Un Modelo para la Homogeneización de las Clases Textuales de la Biblioteca Electrónica Scielo-Chile: la Variabilidad del Artículo de Investigación en Diversas Disciplinas

A Model for the Homogenization of Textual Classes in the Scientific Library Scielo-Chile: The Variability of Research Articles in Different Disciplines

Omar Sabaj, Ken I. Matsuda y Miguel A. Fuentes

Universidad de La Serena, Departamento de Artes y Letras, Facultad de Humanidades, Avenida Raúl Bitrán S/N, Campus Andrés Bello, Colina el Pino, La Serena-Chile (e-mail: omarsabaj@userena.cl, kmatsuda@userena.cl,miguelfuentesc@hotmail.com).

Resumen

En este trabajo se presenta un modelo formal y se aplican diversos índices cienciométricos que permiten analizar la variación de las clases textuales contenidas en los artículos de investigación publicados en las revistas de la biblioteca electrónica Scielo-Chile (http://www.scielo.cl). Para ello, se generó una matriz de 33 criterios formales para clasificar la variada gama de clases textuales que aparecen en las revistas de la biblioteca. Los resultados muestran que existen al menos tres tipos de artículos de investigación, y que su ocurrencia depende de la revista y de la disciplina a la que pertenece: de investigación científica, de investigación o estudio y de investigación teórica. Además, se establece que en algunas revistas la denominación de las clases textuales, por parte de los autores y editores no concuerda con la denominación derivada de la aplicación de los criterios formales. Por último, se estima que los artículos de investigación, en promedio, representan un 70% de todos los registros de Scielo-Chile.

Palabras clave: *comunicación científica, artículos de investigación, clases textuales, Scielo-Chile.*

Abstract

A formal model is presented and various science measurement indexes are applied for determining the variation in the textual classes included in research articles published in journals available in the electronic library Scielo-Chile (http://www.scielo.cl). To this end, a matrix of 33 formal criteria was generated to categorize the wide spectrum of textual classes that appear in the scientific journals of the library. The results show that there exist at least three types of research articles and that the presence of each type depends on the scientific journal and the discipline to which it belongs: scientific research, research and study, and

theoretical research. In addition to that, it is established that, in some scientific journals, the designation of textual classes by authors and editors does not agree with the designation derived from the application of the formal categorization. Finally, it is estimated that research articles represent 70% of all registered material included in Scielo-Chile.

Keywords: *scientific communications, research articles, textual classes, Scielo-Chile.*

INTRODUCCIÓN

Hoy en día, para ningún actor del mundo académico científico es una novedad la importancia de las publicaciones para el desarrollo de la investigación (Valderrama, 2001). Por el contrario, para muchos de los que vivimos en este mundo, publicar se ha vuelto una práctica inherente a nuestra vida profesional. A pesar de ello, desde el punto de vista de la enseñanza de la escritura científica, en Chile no existen programas formales que tiendan a institucionalizar esta práctica; por ejemplo, en la forma de centros de escritura. Por otro lado, los estudios dedicados al análisis de los artículos de investigación en español tienen desarrollos dispersos, que no permiten contar con un programa definido de trabajo. En este artículo se constata primero la falta de uniformidad en la denominación de las clases textuales contenidas en la biblioteca Scielo-Chile y se propone, en segundo término, un modelo para homogenizar la denominación de estas clases.

En primer lugar, mostramos el ámbito general de la investigación, caracterizando y comparando qué aspectos han sido o no estudiados, respecto de los artículos de investigación en español. En segundo término, definimos, sucintamente, algunas nociones clave para la delimitación del problema investigado. A continuación, mostramos los fundamentos metodológicos utilizados en la indagación empírica, junto con una discusión detallada de los procedimientos de obtención de información, los cálculos, los índices y las pruebas estadísticas aplicadas. Después de la exposición metodológica, se muestran los resultados, para finalizar con una discusión y algunas conclusiones del estudio. En general, los resultados muestran que existe una dependencia entre los tipos de artículos investigación y la disciplina.

Cualquiera que pretenda analizar la variación de los artículos de investigación (en adelante AI) a través de las disciplinas, debe reconocer en el libro de Bazerman (1988) una suerte de obra fundante. Por su parte, el famoso libro de Swales (1990), y todo el programa investigativo de Hyland (1998, 1999), también son referencias obligadas en este ámbito. A pesar de los enormes avances que estas obras fundantes han suscitado en las últimas dos décadas, estos estudios no están exentos de críticas y dejan vacíos que la presente investigación pretende completar. Al menos parcialmente.

Como una forma de explicar y definir el ámbito de esta investigación, en la siguiente tabla se muestran algunos vacíos y desafíos que este estudio pretende afrontar.

Tal como sucede en muchas otras áreas de la ciencia, los estudios en español son en este ámbito muchos más escasos que en inglés (punto 1). Esto causa una falta de información que deja a nuestro idioma en una condición clara de desventaja

respecto de los avances en otras lenguas. Tradicionalmente, además, y gracias a los avances en el ámbito del denominado Inglés con fines específicos y el análisis de género a través de las movidas retóricas (Swales, 1990), la parte que más se ha estudiado en el artículo de investigación corresponde a las introducciones y a los resúmenes (punto 2), pudiéndose encontrar solo a partir de la última década, estudios que pesquisan el AI como un todo. Así también (punto 3), es común que los trabajos en esta línea de investigación consideren un rango limitado de disciplinas, siendo aquellas con un mayor prestigio social y mayor impacto aplicado (ingeniería, ciencias exactas y naturales) las que favorablemente han sido estudiadas, quedando otras (i.e. humanidades y ciencias sociales) muy relegadas respecto de esas. Otro aspecto evidente en un rastreo de este tipo de trabajos, es el que se señala en el punto 4, a saber, que los estudios solamente se centran en AI altamente estructurados, con esquemas determinados y con partes superestructurales definidas (título, resumen, introducción, antecedentes teóricos, etc.). Esto oculta el hecho de que este tipo de textos no es tan homogéneo como se presenta en esos trabajos, omitiendo que, en efecto, se trata de un tipo textual cuya estructura varía de acuerdo a la disciplina, tal como ha sido descrito, entre otros, por Holmes (1997). Un punto especialmente relevante para el presente trabajo es aquel que se señala en el punto 5. Todos los trabajos que hemos revisado llevan a cabo la recolección de los AI a investigar, usando criterios (importancia de la disciplina, factor de impacto de la revista, entre otros) que no tienen en cuenta el contexto natural en los que este tipo de textos ocurren, a saber, las revistas especializadas. Esto conlleva a la falsa inferencia de que todas las disciplinas consideran a las revistas solamente como un medio para presentar AI, ignorando que la proporción de estos textos es altamente variable en cada revista, las que además contienen otros tipos de textos (editoriales, homenajes, reseñas, etc.), lo que, en definitiva, no nos permite saber cuál es la verdadera presencia de los AI en las disciplinas de cada revista. Al contrario, en el presente trabajo se busca analizar estos textos dentro del contexto general de una librería científica, de libre acceso en línea (http://www.scielo.cl) que contiene revistas especializadas de distintas disciplinas. En este sentido, se podrá determinar la presencia de este tipo de textos en el marco de la práctica científica chilena. Por último, en el punto 6, se presenta uno de los aspectos más criticados respecto de la investigación en esta área. Específicamente, que los trabajos no cumplen con la validación empírica de los procedimientos analíticos, tratándose muchas veces de modelos formales abstractos y normativos que no se contrastan con análisis empíricos de datos. En síntesis, a partir de la revisión de la literatura pertinente, y aun cuando se pueden presentar excepciones para cada uno de los vacíos señalados en el tabla 1, no se han encontrado investigaciones que integren y enfrenten de forma global los aspectos presentados.

Tabla 1: Algunos vacíos en el estudio de los artículos de investigación

	Vacío	Excepción
1	Se centran mayoritariamente en artículos escritos en idioma inglés	(Soto, 2002 & 2005; Soto & Zenteno, 2003; Soto, Martínez & Sadowsky, 2005)
2	Las investigaciones se han focalizado exclusivamente en partes estructurales específicas (i.e. Resúmenes e introducciones)	(Ngozi, 1997; Posteguillo, 1998; Swales, 2004; Kanoksilapatham, 2005; Pho, 2008)
3	Los trabajos consideran un rango limitado de disciplinas, siendo algunas siempre más estudiadas que otras	(Falahati, 1999; Samraj, 2001; Hyland, 2008)
4	Las investigaciones solo consideran en sus *corpora* artículos de investigación altamente estructurados	(Holmes, 1997)
5	La recolección de los *corpora* usados en las investigaciones no consideran la ocurrencia de los AI en los contextos naturales en que aparecen (i.e. Revistas especializadas), específicamente en relación a la ocurrencia de otras clases textuales que también aparecen en esas revistas (Editoriales, reseñas, homenajes, etc.)	
6	Falta de validación empírica de los procedimientos analíticos	(Fleiss, 1971; Grookes, 1986; Kanoksilapatham, 2005; Hayes & Krippendorff, 2007)

En este contexto, el propósito de este trabajo es doble. Por una parte se busca comprobar falta de uniformidad en la denominación de las clases textuales contenidas en la biblioteca Scielo-Chile. Por otra parte, se realiza una propuesta metodológica para uniformar dicha denominación. Tal como mostraremos más adelante, este acercamiento permite -por una parte- considerar la variedad de textos que pueden agruparse bajo el rótulo de AI, y -por otra- tener en cuenta el contexto natural en los que estos tipos de textos ocurren, a saber, las revistas especializadas.

Uno de los mayores desafíos que desde siempre ha enfrentado la lingüística del texto, es el denominado problema de los tipos textuales, principalmente debido a la proliferación de términos que, en muchas ocasiones, no concuerdan entre ellos. En este breve apartado, exponemos concisamente este problema y presentamos algunas decisiones y precisiones terminológicas que resultan clave para la comprensión de los procedimientos llevados a cabo en este trabajo. Algunos de los términos usuales que se han utilizado en lingüística textual para referirse a los tipos textuales son, entre otros: la noción de género (Ciapuscio, 2005), las clases textuales (Ciapuscio, 1994), los tipos textuales (Ciapuscio, 1994) y el concepto de registro (Biber, 1988). Mientras que la noción de género, con una larga tradición en los estudios literarios, se centra en el análisis de los tipos textuales como la configuración textual que adoptan situaciones comunicativas típicas con propósitos determinados o prácticas sociales establecidas socio-culturalmente; la idea de registro (Biber, 1988) se focaliza en los rasgos lingüísticos que caracterizan, en la forma de una constelación o haz de rasgos, a un tipo de texto particular. Teniendo en cuenta que estas dos nociones (género y registro) son de utilización común, y reconociendo su valor en esos términos, para el propósito de esta investigación utilizaremos, sin embargo, la distinción definida por Ciapuscio (1994) entre clases y tipos textuales. Una clase textual se entiende como la denominación natural que los

tipos textos reciben por parte de los usuarios. Un tipo textual, en cambio, es la denominación que recibe un texto a partir de criterios formales derivados de un análisis científico. En definitiva, como usuarios del lenguaje clasificamos los textos que usamos a partir de criterios naturales, es decir, a partir de nuestro conocimiento como hablantes nativos de un lengua. En cambio, la denominación científica de un texto (lo que se denomina tipología o taxonomía), proviene de nuestro conocimiento como expertos en el análisis textual. De más está decir que, debido a que ambas denominaciones se derivan de criterios de distinta naturaleza (naturales en un caso y formales científicos en el otro), estas pueden, en muchas ocasiones, no coincidir. En otras palabras, si le pedimos a un usuario que nos diga qué tipo de texto usa y luego tipologizamos ese texto, los resultados pueden ser divergentes. Estas nociones son especialmente funcionales a este trabajo, y se describirán de forma más detallada cuando en la siguiente sección definamos los criterios de inclusión y exclusión del corpus de análisis.

MATERIALES Y MÉTODOS

El corpus: La biblioteca Scielo-Chile

Debido a que la presente investigación forma parte de un proyecto mayor (FONDECYT 11080097), cuyo objetivo es analizar la variación de los Al a través de las disciplinas, la elección del universo a investigar estaba predeterminada por el objetivo de este proyecto mayor. Uno de los objetivos centrales de ese proyecto es la recolección de un corpus ecológico que permita reflejar la práctica científica chilena, y para esos efectos, la iniciativa de la biblioteca electrónica científica en línea (SCIELO, por sus siglas en inglés) resulta ser especialmente adecuada. El sistema Scielo es una iniciativa brasilera e iberoamericana para el desarrollo de una metodología común, para el almacenamiento y diseminación de la ciencia (Scielo, 2009a). En la mayoría de los países en los que el sistema está vigente (Argentina, Brasil, Colombia, Chile, Cuba, España, Venezuela, entre otros), han sido las agencias científicas gubernamentales las que se han encargado de la implementación y mantención del sistema. Consecuentemente, la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) ha implementado la iniciativa en nuestro país. La colección chilena contiene (al mes de octubre del año 2009) 81 títulos, divididos en ocho áreas temáticas (ciencias agrícolas, ciencias biológicas, ciencias de la salud, ciencias exactas y de la tierra, ciencias sociales aplicadas, humanidades, ingeniería y lingüística, letras y artes) que corresponden, en parte, a la clasificación de las ciencias de la UNESCO. Las 81 revistas contenidas en la base han sido evaluadas según criterios comunes (Scielo, 2009b), y esta base es reconocida oficialmente por los sistemas institucionales como un indicador válido para la postulación a fondos concursables, acreditación de universidades y la evaluación del curriculum de investigadores consolidados y en formación. El valor investigativo de esta biblioteca Scielo-Chile se basa en el hecho de que en él se incluyen un amplio espectro de revistas, ciencias y disciplinas, por lo que se constituye como un reflejo fiel de la producción científica nacional.

El problema de la delimitación del corpus: Los criterios de inclusión y exclusión

Debido al tamaño de la base de datos, que cuenta con más de 14.000 registros, debimos establecer previamente algunos criterios que nos permitieran construir a partir de esta base un corpus representativo de artículos de investigación. Para ello, determinamos los siguientes criterios:

A)criterios excluyentes:

a.1) se dejaron fuera todos los Al que estaban escritos en un idioma distinto al español
a.2) Se excluyeron del estudio todos los trabajos publicados antes del año 2000 y después del año 2008.
a.3) se excluyeron del estudio todos aquellos escritos que no correspondan con el tipo textual foco de esta investigación (reseñas, obras, proyectos, etc.).

Una vez aplicados estos criterios excluyentes, se procedió con los siguientes criterios:

B)criterios incluyentes:

b.1) se consideraron todas las disciplinas de la base Scielo Chile
b.2) Se incluyeron Al de cada una de las revistas de la base Scielo Chile

Cabe señalar que los criterios incluyentes se aplicaron después de los excluyentes, por lo que -en algunos casos- estos no pudieron aplicarse; por ejemplo, hubo revistas que no se pudieron incluir (b.2) en el estudio debido a que solo publicaban trabajos en inglés (cfr. a.1).

La dificultad en la aplicación de uno de los criterios excluyentes (a.3) fue lo que motivó el estudio que se presenta en este artículo. El problema radicaba en el hecho de determinar qué tipo de texto correspondía al objeto de estudio de la investigación, a saber, cómo responder la pregunta respecto de qué era y qué no era un Al. En una exploración previa de la Base Scielo Chile, nos percatamos de que no podíamos utilizar como criterio la denominación (clases textuales) que las revistas otorgaban a los textos que en ellas se publicaban. Fue por esta razón que diseñamos una forma de tipologizar, con criterios formales, las clases textuales que aparecían en las revistas.

Cabe señalar que una vez aplicado el criterio temporal (a.2), es decir, considerando solo los artículos publicados entre los años 2000 y 2008, el universo a determinar se redujo de 81 a 72 títulos.

Técnicas de exploración y resultados preliminares

Para enfrentar el desafío de tipologizar las clases textuales contenidas en las revistas de la Base Scielo, se entrenó a un conjunto de 8 colaboradores para que exploraran (identificaran y cuantificaran) las clases textuales contenidas en cada una de las 72 revistas restantes que quedaron, luego de la aplicación del criterio temporal (a.2). Estos datos fueron traspasados a hojas de cálculo, tal como se muestra en la Tabla 2:

Tabla 2: Hoja de cálculo utilizada en la fase exploración

Nombre de la revista	Año	Volumen	número	Nombre de la clase 1	Nombre de clase 2	Nombre de la clase n

En esta primera fase de exploración, se realizaron los siguientes hallazgos:

1. Un total de 29 revistas contenían solo clases textuales y estas revistas tendían a representar solo una disciplina (Revistas Tipo A)
2. Un total de 43 revistas contenían además de clases textuales, divisiones temáticas y más de una disciplina (Revistas Tipo B)

Así también, esta primera fase de exploración permitió constatar en primer término que muchas revistas usan nombres poco transparentes para denominar las clases textuales, por ejemplo, "Documento" o bien "Artículo". Por otro lado, lo que en una revista se denomina artículo no necesariamente corresponde al mismo tipo de texto que en otra revista. En definitiva, como ya adelantamos, si queríamos recolectar los Al contenidos en la Base Scielo no podíamos utilizar el nombre de las clases que las revistas contenían.

La construcción del libro de códigos

Para abordar la tipologización de la base de datos se procedió, en primer lugar, a utilizar una metodología abductiva, descrita ampliamente en el marco de la Teoría Empíricamente Fundada (Strauss & Corbin, 1990; Strauss & Corbin, 1997; Coffey & Atkinson, 2003), que consiste en la creación de un modelo a partir de un proceso de exploración y codificación en un análisis previo de los datos. Para ello, se creó el libro de códigos a partir de una exploración y codificación de las clases textuales contenidas en las 29 revistas Tipo A. La primera etapa, llevada a cabo por dos investigadores en conjunto, consistió en establecer o "levantar" criterios desde el análisis de las 232 clases textuales contenidas en estas 29 revistas. Los 33 criterios levantados se pueden agrupar según su naturaleza: la estructura del texto, su función o propósito comunicativo y el contenido de los textos analizados. Estos criterios se fueron afinando y homologando a medida que se completaba el análisis. A continuación, todas las clases textuales fueron codificadas con uno y solo uno de los 33 criterios. Estos criterios están asociados conceptualmente a un tipo de texto que es la denominación por parte de los investigadores de la aplicación del criterio. A su vez, estos tipos textuales se pueden agrupar macro-tipos textual que es una categoría más general en la que se obvian algunas de las diferencias entre los tipos textuales. A modo de ejemplo, a continuación, en laTabla 3, se presenta un extracto del libro de códigos, específicamente la parte de la Revista Acta Bioética (AB). Tal como se observa, el libro de código permite homogeneizar para comparar a partir de la determinación del criterio y del tipo, clases textuales diversas. Por ejemplo, si se observa las columnas 2 a la 6, podemos establecer que la revista denomina de 4 formas distintas (Colectánea, Documentos, editorial, introducción) a un mismo tipo de texto (Comentarios). Asimismo, en la revista se utilizan algunas de estas clases textuales (Editorial, Introducción) para otros tipos de textos (Presentaciones). En definitiva, la aplicación del libro de códigos a las 29 revistas permite contar con criterios formales para poder tipificar y comparar las clases textuales de las revistas contenidas en la Biblioteca electrónica Scielo.

Tabla 3: Extracto del libro de códigos

	CR	Clase	Tipo	Macro tipo	Criterio
1	AB	Recensiones	Reseña	Discurso destinado al comentario de obras	Comentario a otra obra
2	AB	Colectánea	Comentarios (de actualidad) generales sobre temas de la disciplina	Comentarios	Comentario general acerca de un tema de la disciplina
3	AB	Documentos	Comentarios (de actualidad) generales sobre temas de la disciplina	Comentarios	Comentario general acerca de un tema de la disciplina
4	AB	Editorial	Comentarios (de actualidad) generales sobre temas de la disciplina	Comentarios	Comentario general acerca de un tema de la disciplina
5	AB	Introducción	Comentarios (de actualidad) generales sobre temas de la disciplina	Comentarios	Comentario general acerca de un tema de la disciplina
6	AB	Special Document	Comentarios (de actualidad) generales sobre temas de la disciplina	Comentarios	Comentario general acerca de un tema de la disciplina
7	AB	Programas de formación bioética	Información sobre programas académicos	Discurso destinado a la difusión de eventos de la comunidad científica que la revista representa	Estado de un programa de maestría de la Universidad Nacional de Cuyo
8	AB	Editorial	Presentaciones	Discurso de la revista	Se presentan los trabajos de la revista
9	AB	Introducción	Presentaciones	Discurso de la revista	Se presentan los trabajos de la revista
10	AB	Preámbulo	Presentaciones	Discurso de la revista	Se presentan los trabajos de la revista

Tres tipos de artículos de investigación

Como resultado preliminar, a partir del libro de códigos, se determinó la existencia de tres tipos de AI que serán foco de esta investigación. Los nombres y los criterios definitorios de estos tres tipos de AI, se muestran a continuación en la Tabla 4:

Tabla 4: tres tipos de AI

Nombre	Sigla	Criterio definitorio
Artículo de investigación científica	AIC	Es un texto altamente estructurado, con categorías superestructurales identificables y obligatorias (Título, Resumen, Palabras Clave, Introducción, Marco Teórico, Metodología, Resultados, Discusión, Conclusión y Referencias Bibliográficas).
Artículo de investigación o estudio	AI o E	Es un texto menos estructurado que el AIC. En este tipo de trabajos, también se analizan datos, pero no aparece de forma explícita la sección metodológica. En este sentido, aunque también contiene partes superestructurales obligatorias (como el Título, el Resumen, las Palabras clave y la Introducción), el resto de ellas pueden variar en su denominación, o bien no aparecer.
Artículo de investigación teórica	AIT	Es un texto cuyo propósito central es hacer una revisión crítica, generalmente comparando distintas perspectivas, respecto de un concepto, una teoría o una metodología. Sus partes superestructurales son de corte temático. En términos generales, en este tipo de escritos se delinean las ventajas y desventajas de adoptar una perspectiva determinada para la comprensión del concepto, teoría o metodología, de modo que se proporcionan nuevas luces respecto del término tratado.

Entrenamiento de codificadores

Luego de la creación del libro de códigos, se procedió a entrenar a 8 asistentes para su aplicación a todas las revistas de la Biblioteca Scielo Chile. Para ello, se implemento el libro de códigos en una hoja de cálculo, utilizando un formulario programado en lenguaje Visual Basic. Los codificadores registraron la información en estos formularios, consultando cada uno de los registros contenidos en el sitio web de cada revista y eligiendo para cada registro uno de los 33 criterios establecidos en el libro de códigos. Se realizaron dos sesiones de dos horas cada una para explicar el funcionamiento de este formulario y las instrucciones para ingresar la información.

índice Fleiss Kappa para el cálculo de confiabilidad inter-codificador

Para probar la consistencia teórica y metodológica del libro de códigos, se calculó el índice de confiabilidad inter-codificador. Este índice permite no solo saber cuánto acuerdo existe en la tarea de codificación, es decir, validar empíricamente su construcción, sino también permite evaluar la calidad del modelo establecido. Así, para que los datos sean empíricamente significativos, se debe demostrar un alto grado de fiabilidad inter-codificador. La fiabilidad (reliability) o confiabilidad se mide por la proporción de acuerdo de todas las categorías de modo idéntico, codificadas por codificadores diferentes. A pesar de que el índice de confiabilidad es ampliamente usado en distintas disciplinas, como en la psiquiatría, psicología o en el diagnóstico médico, su utilización en lingüística es relativamente reciente y, desde un punto de vista general, escasa.

Para medir la fiabilidad, uno de los métodos más simples y frecuentemente utilizado es el acuerdo porcentual, el que reporta el número de acuerdos con relación al total de unidades de análisis categorizadas; sin embargo, este método no da cuenta de los posibles acuerdos debido al azar. Para solucionar esto, Scott (1955) propuso un coeficiente que denominó Pi, a partir de cual que se calcula la fiabilidad inter-codificador para datos nominales, corrigiendo el nivel de acuerdos observados, al incluir el nivel de acuerdo debido al azar, basándose, además, en el supuesto de homogeneidad en los marginales en la matriz de datos que representan los acuerdos. Por otra parte, Cohén (1960) propuso la estadística kappa, un coeficiente similar a la Pi de Scott, pero sin la suposición de homogeneidad, con el propósito de extenderlo para considerar ponderaciones sobre los criterios de clasificación. El empleo de kappa y kappa ponderado está restringido a los casos donde el número de codificadores es dos. Posteriormente Fleiss (1971), generalizó el concepto de Scott (1955) ampliandolo a más de dos codificadores para datos dicotomicos o escalas nominales. Krippendorff (1990) propone un coeficiente para todo propósito, que denomina alfa, para el cálculo de la confiabilidad inter-codificadores para el análisis de contenido.

Para los propósitos del presente estudio, se implemento una macro en una hoja de cálculo, para estimar el coeficiente Kappa de Fleiss. Se usó este indicador porque además de calcular la confiabilidad inter-codificadores global, permite estimar la confiabilidad entre los ítems. Conceptualmente, el cálculo para Kappa de Fleiss se puede expresar por medio de las siguientes ecuaciones:

Sea N el total de sujetos, sea n el número de rating por sujetos, y sea k el número de categorías. Lo sujetos son indizados por i=1,...,N y las categorías por j=1, ...,n. Los n_{ij} representa el número de codificadores que asignaron el i-ésimo artículo en la j-ésima categoría. El algoritmo para calcular Kappa consta de los siguientes pasos:

En primer lugar, se calcula p_j que es la proporción de todas las asignaciones las que se asignó en la j-ésima categoría:

$$p_j = \frac{1}{Nn}\sum_{i=1}^{N} n_{ij} \qquad 1 = \frac{1}{n}\sum_{j=1}^{k} n_{ij} \tag{1}$$

En segundo lugar, se calcula p_i, que representa la proporción de los acuerdos para el i-ésimo artículo:

$$P_i = \frac{1}{n(n-1)}\sum_{j=1}^{k} n_{ij}(n_{ij}-1) = \frac{1}{n(n-1)}\sum_{j=1}^{k}(n^2_{ij} - n_{ij}) = \frac{1}{n(n-1)}[(\sum_{j=1}^{k} n_{ij}^2) - n] \tag{2}$$

En tercer lugar, se calcula $\overline{P_e}$, que es el promedio de los Pi, o valor esperado

$$\overline{P} = \frac{1}{N}\sum_{i=1}^{N} P_i = \frac{1}{Nn(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{k}(n^2_{ij} - Nn) \tag{3}$$

$$\overline{P_e} = \sum_{j=1}^{k} P_j^2$$

En síntesis, se puede decir que $\overline{P}$ representa la proporción de frecuencias observadas y $\overline{P_e}$ las proporciones esperadas; entonces, el coeficiente Kappa de Fleiss se calcula como:

$$k = \frac{\overline{P} - \overline{P_e}}{1 - \overline{P_e}} \tag{4}$$

A pesar de que no existe un acuerdo universal para la interpretación de los valores de Kappa, a partir de los estudios de Landis y Koch (1977) se establece que sobre un 0,6 existe un acuerdo sustancial, y que sobre 0,8 se obtiene un acuerdo casi perfecto.

índice de variación relativa

Tal como lo hemos sostenido, los usuarios de las revistas (editores, autores) denominan a los artículos según criterios propios, es decir, en clases textuales, las que a menudo no corresponden a criterios formales. Producto del análisis realizado en la confección del libro de códigos, se constató que existe una discrepancia entre las clases y los tipos textuales. La situación descrita para la relación entre clases y tipos textuales, se puede modelar por medio de una distribución binomial. Para ello, se definirá la variable aleatoria X como el número de tipos textuales, entonces la expresión probabilística es:

$$P(X_i = k_i) = \binom{n_i}{k_i} p_i^{k_i} q_i^{n_i - k_i} \tag{5}$$

Donde n_i representa el número de clases presentes en la revista i-ésima, k_t representa el número de tipos textuales, p_i es la probabilidad de ocurrencia del tipo textual en la revista i-ésima, finalmente, q_i es la probabilidad de no ocurrencia del tipo textual en la revista i-ésima, es decir, la probabilidad de ocurrencia de la clase textual.

Desde esta perspectiva, teóricamente sabemos que la esperanza matemática o promedio de ocurrencia de los tipos textuales y la varianza están dada por:

$$E[X] = n_i p_i \quad (6)$$

Y la varianza,

$$\sigma^2 = V[X] = n_i p_i q_i \quad (7)$$

Para hacer la estimación de este coeficiente, se utilizó una muestra del corpus total que corresponde a las mismas 29 revistas con las cuales se construyó el libro de códigos.

Desde una perspectiva metodológica, proponemos una triangulación para corroborar este cálculo, utilizando un índice de variación porcentual, que tiene la siguiente fórmula $IV = \frac{C_i - T_i}{C_i}$,donde C_i representa el número de clases textuales presentes en la revista i-ésima, T_i representa el número de Tipos textuales presentes en la revista i-ésima. Es interesante destacar que los resultados obtenidos utilizando esta fórmula más simple son coincidentes con aquellos que se obtienen, utilizando la distribución binomial.

RESULTADOS Y DISCUSIÓN

Confiabilidad inter-codificador en la aplicación del libro de códigos

A continuación, en la Tabla 5 se presentan los resultados del índice Fleiss-Kappa de la confiabilidad inter-codificador en la aplicación del libro de códigos:

Tabla 5: índice Fleiss-Kappa

Revista	Número Artículos	Número Categorías	Número Codificadores	$\overline{P}$	$\overline{P_e}$	Kappa
Tipo A	391	11	8	0,94656558	0,27212648	0,92658832
Tipo B	439	11	8	0,92523593	0,2568295	0,89939849

El cálculo del índice Fleiss-Kappa se realizó por separado para los dos grandes tipos de revistas determinadas en la fase de exploración preliminar (ver apartado 'Técnicas de exploración y resultados preliminares'), a saber, las revistas Tipo A y tipo B. Tal como se muestra en la Tabla 5, en ambos casos el índice calculado supera el 0,7, establecido en esta investigación como un puntaje de corte para validar la confiabilidad del instrumento utilizado. Este resultado indica que el instrumento construido y los criterios definidos son consistentes teóricamente y empíricamente válidos para la clasificación de las clases textuales y que, por lo tanto, su aplicación no depende del azar. Se puede inferir, a partir de los valores de kappa calculados para ambos tipos de revistas, que existe (según la escala de Landis y Koch,1977), un acuerdo casi perfecto.

índice de correspondencia relativa

A continuación, en la Tabla 6, se presentan los resultados del índice de variación entre las clases y los tipos textuales.

Tabla 6: índice de variación de las clases textuales en 29 revistas de la Biblioteca Scielo Chile

Revista	N° clases	N° tipos	p_i	q_i	$n_i p_i$	$n_i p_i q_i$	Desv Sta.
Revista Chilena de Historia Natural	28	11	0,39	0,61	11,00	6,68	2,58
Revista Musical Chilena	22	10	0,45	0,55	10,00	5,45	2,34
Acta bioética	17	9	0,53	0,47	9,00	4,24	2,06
Bosque	15	7	0,47	0,53	7,00	3,73	1,93
Revista Chilena de Literatura	10	3	0,30	0,70	3,00	2,10	1,45
Parasitología al día	9	6	0,67	0,33	6,00	2,00	1,41
Cuadernos de Economía	10	8	0,80	0,20	8,00	1,60	1,26
Maderas	8	6	0,75	0,25	6,00	1,50	1,22
Archivos de Medicina Veterinaria	7	5	0,71	0,29	5,00	1,43	1,20
Boletín de química	7	5	0,71	0,29	5,00	1,43	1,20
Revista Chilena de Derecho (Valdivia)	7	5	0,71	0,29	5,00	1,43	1,20
Idesia	6	4	0,67	0,33	4,00	1,33	1,15
Revista Agricultura técnica	6	4	0,67	0,33	4,00	1,33	1,15
Revista Psyke	7	6	0,86	0,14	6,00	0,86	0,93
Acta Literaria	6	5	0,83	0,17	5,00	0,83	0,91
Estudios pedagógicos	6	5	0,83	0,17	5,00	0,83	0,91
Magallania	6	5	0,83	0,17	5,00	0,83	0,91
Revista de ciencias del suelo y nutrición vegetal	6	5	0,83	0,17	5,00	0,83	0,91
Revista de filosofía	5	4	0,80	0,20	4,00	0,80	0,89
Ciencia e investigación Agraria	4	3	0,75	0,25	3,00	0,75	0,87
Revista Historia	4	3	0,75	0,25	3,00	0,75	0,87
Terapia Psicológica	3	2	0,67	0,33	2,00	0,67	0,82
Ciencia y enfermería	4	4	1,00	0,00	4,00	0,00	0,00
Ingeniare	5	5	1,00	0,00	5,00	0,00	0,00
International Journal of Morphology	4	4	1,00	0,00	4,00	0,00	0,00
Revista de geografía norte grande	4	4	1,00	0,00	4,00	0,00	0,00
Revista de Ingeniería en construcción	3	3	1,00	0,00	3,00	0,00	0,00
Revista facultad de ingeniería Universidad de Tarapacá	6	6	1,00	0,00	6,00	0,00	0,00
RLA	7	7	1,00	0,00	7,00	0,00	0,00
	232	**154**	**0,66**	**0,34**	**154,00**	**51,78**	**7,20**

Para interpretar los datos de la Tabla 6, se debe considerar que a mayor valor del índice (desviación estándar), mayor es la variación de la correspondencia entre los tipos y las clases textuales. Así, se puede sostener que aquellas revistas en las cuales los valores son igual a 0, existe una alta correspondencia entre las clases y tipos. Esto supone que los usuarios (autores y editores) de estas revistas en las que el valor es igual a 0, utilizan, aunque no conscientemente, criterios formales en la denominación de los textos que aparecen en las revistas. Si establecemos a partir de los resultados un rango de variación, se pueden establecer los siguientes criterios para determinar el grado de variación, los cuales se muestran en la siguiente Tabla 7:

Tabla 7: Rangos para determinar el grado de variación entre clases y tipos textuales

Rango	Grado de variación
0,00 – 0,86	Baja
0,87 – 1,72	Media
1,73 – 2,58	Alta

A partir de los datos de las Tabla 6 y 7, podemos observar que un total de 4 revistas, se encuentran en el rango de un alto grado de variación. En estas revistas, los nombres de las clases textuales no siguen ningún criterio formal y se utilizan indistintamente los mismos nombres para tipos de textos distintos (ver Tabla 3). Asimismo, existen 17 revistas con una grado medio de variación. Se puede sostener al respecto, que en estas revistas se utilizan algunos criterios formales en la denominación de los textos contenidos en ellas. Por último, 8 revistas con un bajo grado de variación, lo que indica una mayor correspondencia entre las clases y los tipos.

La presencia de los Artículos de Investigación en las revistas Scielo Chile

En la Figura 1, se presentan los resultados de la aplicación del libro de códigos a la totalidad de los registros de la biblioteca Scielo Chile, comprendidos entre los años 2000 y 2008:

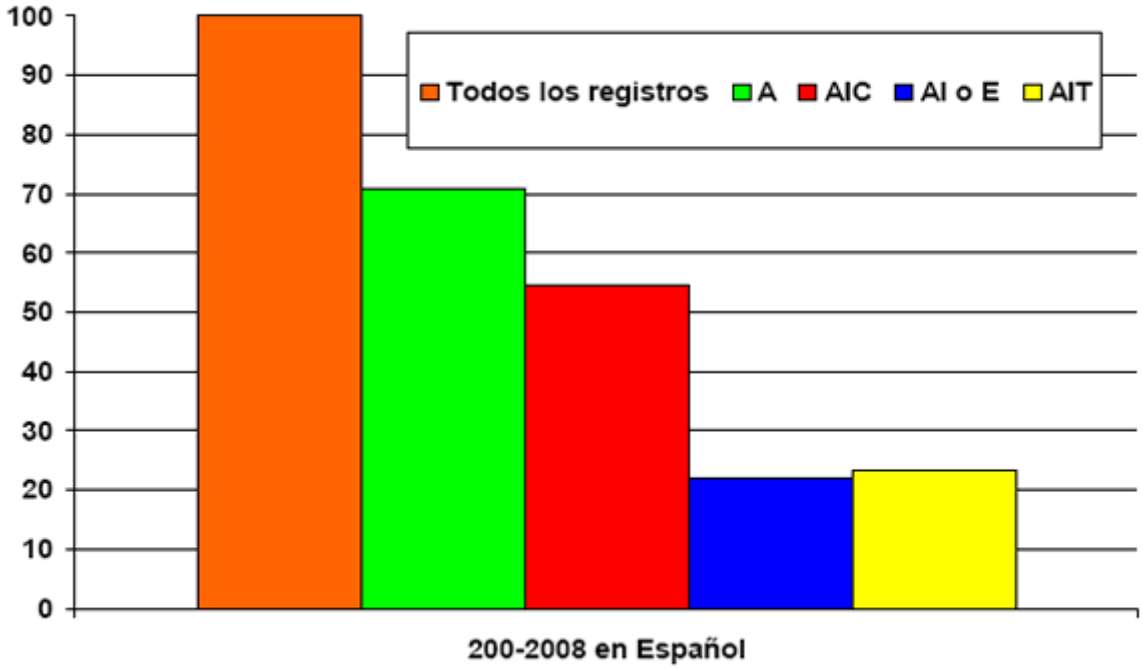

Fig. 1: presencia de los AI en la biblioteca Scielo Chile

Tal como se muestra en Figura 1, en el periodo estudiado, la biblioteca Scielo Chile contiene 14.398 registros. De estos registros, un 70 % (10.183) corresponde a Artículos (A), que se presentan desagregados según su tipo en los colores rojo, azul y amarillo. Este resultado general indica que el resto de los tipos textuales (editoriales, reseñas, homenajes, etc.) representan un 30% de los registros de toda la base. Respecto de los tipos de Artículos, se muestra que un 55 % de ellos (5.574) corresponden al Artículo de Investigación Científica (AIC). Por otra parte, aproximadamente un 22% (2.259) de los registros, corresponden a Artículos de Investigación o Estudios (AI o E). Asimismo, con un porcentaje levemente superior

(23%) aparecen los Artículos de Investigación Teórica (AIT), que constituyen 2.350 registros de toda la base analizada. A partir de estos resultados generales, se puede establecer que el Artículo de Investigación Científica (AIC) es, en definitiva, el tipo de texto prototípico por medio del cual se transmite conocimiento en la biblioteca Scielo Chile.

Para describir en detalle, a saber, considerando la variación a través de las áreas de la ciencia, desagregamos los datos de la Figura 1, en la Figura 2:

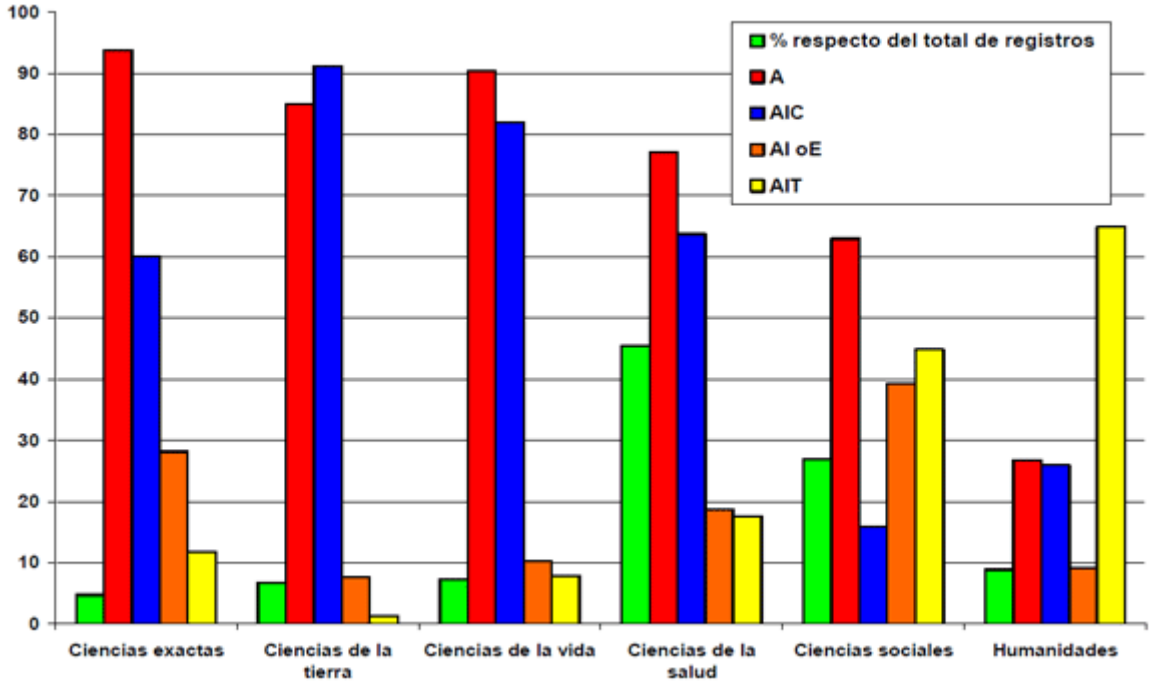

Fig. 2: Tipos de Artículos a través de las áreas de la ciencia

En la Figura 2, se muestran diversos aspectos relevantes que dan cuenta de la producción científica chilena, a partir del análisis de la Biblioteca Scielo Chile. En primer término, se aprecia en color verde la presencia de las áreas de la ciencia en la base estudiada. Tal como se muestra, las ciencias de la salud constituyen el 46% de todo lo que se escribe en Scielo Chile. A continuación, las ciencias sociales aparecen con un 27%. El resto de las áreas de las ciencias no superan el 10% de todos los registros: humanidades con un 9%, ciencias de la tierra y ciencias de la vida con un 7% y las ciencias exactas con un 5% del total de los registros. Ahora bien, la presencia de Artículos (A) (independientemente de los subtipos), es claramente variable entre las distintas áreas de la ciencia. Este porcentaje (en color rojo) representa la importancia que le dan las revistas de cada una de estas áreas a la transmisión de conocimiento en forma de investigaciones, y es inversamente proporcional a la importancia que se le da a otros tipos de textos que típicamente aparecen en las revistas especializadas (i.e., editoriales, reseñas, homenajes, etc.), que no tienen como función primordial transmitir conocimiento, sino informar sobre eventos de la comunidad, presentar los trabajos de la revista, homenajear a personajes relevantes de la disciplina, entre otras funciones. Así, vemos, por ejemplo, que tanto en las ciencias exactas como en las ciencias de la vida la presencia de artículos supera el 90%. La presencia de artículos en las ciencias de la tierra y las ciencias de la salud representan un 85% y 77% respectivamente, superando el promedio general (ver Figura 1). En las ciencias sociales y en las humanidades, en cambio, el valor no supera el promedio general, con un 63% en el

caso de las ciencias sociales y solo un 27 % en el caso de las humanidades. Este último resultado implica que en las revistas de humanidades se da mucho más espacio a difundir información sobre la comunidad científica que la revista representa, a presentar reseñas, homenajes o a hacer comentarios generales sobre la disciplina. Estos textos en esas áreas son más importantes, al menos en frecuencia, que los artículos de investigación focos del presente estudio. Así también, se aprecia una distribución dispar en los tipos específicos de artículos, que se presentan desagregados: en azul, los Artículos de Investigación Científica (AIC), en naranja los Artículos de Investigación o Estudios(AI o E) y, en amarillo, los Artículos de Investigación Teórica (AIT). Al respecto, se constata un patrón común entre las ciencias exactas, ciencias de la tierra, ciencias de la vida y ciencias de la salud, a saber, en estas áreas, los AIC superan a los Al o E y, estos a su vez, superan a los AIT. La importancia de los AIC en estas áreas se debe principalmente a que el enfoque metodológico predominante de los objetos de estudio de las disciplinas de cada una de estas áreas es de carácter empírico experimental. En cambio, tanto en las ciencias sociales como en las humanidades, es el AIT el más frecuente, pero ambas áreas de la ciencia también difieren en la ocurrencia de los otros tipos de artículos. Mientras en las humanidades el segundo tipo de artículos más frecuente es el AIC, en las ciencias sociales este es el menos frecuente, siendo el segundo más importante los Al o E.

CONCLUSIONES

A partir de la creación y aplicación de un libro de códigos, en esta investigación se logró tipologizar la gran variedad de clases textuales contenidas en la Biblioteca Scielo Chile. Este libro de códigos fue validado empíricamente con un alto nivel de confiabilidad, lo que permite sostener su consistencia teórica y utilidad descriptiva. Este libro de códigos resulta ser una herramienta metodológica útil para uniformar la variabilidad de la denominación de las clases textuales de las revistas que componen la Biblioteca Scielo-Chile.

Así también, la investigación realizada permitió establecer un índice de variación relativa que da cuenta de la relación entre clases y tipos textuales. Al respecto, se concluye que mientras algunas revistas utilizan nombres específicos para los textos que presentan, otras revistas no siguen un criterio formal para la denominación de estos textos, utilizando una misma etiqueta para textos cuya función, estructura y contenido son variados.

De la aplicación del libro de códigos al total de los registros de la biblioteca Scielo Chile, se constató que los artículos constituyen un 70% del total de los registros de la biblioteca y que la distribución de los tipos específicos de artículos es sensible a las áreas de la ciencia. Así, mientras el Artículo de Investigación Científica es característico de las ciencias de la salud, de las ciencias exactas, de las ciencias de la vida y de la tierra, el Artículo de Investigación Teórica y el Artículo de Investigación o Estudio son característicos de las ciencias sociales y las humanidades.

Tanto los resultados como los procedimientos derivados de esta investigación, pueden ser útiles para contrastar o describir otras bibliotecas electrónicas similares a la analizada en esta investigación.

AGRADECIMIENTOS

Los autores agradecen el apoyo de CONICYT a través del proyecto FONDECYT 11080097: "El artículo de investigación a través de las disciplinas: el caso del indexador Scielo Chile".

REFERENCIAS

Bazerman, C; *Shaping written knowledge: The genre and activity of the experimental article in science,* 1ª edición. The University of Wisconsin Press, Londres, Inglaterra (1988). [Links]

Biber, D.; *Variation across speech and writing,* 1ª edición. Cambridge University Press, Cambridge, Inglaterra (1988). [Links]

Ciapuscio, G.; *Tipos textuales,* 1ª edición. Eudeba, Buenos Aires, Argentina (1994). [Links]

Ciapuscio, G.; *La noción de género en la Lingüística Sistémico Funcional y en la Lingüística Textual.* Revista Signos: 38 (57), 31-48 (2005). [Links]

Cohén, J.; *A coefficient of agreement for nominal scales.* Educational and Psychological Measurement: 20 (1). 37-46 (1960). [Links]

Coffey, A. y P. Atkinson; *Encontrar el sentido a los datos cualitativos,* 1ª edición. Editorial Universidad de Antioquia, Antioquia, Colombia (2003). [Links]

Falahati, R.; *The Use of Hedging across Different Disciplines and Rhetorical Sections of Research Articles* (1999),http://www.sfu.ca/gradlings/NWLC_Proceedings/falahati99-112.pdf. Acceso: 15 de noviembre (2009). [Links]

Fleiss, J.; *Measuring Nominal Scale Agreement among many Rater.* Psychological Bulletin: 76 (5), 378-382(1971). [Links]

Grookes, G.; *Towards a Validated Analysis of Scientific Text Structure.* Applied Linguistics: 7, 57-70 (1986). [Links]

Hayes, A. y K. Krippendorff; *Answeríng the Cali for a Standar Reliability Measure for Coding Data.* Communication Method and Measures: 1 (1), 77-89 (2007). [Links]

Holmes, R.; *Genre Analysis, and the Social Sciences: An Investigation of the Structure of Research Article Discussion Sections in Three Disciplines.* English for Specific Purposes: 16 (4) 321-337 (1997). [Links]

Hyland, K.; *Hedging in Scientific Research Articles,* 1ª edición. Benjamins, Amsterdam, Holanda (1998). [Links]

Hyland, K.; *Disciplinary discourses: Writer stance in research articles,* in Writing: Texts, processes and practices: Longman. pp. 99-121. London (1999). [Links]

Hyland, K.; *As can be seen: Lexical Bundles and Disciplinary Varíation.* English for Specific Purposes: 27, 4-21 (2008). [Links]

Kanoksilapatham, B.; *Rhetorical structure of biochemistry research articles.* English for Specific Purposes: 24, 269-292 (2005). [Links]

Krippendorff, K. *Metodología de análisis de contenido. Teoría y práctica,* 1ª edición, Paidós, Barcelona, España (1990). [Links]

Ngozi, K.; *The medical Research Papen Structures and Function.* English for Specific Purposes: 16 (2), 119-138(1997). [Links]

Pho, P. D.; *How can Learning about the Structure of Research Articles Help International Students?* (2008),http://www.isana.org.au/files/2008%20Conference%20Proceedings/paper_Dzung.pdf Acceso: 15 de noviembre (2009). [Links]

Posteguillo, S.; *The Schematic Structure of Computer Science Research Article.* English for Specific Purposes: 18 (2), 139-160 (1998). [Links]

Samraj, B.; *Introductions in research articles: varíations across disciplines.* English for Specific Purposes: 21, 1-17(2001). [Links]

Scielo; *Scielo Chile* (2007a), http://www.scielo.cl/scielo.php. Acceso: 15 de noviembre (2009). [Links]

Scielo; *Requisitos de postulación a Scielo-Chile* (2007b),http://www.scielo.cl/criterios/sp/. Acceso: 15 de noviembre (2009). [Links]

Scott, W. ; *Reliability of contentí analysis: the case of nominal scale coding.* Public Opinión Quarterly: 19. 127-141 (1955). [Links]

Soto, G.; *Los géneros discursivos como tecnologías cognitivas.* Revista de la Sociedad Argentina de Lingüística: 1. 37-51 (2005). [Links]

Soto, G.; *La argumentación en artículos científicos escritos en español.* Actas del Congreso Internacional La Argumentación: Lingüística, Retórica, Lógica 1262-1269, Buenos Aires, Argentina 10 al 12 de Julio (2002). [Links]

Soto. G., R. Martínez y S. Sadowsky; *Verbos y sustantivos en textos científicos. Análisis de variación en un corpus de textos de ciencias aplicadas, naturales, sociales y humanidades.* Philologia Hispalensis, 169-187 (2005). [Links]

Soto, G. y C. Zenteno; *Los sintagmas nominales en textos científicos en español,* XV Congreso de la Sociedad Chilena de Lingüística. Santiago, Chile, 13 al 17 de octubre (2003). [Links]

Strauss, A. y J. Corbin; *Basics of qualitative research: grounded theory procedures and techniques,* 1ª edición. Sage, London, Newbury Park (1990). [Links]

Strauss, A. y J. Corbin (Eds.); *Grounded theory in practice,* 1ª edición. Sage, Londres, Inglaterra (1997). [Links]

Swales, J. *Genre analysis. English in academic and research settings,* 1ª edición, Cambridge University Press, Cambridge, Inglaterra (1990). [Links]

Swales, J. ; *Research genres. Explorations and application,* 1ª Edición. Cambridge University Press, Cambridge , Inglaterra, (2004). [Links]

Valderrama, J.; *La publicación en revistas especializadas, un recurso estratégico en la investigación científica y tecnológica,* en A. Cellino (Ed.) Los laberintos del futuro. Ciencia y tecnología en América Latina: Editorial de la Universidad Nacional del Litoral. Capítulo 4. Santa Fe - Argentina (2001). [Links]

Recibido Dic. 17, 2009; Aceptado Feb. 11, 2010; Versión Final recibida Mar. 08, 2010.