La similitud léxico-semántica en artículos de investigación científica en español: Una aproximación desde el Análisis Semántico Latente

Lexical-semantic similarity in scientific research articles in Spanish: An approach to Latent Semantic Analysis

René Venegas

Pontificia Universidad Católica de Valparaíso

Chile

Dirección para correspondencia

RESUMEN

Esta investigación es un estudio comparativo de la relación de similitud léxico-semántica entre tres variables textuales (palabras clave, resumen y el contenido de artículos de investigación científica). Además, se comparan a partir de los valores de similitud léxico-semántica dos áreas científicas (ciencias biológicas y ciencias sociales). El estudio se realiza utilizando una muestra estratificada representativa correspondiente a 22 artículos de investigación científica de ambas áreas científicas, incluidos en un corpus de 675 artículos científicos. Para la determinación de la similitud léxico-semántica entre las variables, se utiliza un método estadístico-computacional denominado Análisis Semántico Latente. Los resultados nos permiten establecer, por una parte, que en la muestra investigada el resumen □macrosemantiza□ mejor el contenido semántico global del artículo que las palabras clave. Por otra parte, no se presentan diferencias significativas entre los promedios de similitud léxico-semántica entre las áreas científicas estudiadas. Estos resultados se explican en función de los complejos procesos de estandarización que tienden a homogeneizar la producción científica.

Palabras Clave: Escritura científica, similitud semántica, análisis semántico latente, artículo de investigación científica.

ABSTRACT

This research focuses on lexical-semantic similarities found in three text variables (key words, abstract, and content in scientific research articles). Two scientific areas (biological sciences and social sciences) from the values of lexical-semantic similarity are compared. The study employs a representative stratified sample of 22 scientific research articles in these two areas, which have been included in a corpus of 675 scientific articles. To determine lexical-semantic similarities among the variables, a computer-statistical method is employed, called Latent Semantic Analysis. The findings help assert, on the one hand, that the abstract □macro-

semanticizes□ better the global semantic content of the article than do the key words. On the other hand, no meaningful differences among the averages of lexical-semantic similarity in the scientific areas studied are revealed. These findings are accounted for in terms of the generally complex standardization processes which tend to homogenize the production of this type of articles.

Key Words: Scientific writing, semantic similarity, latent semantic analysis, scientific research article.

INTRODUCCIÓN

La investigación que presentamos a continuación se encuentra enmarcada en los estudios del discurso especializado, específicamente, en lo que respecta a la escritura del artículo de investigación científica. Si bien el término □discurso especializado□ se encuentra, en la actualidad, ampliamente aceptado por los estudiosos del lenguaje, se debe reconocer que su utilización no surgió sino solo hace unos pocos años. En general, la noción de discurso especializado se concibe de forma amplia y globalizadora (Parodi, 2005), reconociéndose al interior del concepto un *continuum*, en el que se alinean textos que van desde una alta hasta una baja especialidad, pudiendo estos corresponder a variados tipos.

Acorde con lo anterior, el tipo de texto tradicionalmente estudiado en el ámbito del discurso especializado de la ciencia es el artículo de investigación científica, siendo considerado como el texto prototípico de este tipo de discurso (Sager, Dungworth & McDonald, 1980; Bazerman, 1988; Swales, 1990, 2004; Halliday & Martin, 1993; Hyland, 1998, 1999, 2000; Salager-Meyer, 1991, 1992; Martin & Rose, 2003). En español este tipo textual ha cobrado interés más recientemente y ha sido estudiado desde diversas perspectivas, por ejemplo, Calsamiglia (1998), Bolívar (2000), Ciapuscio (2000, 2003), Cassany, López y Martí (2000), Moyano (2000), Ciapuscio y Otañi (2002), López (2002), Mogollón (2003), Martín (2003), Gotti (2003).

En general, estos autores, conciben el artículo de investigación científica como el texto escrito, publicado en una revista especializada, que tiene como finalidad informar a la comunidad científica los resultados de un trabajo de investigación realizado mediante la aplicación del método científico, lo que exige una clara estructuración retórica, adhiriendo comúnmente al modelo IMRD (Introducción, Método, Resultados, Discusión) propuesto por Swales (1990). Sin embargo, como el mismo Swales (2004) plantea, esta estructura variará según las características propias de cada disciplina científica.

Las investigaciones en este ámbito han sido realizadas tradicionalmente desde perspectivas lingüístico-textuales, retóricas y sociocognitivas, utilizando, en la mayoría de los casos, muestras ejemplares de textos y criterios preferentemente cualitativos. El enfoque de análisis más desarrollado ha sido la comparación interlenguas para propósitos especiales y didácticos y en áreas disciplinares muy particulares (medicina, psicología, economía, leyes, por mencionar las más estudiadas).

Ahora bien, desde la realidad chilena, observamos que la escritura científica presenta indicadores que no son muy favorables. Así, por ejemplo, en los últimos años el presupuesto anual de la Comisión Nacional de Ciencia y Tecnología de Chile se ha duplicado respecto del año 1999, alcanzando un monto cercano a los sesenta

millones de dólares, aun así, la producción científica en términos de publicaciones alcanza apenas un 10,5% de las publicaciones científicas de los países latinoamericanos más productivos, ubicándose en este ranking bajo Brasil, Argentina y México y superando en apenas en un 5% a Venezuela. Este valor, en términos internacionales, cae a un 0,18% de publicaciones científicas registradas en ISI (International Scientific Index) entre 1981 y 2002 (CONICYT, 2004).

Otro dato que confirma esta situación poco afortunada en cuanto al desarrollo científico chileno es el hecho de que a partir del año 2000 las universidades han invertido en educación y desarrollo, en promedio, casi el doble del dinero que se invirtiera en las décadas de los ochenta y los noventa (de 62.510 a 110.193 millones de pesos), lo cual ha redundado en un aumento promedio de casi dos veces la cantidad de postgraduados. Sin embargo, el índice de productividad de artículos por postgraduado en promedio no ha variado, siendo incluso algo menor, manteniéndose en un 0,28. En otros términos esto implica que por artículo existen entre tres a cuatro postgraduados (CONICYT, 2004).

Todo lo anterior indica claramente que a pesar de la implementación de nuevas políticas de desarrollo científico, el impacto en la escritura científica en términos de artículos de investigación publicados en revistas de corriente principal es muy escaso. Esto, en parte, debido a que escribir un artículo de investigación puede resultar una tarea muy compleja para el investigador, sobre todo para quienes se inician en el ámbito científico.

Sin duda, existe aquí una problemática que en Chile aún no se ha tratado lo suficiente y que exige un tratamiento exhaustivo desde la perspectiva de las políticas de educación superior y desde el desarrollo científico nacional, en particular desde disciplinas como la lingüística y la psicolingüística.

En un esfuerzo por aportar a la discusión respecto del problema de la escritura científica en español, en este trabajo llevamos a cabo un estudio cuyos objetivos son: a) comparar, utilizando una herramienta computacional de análisis vectorial denominada Análisis Semántico Latente (LSA, por su sigla en inglés), la relación léxico-semántica entre tres variables textuales presentes en artículos de investigación científica, estas son: palabras clave, resumen y contenido del artículo y b) comparar, a partir de los valores de similitud léxico-semántica de las variables textuales, una muestra de artículos de investigación científica de dos áreas de la ciencia (ciencias biológicas y ciencias sociales).

Para llevar a cabo estos objetivos, se presentan, en la primera parte, los antecedentes teóricos que sustentan la investigación, estos son: discurso especializado de la ciencia, semántica colocacional y análisis semántico latente. En la segunda parte se presenta el marco metodológico del trabajo, así como la interpretación de los resultados. Por último, se cierra este artículo con las conclusiones obtenidas en esta investigación.

1. ANTECEDENTES TEÓRICOS

1.1 El artículo de investigación científica: Un tipo de discurso especializado

Sabemos que en la ciencia no existe siempre consenso en la denominación de los objetos de estudio, debido fundamentalmente a la focalización y delimitación que deben realizar los autores al intentar conceptualizarlos. Normalmente, los abordajes

son múltiples en razón de supuestos teóricos divergentes. El concepto de □discurso especializado□no es la excepción. Este ha sido denominado de múltiples maneras, por ejemplo: discurso académico, discurso especial, discurso profesional, discurso técnico, discurso institucional, etc.

Alcanzar un relativo orden terminológico y lograr una visión más o menos homogénea tampoco resulta fácil (Ciapuscio, 2000; López, 2002), de modo que determinar de forma discreta si un texto se clasifica como de especialidad o de tipo general es, sin duda, un problema teórico y descriptivo (Schröder, 1991; Parodi, 2004). Hoy en día, la postura predominante está en favor de un *continuum* de textos que se distribuyen de manera progresiva desde un dominio altamente especializado hasta otro extremo mucho más divulgativo y general (Gläser, 1982; Schröder, 1991; Halliday & Martin, 1993; Jeanneret, 1994; Peronard, 1997; Ciapuscio, 1994, 2000; Cabré, 2002; Parodi, 2004), aceptando que la realidad no se circunscribe a la idea de límites discretos sino más bien difusos (Lakoff, 1987).

Por su parte, Gotti (2003), siguiendo también la idea del *continuum* plantea, en relación con la naturaleza multidimensional del discurso especializado, que no existe homogeneidad entre los diferentes lenguajes especializados. Este autor argumenta que las variaciones disciplinares producen no solo connotaciones léxicas especiales, sino que también a menudo influencian otras opciones (morfosintácticas, textuales y pragmáticas), teniendo además repercusiones en las peculiaridades epistemológicas, semánticas y funcionales de una variedad de discurso especializado.

De esta manera, las diferencias entre los discursos permiten reconocer diferencias de nivel en el discurso especializado, ya que, por ejemplo, la sola presencia de un especialista no es suficiente para asegurar el uso especializado del lenguaje. De hecho, Gotti (2003) distingue al menos tres niveles diferentes en los cuales el experto podría referirse a un tópico relacionado con su profesión. En esta investigación, nos interesa destacar el primer nivel planteado por Gotti (2003), esto es, el nivel en el que el especialista se comunica con otros especialistas para describir un proyecto de investigación, para presentar sus resultados, para explicar el uso de ciertos equipos, e incluso, para debatir tópicos relacionados con el campo disciplinar, etc. En este tipo de comunicación, materializada particularmente en el artículo de investigación científica, los lectores comparten una cantidad considerable de conocimiento, por lo que el hablante especialista puede hacer un uso frecuente de terminología especializada cuyo sentido tiende a estar garantizado (Gläser, 1982, 1993; Halliday & Martin, 1993; Cabré, 1999, 2002; Ciapuscio, 2003; Gotti, 2003).

En atención a lo anterior, concebimos al artículo de investigación científica (AIC) como el texto escrito, generalmente publicado en una revista especializada, que tiene como finalidad informar a la comunidad científica los resultados de un trabajo de investigación realizado mediante la aplicación del método científico, según las características de cada disciplina de la ciencia. Su estructura es bastante rígida (al menos como se presenta en algunas disciplinas empíricas) y expone en general los siguientes apartados: Introducción, Materiales y Métodos, Resultados, Discusión y Conclusiones (IMRD) (Swales, 1990). Estas secciones están precedidas por un título, por la mención de los autores y de las instituciones en las que ellos se desempeñan como investigadores y por un resumen, destinado a informar sucintamente a los lectores acerca del contenido de todo el AIC para que ellos decidan si les resulta útil la lectura completa del texto (Moyano, 2000).

Swales (1990, 2004) sostiene que hay características que se repiten

suficientemente en los artículos científicos de un extenso rango de disciplinas como para considerar la existencia de un Macro-Género. Sin embargo, sabemos que los artículos varían de una disciplina a otra en grados de estandarización y estilo: las ciencias conocidas como □duras□, □exactas□ o □físicas□ siguen un modelo más rígido, mientras que en las ciencias sociales existen grupos que han intentado adaptarse a ese modelo con diferentes grados de éxito, mientras que otros se resisten a establecer reglas fijas para sus textos (Moyano, 2000; Mogollón, 2003, Swales, 2004). Por otra parte, cabe hacer notar que en este sentido las revistas científicas, cada vez más, buscanestandarizar la producción de los artículos con el fin de incorporarse o mantenerse en los indexadores internacionales de artículos científicos, dado que esta estandarización en alguna medida facilitaría tanto la redacción como la lectura de los mismos, permitiendo, entre otras actividades, la replicación de la investigación.

Diversos estudios se han llevado a cabo en función de cada una de las partes del artículo de investigación científica. Por ejemplo, la introducción ha sido profundamente trabajada por Swales (1990), la introducción y conclusión por Gnutzmann y Oldenburg (1991), la conclusión por Ciapuscio y Otañi (2002), el resumen por Salager-Meyer (1991) y Bolívar (2000), el resumen y la introducción por Martín (2003), las secciones introducción y discusión por Dudley-Evans (1986). Todas estas investigaciones han sido llevadas a cabo fundamentalmente desde perspectivas lingüístico-textuales, retóricas y sociocognitivas y, la mayoría de ellas, desde un enfoque comparado interlenguas en muestras ejemplares de textos.

En lo que sigue nos referiremos, con algo más de detalle, a dos aspectos del artículo de investigación científica, estos son: el resumen y las palabras clave.

1.1.1 El resumen en el AIC

Como se sabe, el resumen es un texto breve que sirve para que el lector identifique de forma rápida y precisa el contenido central del artículo. Este texto se ubica entre el título y la introducción. Se debe destacar que el resumen representa en la actualidad un papel muy importante, pues existen sistemas de información bibliográficos que al realizar una búsqueda de artículos, además del título y los autores, muestran también el resumen (Maruhenda, 2003).

El resumen, en términos generales, debiera presentar la misma estructura lógica que el artículo. Tradicionalmente no lleva subtítulos (con la excepción de algunos artículos en medicina) y carece de discusión, así como de citas bibliográficas, cuadros, tablas o figuras. Así, en una extensión no superior a las 300 palabras, se tiene que poner en evidencia el rigor científico del trabajo de investigación, por lo que expone brevemente los objetivos, la información imprescindible acerca de los materiales y los métodos de investigación utilizados y la conclusión más relevante del estudio (Ratteray, 1985; Swales, 1990; López, 1997; Moyano, 2000).

Desde el punto de vista semántico el resumen corresponde a una globalización (condensación de la información en unidades menores) y a una conceptualización de la red de contenidos del texto. En este sentido, proponemos denominar a este proceso de globalización como □macrosemantización□, ya que el resumen es la textualización de un significado que representa de modo abstracto el significado total del contenido del artículo.

Según van Dijk y Kintsch (1983), el hecho de que el resumen se encuentre al inicio de la lectura, en un texto cualquiera, ayuda al lector a formarse una hipótesis sobre

el tópico del discurso o del episodio, de tal manera que las oraciones siguientes pueden procesarse de manera arriba-abajo para tales macroproposiciones. Cuando en su confección se sigue este tipo de procesamiento cognitivo, el producto que se obtiene resulta en ocasiones hasta más claro y coherente que el propio trabajo sometido al proceso de análisis y síntesis. Tal afirmación se explica por sí misma, pues el resumen como producto no es más que el resultado de una abstracción, en la que se sintetiza la información que ofrece el documento de origen manteniendo sus partes esenciales (López, 1997).

Una forma de estudio más actualizada, aunque poco conocida en español, es la que permite estudiar el resumen desde una perspectiva computacional. En este sentido destacan los trabajos realizados por los investigadores del Instituto de Ciencias Cognitivas de la Universidad de Colorado, Boulder, en Estados Unidos. En este centro, se han desarrollado estudios teóricos y empíricos orientados fundamentalmente a probar que los modelos vectoriales, en particular el LSA, representan el procesamiento inductivo de los significados que realizan los seres humanos (http://lsa.colorado.edu/) (Kintsch, Steinhart, Stahl, LSA Research Group, Matthews & Lamb, 2000). También destacan los trabajos realizados por los miembros del IIS (*Institute of Intelligent Systems*) de la Universidad de Memphis, quienes han utilizado el LSA en el tutor virtual inteligente, denominado AutoTutor (www.autotutor.org) (Graesser, Person, Harter & Tutoring Research Group, 2001).

En virtud de las anteriores consideraciones, viene al caso precisar que el resumen ha de cumplir no solo la función de proporcionar elementos que estimulen la consulta del documento original, sino, más aún, debe facilitar la obtención de un primer nivel de asimilación del problema que se aborda y propiciar un precedente informativo sólido. En términos más cognitivos, el resumen debe entregar al lector, en este caso un lector especializado, los elementos necesarios para construir una representación del modelo de situación, que le permita comprender los procedimientos básicos y resultados que se consignarán en el artículo.

1.1.2. Palabras clave

Muchas revistas científicas piden a los autores que agreguen, luego del resumen, un listado de palabras clave a su artículo. Sin embargo, escribir tales palabras clave no es tarea fácil, pues ellas, junto al título y al resumen, constituyen rasgos que describen el contenido semántico global de un texto en grados variables de detalle y de abstracción.

Las palabras o frases clave (Turney, 1997, 1999) pueden cumplir principalmente dos objetivos: a) permiten resumir el contenido de un artículo de investigación. En este caso, las palabras clave son una forma altamente abstracta de resumir los significados más relevantes del contenido de un documento. Ellas le permiten al lector determinar rápidamente si el artículo está o no en su campo de interés; b) permiten indexar artículos según una temática determinada. Además, facilitan una búsqueda rápida de un artículo relevante para el lector, cuando existe una necesidad específica (Turney, 1999).

Para Hartley y Kostoff (2003), las palabras clave indican los conceptos principales y delimitan el campo de interés de la investigación. De esta manera, las palabras clave permiten al lector decidir cuándo un artículo contiene o no material relevante para su interés, proveen a los lectores de un grupo de términos convenientes para ser usados en búsquedas en internet para localizar otros materiales en el tópico, ayudan a los editores/indizadores a agrupar materiales relacionados, permiten a los investigadores intercambiar documentos en los temas de una disciplina y vinculan

tópicos específicos de preocupación con tópicos en metaniveles más altos. Esto último, referido a las listas de palabras clave que pueden identificar un área de investigación en particular. Por ejemplo, si se hace una lista de las palabras clave de los artículos de investigación científica de una disciplina particular, las más altas frecuencias indicarían cuáles son los tópicos más comúnmente tratados en esa disciplina, lo que indicarían niveles de abstracción más altos y darían cuenta de los metatópicos del área (Hartley & Kostoff, 2003).

Por lo general, el estudio que se hace de las palabras clave está en relación con los estudios que tienen por interés la extracción de información desde los textos, a través de herramientas computacionales que funcionan, principalmente, con métodos matemático-estadísticos y con modelos algorítmicos. Estos estudios tienen el propósito de obtener documentos relevantes ante la necesidad de un usuario de obtener cierta información. Por ejemplo, en el caso de un artículo científico que debe ser indexado y no posee palabras clave o resumen. Los distintos métodos o modelos permitirán, a través de distintas técnicas, entregar, bien un listado de las palabras más relevantes del texto o bien un texto que sintetice la información del documento de origen (ver Turney, 1997, 1999).

1.2. Estudio del significado asociativo colocacional por similitud semántica

El análisis de las relaciones léxico-semánticas que realizaremos en esta investigación se sustenta en la noción de similitud semántica, esta se remonta a los aportes realizados por Russell (1937) con su □teoría de clases y similitud□. Hoy en día, esta noción se implementa como medida probabilística o como el grado de intercambiabilidad de una palabra por otra en un contexto particular, dado que la similitud semántica se concibe bajo el supuesto de que palabras semánticamente similares se comportarán de manera similar (Manning & Schütze, 2003; Matsumoto, 2003).

Ahora bien, mientras para algunos autores la similitud semántica se entiende como una extensión de sinonimia, otros la entienden en el sentido de que dos palabras comparten un mismo dominio semántico o tópico (Manning & Schütze, 2003). En esta lógica, las palabras son similares si ellas refieren a entidades en el mundo que tienen muchas posibilidades de co-ocurrir, como en □doctor□, □enfermera□, □fiebre□ e □intravenosa□ refiriendo estas palabras a diferentes entidades e, incluso, pudiendo pertenecer a categorías sintácticas diferentes. Esto ocurre debido a que los modelos algorítmicos usados para extracción de información están basados en una interpretación extrema de los principios de la semántica composicional (Jurafsky & Martin, 2000). En estos sistemas, el significado de los documentos reside solamente en las palabras que están contenidas en ellos. Es decir, el orden y la constitución de las palabras que forman las oraciones, que a su vez constituyen el texto, no tienen importancia en la determinación de su significado. Dado que se ignora la información sintáctica, este tipo de aproximación es comúnmente denominado como método de □depósito de palabras□ (*bag of words*) (Jurafsky & Martin, 2000).

Desde una perspectiva netamente lingüística, esta noción es compatible con la de significado colocacional desarrollada por la escuela funcionalista inglesa (Palmer, 1980). Como sabemos, esta noción se basa en la teoría contextual del significado, en la cual una palabra adquiere significado por las palabras que la acompañan (Palmer, 1980; Stubbs, 1996, 2001). Esta concepción del significado ha dado origen a la semántica de corpus computacional, la cual, a través del uso de herramientas computacionales permite llevar a cabo estudios empíricos del significado utilizando grandes corpora textuales (Halliday, 1991, 1992; Sinclair,

1991; Stubbs, 1996, 2001).

Esta focalización de la lingüística de corpus, planteada indudablemente como un resurgimiento de los estudios empiricistas, esta vez apoyados en una tecnología más poderosa y en el libre acceso a incontables textos en formato electrónico, permite ahondar en los estudios del lenguaje y, a través de este, acercarse, en alguna medida, a la comprensión de la mente humana, esto es, indagar en la naturaleza del lenguaje como manifestación de la mente (Osgood, 1952; Osgood, Suci & Tannenbaum, 1976; Chafe, 1994).

Con relación al estudio del significado, Stubbs (2001) plantea la posibilidad de estudiar las relaciones léxico-semánticas, a partir del estudio de colocaciones y frecuencias de palabras. Este autor, basándose en ejemplos concretos, promueve los métodos observacionales de una semántica de corpus, argumentando que los datos obtenidos de los corpus proveen evidencia respecto del significado denotativo y connotativo. Sin embargo, se ha criticado que la mayoría de los estudios de frecuencias en lingüística de corpus se limita al recuento aislado de las unidades más frecuentes, ocultando diversos aspectos interesantes que dicen relación con unidades de frecuencia nula, mínima o media (Rojo, 2002).

Teniendo en cuenta los aspectos mencionados, planteamos que para estudiar cuantitativamente relaciones léxico-semánticas en corpus no podemos centrarnos solo en las más altas frecuencias sino en todo el rango de frecuencias de ocurrencias, e incluso más, para un estudio completo se requiere considerar, además, las co-cocurrencias entre palabras y otras unidades textuales, así como las relaciones que aparecen debilitadas por el tamaño de los corpus.

Dado lo anterior optamos por un método vectorial basado en un análisis semántico latente, que permite reconocer mediante la reducción de dimensionalidad las similitudes semánticas existentes entre las unidades lingüísticas o textuales a partir de las correlaciones entre, no solo, palabras, sino que entre palabras y documentos. En síntesis, la idea más relevante en relación con la similitud semántica es que los resultados pueden ser explicados por el grado de intercambiabilidad contextual o el grado en el cual una palabra puede ser substituida por otra en un contexto dado. Desde una perspectiva algorítmica, la medición de la similitud semántica es conceptualizada por los modelos de tipo vectorial como una medida de similitud de vectores para determinar la similitud de dos palabras que son representadas como vectores en un espacio multidimensional o multivectorial. Para llevar a cabo esto, se construye una matriz en la cual se representa numéricamente la co-ocurrencia de las palabras por una unidad mayor, denominada □documento□ (normalmente oraciones o párrafos).

1.3. El Análisis Semántico Latente

El LSA es un método de análisis vectorial que permite extraer e inferir relaciones del uso contextual de palabras en documentos. Esto se realiza a través de la implementación del algoritmo de similitud semántica y la reducción dimensional. El LSA toma como datos de entrada únicamente la segmentación del texto en palabras, frases, oraciones o párrafos (Landauer & Dumais, 1996, 1997; Landauer, Foltz & Laham, 1998).

Debemos destacar que el LSA no usa ninguna información lingüística previa o conocimiento perceptual, esto significa que está solamente basado en un método de aprendizaje matemático general que logra efectos inductivos poderosos,

extrayendo un adecuado número de dimensiones para representar objetos y contextos (Landauer et al., 1998).

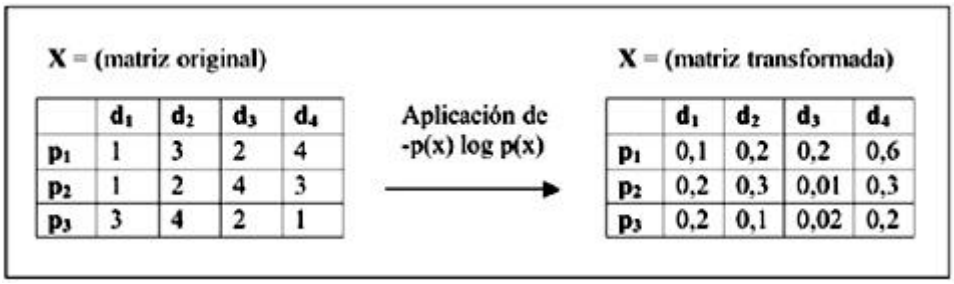
Este método extrae sus representaciones de significado de palabras y párrafos, exclusivamente a partir del análisis matemático-estadístico del texto. Nada de su conocimiento viene desde la información perceptual sobre el mundo físico, del instinto, o de la experiencia generada por funciones corporales, sentimientos y/o intenciones. Así, su representación del significado es parcial y limitada, puesto que no hace uso de relaciones sintácticas ni lógicas, ni morfológicas. A pesar de lo anterior, Landauer (2002) explica, al menos para la lengua inglesa, que el 80% de la información potencial en el lenguaje está en la elección de palabras sin tener en cuenta el orden en el que ellas aparecen.

Junto con esta idea de representación, sin sintaxis, aparece la idea de que en estas grandes cantidades de*corpora* existen interrelaciones semánticas débiles entre palabras que son potenciadas por el método de reducción de dimensiones denominado Descomposición en Valores Singulares (SVD, por su sigla en inglés). En este sentido, la metáfora que subyace al término □latente□ es que por medio de la reducción de dimensionalidad que se realiza usando el SVD, se obtiene una representación adecuada de las relaciones existentes entre las palabras en un corpus textual, en el cual estas relaciones son muy débiles debido al gran número de palabras (Landauer & Dumais, 1996, 1997; Landauer et al, 1998).

El procedimiento que se lleva a cabo al utilizar el LSA para la representación de los textos en espacios semánticos multidimensionales es el propuesto para el método de Indexación Semántica Latente (LSI por su sigla en inglés) por Deerwester, Dumais, Furnas, Landauer y Harshman (1990). Este método, aplicado originalmente en el área de la recuperación de información, ha sido utilizado en los últimos años en psicolingüística con fines teóricos y metodológicos (Deerwester et al, 1990; Foltz, 1990; Landauer & Dumais, 1996, 1997; Landauer et al., 1998; Kintsch, 1998, 2000, 2001; Landauer, 2002; Quesada, Kintsch & Gómez, 2002; Quesada, 2003). Mayor información sobre la discusión teórica y metodológica que ha suscitado este método, en español, se puede encontrar en Venegas (2003, 2005) y Gutiérrez (2005).

En términos generales este método, implementado computacionalmente, funciona del siguiente modo: El primer paso es construir a partir de los textos una matriz de co-ocurrencias (X) en la cual cada columna representa a un documento o cotexto (d) y cada fila representa una palabra del texto (p). Cada celda contiene la frecuencia con la cual la palabra de la fila aparece en el pasaje de texto denotado por su columna (p, d). La entrada en la celda está sujeta a una transformación doble por la cual cada celda es □pesada□ por una función que expresa tanto la importancia de un pasaje particular de texto como el grado por el cual el tipo de palabra transporta información en el dominio del discurso en general. Esto quiere decir que la frecuencia de la palabra en cada celda es convertida en su logaritmo (log). Luego, se calcula para cada palabra una medición de la teoría de información, entropía, esto es, -p(x) log p(x) para todas entradas en la fila, después cada entrada en la celda de la matriz es dividida por el valor entrópico de cada fila. Esta transformación es importante, porque se enfatizan las palabras portadoras de significado específico, reduciendo la influencia de términos que ocurren en una gran cantidad de documentos y ponderando aquellas que están más asociadas con un tipo particular de documento (ver Figura 1)[1].

Figura 1. Primer paso del método LSA.

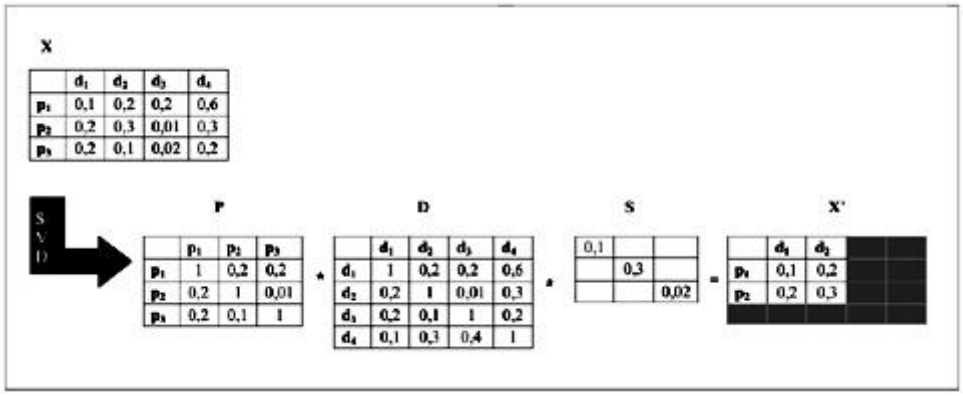


	d_1	d_2	d_3	d_4
p_1	1	3	2	4
p_2	1	2	4	3
p_3	3	4	2	1

	d_1	d_2	d_3	d_4
p_1	0,1	0,2	0,2	0,6
p_2	0,2	0,3	0,01	0,3
p_3	0,2	0,1	0,02	0,2

El segundo paso, es aplicar a la matriz resultante el método SVD. Este descompone la matriz rectangular (aquella que considera diferentes entidades en las filas y en las columnas, por ejemplo, términos por documentos) en el producto de otras tres matrices (palabras por palabras (P), documentos por documentos (D) y una de valores singulares (S)), esto es en una matriz de menores dimensiones (X□), que representa a la matriz original (ver Figura 2).

Figura 2. Segundo paso del método LSA: Aplicación de SVD.



Para propósitos explicativos es útil interpretar la SVD en términos geométricos. Esto significa que los valores de las filas y columnas de la matriz reducida son tomados como coordenadas de puntos que representan los documentos y los términos en un espacio multivectorial o multidimensional de *k*-dimensiones (donde *k*significa que las dimensiones son menores y representan a las dimensiones originales de la matriz de co-ocurrencias entre palabras y documentos). La cantidad de dimensiones (*k*) es la dimensionalidad a usar para calcular las similitudes entre las unidades textuales a comparar, esta está correlacionada con la ocurrencia de los términos en la matriz original, y a partir de ella se puede segmentar el espacio semántico en un buennúmero de categorías subsimbólicas que pueden ser combinadas significativamente. Normalmente las dimensiones a utilizar varían entre 50 y 400 dimensiones (Landauer et al., 1998; Wiemer-Hasting, 2004).

Por último, la similitud entre vectores es calculada usando medidas de coseno, cuyos valores van de 1 para vectores con la misma dirección (esto significa que lo medido es igual) a 0 para aquellos vectores ortogonales (perpendiculares en el espacio multivectorial, es decir, que lo medido es completamente distinto). Los valores deben ser normalizados, para hacer más efectiva la comparación entre ellos, ya que si no se hace, vectores más largos (correspondiente a documentos más largos) podrían tener una ventaja injusta respecto de los vectores más cortos.

Además, la normalización de los valores de coseno permite que estos sean calculados como un producto simple (multiplicación de los vectores) (Deerwester et al, 1990; Landauer et al, 1998; Manning & Schütze, 2003).

2. MARCO METODOLÓGICO

2.1. Tipo de estudio y variables

Como planteamos en la introducción de este trabajo, los objetivos de esta investigación son: a) comparar, utilizando una herramienta computacional de análisis vectorial denominada LSA, la relación léxico-semántica entre tres variables textuales presentes en artículos de investigación científica, estas son: palabras clave, resumen y contenido del artículo; y b) comparar, a partir de los valores de similitud léxico-semántica de las variables textuales, una muestra de artículos de investigación científica de dos áreas de la ciencia (ciencias biológicas y ciencias sociales).

Para cumplir con estos objetivos realizamos una investigación de carácter no experimental exploratorio-descriptiva, enmarcada en una metodología cuantitativa. Es exploratoria puesto que no existen estudios respecto de la escritura científica realizados en español utilizando herramientas computacionales de cálculo de similitud léxico-semántica (SLS) con LSA. Es descriptiva, pues se pretende indagar la incidencia y los valores en que se manifiestan las relaciones entre las variables textuales a investigar y, de este modo, caracterizar los artículos de investigación científica de las dos áreas en estudio a partir de su contenido léxico.

Las variables que se consideran en este estudio son de dos tipos: textual y disciplinaria. Las variables textuales son: a) palabras clave: un grupo de palabras o frases nominales, que de forma altamente abstracta compactan el significado principal del artículo de investigación científica con el que están asociadas; b) resumen: texto breve (de 300 palabras aproximadamente) que compacta léxico-semánticamente el significado global del artículo de investigación científica del cual procede; c) contenido: un texto compuesto por varios apartados retórico-estructurales en los cuales se textualizan los procedimientos lógicos, seguidos por uno o varios investigadores en una investigación científica y los argumentos que sirven para dar cuenta e interpretar los hallazgos de dicha investigación.

La variable disciplinaria corresponde al área científica, más específicamente a los artículos de investigación científica publicados en revistas científicas indexadas, agrupados a priori en dos dominios: ciencias biológicas y ciencias sociales.

2.2. Hipótesis del estudio

Las hipótesis de estudio de esta investigación son las siguientes:

H_1: Al comparar los índices de similitud léxico-semántica entre las variables textuales (palabras clave-resumen, palabras clave-contenido y resumen-contenido) de los artículos de investigación científica de dos áreas del conocimiento la relación resumen-contenido presentará valores más altos de similitud que las relaciones de las palabras clave-contenido y palabras clave-resumen ($\alpha < 0,05$).

H_2: Al comparar los índices de similitud léxico-semántica entre las variables textuales (palabras clave-resumen, palabras clave-contenido y resumen-contenido) de los artículos de investigación científica en un espacio semántico especializado se

presentan diferencias significativas entre las dos áreas del conocimiento investigadas ($\alpha < 0,05$).

2.3. El corpus

Para llevar a cabo la cuantificación de las SLS entre las variables textuales: palabras clave-resumen, palabras clave-contenido y resumen-contenido se determinó estadísticamente un corpus de investigación de 22 artículos de investigación científica: 12 de ciencias biológicas y 10 de ciencias sociales.

Este corpus de investigación corresponde a una muestra estratificada representativa, con un 97% de confianza, de un corpus mayor denominado ARTICO (Artículos de Investigación Científica Originales) (Venegas, 2005). ARTICO está constituido por 675 artículos pertenecientes a un limitado rango de ámbitos de la ciencia, a saber: ciencias exactas (ciencias químicas, física e ingeniería química), ciencias biológicas (ciencias veterinarias, zoología y oceanología, ciencias morfológicas, ecología y sus subdisciplinas, infectología y microbiología) y ciencias sociales (ciencias de la comunicación, ciencias de la información documental, antropología social y cultural, arqueología, economía, sociología, políticas culturales y sociales), publicados entre los años 2000 y 2003 en español por revistas de corriente principal en cada una de las áreasde la ciencia, esto es, revistas puestas a disposición para los investigadores en los indexadores electrónicos ScIELO (Scientific Electronic Library Online) y Latindex (Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal) que cumplen con los criterios internacionales de indexación (por ejemplo, ver www.scielo.org o www.latindex.unam.mx). Además, son textos que presentan las variables textuales requeridas para la investigación. En el Anexo 1 se especifica el corpus ARTICO, considerando las revistas, la cantidad de artículos seleccionados y el total de palabras, según el área científica en estudio.

En la Tabla 1, se presenta el corpus de investigación utilizado para llevar a cabo la cuantificación de las similitudes semánticas. Para mayores detalles bibliográficos ver Anexo 2 (Codificación y referencias bibliográficas de la muestra de investigación).

Tabla 1. Corpus de investigación.

	ciencias biológicas	Nº palabras	ciencias sociales	Nº palabras
1	AMV3-2002 34(1)	4908	AMB1 2001 6	5550
2	GC3 2002 66(2)	3086	AMB11 2001 6	8395
3	GC3 2003 67 (1)	4982	AMB14 2001 6	4121
4	RCHA4 2001 19(3)	3150	AD14 2002 5	6156
5	RCHA5 2001 19(3)	2735	CHU2 2002 34 (1)	13419
6	RCHA5 2002 20(2)	1780	INE4 2001 25(2)	8876

7	RCHA2 2002 20(3)	4251	INE2 2002 26(2)	8034
8	RCHN6 2000 73(4)	5010	NS3 2003 184	6701
9	RCHN7 2000 73(4)	4407	NS7 2003 186	7150
10	RCHN1 2002 75(2)	5522	NS10 2003 188	5436
11	AMV2 2001 33(1)	2609		
12	AMV15 2002 34(2)	3519		
suma		45959		73838
promedio		3829,9		7383,8

2.4. Conformación del espacio semántico ES-ARTICO

Con el fin de llevar a cabo los análisis de SLS entre las variables a investigar, se ha utilizado el LSA tal como ha sido descrito anteriormente. De este modo, a partir del corpus ARTICO, se construyó un espacio semántico especializado de 294 dimensiones, denominado ES-ARTICO (Venegas, 2005). Cabe señalar que la versión computacional utilizada ha sido generosamente facilitada por el equipo de investigadores del IIS (*Institute of Inteligent Systems*) de la Universidad de Memphis, dirigido por Arthur C. Graesser.

Luego se procedió a calcular para cada artículo del corpus de investigación, en contraste con ES-ARTICO, la SLS de cada una de las palabras clave respecto del resumen del artículo, así como la SLS de cada una de las palabras clave respecto de cada párrafo del contenido del artículo. Así también, se calculó la SLS del resumen respecto de cada párrafo del contenido de cada artículo de investigación científica. Por último, se procedió a realizar pruebas estadísticas de diferencias de medias para comprobar si existían o no diferencias significativas entre la relación de variables y entre las áreas de la ciencia.

3. RESULTADOS

3.1. Análisis de los índices de similitudes semánticas entre las variables

Los datos que se consideran para el análisis corresponden a los promedios totales de las SLS entre las variables textuales. Con estos promedios se realizó un procedimiento estadístico de segmentación en cuartiles con el fin de establecer un parámetro que nos permitiera evaluar por área y comparar entre las áreas los índices de similitud léxico-semántica (I-SLS). En la Tabla 2 se observa la relación entre cada rango porcentual del cuartil con el grado de similitud léxico-semántica (Gº-SLS) y el valor umbral obtenido por todos los I-SLS entre variables textuales, calculados para todos los artículos del corpus de investigación.

Tabla 2. Valores de segmentación en cuartiles para evaluación de los índices de similitud léxico-semántica.

Cuartiles	**1**	**2**	**3**	**4**
Rango Porcentual	0-25%	25-50%	50-75%	75-

				100%
Grado de similitud léxico-semántica (Gº-SLS)	Bajo	Medio Bajo	Medio Alto	Alto
Valor de similitud léxico-semántica para cada rango	0,0134	0,1781	0,2915	0,4976

De este modo, aquellos valores que se encuentren en el rango que va de 0,0134 a 0,1780 (para la presentación de los resultados se consideran cuatro posiciones decimales) corresponderán a un bajo Gº-SLS. Aquellos que se encuentren entre 0,1781 y 0,2914 tendrán un Gº-SLS medio bajo. Los I-SLS que se ubiquen entre 0,2195 y 0,4975 presentarán un Gº-SLS medio alto. Por último, aquellos I-SLS que sobrepasen el valor 0,4976 corresponderán a un Gº-SLS alto.

3.2. Análisis de las similitudes semánticas de las variables en ciencias biológicas

En la Tabla 3 se sintetizan los resultados obtenidos para la comparación de SLS entre las variables textuales en ciencias biológicas.

Con relación a las variables palabras clave y resumen a comparar en esta área, cabe señalar que RCHA5 2001 19(3) es el artículo que presenta el menor I-SLS (0,1468), correspondiendo este índice a un bajo Gº-SLS. Por el contrario, el artículo que presenta mayor I-SLS entre estas variables es GC3 2003 67(1) (0,3569), valor que según nuestra segmentación en cuartiles corresponde a un Gº-SLS medio alto. En cuanto al índice promedio entre las variables, su valor alcanza a 0,2595 correspondiendo a un Gº-SLS medio bajo.

Tabla 3. Índices de similitud léxico-semántica entre variables textuales en ciencias biológicas.

Artículos ciencias biológicas	**Similitudes entre variables**		
	P + R	**R+C**	**P + C**
AMV3-2002 34(1)	0,2259	0,6196	0,1446
GC3 2002 66(2)	0,2928	0,6218	0,2054
GC3 2003 67 (1)	**0,3569**	0,3472	0,1412
RCHA4 2001 19(3)	0,2856	0,5635	0,2003
RCHA5 2001 19(3)	0,1468	0,4022	0,0661
RCHA5 2002 20(2)	0,2058	0,4324	0,1025
RCHA2 2002 20(3)	0,2698	0,5155	0,1845
RCHN6 2000 73(4)	0,2180	0,4207	0,0834
RCHN7 2000 73(4)	0,2670	**0,6366**	**0,2059**
RCHN1 2002 75(2)	0,2915	0,5062	0,1602
AMV2 2001 33(1)	0,2546	0,4601	0,1546
AMV15 2002 34(2)	0,2988	0,4350	0,1418
Promedio	**0,2595**	**0,4967**	**0,1492**

En cuanto a la comparación entre las variables resumen y contenido en esta área es posible establecer que el artículo GC3 2003 67(1) presenta el I-SLS más bajo (0,3472), correspondiendo a un Gº-SLS medio alto entre las variables. Por el

contrario, el artículo de investigación que presenta el mayor I-SLS entre estas variables es RCHN7 2000 73(4) obteniendo un valor de 0,6366, correspondiendo a un alto Gº-SLS. En cuanto al valor promedio de I-SLS de estas variables en ciencias biológicas, este alcanza a 0,4967 correspondiendo a un Gº-SLS medio alto.

Como podemos observar en la comparación entre palabras clave y contenido en esta área, el artículo RCHA5 2001 19(3) es el que presenta el menor I-SLS entre las variables con un valor de apenas 0,0661, correspondiendo a un Gº-SLS bajo. Por otra parte, el artículo RCHN7 2000 73(4) presenta el índice más alto en esta comparación con 0,2059, siendo esto un Gº-SLS medio bajo. En promedio los artículos de esta área presentan entre las palabras clave y el resumen un índice de 0,1492, lo cual correspondería a un Gº-SLS bajo. Este valor promedio entre estas variables ubica a los artículos de ciencias biológicas como los de menor I-SLS entre los artículos investigados para ambas áreas de la ciencia. En este sentido, tanto resumen como palabras clave, como comprobaremos más adelante, se relacionan menos en los artículos de esta área que en los de ciencias sociales. Esto puede deberse a que existe menor rigurosidad en la redacción del resumen, no considerando algunos significados relevantes del contenido y a que la elección de las palabras clave se utilizarían menos para macrosemantizar el contenido semántico global del artículo que para dar cuenta de aspectos más contextuales o metatopicales de la investigación (por ejemplo, referencia al territorio temático de la investigación, descripción de lugares geográficos, etc.).

En general estos bajos I-SLS en los artículos de la muestra en ciencias biológicas, pueden deberse a que el método de cuantificación de las similitudes no capta las estrategias de macrosemantización que exceden el contenido textual. Es posible pensar, en este sentido, que las palabras clave sí macrosemantizan el significado global del texto, pero no en términos de las relaciones léxicas intratextuales, sino que en función de relaciones de macrosemantización más abstractas aún, de tipo exógenas al texto y de carácter eminentemente intertextual e incluso interdiscursivo.

Otro argumento posible, que puede explicar esta menor SLS, particularmente en cuanto al resumen, es que los artículos escogidos al azar en la muestra presentan una mayor tendencia hacia una construcción descriptiva más que informativa del resumen. De este modo, se ofrecería en estos resúmenes los enunciados fundamentales del trabajo original, pero no en cuanto a resultados concretos de las reflexiones o de los estudios expuestos en el artículo (ver Ratteray, 1985).

3.3. Análisis de las similitudes semánticas de las variables en ciencias sociales

En la Tabla 4 se presentan los resultados obtenidos para la comparación de SLS entre las variables en estudio en ciencias sociales.

Los resultados para la comparación entre las variables textuales palabras clave y resumen muestran que el artículo AD14 2002 5 es el que obtiene el menor I-SLS (0,0891), correspondiendo a un Gº-SLS bajo. Por el contrario, el artículo de investigación INE2 2002 26(2) es el que presenta el mayor I-SLS (0,4762), correspondiendo a un Gº-SLS medio alto. Cabe hacer notar, que este artículo es el que presenta el mayor I-SLS entre las variables palabras clave y resumen de todos los artículos investigados en las dos áreas de la ciencia. El resultado promedio de I-SLS entre estas variables alcanza a un valor de 0,2532, correspondiendo a un Gº-SLS medio bajo.

Tabla 4. Índices de similitud léxico-semántica entre variables textuales en ciencias sociales.

Artículos ciencias sociales	Similitudes entre variables		
	P + R	R+C	P + C
AMB1 2001 6	0,2479	0,1693	0,0507
AMB11 2001 6	0,2256	0,6572	0,2593
AMB14 2001 6	0,1020	0,3830	0,1266
AD14 2002 5	0,0891	0,5106	0,0707
CHU2 2002 34 (1)	0,3580	0,4831	0,2003
INE4 2001 25(2)	**0,3224**	0,5900	**0,2608**
INE2 2002 26(2)	0,4762	0,6307	0,3762
NS3 2003 184	0,2183	0,5019	0,1470
NS7 2003 186	0,2405	0,5965	0,2128
NS10 2003 188	0,2517	**0,7268**	0,2730
promedio	**0,2532**	**0,5249**	**0,1977**

En cuanto a la comparación entre resumen y contenido en los artículos de ciencias sociales, podemos observar que el artículo que presenta el menor I-SLS es AMB1 2001 6 con 0,1693, lo cual corresponde a un Gº-SLS bajo. Por otra parte, el artículo NS10 2003 188 presenta el mayor I-SLS entre las variables resumen y contenido en esta área, alcanzando un valor de 0,7628. En promedio, los I-SLS entre las variables alcanzan un valor de 0,5249, lo que corresponde a un alto Gº-SLS promedio.

Con relación a la comparación entre las variables palabras clave y contenido es posible observar que el artículo AMB1 2001 6 presenta el menor I-SLS entre las variables (0,0507), correspondiendo a un Gº-SLS medio bajo. Por el contrario, el artículo INE2 2002 26(2) es el que presenta el mayor I-SLS (0,3762) entre las variables palabras clave y contenido en ciencias sociales. En cuanto al valor promedio de los I-SLS entre estas variables es posible señalar que alcanzan un valor correspondiente a 0,1977, lo que implica un Gº-SLS medio bajo.

3.4. Comparación de las variables textuales según áreas de la ciencia

A continuación, presentamos comparativamente los resultados de SLS entre las variables textuales según las dos áreas de la ciencia en estudio.

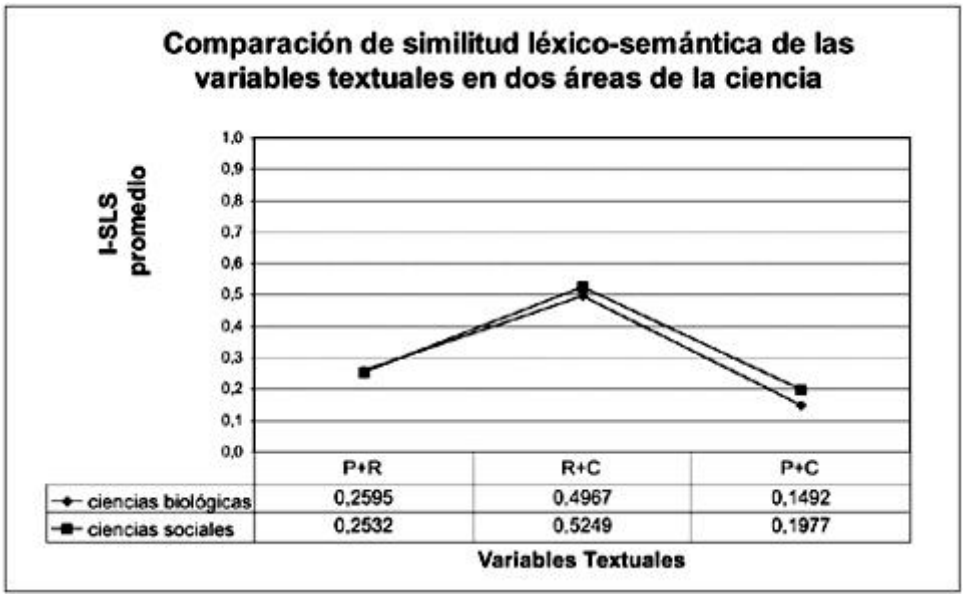


Gráfico 1. Comparación de los índices promedios de similitud léxico-semántica entre las variables textuales de las dos áreas.

Como se puede observar en el Gráfico 1 en los promedios de SLS entre las variables textuales se configura un patrón de relaciones de similitud semántica común a ambas áreas científicas, así las variables palabras clave-resumen y palabras clave-contenido presentan en todos los artículos investigados, independientemente de la ciencia a la cual pertenecen, un valor promedio menor que las relaciones de similitud semántica entre resumen-contenido. En términos más específicos, se observa un índice promedio levemente mayor en la relación entre las palabras clave y el resumen que en las relaciones entre palabras clave-contenido, lo que permite suponer una tendencia a relacionar las palabras clave más con el resumen que con el contenido textual.

La contrastación estadística de estas relaciones nos permite confirmar la idea anterior. Así, al comparar la SLS entre las variables textuales (palabras clave-resumen, palabras clave-contenido y resumen-contenido), utilizando la prueba U de Mann-Whitney y considerando un 5% de error, podemos establecer que en ciencias biológicas existe diferencia estadística entre las tres relaciones, es decir, que la SLS entre palabras clave y contenido es menor que la relación de SLS entre palabras clave y resumen. A su vez ambas relaciones son menores que la SLS existente entre el resumen y el contenido. En ciencias sociales la prueba estadística también mostró diferencias entre las tres relaciones de variables. Así, la SLS entre palabras clave-resumen es menor que la SLS entre palabras clave-contenido, y ambas son menores que la SLS existente entre el resumen y el contenido de los artículos investigados.

Estos resultados nos permiten confirmar nuestra primera hipótesis de investigación, dado que en ambas áreas el resumen macrosemantiza mejor que las palabras clave el contenido semántico global de los AIC. También podemos identificar un patrón de SLS, en ambas áreas, en el que se presenta una jerarquía de macrosemantización. Así la mayor macrosemantización está dada en la relación resumen-contenido, seguida por palabras clave-resumen, siendo la menor macrosemantización la que se presenta entre palabras clave y contenido.

Un argumento posible en torno a estos resultados es que los artículos seleccionados, independiente del área de la ciencia a la cual representan, responden a procesos similares de producción, esto es, son artículos que dan cuenta de investigaciones o reflexiones científicas, producidos por escritores científicos y publicados en revistas indexadas. Esto supone un exigente proceso de estandarización en cuanto a su forma y contenido, y por ello en cuanto a las relaciones de similitud léxico-semántica entre las variables estudiadas. Esta estandarización se concreta gracias a los procesos editoriales que tienen por fin publicar artículos coherentes e informativos.

3.5. Comparación de las áreas científicas

Los resultados anteriores nos permiten establecer a partir de las SLS de las variables textuales un patrón de comportamiento similar para ambas ciencias. Con el fin de comprobar si este patrón diferencia o no las áreas en estudio se comparan estadísticamente los valores promedios de todas las relaciones de SLS (ver Gráfico 2).

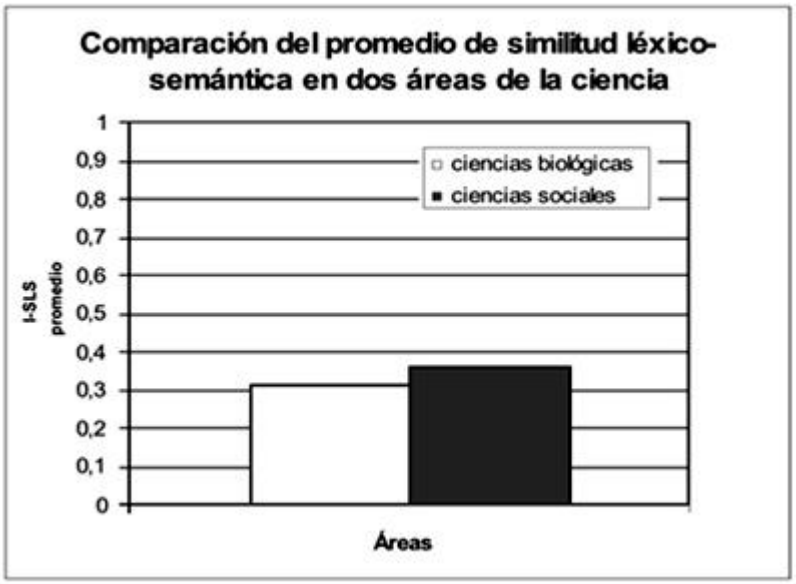


Gráfico 2. Comparación del promedio de similitud léxico-semántica de las dos áreas según ES-ARTICO.

De este modo, se aplicó el test no paramétrico de Wilcoxon, según el cual se obtuvo un valor alfa igual a 0,4510. Con este resultado, nuestra segunda hipótesis de investigación se refuta, esto quiere decir que al analizar los I-SLS entre las variables textuales (palabras clave-resumen, palabras clave-contenido y resumen-contenido) de los artículos de investigación científica, en un espacio semántico especializado, no se presentan diferencias significativas (con un error de un 5%) entre las tres áreas del conocimiento investigadas.

Una interpretación posible de este resultado es que, en términos generales, los artículos de investigación científica usados en la muestra en estudio son el producto material de un complejo proceso de producción científica, en el que múltiples agentes conjugan sus competencias discursivas y disciplinares para lograr un artículo de investigación científica que tenga la calidad suficiente para ser publicado en una revista científica indexada, acorde con estándares internacionales de calidad

científica. Así, pareciera que los artículos de investigación tienden hoy en día a una mayor homogeneización retórico-estructural, y por ello también a una SLS, en términos muy generales, similar entre las áreas. Todo lo anterior, pensamos, en función del mejoramiento de la calidad comunicativa de las investigaciones realizadas en las distintas áreas de la ciencia.

CONCLUSIONES

Según los resultados obtenidos en este trabajo podemos afirmar, en primer lugar, que la relación resumen-contenido en todos los AIC de nuestro corpus es mayor que las relaciones palabras clave-resumen y palabras clave-contenido. Estos resultados confirman la función macrosemantizadora del resumen, así como también la idea de que el resumen ayuda al lector a formarse una hipótesis sobre los tópicos centrales del texto. En nuestro caso particular, hemos comprobado que el grado de SLS del resumen es estadísticamente alto con relación al contenido, por lo que para este tipo de textos, en las áreas investigadas, las hipótesis de los lectores serían cumplidas probablemente en un alto grado.

Otra conclusión, a partir de estos datos, permite establecer que las palabras clave, al menos en los AIC estudiados, no tienen una clara función macrosemantizadora del significado global del contenido. Siendo esto así, se puede argumentar que lo más probable es que las palabras clave funcionen situando el artículo de investigación en un campo disciplinar, temático o procedimental, haciendo muy poca referencia a ello en el interior del artículo. Cabe señalar, que el método de cuantificación de las similitudes se funda en la colocación cotextual de las palabras en el texto, siendo de este modo altamente probable que esta función de las palabras clave no sea captada por los valores de SLS que entrega el LSA. De este modo, se puede dar el caso de que las palabras clave sí macrosemanticen el significado global del texto, pero no en términos de las relaciones léxicas intratextuales, sino que en función de relaciones de macrosemantización más abstractas aún, de tipo metatopicales y eminentemente intertextuales. Cabe también argumentar que las palabras clave cumplirían con otra función específica. Esta sería de corte más persuasivo que informativo, esto es, los escritores o incluso los comités editoriales utilizan o sugieren palabras o frases clave que, si bien no se relacionan fuertemente con el texto, sí despiertan el interés de un posible lector, cumpliéndose con ello un primer paso de acercamiento a la lectura del texto por parte de comunidades científicas en las cuales se tenga un particular interés.

Concluimos también que no se presentan diferencias significativas entre los textos del corpus de investigación, provenientes de las áreas de ciencias biológicas y ciencias sociales, según las variables textuales analizadas. Como se comprende, este resultado no responde a lo proyectado, dado que nocionalmente se esperaba que en ciencias biológicas existiera una mayor rigurosidad y estandarización en la escritura del artículo de investigación científica, fundamentalmente entre las partes del artículo científico y en el uso terminológico del léxico, en contraste con los textos provenientes de las ciencias sociales, donde suponíamos una menor estandarización retórico-estructural, variabilidad léxica e inestabilidad conceptual. Una explicación posible para esta homogeneidad, detectada a partir de los datos, es que si bien es posible reconocer en las diferentes disciplinas distintos campos temáticos, cuya textualización se realiza por medio de recursos léxicos y terminológicos e incluso en muchos casos retórico-estructurales propios de la disciplina, en lo que concierne a las relaciones semánticas de tipo léxico asociativo que se establecen en el nivel textual global de los artículos no existirían diferencias significativas en la forma de establecer relaciones léxico-semánticas.

Lo anterior nos permite pensar que los escritores científicos, autores de los artículos investigados, pueden presentar una competencia textual similar, la que les permitiría construir relaciones léxico-semánticas al interior de sus textos en términos más o menos parecidos, independientemente del área en la que se especializan.

Otro argumento complementario al anterior tiene que ver con que los AIC que constituyen el corpus de investigación corresponden a artículos publicados en revistas científicas indexadas acorde con exigentes estándares de calidad internacional. En este sentido, el artículo producido por el autor o autores está sometido a un complejo proceso editorial. A través de él, muchos productores/comprendedores científicos conjugan sus conocimientos disciplinares y sus competencias textuales-discursivas para co-construir un artículo de investigación científica que tenga no solamente calidad en el contenido, sino que también presente calidad en su organización, acorde con la estructura retórica exigida por la revista. Así, es posible que los artículos de investigación tiendan hoy en día, en la mayoría de las disciplinas, a una mayor homogeneización retórico-estructural y, por ello, también a una SLS similar. De este modo, se tiende hacia una paulatina semejanza en la cual se pierden las diferencias disciplinares propias, al menos en cuanto a las diferencias en los procesos de macrosemantización entre las variables investigadas en las muestras de textos de ciencias biológicas y ciencias sociales.

Acorde con lo anterior, es posible inferir que un investigador que quiera integrarse a una comunidad discursiva científica debe aprender, entre otras cosas, a comunicar su investigación, según las normas semántico-textuales asociadas a este tipo de texto y a las normas disciplinares propias de las revistas de su especialidad, asistido en este proceso de aprendizaje por las instancias editoriales de la revista en la cual se desea publicar. Este proceso incluye múltiples evaluaciones y sugerencias que hacen tanto los pares científicos como el comité editorial y/o editor de la revista. En ellos recae la responsabilidad final del artículo publicado, constituyéndose por ello todo el proceso de escritura científica en una co-construcción semántico-textual, orientada hacia un producto discursivo constructor de conocimiento disciplinar. De este modo, resulta interesante la idea de que el artículo de investigación científica, finalmente publicado, en algunos casos, pueda ser un producto de gran interacción. Esto quiere decir que, para llegar al formato final, el escritor debe atender a múltiples voces que en definitiva le podrían hacer cambiar no solo su formato y contenido, sino que también su propósito comunicativo original.

Finalmente, debemos poner énfasis en el hecho de que los resultados obtenidos a través del LSA, nos permiten confirmar empíricamente la función del resumen en el AIC y determinar que las variables textuales macrosemantizan el contenido de los AIC del mismo modo, independiente del área disciplinar. En otras palabras el LSA, utilizando solo datos matemáticos-estadísticos a partir de los textos seleccionados, permite de una manera bastante precisa y económica, en términos informáticos, establecer las fuerzas de relación entre los componentes léxicos, asignándoles valores de similitud semántica y entregando como producto el grado de similitud léxico-semántica entre componentes textuales e incluso retórico-estructurales, como ha sido lo realizado en nuestra investigación. De este modo, resaltamos el valor de la herramienta que hemos utilizado para construir el primer espacio semántico en español y usada en una investigación de corte lingüístico-textual ya que, comparativamente con modelos simbólicos o basados en análisis lógico-

proposicionales, el LSA es mucho más económico en términos de procesamiento y eficiente en términos de tiempo y costos de programación.

NOTA

[1] Los números de las Figuras 1 y 2 se presentan a modo de representación de los procesos, no correspondiendo a valores obtenidos por los procesos matemático-estadísticos.

REFERENCIAS BIBLIOGRÁFICAS

Bazerman, Ch. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: The University of Wisconsin Press. [Links]

Bolívar, A. (2000). Homogeneidad versus variedad en la estructura de los resúmenes de investigación para congresos. *Akademos*, 2,121-138. [Links]

Cabré, M.T. (1999). El discurs especialitzat o la variació funcional determinada per la temàtica: Noves perspectives. En T. Cabré (Ed.), *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos* (pp. 151-173). Barcelona: IULA. [Links]

Cabré, M.T. (2002). Textos especializados y unidades de conocimiento: Metodología y tipologización. En J. García & M. Fuentes (Eds.), *Texto, terminología y traducción* (pp. 122-187). Barcelona: Almar. [Links]

Calsamiglia, H. (Coord.) (1998). *Análisis discursivo de la divulgación científica* [en línea]. Disponible en:http://www.upf.es/dtf/personal/danielcass/anali.htm [Links]

Cassany, D., López, C. & Martí, J. (2000). La transformación divulgativa de redes conceptuales científicas: Hipótesis, modelo y estrategias. *Discurso y Sociedad, 2*(2), 73-103. [Links]

Ciapuscio, G. (1994). *Tipos textuales*. Buenos Aires: EUDEBA. [Links]

Ciapuscio, G. (2000). Hacia una tipología del discurso especializado. *Discurso y Sociedad, 2*(2), 39-71. [Links]

Ciapuscio, G. (2003). *Textos especializados y terminología*. Barcelona: IULA. [Links]

Ciapuscio, G. & Otañi, I. (2002). Las conclusiones de los artículos de investigación desde una perspectiva contrastiva. *RILL*, 15, 117-133. [Links]

CONICYT (2004). *Indicadores científicos y tecnológicos* [en línea]. Disponible en:http://www.conicyt.cl/bases/indicadores/ [Links]

Chafe, W. (1994). *Discourse, consciousness and time*. Chicago: The University of Chicago Press. [Links]

Deerwester, S., Dumais, S., Furnas, G., Landauer, T. & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391-407. [Links]

Dudley-Evans, T. (1986). Genre analysis: An investigation of the introduction and discussion sections of MSc. dissertation. En M. Coulthard (Ed.), *Talking about text* (pp. 128-145). Birmingham: English Language Research, Birmingham University. [Links]

Foltz, P. (1990). Using latent semantic indexing for information filtering. En R. Allen (Ed.), *Actas de the Conference on Office Information Systems* (pp. 40-47). Cambridge, MA: MIT Press. [Links]

Gläser, R. (1982). *The problem of style classification in LSP (ESP)*. Ponencia presentada en the 3rd European Symposium on LSP, Copenhague, Dinamarca. [Links]

Gläser, R. (1993). A multi-level model for a typology of LSP genres. *Fachsprache. International Journal of LSP*, *15*(1-2), 18-26. [Links]

Gnutzmann, C. & Oldenburg, H. (1991). Contrastive text linguistics in LSP-research: Theoretical considerations and some preliminary findings. En H. Schröder (Ed.), *Subject-oriented texts. Language for special purposes and text theory* (pp. 103-136). Berlin: W. de Gruyter. [Links]

Gotti, M. (2003). *Specialized discourse. Linguistic features and changing conventions*. Bern: Peter Lang. [Links]

Graesser, A., Person, N., Harter, D. & Tutoring Research Group (2001). Teaching tactics and dialog in Autotutor. *International Journal of Artificial Intelligence in Education*, 12, 257-279. [Links]

Gutiérrez, R. (2005). Análisis semántico latente: ¿Teoría psicológica del significado? *Revista Signos*, 38(59), 303-323. [Links]

Halliday, M. (1991). Corpus studies and probabilistic grammars. En K. Aijmer & Altenberg, B. (Eds.), *English corpus linguistics. Studies in honour of Jan Svartvik* (pp. 31-43). London: Longman. [Links]

Halliday, M. (1992). Language as a system and language as a instance: The corpus as a theoretical construct. En J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 61-77). New York: W. de Gruyter. [Links]

Halliday, M. & Martin, J. (1993). *Writing science: Literacy and discursive power*. London: Falmer. [Links]

Hartley, J. & Kostoff, R. (2003). How useful are □key words□ in scientific journals? *Journal of Information Science*, *29*(5), 433-438. [Links]

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: Benjamins. [Links]

Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. En C. Candlin & K. Hyland (Eds.), *Writing texts, processes and practice* (pp. 99-121).

London: Longman. [Links]

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman. [Links]

Jeanneret, Y. (1994). *Écrire la science. Formes et enjeux de la vulgarisation*. Paris: PUF. [Links]

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

Kintsch, E., Steinhart, D., Stahl, G., LSA Research Group, Matthews, C. & Lamb, R. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments*, *8*(2), 87-109. [Links]

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press. [Links]

Kintsch, W. (2000). Metaphor comprehension: A computational Theory. *Psychonomic Bulletin & Review*, *7*(2), 257-266. [Links]

Kintsch, W. (2001). Predication. *Cognitive Science*, *25*(2), 173-202. [Links]

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: Chicago University Press. [Links]

Landauer, T. (2002). On the computational basis of learning and cognition: Arguments from LSA. *Psychology of Learning and Motivation*, 41, 43-84. [Links]

Landauer, T. & Dumais, S. (1996). How come you know so much? From practical problem to theory. En D. Hermann, C. McEvoy, M. Jonson & P. Hertel (Eds.), *Basic and applied memory: Memory in context* (pp. 105-126). Mahwah, NJ: Erlbaum. [Links]

Landauer, T. & Dumais, S. (1997). A solution to Plato□s problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211-240. [Links]

Landauer, T., Foltz, P. & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284. [Links]

López, C. (2002). Aproximaciones al análisis de los discursos profesionales. *Revista Signos*, *35*(51-52), 195-215. [Links]

López, J. (1997). El resumen como fuente de información y medio de comunicación. *RESUMED*, *10*(3), 103-106. [Links]

Manning, C. & Schütze, H. (2003). *Foundations of statistical natural language processing*. Cambridge: MIT Press. [Links]

Martin J. & Rose, D. (2003). *Working with Discourse. Meaning beyond the clause*. London: Continuum. [Links]

Martín, P. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, 22, 25-43. [Links]

Maruhenda, Y. (2003). *Estructura de un artículo científico* [en línea]. Disponible en:http://gidovi.umh.es/doctorado/internetInvestigacion/estructuraArticulo.html [Links]

Matsumoto, Y. (2003). Lexical knowledge acquisition. En R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 396-409). New York: Oxford University Press. [Links]

Mogollón, G. (2003). Paradigma científico y lenguaje especializado. *Revista de la Facultad de Ingeniería de la Universidad Central de Venezuela*, *18*(3), 5-14. [Links]

Moyano, E. (2000). *Comunicar ciencia*. Buenos Aires: Secretaría de Investigaciones. Universidad Nacional de Lomas de Zamora. [Links]

Osgood, Ch. (1952). The nature and measurement of meaning. *Psychological Bulletin*, 49, 197-237. [Links]

Osgood, Ch., Suci, G. & Tannenbaum, P. (1976). *La medida del significado*. Madrid: Gredos. [Links]

Palmer, F. (1980). *La semántica*. Madrid: Siglo XXI. [Links]

Parodi, G. (2004). Textos de especialidad y comunidades discursivas técnico-profesionales: Una aproximación basada en corpus computarizado. *Estudios Filológicos, 39*(39), 7-36. [Links]

Parodi, G. (2005). Lingüística de corpus y análisis multidimensional: Exploración de la variación en el corpus PUCV-2003: Una aproximación multiniveles. En G. Parodi (Ed.), *Discurso especializado e instituciones formadoras* (pp. 83-125). Valparaíso: Ediciones Universitarias de Valparaíso. [Links]

Peronard, M. (1997). ¿Qué significa comprender un texto escrito? En M. Peronard, L. Gómez, G. Parodi & P. Núñez (Comps.), *Comprensión de textos escritos: De la teoría a la sala de clases* (pp. 55-78). Santiago: Andrés Bello. [Links]

Quesada, J. (2003). Latent problem solving analysis (LPSA): A computational theory of representation in complex, dynamic problem solving tasks. Tesis doctoral, Universidad de Granada, España [en línea]. Disponible en: http://www.andrew.cmu.edu/user/jquesada//dissertation/ [Links]

Quesada, J., Kintsch, W. & Gómez, E. (2002). A theory of complex problem solving using latent semantic analysis. En W. Gray & C. Schunn (Eds.), *Actas de the 24th Annual Conference of the Cognitive Science Society* (pp. 750-755). Mahwah, NJ: Erlbaum. [Links]

Ratteray, O. (1985). Expanding roles for summarized information. *Written Communication*, *2*(4), 457-472. [Links]

Rojo, G. (2002). *Sobre la lingüística basada en análisis de corpus* [en línea]. Disponible en:http://www.uzei.com/corpusajardunaldia/01_grojo.pdf[

[Links]STANDARDIZEDENDPARAG]

Russell, B. (1937). *Principles of mathematics*. London: George Allen & Unwin. [Links]

Sager, J., Dungworth, D. & McDonald, P. (1980). *English special languages*. Wiesbaden: Oscar Brandstetter. [Links]

Salager-Meyer, F. (1991). *A text-type based discourse analysis of medical English. Abstracts internal structuring*. Ponencia presentada en the Second Latin American ESP Colloquium, Santiago de Chile, Chile. [Links]

Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *ESP Journal*, *11*(2), 93-113. [Links]

Schröder, H. (1991). Linguistic and text-theorical research on languages for special purposes. A thematic and bibliographical guide. En H. Schröder (Ed.), *Subject-oriented texts. Languages for special purposes and text theory* (pp. 1- 48). Berlin: W. de Gruyter. [Links]

Sinclair, J.M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press. [Links]

Stubbs, M. (1996). *Text and corpus analysis. Computer-assisted studies of language and culture*. Malden, MA: Blackwell Publishers. [Links]

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers. [Links]

Swales, J. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press. [Links]

Swales, J. (2004). *Research genres. Explorations and applications*. Cambridge: Cambridge University Press. [Links]

Turney, P. (1997). *Extraction of keyphrases from text: Evaluation of four algorithms*. Ottawa: National Research Council Canada. Technical Report ERB-1051. [Links]

Turney, P. (1999). *Learning to extract keyphrases from text*. Ottawa: National Research Council, Institute for Information Technology. Technical Report ERB-1057. [Links]

van Dijk, T. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press. [Links]

Venegas, R. (2003). Análisis semántico latente: Una panorámica de su desarrollo. *Revista Signos*,*36*(53),121-138. [Links]

Venegas, R. (2005). Las relaciones léxico-semánticas en artículos de investigación científica: Una aproximación desde el análisis semántico latente. Tesis doctoral, Pontificia Universidad Católica de Valparaíso, Chile. [Links]

Wiemer-Hasting, P. (2004). *The FAQ for using LSA at Umemphis* [en línea].

Disponible en:http://www.msci.memphis.edu/~wiemerhp/trg/lsa-followup.html [Links]

Warta, V. (1996). Embedded case reports: A genre-analysis issue in teaching English for medical purposes. Tesis de Magíster, Aston University, Aston, Estados Unidos. [Links]

Dirección para correspondencia: René Venegas (rene.venegas@ucv.cl). Tel:(56-32) 273388. Fax:(56-32) 273448. Instituto de Literatura y Ciencias del Lenguaje, Pontificia Universidad Católica de Valparaíso. Av. Brasil 2830, piso 9, Valparaíso, Chile.

Recibido: 20-VI-2005 Aceptado: 3-XI-2005

ANEXO 1: Corpus ANTARTIC

CÓDIGO	Corpus ARTICO	Art.	Palabras
	CIENCIAS EXACTAS		
BSCQ	BOLETÍN DE LA SOCIEDAD CHILENA DE QUÍMICA http://www.scielo.cl/scielo.php?pid=0366-1644&script=sci_serial	78	275981
RCQ	REVISTA COLOMBIANA DE QUÍMICA http://www.icfes.gov.co/revistas/recolqui/	29	91025
RPQ	REVISTA PERUANA DE QUÍMICA E INGENIERÍA QUÍMICAhttp://sisbib.unmsm.edu.pe/bibvirtual/publicaciones/ing_quimica/Reglamento.htm	6	20841
AAL	ACTAS DE LA ACADEMIA LUVENTICUS http://www.luventicus.org/Actas/	6	21615
ITERC	INTERCIENCIA http://www.interciencia.org/	6	36615
ACV	ACTA CIENTÍFICA VENEZOLANA http://acta.ivic.ve/	7	32831
		132	**478.908**
	CIENCIAS BIOLÓGICAS		
GC	GAYANA CONCEPCIÓN http://www.scielo.cl/scielo.php?pid=0717-6538&script=sci_serial	40	159070
RCHA	REVISTA CHILENA DE ANATOMÍA http://www.scielo.cl/scielo.php?pid=0716-9868&script=sci_serial	66	219685
RCI	REVISTA CHILENA DE INFECTOLOGÍA http://www.scielo.cl/scielo.php?pid=0716-1018&script=sci_serial	34	136299
RCHN	REVISTA DE HISTORIA NATURAL http://www.scielo.cl/scielo.php?pid=0716-078X&script=sci_serial	101	657749
AMV	ARCHIVOS DE MEDICINA	57	263781

	VETERINARIA http://www.scielo.cl/scielo.php?pid=0301-732X&script=sci_serial		
		298	**1.436.584**
	CIENCIAS SOCIALES		
AMB	ÁMBITOS. REVISTA INTERNACIONAL DE COMUNICACIÓNhttp://www.ull.es/publicaciones/latina/ambitos/ambitos.htm	74	497473
CHU	CHUNGARA http://www.scielo.cl/scielo.php?pid=0717-7356&script=sci_serial	57	351956
AD	ANALES DE DOCUMENTACIÓN http://www.um.es/fccd/anales/	54	379925
NS	NUEVA SOCIEDAD http://www.nuevasoc.org.ve/home/	26	152501
INE	INVESTIGACIONES ECONÓMICAS http://www.funep.es/invecon/sp/sie.asp	34	333777
		245	**1.715.632**
	TOTAL	**675**	**3.631.124**

ANEXO 2: Referencias bibliográficas del corpus de investigación

CIENCIAS BIOLÓGICAS		
ID	**Código**	**Referencia Bibliográfica**
1	AMV3-2002 34(1)	Sievers, G., Jara, M., Cárdenas, C. & Núñez, J. (2002). Estudio anual de la eliminación de huevos y ooquistes de parásitos gastrointestinales y larvas de nemátodos pulmonares en ovinos de una estancia en Magallanes, Chile. Archivos de Medicina Veterinaria, 34(1), 37-47.
2	GC3-2002 66(2)	Lara, G., Parada, E. & Peredo, S. (2002). Alimentación y conducta alimentaria de la almeja de agua dulce Diplodon chilensis (bivalvia: hyriidae). Gayana (Concepc.), 66(2), 107-112.
3	GC3-2003 67(1)	Alarcón, M. (2003). Sifonapterofauna de tres especies de roedores de Concepción, VI Región, Chile. Gayana (Concepc.), 67(1), 16-24.
4	RCHA4-2001 19(3)	Vásquez, B. (2001). Presencia de CBG en el estroma ovárico de mamíferos. Revista Chilena de Anatomía, 19(3), 279-284.
5	RCHA5-2001 19(3)	Castro, A., Ghezzi, M., Alzota, R., Lupidio, M. & Rodríguez, J. (2001). Morfología del hígado de llama (Lama glama). Revista Chilena de Anatomía, 19(3), 291-296.
6	RCHA5-2002 20(2)	Briones, F., Calderón, M., Muñoz, J., Venegas, F. & Araya, N. (2002). El anticuerpo monoclonal Ki-67 como elemento de valor diagnóstico y pronóstico en neoplasias mamarias caninas. Revista Chilena de Anatomía, 20(2), 165-168.
7	RCHA2-2002 20(3)	Babinski, M., Chagas, M., Costa, W. & Pereira, M. (2002). Morfología y fracción del área del lumen glandular de la zona de transición en la próstata humana. Revista Chilena de Anatomía, 20(3), 255-262.
8	RCHN6-	Véliz, D. & Vásquez, J. (2000). La Familia Trochidae (Mollusca:

	2000 73(4)	Gastropoda) en el norte de Chile: consideraciones ecológicas y taxonómicas. Revista Chilena de Historia Natural, 73(4), 757-769.
9	RCHN7-2000 73(4)	Martínez, G. & Montecino, V. (2000). Competencia en Cladocera: implicancias de la sobreposición en el uso de los recursos tróficos. Revista Chilena de Historia Natural, 73(4), 787-795.
10	RCHN1-2002 75(2)	Canals, M., Atala, C., Olivares, R., Novoa, F. & Rosenmann, M. (2002). La asimetría y el grado de optimización del árbol bronquial en Rattus norvegicus y Oryctolagus cuniculus. Revista Chilena de Historia Natural, 75(2), 271-282.
11	AMV2-2001 33(1)	Díaz, D., Picco, E., Encinas, T. Rubio, M. & Litterio, N. (2001). Residuos tisulares de nicotinato de norfloxacina administrado por vía oral en cerdos. Archivos de Medicina Veterinaria, 33(1), 37-42.
12	AMV15-2002 34(2)	Perfumo, C., Sanguinetti, H., Giorgio, N., Armocida, A., Machuca, M., Massone, A., Risso, M., Aguirre, J. & Idiart, J. (2002). Constrictura rectal en cerdos necropsiados en una granja de ciclo completo en confinamiento. Consideraciones sobre su prevalencia, hallazgos anatomopatológicos y etiopatogenia. Archivos de Medicina Veterinaria, 34(2), 245-252.
CIENCIAS SOCIALES		
1	AMB1-2001 6	Emanuelli, P. (2001). Dominante cultural y productos televisivos: Géneros que homogenizan preferencias. Ámbitos, 6, 7-20.
2	AMB11-2001 6	Barrero, A. (2001). Juicios paralelos y Constitución: Su relación con el Periodismo. Ámbitos, 6, 171-189.
3	AMB14-2001 6	Egea, C. (2001). La carrera por la comunicación local (1998-2000) ☐Los grandes☐ se atreven con ☐lo pequeño☐. Ámbitos, 6, 237-260.
4	AD14-2002 5	Moreira, J. (2002). Aplicaciones al análisis automático del contenido provenientes de la teoría matemática de la información. Anales de Documentación, 5, 273-286.
5	CHU2-2002 34(1)	Schiappacasse, V. & Niemeyer, H. (2002). Ceremonial Inca provincial: El asentamiento de Sagura (cuenca de Camarones). Chungará, Revista de Antropología Chilena, 34(1), 53-84.
6	INE4-2001 25(2)	Goicolea, A., Lisandro, O. & Maroto, R. (2001). Picos de inversión y productividad del trabajo en los establecimientos industriales madrileños. Investigaciones Económicas, 25(2), 255-288.
7	INE2-2002 26(2)	Del Río, C. (2002). Desigualdad intermedia paretiana. Investigaciones Económicas, 26(2), 299-321.
8	NS3-2003 184	Hualde, A. (2003). ¿Existe un modelo maquilador? Reflexiones sobre la experiencia mexicana y centroamericana. Nueva Sociedad, 184, 86-101.
9	NS7-2003 186	Giacalone, R. (2003). Integración Norte/Sur y tratamiento especial y diferenciado en el contexto regional. Nueva Sociedad, 186, 69-85.
10	NS10-2003 188	Costa, S. (2003). Derechos humanos en el mundo posnacional. Nueva Sociedad, 188, 52-65