> <li>43. Atenuadores</li> <li>44. Enfáticos</li> </ol> <p><b>L. Adverbios</b></p> <ol style="list-style-type: none"> <li>45. De lugar</li> <li>46. De tiempo</li> <li>47. De modo</li> <li>48. De cantidad</li> </ol> <p><b>M. Marcadores de subordinación</b></p> <ol style="list-style-type: none"> <li>49. Subordinadas sustantivas con "que"</li> <li>50. Subordinadas adjetivas pron. relativo</li> <li>51. Subordinadas adverbiales de razón o c/e</li> <li>52. Subordinadas adverbiales de concesión</li> <li>53. Subordinadas adverbiales condicionales</li> <li>54. Subordinadas adverbiales de tiempo</li> <li>55. Frases infinitivas en función nominal</li> </ol> <p><b>N. Frases preposicionales y adjetivos</b></p> <ol style="list-style-type: none"> <li>56. Frases prep. (compl. del nombre)</li> <li>57. Adjetivos atributivos (calificativos)</li> <li>58. Adjetivos predicativos</li> <li>59. Adjetivos demostrativos</li> <li>60. Participio función adjetiva</li> </ol> <p><b>Ñ. Marcadores de Coordinación</b></p> <ol style="list-style-type: none"> <li>61. Conjunciones adversa., adit. y disyun.</li> </ol> <p><b>O. Marcadores de negación</b></p> <ol style="list-style-type: none"> <li>62. Adverbio de negación</li> <li>63. Adverbios de tiempo</li> <li>64. Conjunción de negación</li> <li>65. Pronombres de negación</li> </ol>
--	--

Tal como se dijo en la introducción de este artículo, en los apartados precedentes se ha descrito el marco referencial teórico y metodológico y los objetivos generales del programa de investigación en curso. 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.

*3.1. 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 técnico-profesionales de la ciudad de Valparaíso, Chile. El detalle de esta información se entrega en la [tabla 4](#):

**Tabla 4**  
**Constitución del CTC**

Ambito 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%)
<b>Totales</b>	<b>74 (100%)</b>	<b>626.790 (100%)</b>

Los datos de la [tabla 4](#) permiten señalar que no existe una relación directa entre el número de textos por área temática de especialización y el correspondiente número de palabras. Esto quiere decir que, por ejemplo, en el ámbito Marítimo (Especialidad Operación Portuaria) si bien se registra la mayor cantidad de textos (49% del total), se detecta el menor número de palabras (25% del total); por el contrario, y de manera interesante, en el área técnica Metalmeccánica (Especialidad Mecánica Industrial) se recolectó el grupo más reducido de textos (sólo 18 y que constituyen el 24% de lo recopilado), no obstante lo cual conforman la muestra más grande en cuanto al número de palabras (39% del total). El área de Contabilidad de Comercio, por su parte, arroja cifras similares a las anteriormente descritas: un total de 20 textos (27% del total) con un número elevado de palabras (36% del total). Por una parte, las cifras entregadas revelan una cierta heterogeneidad respecto a la configuración del corpus según cada uno de los tres ámbitos, y, por otra, demuestran 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 pudiera 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.

*3.2. Los tipos textuales en el CTC.* A partir de los doce tipos de textos técnico-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.

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

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 (da 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).

**Tabla 5**  
**Ocurrencia de las clases textuales en los tres corpora del CTC**

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	3.286	2	5.719	9*	17.864
Ley	-	-	-	-	3	67.905	3	67.905
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
<b>Totales</b>	<b>36</b>	<b>155.160</b>	<b>18</b>	<b>246.374</b>	<b>20</b>	<b>225.256</b>	<b>74</b>	<b>626.790</b>

\* Tipos de textos con mayor frecuencia de ocurrencia en un área de especialidad.

Según se observa, la mayor heterogeneidad tipológica se registra en el área Marítima 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). 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.

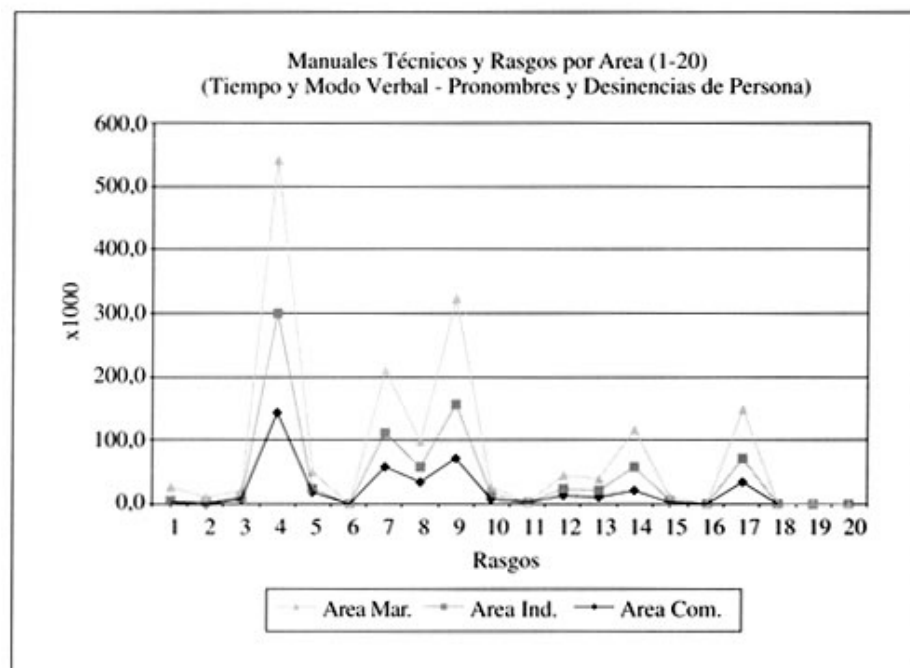
3.3. *El Análisis Multirrasgos (AMR) en los manuales del CTC.* Ahora bien, con el fin de mostrar preliminarmente la utilidad de contar con un número importante de rasgos gramaticales para la indagación profunda de las características de los textos recolectados, a continuación se entregan tres gráficos que muestran la frecuencia de ocurrencia de las sesenta y cinco variables lingüísticas desagregadamente. Dado 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 anteriormente, se da cuenta de la relación

Con el fin de que la comparación de los datos sea representativa, estos 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. Para efectos de una presentación visualmente más comprensible, se han 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 (presentes, pretéritos y futuros) y de modo verbal, de desinencias verbales de persona (primera, segunda 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/imperativo y subjuntivo/imperativo, 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 puntaje 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 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 multiniveles está caracterizado con un foco referencial y una posible función secundaria de índole apelativa. Algunos de estos rasgos apuntan también a una estructura textual predominante de tipo expositiva-normativa.

Gráfico 1  
Primeros 20 rasgos por área de especialidad

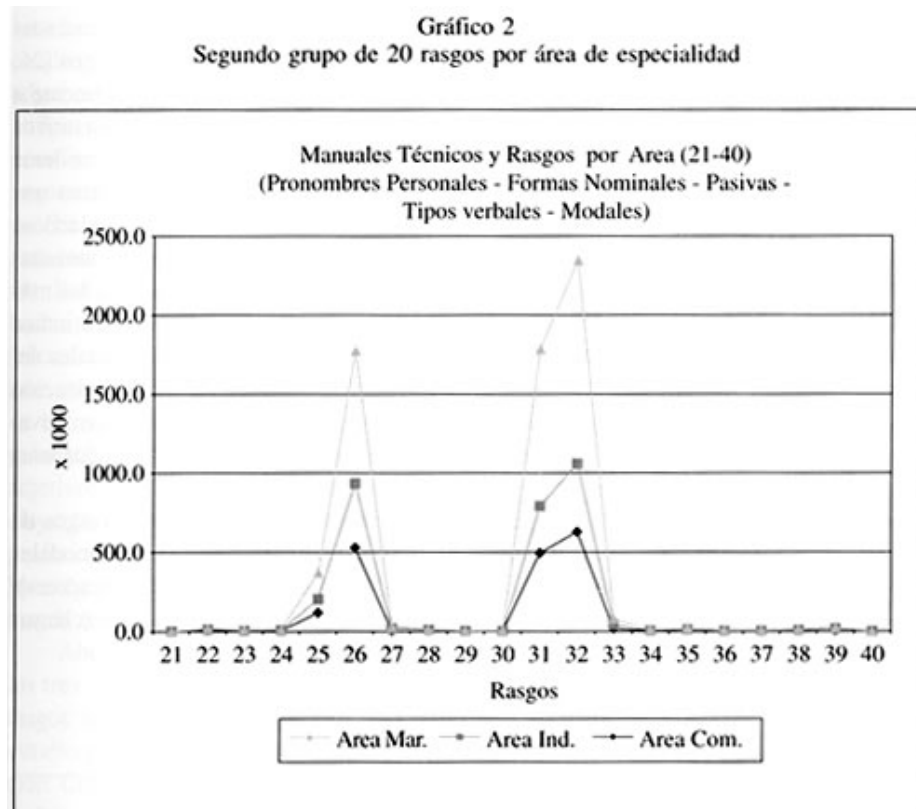




frecuencias de ocurrencia de los pronombres personales de primera (singular y plural) y de segunda (singular). No es extraño que en este tipo de texto sea regularmente escasa la presencia explícita de pronombres personales (rasgos 18, 19 y 20); bien sabido es que en el español escrito académico la tendencia 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 personales explícitos (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).

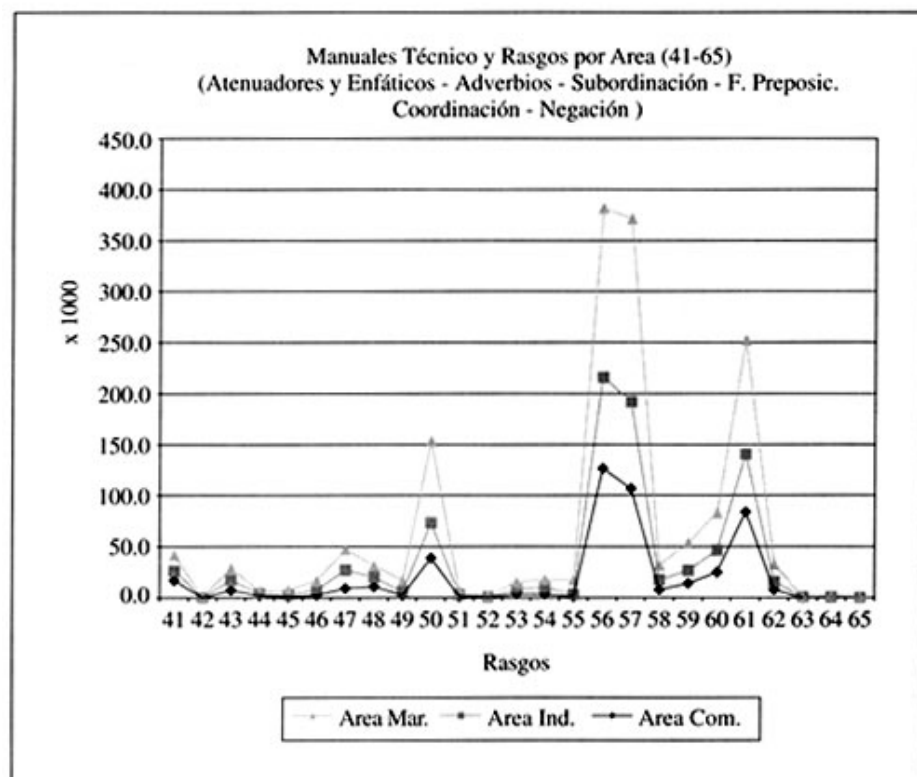


Como se anticipara, los datos de este gráfico revelan una distribución bastante sistemática en las frecuencias de los rasgos involucrados, hecho que, por un lado, entrega información relevante acerca del tipo de texto especializado que estoy describiendo y, por otro, permite inferir un impacto mayor en el área técnico-profesional de la clase textual "manual". 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 ([Lassen 2003](#)). Por supuesto que todo ello habrá de comprobarse, por un lado, con herramientas estadísticas y

Junto a lo anteriormente dicho, no es menor que 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/tipo 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 destaquen 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 volición), 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,



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; solo 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, las frases preposicionales como complemento del nombre, los adjetivos atributivos y las conjunciones en tres tipos (adversativas, aditivas y disyuntivas). Son ellas las que producen esta distinción entre los ámbitos técnico-profesionales. 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 variadas partes del mundo, y Latinoamérica no es una excepción. A pesar de ello, tal como puntalicé en la introducción a este artículo, 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.

Desde este análisis, espero que información como la aportada en este informe científico contribuya paulatinamente a elaborar un perfil lingüístico de los alumnos de áreas técnico-profesionales en la educación diferenciada secundaria. Por cierto que en este estudio he aportado información parcial basada en apenas tres establecimientos educativos y sólo correspondiente a 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 homogeneizador es peligroso. En este sentido, espero que los resultados preliminares que entrego constituyan un pequeño 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

La metodología utilizada en la determinación de las clases textuales del corpus PUCV-2003 desde una óptica 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 ([Parodi y Gramajo 2003](#)). Por otra, la aproximación multirrasgos aplicada de manera preliminar 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 aglutinarse con un comportamiento relativamente homogéneo y con ello prueban que constituyen lo que inicialmente y, en base a otros criterios, se definió como un tipo textual particular.

La aparente distinción por área temático-profesional al interior de la clase MT, producida por la mayor ocurrencia y posible significatividad estadística de diez rasgos lingüísticos en el ámbito Marítimo de especialidad Operación Portuaria, apoya la idea de que el ámbito de especialización podría incidir al interior de una clase textual tal como la del MT. Aunque son un número menor en relación con el total de los sesenta y cinco rasgos lingüísticos explorados, estos diez rasgos tendrían relevancia en la detección de variabilidad emergente en comparación con las otras dos áreas de especialización (comercial e industrial) cuyo comportamiento es 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 computacional de corpus, 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 trabajadas simultáneamente como ha sido el caso de esta investigación brindan una mayor confiabilidad y "robustez" a las conclusiones propuestas y en definitiva 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 suyo 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.

Finalmente, quiero destacar que un objetivo emergente a partir de este artículo lo constituyó la ampliación de los corpus del proyecto original. Con el fin de alcanzar una descripción acuciosa del corpus técnico-científico (CTC), se optó por recolectar y marcar estructuralmente 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 basados en 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-profesionales 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 aportará de manera

## OBRAS CITADAS

Aarts, J. 1991. "Intuition-based and observation-based grammars". *English Corpus Linguistics. Studies in honor of Jan Svartvik*, eds. K. Aijmer & B. Altenberg. London: Longman. 44-62.

Aijmer, K. & B. Altenberg, eds. 1991. *English Corpus Linguistics. Studies in honor of Jan Svartvik*. London: Longman.

Biber, D. & E. Finegan. 1986. "An initial typology of English text types". *Computer linguistics. Recent developments in the use computer corpora in English language research*. Eds. J. Aarts & W. Meijs. Amsterdam: Rodopi. 19-46.

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

Biber, D. 1994. "Using register-diversified corpora for general language studies". *Using large corpora*. Ed. S. Armstrong. Cambridge, Massachusetts: The MIT Press. 180-201.

Biber, D., R. Reppen, V. Clark & J. Walter. 2001. "Representing spoken language in university settings: The design and construction of the spoken component of the T2K-SWAL Corpus". *Corpus Linguistics in North America. Selections from the 1999 Symposium*. Eds. R. Simpson & J. Swales. Ann Arbor: The University of Michigan Press. 48-57.

Biber, D., S. Conrad & R. Reppen. 1998. *Corpus Linguistics Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Bod, R. 2003. "Introduction to Elementary Probabilistic Theory and Formal Stochastic Language Theory". *Probabilistic Linguistics*. Eds. R. Bod, J. Hay & S. Jannedy. London: The MIT Press. 11-37.

Brinker, K. 1988. *Linguistische Textanalyse*. Berlin: E. Schmidt.

Briscoe, T. & J. Carroll. 1994. "Generalized probabilistic LR parsing of natural language (corpora) with unification-based grammars". *Using large corpora*. Ed. S. Armstrong. Cambridge, Massachusetts: The MIT Press. 25-59.

Bunt, H. & A. Nijholt. 2000. "New parsing technologies". *Advances in probabilistic and other parsing technologies*. Eds. H. Bunt & A. Nijholt. The Netherlands: Kluwer Academic Publishers. 1-12.

Cabré, M. 2002. "Textos especializados y unidades de conocimiento: metodología y tipologización" (I). *Texto, terminología y traducción*. Eds. J. García & M T. Fuentes. Barcelona: Almar. 122-187.

Cademartori, Y., A. Gramajo & g. Parodi. 2003. *Texts classes in three specialized areas in technical secondary education*. Meeting of the Society for Text & Discourse, Madrid, (July). Madrid: Universidad Autónoma de Madrid.

Calsamiglia, E. 2000. "Decir la ciencia: Las prácticas divulgativas en el punto de mira". *Revista Iberoamericana del Discurso y Sociedad* 2.2: 39-71.

Cameron, L. 2003. *Metaphor in educational discourse*. London: Continuum.

Cassany, D., C. López & J. Martí. 2000. "La transformación divulgativa de redes conceptuales científicas. Hipótesis, modelos y estrategias". *Revista Iberoamericana del Discurso y Sociedad* 2.2: 73-103.

Ciapuscio, G. 1994. *Tipos textuales*. Buenos Aires: Eudeba.

Ciapuscio, G. 2000. "Hacia una tipología del discurso especializado". *Revista Iberoamericana*

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

Ciapuscio, G. & I. Kuguel. 2002. "Hacia una tipología del discurso especializado". *Texto, terminología y traducción*. Eds. J. García & M T. Fuentes. Salamanca: Almar. 37-73.

Conrad, S. & D. Biber. 1998. "Multi-dimensional methodology and the dimensions of register variation in English". *Variation in English: Multi-dimensional Studies*. Eds. S. Conrad & D. Biber. London: Longman. 13-42.

Chafe, W. 1992. "The importance of corpus linguistics to understand the nature of language". *Directions in Corpus Linguistics*. Ed. J. Svartvik. Berlin: Mouton de Gruyter. 79-97.

Chafe, W. 1994. *Discourse, consciousness, and time*. Chicago: The University of Chicago Press.

Chomsky, N. 1957. *Syntactic structure*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.

Christie, F. & J. Martin, eds. 1997. *Genre and Institutions. Social processes in the workplace and school*. London: Continuum.

Christie, F. 1998. "Science and apprenticeship. The pedagogic discourse". *Reading science. Critical and functional perspectives on discourse of science*. Eds. J. Martin & R. Veel. London: Routledge. 152-180.

Church, K. & R. Mercer. 1993. "Introduction to the special issue on computational linguistics. Using large corpora". *Computational Linguistics* 9.1: 1-24.

Echeverría, M. 2002. "Programas computacionales para el español como lengua materna". *Signos* 35.51-52: 163-193.

Ferreira, A., D. Campos & E. Ruggeri. 1998. "VERBUM: Una aplicación multimedial para la enseñanza del Latín". *Estudios Clásicos* 114: 121-134.

Ferreira-Cabrera, A. & J. Atkinson-Abudity. 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.

Fillmore, Ch. 1992. "Corpus linguistics and Computer-aided armchair linguistics". *Directions in Corpus Linguistics*. Ed. J. Svartvik. Berlin: Mouton de Gruyter. 35-60.

Francis, N. 1979. "A tagged corpus: problems and prospects". *Studies in English linguistics for Randolph Quirk*. Eds. S. Greenbaum, G. Leech & J. Svartvik. London: Longman. 192-209.

Fries, Ch. 1952. *The structure of English*. New York: Harcourt, Brace and Company.

Gläser, R. 1982. *The problem of style classification in LSP (ESP)*. Paper presented at the 3<sup>rd</sup> European Symposium on LSP: Copenhagen.

Goldman, S. & G. Bisanz. 2002. "Toward a functional analysis of scientific genres: Implications for understanding and learning processes". *The psychology of science text comprehension*. Eds. J. Otero, J. León & A. Graesser. N. J.: Lawrence Erlbaum. 19-50.

Gómez, L. 1998. "Dimensión social de la comprensión verbal". *Comprensión de textos escritos: de la teoría a la sala de clases*. Eds. M. Peronard, L. Gómez, G. Parodi & P. Núñez. Santiago de Chile: Editorial Andrés Bello. 34-58.

Goldman. Mahwah, N.J.: Lawrence Erlbaum. 1-23.

Gunnarsson, B., P. Linell & B. Nordberg, eds. 1997. *The construction of professional discourse*. London: Longman.

Hair, J., R. Anderson, R. Tatham & W. Black. 1999. *Análisis multivariante*. Madrid: Prentice Hall.

Halliday, M. 1975. *Learning how to mean. Explorations in the development of language*. London: Edward Arnold.

Halliday, M. 1985. *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M. 1992. "Language as system and language as instance: The corpus as a theoretical construct". *Directions in Corpus Linguistics*. Ed. J. Svartvik. Berlin: Mouton de Gruyter. 61-77.

Halliday, M. 1993. "On language and physical science". *Writing science. Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Halliday, M. 1994. "On language in relation to the evolution of human consciousness". *Nobel Symposium The relation between language and the mind*. Stockholm, August 8-12.

Halliday, M. 1998. "Things and relations. Regrammaticising experience as technical knowledge". *Reading science. Critical and functional perspectives on discourse of science*. Eds. J. Martin & R. Veel. London: Routledge. 185-235.

Halliday, M. & J. Martin. 1993. *Writing science. Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Jeanneret, I. 1994. *Écrire la science. Formes et enjeux de la divulgation*. Paris: Presse Universitaires de France.

Johansson, S. 1991. "Times change, and so do corpora". *English Corpus Linguistics. Studies in honor of Jan Svartvik*. Eds. A. Aijmer & B. Altenberg. London: Longman. 305-314.

Jurafsky, D. & J. Martin. 2000. *Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition*. New Jersey: Prentice-Hall.

Jurafsky, D. 2003. "Probabilistic Modelling in Psycholinguistics: Linguistics comprehension and production". *Probabilistic Linguistics*. Eds. R. Bod, J. Hay & S. Jannedy. London: The MIT Press. 38-95.

Kennedy, G. 1998. *An introduction to corpus linguistics. Studies in language and linguistics*. London: Longman.

Kuhn, T. 1970. *The structure of scientific revolutions*. Chicago: The University of Chicago Press.

Lakoff, G. 1972. "A study in meaning criteria and the logic of fuzzy concepts". *Chicago Linguistics Society* 8: 183-288.

Lassen, I. 2003. *Accessibility and acceptability in Technical Manuals*. Amsterdam: Benjamins.

Leech, G. 1991. "The state of the art in corpus linguistics". *English Corpus Linguistics. Studies in honor of Jan Svartvik*. Eds. K. Aijmer & B. Altenberg. London: Longman. 8-29.

López, C. 2002. "Aproximaciones al análisis de los discursos profesionales". *Signos* 35.51-52: 195-215.

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.

Martin, J. 1986. "Intervening in the process of writing development". *Writing to mean: teaching genres across the curriculum*. Eds. C. Pointer & J. Martin. Sydney: Applied Linguistics Association of Australia. 11-43.

Martin, J. 1992. *English text. System and structure*. Amsterdam: John Benjamins.

Martin, J.. 1997. "Analyzing genre: functional parameters". *Genre and institutions. Social processes in the workplace and school*. Eds. F. Christie & J. R. Martin. London: Continuum. 3-39.

Martin, J.. 1998. "Discourse of science: Recontextualization, genesis, intertextuality and hegemony". *Reading science. Critical and functional perspectives on discourse of science*. Eds. J. Martin & R. Veel. London: Routledge. 3-14.

Martin, J., F. Christie & J. Rothery, J. 1987. "Social processes in education". *The place of genre in learning*. Ed. F. Reid. Geelong, Victoria: Centre for Studies of Literary Education. 67-149.

Matthiessen, C. 1998. "Construing processes of consciousness: From the commonsense model to the uncommonsense model of cognitive science". *Reading science. Critical and functional perspectives on discourse of science*. Eds. J. Martin & R. Veel. London: Routledge. 327-356.

McEnery, T. & A. Wilson. 1996. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Minker, W., A. Waibel & J. Mariani. 1999. *Stochastically-based semantic analysis*. The Netherlands: Kluwer Academic Publishers.

Moreno, A. 1998. *Lingüística computacional: introducción a los modelos simbólicos, estadísticos y biológicos*. Madrid: Síntesis.

Núñez, P., A. Gramajo & G. Parodi, 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.

Oakes, M. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Otero, J., J. León & A. Graesser, eds. 2002. *The psychology of science text comprehension*. Mahwah: Lawrence Erlbaum.

Parodi, G. 1999. "Conexiones entre comprensión y producción de textos escritos: Una aproximación psicolingüística". *Discurso, cognición y educación. Ensayos en honor a Luis Gómez Macker*. Ed. G. Parodi. Valparaíso: Ediciones Universitarias de Valparaíso de la Universidad Católica de Valparaíso. 59-90.

Parodi, G. 2002. *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: Editorial Universitaria de Valparaíso de la Pontificia Universidad Católica de Valparaíso.



Latinoamericana de Estudios del Discurso (ALED) (octubre). Puebla: Benemérita Universidad Autónoma de Puebla.

Parodi, G. & A. Gramajo. 2003. "Los tipos textuales del corpus PUCV-2003: una aproximación multiniveles". *Signos* 36.54: 207-223.

Parodi, G. & R. Venegas, 1994. *BUCOLICO: un programa para el análisis de corpus lingüístico*. Pontificia Universidad Católica de Valparaíso. (En preparación).

Peronard, M. & L. Gómez, 1985. "Reflexiones acerca de la comprensión lingüística: hacia un modelo". *Revista de Lingüística Teórica y Aplicada* 23: 19-32.

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?". *Comprensión de textos escritos: de la teoría a la sala de clases*. Eds. M. Peronard, L. Gómez, G. Parodi & P. Núñez. Santiago de Chile: Editorial Andrés Bello. 55-78.

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

Rojo, G. 2001. "La explotación de la base de datos sintácticos del español actual (BDS)". *Lingüística con corpus. Catorce aplicaciones sobre el español*. Ed. J. Kock. Salamanca: Universidad de Salamanca. 255-286.

Rose, D. 1997. "Science, technology and technical literacies". *Genre and institutions. Social processes in the workplace and school*. Eds. F. Christie & J. R. Martin. London: Continuum. 40-72.

Santana, O., J. Pérez, F. Carreras, J. Duque, Z. Hernández & G. Rodríguez. 1997. "FLAVER: flexionador y lematizador automático de formas verbales que trata conjugación y pronombres enclíticos". *Lingüística Española Actual* XIX.2: 229-282.

Schröder, H. 1991. "Linguistic and text-theoretical. Research on language for special purposes. A thematic and bibliographical guide". *Subject-oriented texts. Languages for special purposes and text theory*. Ed. H. Schröder. Berlin: Walter de Gruyter. 1-48.

Shannon, C. 1949. "The mathematical theory of communication". *The mathematical theory of communication*. Eds. C. Shannon & W. Weaver. Urbana, Illinois: The University of Illinois Press. 2-34.

Simpson, R. & J. Swales. 2001. "Introduction. North American perspectives on Corpus Linguistics at the Millennium". *Corpus Linguistics in North America. Selections from the 1999 Symposium*. Eds. R. Simpson & J. Swales. Ann Arbor: The University of Michigan Press. 1-14.

Sinclair, J. 1982. "Reflections on computer corpora in English language research". *Computer corpora in English language research*. Ed. S. Johansson. Bergen: Norwegian Computing Centre for the Humanities. 1-6.

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

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

Svartvik, J., ed. 1992. *Directions in corpus linguistics: proceeding of Nobel symposium*. Berlin: Mouton Gruyter.

Veel, R. 1997. "Learning how to meanscientifically speaking: apprenticeship into scientific discourse in the secondary school". *Genre and institutions. Social processes in the workplace*

Veel, R. 1998. "The greening of school science: Ecogenesis in secondary classrooms". *Reading science. Critical and functional perspectives on discourses of science*. Eds. J. Martin & R. Veel. London: Routledge. 114-151.

Véliz, M. 2002. "Desarrollo de estrategias de lectura por medio del computador: evaluación de una experiencia". *Lingüística e interdisciplinariedad: Desafíos del nuevo milenio. Ensayos en honor a Marianne Peronard*. Ed. G. Parodi. Valparaíso: Ediciones Universitarias de la Pontificia Universidad Católica de Valparaíso. 161-178.

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). Universidad Metropolitana de Ciencias de la Educación, Santiago de Chile.

Villaseñor, L. M. Montes, M. Pérez & D. Vaufreydaz. 2002. "Comparación léxica de corpus para generación de modelos de lenguaje". *Procedding del IBERAMIA. Workshop on Multilingual Information Access and Natural Language Processing*, noviembre. Sevilla, España.