



Revista Signos

ISSN: 0035-0451

revista.signos@ucv.cl

Pontificia Universidad Católica de Valparaíso  
Chile

Sabaj, Omar

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

Revista Signos, vol. 37, núm. 56, 2004, pp. 75-89

Pontificia Universidad Católica de Valparaíso  
Valparaíso, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=157013763006>

- 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

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

**Omar Sabaj**

*Pontificia Universidad Católica de Valparaíso  
Chile*

**Resumen:** A pesar de que existe común acuerdo en que los verbos cumplen, como categoría lingüística (i.e. gramatical, morfológica y semántica), un rol central en las lenguas humanas, su estudio ha quedado relegado casi exclusivamente al estudio de la gramática. A partir del uso de herramientas computacionales, el objetivo de este trabajo es describir el comportamiento estadístico de los verbos, entendidos como unidades lexicográficas (lemas) que presentan una ocurrencia establecida en un corpus determinado. El corpus analizado en esta investigación (Corpus PUCV-2003) está constituido por tres registros principales (el corpus técnico científico, CTC, el corpus de literatura latinoamericana, CLL y el corpus de entrevistas orales, CEO) que en total superan el millón de palabras. Del corpus general, se extrajeron el total (3.558) de los tipos de lemas verbales con sus correspondientes casos. Los resultados obtenidos muestran las siguientes tendencias. En cuanto a la especialización de los verbos, existen muy pocos verbos que se pueden considerar especializados. En segundo lugar, existe una mayor comunalidad entre los verbos del CLL y del CEO, en comparación con los verbos del CTC contrastado con cualquiera de los otros registros. Finalmente, se observa que la mayor variabilidad verbal se da en el CLL 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.

**Palabras Clave:** Verbos, especialización, especificidad, comunalidad, estadística léxica.

### **Verbal specificity, specialization and variability: A computational approach in lexical statistic**

**Abstract:** Although there is a common agreement in that verbs fulfill, as a linguistic category (i.e. grammatical, morphological and semantic), a central roll in human languages, its study has been relegated almost exclusively to the grammar. By using computational tools, in this work the verbs are investigated as a lexicographical unit (lemmas) with a determined statistical behavior in a given corpus. The corpus analyzed in this research (Corpus PUCV-2003) is constituted by three main registers (the technical scientific corpus CTC, the corpus of Latin American Literature CLL and the corpus of oral interviews CEO), which altogether overpass a million of words. From the general corpus, the total (3.558) of the lemma types with their corresponding cases was extracted (tokens). The obtained results show the following tendencies. As far as the specialization of the verbs, there are very few verbs that can be considered as specialized. Secondly, there is a greater commonality between the verbs of the CLL and the CEO, in comparison with the verbs of the CTC contrasted with any other registers. Finally, it is observed that the greater verbal variability occurs in the CLL and the smaller variability in the CEO, nevertheless, in this last one there is a greater percentage of verbs by total amount of words. From these results it is possible to conclude that in the oral interviews more verbs are used (than in the other registers) although such verbs tend to be the same.

**Key Words:** Verbs, verbal specialization, verbal specificity, verbal variability, lexical statistics.

**Recibido:** 17 de mayo de 2004

**Aceptado:** 7 de septiembre de 2004

\* Investigación desarrollada en el marco del Proyecto FONDECYT N° 1020786.

**Correspondencia:** Omar Sabaj (omar.sabaj@ucv.cl). Tel.: (056-32) 273388 Fax: (056-32) 273448. Pontificia Universidad Católica de Valparaíso, Av. Brasil # 2830, piso 9, Valparaíso, Chile.

## INTRODUCCIÓN

A pesar de que existe común acuerdo en que los verbos cumplen, como categoría lingüística (i.e. gramatical, morfológica, semántica y pragmático-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 (Lorente, 2002; Markman, 1999).

A partir del uso de herramientas computacionales, en este artículo se describe el comportamiento estadístico de los verbos, entendidos como unidades lexicográficas (lemas) que presentan 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 semi-estructuradas a alumnos de cuarto medio; y c) un corpus de literatura latinoamericana (CLL), 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, reseñaremos algunas pesquisas 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 CLL, 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 CLL 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.

## ANTECEDENTES TEÓRICOS

### 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, Conrad y Reppen (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 toma en cuenta todo el rango de las frecuencias verbales del corpus a analizar. 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 me-

dida 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.

### 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; Quilis, 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). Ahora 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 aprendidos 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, 2001; Svartvik, 1992; Kennedy, 1998; Caravedo, 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 sintética 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):

- a) Atribuir números a hechos y manifestaciones lingüísticas.
- b) 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.
- c) 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 estadio netamente estadístico.

Los resultados se obtienen por medio de cálculos matemáticos que guardan una relación con los fenómenos lingüístico. Desde el punto de vista matemático, dichos resultados son incuestionables. Finalmente, el punto (c) tiene un carácter estrictamente lingüístico, se basa en los resultados matemáticos, solamente en tanto son útiles para dar una explicación lingüística del fenómeno.

### 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.

### 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 las 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) A los actos de habla del tipo 'opinar, comunicar, transmitir, decir'.

- b) A verbos que expresan la finalidad del discurso como 'presentar, convencer, instruir, demostrar'.
- c) 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'.

### 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 *per 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). Otros 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'.

### 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' 'instaurar penicilina', 'producir recursos', 'disminuir costos', etc. Dentro de los verbos fraseológicos también se integran los llamados verbos de soporte, 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 especializado estará limitada, 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 difieran 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 sintagmáticamente el contexto de uso de estas unidades, análisis que, como ya dijimos anteriormente, no constituirá una parte de la metodología de la presente investigación.

### Los verbos-término o verbos especializados

Como parte final de clasificación aparece un tipo de verbos que la misma autora denomina verbos-término. Estos son la clase más básica de verbos de esta

clasificación y corresponden fundamentalmente 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 reiterada, a un ámbito de especialidad. Verbos, como *capialzar* (arquitectura), *eutrofizar* (ecología), *acetificar* (química), son un buen ejemplo, y se caracterizan por estar en correlación con unidades léxicas de categoría nominal o adjetiva (*capialzar*, *capialzado*; *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.

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.

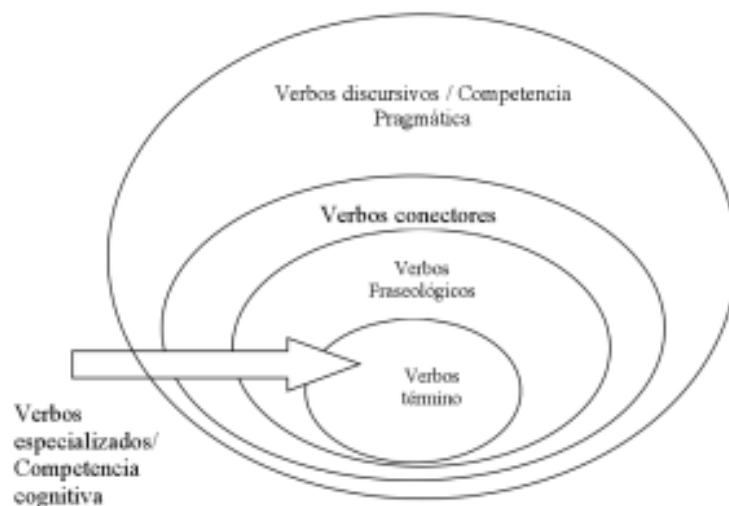


Figura 1: Estructura de los verbos de especialidad. Adaptado de Lorente (2002)

### 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 desarrollada por Johansson (1981). Como ya señalamos, este autor elaboró una fórmula matemática para determinar 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 contrastado 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 aparezcan en los otros *corpora* que sirven de contraste. Opuesta a la idea de especificidad está la idea de comunalidad verbal. En términos estadísticos un corpus presenta un alto grado de comunalidad verbal, en cuanto sus verbos aparezcan en más de uno de los registros o subcorpus que componen el corpus general. Aunque sabemos que, tanto el grado de especificidad como de comunalidad 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.

### 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*), la 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 las siguientes dos fórmulas:

- a) Variabilidad verbal =  $\frac{\text{Tipos de lemas verbales}}{\text{Total de casos verbales}}$
- b) Porcentaje verbal del corpus =  $\frac{\text{Total de casos verbales} * 100}{\text{Número total de palabras}}$

### El estudio

El estudio del que se da cuenta este artículo constituye parte de un proyecto mayor (FONDECYT 1020786), cuyos resultados, en parte, ya han sido publicados, razón por la cual omitiremos en el presente trabajo los pasos metodológicos de descripción y procedimiento de recolección del corpus puesto que ya han sido descritos en otros artículos (Parodi, 2004 y 2005). 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.

Del corpus de análisis, se extrajeron el total (3.558) de los tipos de lemas verbales (*types*) con sus correspondientes casos (*tokens*). Una vez aisladas las unidades verbales se procedió, en primer lugar, a determinar cuáles de los tipos de lemas verbales del

corpus correspondía a unidades especializadas, es decir, unidades semánticas estrechamente vinculadas a un área específica del conocimiento. Para esto se utilizaron distintos diccionarios electrónicos de libre acceso ([www.rae.es](http://www.rae.es), [www.wordreference.com](http://www.wordreference.com), etc.). En este caso, se estableció también, cuánto del porcentaje total de los verbos correspondían a verbos especializados. Luego, se investigó la especificidad de los verbos en cada uno de los corpus. Para ello se trabajó sobre el siguiente supuesto: un verbo X que aparece solo en uno de los registros y no en los otros es específico de ese registro. Inherente a la idea de especificidad está la idea de comunalidad, que también fue estudiada, es decir, cuáles son los verbos que son comunes en uno y en otro registro.

Finalmente, se investigó la variabilidad de los verbos en los distintos registros. Para esto se llevó a cabo un análisis estadístico del clásico coeficiente entre tipos y casos verbales (*verbal type/token ratio*) pero dado que, tal como ha demostrado el estudio de Lebart, Salem y Bécue (2000), dicha tasa de variabilidad depende en un alto grado del número total de palabras, se complementó ese cálculo con una estimación del porcentaje de los verbos por el número total de palabras, es decir, cuánto (porcentualmente hablando) de un registro corresponde a unidades verbales.

### 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?

### 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.

### 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 comunalidad 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.

### 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 en [www.elgrial.cl](http://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](http://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 comunalidad 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 comunalidad). Estos índices implican el uso de una estadística de proporciones representadas en porcentaje.

La última 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.

## RESULTADOS

### 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.

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

**TABLA 1**  
**Los verbos especializados en el Corpus PUCV- 2003**

	Verbo (lema)	Frecuencia		Verbo (lema)	Frecuencia
1	Abarloar	1	24	Granallar	1
2	Abiselar	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	Mandrilar	2
7	Achaflanar	6	30	Mineralizar	1
8	Acodar	2	31	Ministrar	6
9	Aduanar	19	32	Minutar	2
10	Aducir	4	33	Niquelar	4
11	Alicatar	10	34	Nitratar	7
12	Arrufar	1	35	Nitrurar	5
13	basamentar	1	36	Oxidar	36
14	Biselar	8	37	Oxigenar	10
15	Bitar	1	38	Refundir	3
16	Bromar	3	39	Roblar	1
17	Burilar	4	40	Sinterizar	2
18	Calafatear	1	41	Soldar	206
19	Carenar	2	42	Trefilar	1
20	Cintar	16	43	Tronerar	1
21	Desguazar	1	44	Varar	2
22	Estibar	58	45	Zarpar	17
23	Garrar	3	46	Zunchar	2

de tipos de verbos del corpus. Las especialidades que emergieron del análisis del significado de estos verbos se presentan en la Tabla 2.

A partir de la aplicación del Anexo 1, en el que aparecen las entradas léxicas de estos verbos en el diccionario ([www.rae.es](http://www.rae.es)), se agruparon los verbos según 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, usuarios y comunidades discursivas de esas áreas o disciplinas. A partir de los datos de la Tabla 2, podemos deducir que es el área metalúrgica la 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 Parodi (2004 y 2005), este corpus se subdivide en áreas de especialidad (marítima– indus-

**TABLA 2**  
**Verbos por especialidad**

MARÍTIMA	METALURGIA	CONSTRUCCIÓN	LEGAL	ECONOMÍA
Abarloar	Acerar	Abiselar	Abogar	Aduanar
Arrufar	Burilar	Achaflanar	Absolver	Insumir
Bitar	Granallar	Acodar	Abolir	
Calafatear	Mandrilar	Alicatar	Aducir	
Carenar	Mineralizar	Basamentar	Infringir	
Desguazar	Niquelar	Biselar	Ministrar	
Estibar	Nitratar	Bromar	Minutar	
Garrar	Nitrurar	Cintar		
Lastrar	Oxidar	Jarrar		
Tronerar	Oxigenar			
Varar	Refundir			
Zarpar	Roblar			
	Sinterizar			
	Soldar			
	Trefilar			
	Zunchar			

trial metalmecánica- 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 metalmecánica- comercial) aparecen más o menos directamente representadas 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 asociaría a una mayor importancia de los verbos como transmisores del conocimiento especializado de una disciplina.

#### **El índice de especificidad y comunalidad verbal de 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 extrajo el número de los verbos exclusivos de cada corpus. Estos datos se presentan en la Tabla 3.

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

que se daba entre los distintos *subcorpora* (y todas sus agrupaciones posibles) del corpus general. En la Tabla 4 se presentan estos resultados.

En la primera columna de la Tabla 4, se muestran las distintas agrupaciones posibles del corpus PUCV-2003. En la segunda columna, se entrega el número de verbos compartidos en cada grupo. 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 del corpus (3.558). Los resultados muestran

**TABLA 3**  
**Índice de especificidad verbal**

Subcorpus	Número de verbos exclusivos	Porcentaje
<b>CTC</b>	581	26.35%
<b>CLL</b>	1119	39.24%
<b>CEO</b>	66	6.21%

**TABLA 4**  
**Índice de comunalidad verbal entre registros**

Subcorpora	Número de verbos compartidos	Porcentaje
<b>CTC/CLL</b>	798	22.42%
<b>CLL/CEO</b>	170	4.77%
<b>CTC/CEO</b>	61	1.71%
<b>CTC/CEO/CLL</b>	771	21.66%

que los registros más comunes desde 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.

#### **Variabilidad verbal en 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.

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 (Tp/Tk) de esta Tabla. 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.

**TABLA 5**  
**Variabilidad verbal del corpus PUCV- 2003**

	<b>PUCV</b>		<b>CTC</b>		<b>CLL</b>		<b>CEO</b>	
	<b>Tp</b>	<b>Tk</b>	<b>Tp</b>	<b>Tk</b>	<b>Tp</b>	<b>Tk</b>	<b>Tp</b>	<b>Tk</b>
<b>V</b>	3.558	254.794	2.205	84.155	2.852	85.434	1.062	85.205
<b>Tp/Tk</b>	0.139		0.262		0.333		0.124	
<b>%V / N</b>	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 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 tienden a ser los mismos verbos, 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.

## 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 metalmecánica y economía). 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 sí 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 sí tiene una relación directa con una de las divisiones del corpus CTC, a saber, el área comercial, no pre-

senta 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 suscitarían 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 (Parodi, 2004a y 2004b), sabemos que el grado de especialización de un texto más que una categoría fija se constituye como un *continuum*. Este *continuum* 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, sí 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:

- a) El corpus de literatura latinoamericana (CLL) es el más específico y, a la vez, el más variable.
- b) 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.
- c) El corpus de textos técnico 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.

d) 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 (macrosituacionales) 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 divulgativa 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, cuyo 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 redundarían 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 investigan 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 especialidad técnica o científica.

Las nociones de especificidad y comunalidad 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 comunalidad, 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 exclusivas) 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 un 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 dos de las tres 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 comunalidad verbal. Si ampliáramos la muestra, aparecerían otros verbos especializados y emergerían otras áreas de conocimientos asociados a ellos. Así también (y más radicalmente), los índices de especificidad y comunalidad 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 proyectable 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 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 descripción exhaustiva 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 las investigaciones 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.

## REFERENCIAS BIBLIOGRÁFICAS

- Biber, D. (1988). *Variation across speech and writing*. Cambridge: CUP.
- Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, 19, 243-258.
- Biber, D.; Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: CUP.
- Caravedo, R. (1999). *Gramática española: Enseñanza e investigación*. Salamanca: Ediciones Universidad de Salamanca.
- Chafe, W. & Danielewicz, 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.
- García, A. (2004a). Los procedimientos matemáticos en estudios e investigaciones lingüísticas: utilidad y riesgo [En línea]. Disponible en: [http://angarmegia.tresuvesdobles.com/riesgos\\_y\\_beneficios1.htm](http://angarmegia.tresuvesdobles.com/riesgos_y_beneficios1.htm)
- García, A. (2004b). Los métodos estadísticos y la investigación lingüística: breve recorrido histórico [En línea]. Disponible en: <http://angarmegia.tresuvesdobles.com/HistMetS.pdf>
- García, A. (2004c). Beneficios y servidumbres de los métodos estadísticos [En línea]. Disponible en: <http://angarmegia.tresuvesdobles.com/BenServS.pdf>
- Johansson, S. (1981). Word frequencies in different types of English texts. *ICAME NEWS*, 5,1-13.
- Joshi, A. (1999). Computational linguistics. En R. Wilson & F. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 162-164). Massachusetts: MIT Press.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Longman.
- Lebart, L. & Salem, A. (1994) *Statistique textuelle*. París: Dunod.
- Lebart, L.; Salem, A. & Bécue, M. (2000). *Análisis estadístico de textos*. Lleida: Editorial Milenio.
- Lorente, M. (2002). Verbos y discurso especializado [en línea]. Disponible en: <http://elies.rediris.es/elies16/Lorente.html>
- Markman, A.(1999). *Knowledge representation*. New Jersey: Erlbaum.
- Moreno, A. (1998). *Lingüística Computacional: Introducción a los modelos simbólicos, estadísticos y biológicos*. Madrid: Síntesis.
- Navarro, T. (1946). *Estudios de fonología española*. New York: Las Américas Publishing Company.
- Parodi, G. (2004). Textos de especialidad y comunidades discursivas técnico-profesionales: Una aproximación basada en corpus computarizado. *Estudios Filológicos* 39, 7-36.
- Parodi, G. (2005). Lingüística de corpus y análisis multidimensional: Exploración de la variación en el corpus PUCV-2003. *Revista Española de Lingüística*. [En prensa].
- Quilis, A. (1999). *Tratado de fonología y fonética españolas*. Madrid: Gredos.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford: 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/largtext.htm>

Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell Publishers.

Svartvik, J. (Ed.) (1992). *Directions in corpus linguistics*. New York: Mouton de Gruyter.

Swales, J. (1990). *Genre Analysis. English in academic and research settings*. Cambridge: Cambridge University Press.

Valencia, A. & Echeverría, M. (1999). *Disponibilidad léxica en estudiantes chilenos*. Santiago: Ediciones UCH & UDEC.

Wiemer-Hastings, P.; Graesser, A. & Wiemer-Hastings, K. (1998). Inferring the meanings of verbs context. En M. Gernsbacher & S. Derry (Eds.), *Proceedings of the twentyeth annual conference of cognitive science* (pp.1142-1147). Wisconsin: Erlbaum.

## ANEXO

El anexo muestra el total de las entradas del diccionario ([www.rae.es](http://www.rae.es)) de los verbos especializados del corpus PUCV-2003.

- 1 **Abarloar** (paras. de barloar) tr. Situar [un buque] de costado casi en contacto con otro buque o con un muelle, etc. SIN. Barloar.
- 2 **Abiselar** tr. Hacer biseles [en un cristal, una moldura, etc.]. SIN. biselar.
- 3 **Abogar** (lat. *advocare*, llamar cerca de sí) intr. Defender en juicio. 2 fig. Interceder, hablar en favor de alguien: ~ por, o contra, alguien o alguna cosa. CONJUG. como llegar.
- 4 **Abolir** (lat. *abolere*). tr. defect. Derogar, dejar sin vigencia una ley, precepto, costumbre, etc.
- 5 **Absolver** (lat. *-ere*) 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. irreg., 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 las puntas de las armas o herramientas]: ~ un sable. 2 Dar [a un líquido] propiedades medicinales con tintura de acero o apagando en él acero hecho ascua. 3 Dar los grabadores un baño de acero [a las planchas de cobre]. 4 tr.-prnl. fig. Fortalecer, vigorizar: esta 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. *cubitu*, codo) tr.-prnl. 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 codales [en la superficie de una piedra o de un madero] para ver si está plana. 4 Acodillar (doblar). 5 Disponer en ángulo piezas de maquinaria. 6 arq. Acodalar. 7 veter. Clavar mal [los clavos] al herrar, desviándolos sobre las partes sensibles. SIN. 2 Cerchar, ensarmentar, 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 conducir. incor.: aduciste.
- 11 **Alicatar** (ár. *alocat*, 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 arrufa por la popa. 3 Gruñir los perros enseñando los dientes. 4 prnl. Embravecerse, irritarse.
- 13 **Basamentar** tr. arq. Poner los basamentos de una edificación.
- 14 **Biselar** Tr. Hacer biseles [en un cristal, una moldura, etc.]. SIN. Abiselar.
- 15 **Bitar** tr. Amarrar y asegurar [la cadena del ancla] a las bitas, abitar.
- 16 **Bromar** tr. Roer la broma [la madera].
- 17 **Burilar** tr. Grabar [figuras o adornos] en los metales con el buril: ~ en cobre.
- 18 **Calafatear** (gr. mod. *kalaphatein*) tr. Cerrar [las juntas de las maderas de las naves] con estopa y brea para que no entre el agua. 2 p. ext. Cerrar [otras juntas].

- 19 **Carenar** (lat. *carinare* ) tr. Reparar el casco [de una nave]: ~ de firme 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 **Estibar** (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 colocar [mercancías] en un local.
- 23 **Garrar** intr. Cejar un buque arrastrando el ancla por no haber ésta hecho presa.
- 24 **Granallar** tr. Reducir [un metal] a granalla.
- 25 **Infringir** (lat. *-ere* ) tr. Quebrantar [la ley, un convenio, etc.]. CONJUG. como dirigir.
- 26 **Insumir** (lat. *insumere*, gastar) tr. Entre economistas, invertir [dinero y otros bienes] en una producción industrial, agrícola, comercial, etc.
- 27 **Jarrar** Tr. fam. Jaharrar. Jaharrar (probl. del ár. *yeyyar*, encalar, der. de *yir*, cal) tr. Revocar [una pared].
- 28 **Lastrar** tr. mar. Poner lastre [a la embarcación]. 2 Afimar [una cosa] cargándola de peso.
- 29 **Mandrilar** tr. Perforar un metal con un mandril II. 2 Ensachar 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 prnl. Cargarse el agua de sustancias minerales. CONJUG. como realizar.
- 31 **Ministrar** (lat. *-are* ) tr.-intr. Servir [un oficio, o ministerio]: ministra en la Audiencia; ministra la justicia. - 2 tr. Dar, suministrar [una cosa].
- 32 **Minutar** (de *minuta* ) 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.-prnl. 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 prnl. 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, discurso, etc.]. 3 fig. Comprender o incluir: ~ una poesía en un texto; refundirse dos obras. - 4 intr. fig. Redundar: el trabajo refundió a su favor. - 5 prnl. Amér. Extraviarse, perderse. 6 Guat., Méx. Guardar algo con mucho ahínco.
- 39 **Roblar** (v. *roborar* ) tr. Robrar. 2 Doblar o remachar [una pieza de hierro] para que esté firme.
- 40 **Sinterizar** (del lat. *sinter*, escoria, ceniza) tr. Metal. Producir piezas de gran resistencia y dureza calentando, sin llegar a la temperatura de fusión, conglomerados de polvo, gralte. metálicos, a los que se ha modelado por presión. CONJUG. como realizar.
- 41 **Soldar** (v. *solidar* ) tr. Pegar sólidamente [dos cosas] o 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 descuido] con acciones o palabras. CONJUG. como contar.
- 42 **Trefilar** tr. Transformar en hilo o alambre [un metal] pasándolo por la hilera.
- 43 **Tronerar** tr. Atronerar. tronera (de *tronar* ) f. 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 *vara*; en b. lat. *varare* ) intr. Encallar la embarcación. 2 fig. Quedar detenido un negocio. 3 Amér. Quedarse detenido un vehículo por avería. 4 tr. Sacar a la playa [una embarcación]. 5 desus. Echar [un barco] al agua.
- 45 **Zarpar** (it. ant. *sarpare*, de orig. incierto) tr.-intr. Levar anclas, hacerse a la mar: la escuadra zarpó del puerto. - 2 intr. Partir o salir embarcado.
- 46 **Zunchar** tr. Reforzar [una cosa] con zunchos.