

EL ROL DE LA TERMINOLOGÍA DEL ÁREA DE CONSERVACIÓN Y RESTAURACIÓN DE BIENES EN PAPEL EN LA DIVULGACIÓN DEL CONOCIMIENTO ESPECIALIZADO: ETAPAS METODOLÓGICAS

Cleci Regina Bevilacqua

Profesora del Departamento de Lenguas Modernas y del Programa de Posgrado del Instituto de Letras; Coordinadora del grupo de investigación Proyecto Terminológico Cone Sul (TERMISUL), Universidad Federal de Rio Grande do Sul (UFRGS). Posdoctoranda en el marco del GTERM, Instituto de Información, FIC, UdelAR.

1 INTRODUCCIÓN

Este texto tiene el objetivo de describir las etapas metodológicas del proyecto de posdoctorado, titulado *La divulgación del conocimiento especializado: análisis del rol de la terminología en el área de Conservación y Restauración de Bienes en Papel*, realizado en la Facultad de Información, en el ámbito del grupo de investigación Terminología y Organización del Conocimiento (GTERM), bajo la dirección del profesor Dr. Mario Barité. El proyecto trata del papel de la terminología en la divulgación del conocimiento científico y establece una interfaz entre la Ciencia de la Información, la Comunicación y la Terminología, aplicada al área de Conservación y Restauración de Bienes Culturales en Papel.

Su objetivo principal es analizar los términos y su uso en textos de distintos niveles de especialización -textos académicos y textos de divulgación- y su función en la representación y transmisión del conocimiento teniendo como foco el área referida anteriormente. Los textos analizados incluyen el portugués de Brasil y el español de Uruguay. Entre sus objetivos específicos se encuentran los siguientes:

- a) describir, en sus distintos niveles lingüísticos, los textos que conforman los corpus relativos a la temática en portugués de Brasil y en español del Uruguay, a fin de caracterizarlos y obtener los elementos necesarios para identificar y analizar los términos;
- b) identificar y analizar los términos de los corpus de estudio para poder verificar si hay variación terminológica y el papel que ejerce en la representación y transmisión del conocimiento especializado;
- c) identificar semejanzas y diferencias entre los textos y la terminología utilizada en cada una de las

lenguas.

Para llevar a cabo el proyecto, se parte de los principios de la Teoría Comunicativa da Terminología (TCT, CABRÉ, 1999), principalmente cuando se afirma que los términos se caracterizan como núcleos conceptuales que representan y transmiten conocimiento de un área específica y que su valor de unidad especializada está determinado por aspectos pragmáticos, es decir, su uso en una situación comunicativa relativa a una temática específica. Al formar parte de la lengua natural, los términos admiten variación (denominativa y conceptual) para adecuarse a las distintas situaciones comunicativas en que se utilizan.

Igualmente, a partir de la propuesta de Calsamiglia (1997, p. 1), se entiende la divulgación de la ciencias como el proceso por el que se hace llegar al público no especializado y amplio el saber producido por especialistas de las disciplinas científicas. En este sentido, se entiende que la ciencia forma parte del patrimonio cultural de las sociedades modernas (Jiménez Aleixandre, 2003), e informar sobre la ciencia significa posibilitar el acceso al público lego sobre los avances científicos.

La investigación se fundamenta aún en el análisis textual de Ciapuscio (2003), que propone la descripción de los textos en cuatro niveles (funcional, situacional, de contenido semántico y formal-gramatical) y en la Lingüística de Corpus (Berber Sardinha, 2004; Biber, 1998).

La justificación que sostiene el proyecto se debe al desarrollo acelerado de la investigación en todos los ámbitos del saber y la consecuente producción de conocimiento especializado y la necesidad de su divulgación a un público amplio. Eso ha hecho que muchos investigadores de las áreas de Lingüística Textual, Lingüística de Corpus, Terminología, Ciencias de la Información y de la Comunicación se vuelquen al análisis de textos especializados y de divulgación científica para buscar entender su constitución y su papel en la representación, transmisión y divulgación de este tipo de conocimiento. En este contexto, es posible identificar dos tipos de situaciones comunicativas: a) una especializada, que incluye los especialistas que producen textos para sus pares o para la formación de futuros especialistas (aprendices), y b) una situación de menor grado de especialización, que involucra especialistas y, más frecuentemente, mediadores -periodistas y traductores- y el público lego. Para dar cuenta de los dos tipos de situaciones comunicativas es que se analizaron textos académicos y de divulgación.

A esa justificación se asocia otra relativa a la investigación actual llevada a cabo por el Grupo Terminológico Cone Sul (TERMISUL, UFRGS), que busca describir prácticas textuales del área de Conservación y Restauración de Bienes Culturales en papel y reconocer su constitución y su terminología. En Brasil, esta área aún está en conformación y requiere tanto la descripción de textos como la recogida y sistematización de su terminología. Es en el sentido de poder contribuir con esta investigación más amplia, que la investigación posdoctoral se propone a analizar textos científicos y de divulgación relativos al área. Puesto que la investigadora actúa en la formación de profesores y traductores en lengua española, se analizaron textos en portugués (variante brasilera) y en español (variante uruguaya). Este hecho, sumado al intercambio ya existente entre los investigadores de ambos proyectos, también justifica la elección de la FIC para desarrollar la investigación.

A continuación, se presentan las etapas metodológicas que permitieron describir los textos del corpus

de estudio y seleccionar los términos que serán analizados.

2 ETAPAS METODOLÓGICAS

La metodología se desarrollo en varias etapas: construcción de los corpus de estudio -recogida y preparación de los textos-; análisis de los textos; e identificación y selección de los términos. A continuación, presentamos brevemente cada una de estas etapas.

2.1 CONSTRUCCIÓN DE LOS CORPUS DE ESTUDIO Y DE REFERENCIA

Para construir los corpus de estudio y de referencia se siguieron los principios de la Lingüística de Corpus (Berber Sardinha, 2004; Biber, 1998). Los principios seguidos fueron: autenticidad (los textos deben ser naturales, es decir, no son producidos con el fin específico del estudio); naturalidad (deben ser producidos por hablantes nativos), representatividad (deben ser prototípicos de determinada categoría o género); formato (deben estar en formato electrónico para ser procesados por herramientas informáticas); propósito (deben atender los propósitos de la investigación y, por ello, deben establecerse criterios de selección de los textos).

Se construyeron dos corpus: uno de estudio y otro de referencia. El corpus de estudio es el corpus a partir del que se identificaron los términos y está constituido de: a) textos especializados en español y portugués; b) textos de divulgación en español y portugués El corpus de referencia, es un corpus de lengua general y se contrasta con el corpus de estudio para identificar los términos más específico de este último. Está conformado por textos periodísticos de Brasil (*Folha de São Paulo*) y del Uruguay (*El País*).

Para la búsqueda de los textos especializados y de divulgación se utilizaron las palabras clave, indicadas por especialista del área¹ o identificadas en las bases de datos de universidades (por ejemplo, BIUR). Algunas de ellas son: *bienes culturales en (formato/soporte) papel / bens culturais em (formato/suporte) papel conservación / conservação, documento / documento, documentación / documentação, libro / livros; museos / museus, papel / papel, patrimonio / patrimônio, preservación / preservação, restauración / restauração / restauro*. Los filtros utilizados en Google para restringir la búsqueda a los dos países fueron: lengua (español y portugués) y dominio (Uruguay y Brasil).

Es corpus de estudio construido a partir de los parámetros anteriores tienen las características siguientes:

- Fecha: corpus sincrónico y cerrado (1990-2007);
- Lengua: corpus del español de Uruguay (uy) y del portugués de Brasil (pt);
- Género: académico (textos científicos y técnicos) y de divulgación (sitios web de museos, fundaciones culturales, blogs, etc.);
- Modo: textos escritos;

¹ La especialista es Silvana Bojanoski, profesora de la carrera de Conservación y Restauración de la Universidad de Pelotas e investigadora colaboradora de TERMISUL. Su tesis de doctorado tiene como objetivo construir un glosario del área para estudiantes de dicha carrera.

- Contenido: temática de la conservación y restauración de bienes en papel;
- Uso en la investigación: corpus de estudio;
- Autoría: producidos por hablantes nativos;
- Tamaño: pequeño.

Es importante destacar aún que el corpus del portugués también proviene del corpus construido por el grupo TERMISUL (un corpus con 1 millón de palabras en portugués) del cual seleccionamos textos semejantes para obtenerse un corpus comparable al del Uruguay. Así, fueron seleccionados textos de los géneros académico y técnico y de temáticas semejantes en las dos lenguas.

El corpus de referencia, según los principios de la LC, debe ser, como mínimo, cinco veces más grande que el corpus de estudio. Los corpus de referencia compilados tienen cerca de 250 mil y 270 mil palabras para el español y portugués, respectivamente.

Para el español, los textos fueron recogidos del diario *El País*. La elección de este diario se debió a que permite el acceso con la inclusión de una cuenta de e-mail o *Facebook*. Además, es un diario representativo en el país lo que permite compararlo al diario elegido en portugués, *Folha de São Paulo*. La recogida de los textos fue de 8 días entre las fechas de 25 de septiembre al 02 de octubre de 2017 y las secciones utilizadas fueron: deportes, economía/negocios, información, mundo y vida actual. Se buscó recoger un número semejante para cada temática, pero eso dependía de los temas tratados y de los días de la semana.

Los textos de divulgación en portugués, como ya se dijo, fueron recogidos de *Folha de São Paulo*, a la que se tiene acceso por suscripción. Las fechas de recogida de los textos fueron las mismas que para el español y los temas también fueron semejantes: cotidiano, economía/negocios, deportes, mundo y poder. Igualmente se buscó recoger un número semejante de noticias siempre que posible. Es importante aclarar que los textos de ambos diarios se utilizaron únicamente con fines de investigación y no estarán disponibles.

De forma sintética, el tamaño de los corpus de estudio y de referencia es el siguiente:

Tabla 1 - Tamaño (tokens) de los corpus de estudio y referencia Uy y Pt

Corpus	Uruguay	Brasil
Especializado	20.876	22.466
Divulgación	20.578	21.788
De referencia	272.751	257.474

Fuente: Autora.

Para obtenerse resultados más precisos con el uso de herramientas de extracción de información lingüística, es necesario que los textos estén convertidos al formato <.txt> y que se retiren informaciones que interfieren en los resultados, es decir, que pueden generar ruido (sumario, resúmenes, figuras, tablas, referencias)².

² Para los corpus académicos de ambas lenguas, este trabajo se hizo dentro del proyecto TERMISUL con la colaboración de las becarias Sabrina Gomes, Thaís Bueno e Vitória Roso.

Además de esta etapa de limpieza, los textos de los corpus de estudio fueron catalogados siguiendo el modelo de catalogación de TERMISUL que se aplicó también a los textos de divulgación. La catalogación tiene el patrón: indicación del país (uy y br), la indicación de la temática del corpus (PP), o sea, corpus sobre conservación y restauración de bienes en papel; el número del texto (001, 002, etc.). Para los textos de divulgación, después de la indicación del país, se agregó una d(ivulgación). A seguir se presentan ejemplos de la catalogación de textos de Uruguay.

Cuadro 1 - Muestra de catalogación de los textos de los corpus de estudio de Uruguay

Código	Referencias	Disponible	Acceso
uyPP001	FRIEDMAN, Virginia; DUARTE, Mirtha; ESCOBAR, Erika. Conservación en el Archivo Literario. 2013. Disponible em: < http://www.bibliotecadelbicentenariogubuyinnovaportal/file/39767/1/conservacion_en_el_archivo_literario.pdf >. Acceso em: 10 jul. 2017.	< http://www.bibliotecadelbicentenariogubuyinnovaportal/file/39767/1/conservacion_en_el_archivo_literario.pdf >	10 jul. 2017
uydPP001	TRUJILLO, Valentin. La maravillosa manía de guardar. 2017. Disponible em: < http://www.elobservador.com.uy/la-maravillosa-mania-guardar-n287173 >. Acceso em: 19 set. 2017.	< http://www.elobservador.com.uy/la-maravillosa-mania-guardar-n287173 >	19 set. 2017

Fuente: Autora.

2.2 ANÁLISIS DE LOS TEXTOS ESPECIALIZADOS Y DE DIVULGACIÓN

Para el análisis de los textos seleccionados de los corpus de estudio (especializado y divulgación), se siguió la propuesta de Ciapuscio (2003) que incluye los siguientes niveles de análisis:

a) Nivel funcional: se refiere a las funciones textuales, entendidas como el efecto de los textos en el contexto de interacción social, teniendo en cuenta los objetivos y actitudes de los interlocutores. La autora distingue las siguientes funciones básicas: expresarse, autopresentarse, contactar, informar y dirigir (convencer).

b) Nivel situacional: corresponde a los conocimientos prototípicos relativos a determinado tipo de situación comunicativa. Incluye los siguientes aspectos: comunicación interna o externa al área o disciplina especializada; interlocutores: (especialista-especialista, especialista-semilego, especialista-lego, semilego-semilego, semilego-lego); relación entre los interlocutores (simétrica o asimétrica); número de interlocutores (diálogo, grupo pequeño o grande); y parámetros espacio-temporales (comunicación personal, gráfica, televisiva, virtual, etc.).

c) Nivel de contenido semántico: incluye el qué (la selección de la información semántica) y el cómo (la organización de la información a lo largo del texto). Se relaciona a la macroestructura y a la coherencia textual y abarca: el tema del texto; las actitudes temáticas; la perspectiva sobre el tema; las formas textuales primarias o derivadas; la partes textuales; tipos de despliegues temáticos (descriptivos, narrativos, argumentativos, expositivos).

d) Nivel formal-gramatical: se refiere a la selección y a la combinación de los recursos verbales y no verbales y abarca: las máximas de formulación relativas al género textual; formas lingüísticas (terminología, fraseologías, expresiones recurrentes) y no lingüísticas (figuras, gráficos, etc.);

aspectos gramaticales.

A partir de dicha propuesta, el análisis se hizo para cada uno de los textos del corpus de estudio y se organizó en tablas que contenían las informaciones presentadas en la figura 1 (texto uyPP001).

Figura 1 - Análisis texto especializados Uruguay (uyPP001)

Fuente: Autora.

A partir del análisis de todos textos para las dos lenguas, se hizo una tabla con una síntesis general para sistematizar los datos e identificar las semejanzas y diferencias entre los textos de las dos lenguas. Una muestra de los resultados para los niveles funcional y situacional se puede ver en la figura 2.

Figura 2 - Muestra de los niveles funcional y situacional del corpus de estudio

Niveles	Texto especializados	Textos de divulgación
Funcional:	informar y dirigir	Informar
Situacional	Interlocutores: especialista-especialista. Relación: simétrica; comunicación interna a la disciplina. Nº interlocutores: amplio. Parámetros espacio-temporales: comunicación gráfica; textos entre 8 y 10 páginas disponibles en internet.	Interlocutores: especialistas-legos. Relación: asimétrica; comunicación externa a la disciplina. Nº interlocutores: muy amplio. Parámetros espacio-temporales: comunicación gráfica; textos entre 1 y 5 páginas disponibles en internet.

Fuente: Autora.

2.3 EXTRACCIÓN E IDENTIFICACIÓN DE LOS TÉRMINOS

Para la extracción de los datos, se usaron las herramientas AntConc³ y Sketch Engine⁴ (SE). Con los datos obtenidos y a partir de determinados criterios, seleccionamos los términos y los organizamos en tablas Excel.

El AntConc es un programa gratuito creado por Laurence Anthony y de las herramientas que ofrece se utilizaron:

- la lista de palabras (*wordlist*): genera lista de palabras por orden de frecuencia (también permite el orden alfabético e inverso). Para generarla se creó y aplicó una lista de exclusión (*stoplist*) que contiene palabras gramaticales que no interesaban en la investigación (pronombres, conjunciones, artículos, preposiciones, números, etc.);
- la lista de palabras clave (*keywords*): se genera a partir de la *wordlist* anterior en contraste con la lista de palabra del corpus de referencia. El resultado es una lista de palabras clave con la indicación del índice de *keyness* (*clavicidad*).
- las concordancias: a partir de un término se generan sus contextos de uso, lo que permite

³ Disponible en <<http://www.laurenceanthony.net/software.html>>. Accedido el: 28 de nov. 2017.

⁴ Disponible en <https://www.sketchengine.co.uk/>. Accedido el: 28 de nov. 2017.

identificar: términos sintagmáticos, especificidades de uso, etc.

- *clusters*: genera un conjunto de coocurrentes a partir de un término, lo que permite identificar fraseologías relacionadas al término, por ejemplo.

A continuación, presentamos algunas figuras que ilustran las herramientas del programa. En la figura 3, se puede ver la lista de palabras clave del corpus académico de Uruguay. La primera columna muestra los textos del corpus; la segunda indica el orden de aparición (1, 2, 3, 4,...); la tercera muestra la frecuencia de la palabra en el corpus (221, 168...); la cuarta, la *keyness* (*clavidad*), es decir, el score de la palabra del corpus de estudio en contraste con su ocurrencia en el corpus de referencia (988.28, 817.003, etc.); finalmente se presenta la *keyword* (palabra clave), o sea, la palabra a que corresponde el índice de *clavidad* (*documentos, biblioteca, conservación, etc.*).

Figura 3 - Lista de palabras clave del corpus académico uy.

Fuente: AntConc

En la figura 4, se muestran los contextos (concordancias) para el término *documentos* y los textos en que aparece cada uno de ellos (uyPP001, uyPP003...).

Figura 4 - Concordancias de la palabra *documentos* del corpus académico uy.

Fuente: AntConc

A su vez, el SE es un programa desarrollado inicialmente por Adam Kilgarriff (2003); es paga, pero ofrece una versión demo para uso por un periodo de 30 días. De sus herramientas, se utilizaron la *keywords* (genera una lista de posibles candidatos a términos simples y sintagmáticos; figura 5) y el *wordketch* (ofrece informaciones sobre la palabra buscada como las relaciones que establece con otras palabras, su función sintáctica, etc., conforme se ve en la figura 6 para el término *conservación*).

Figura 5 - Muestra de la lista de candidatos a términos y multiwords (términos sintagmáticos)

Fuente: Sketch Engine

Figura 6 - *Wordsketch* para *conservación*.

Fuente: Sketch Engine

2.4 EXTRACCIÓN DE LOS CANDIDATOS A TÉRMINOS

Para cada uno de los corpus de estudio se subió una *stoplist* y el corpus de referencia tanto en el AntConc como en el SE. Como se afirmó, la *stoplist* permite excluir palabras gramaticales que pueden generar ruido en la extracción. Por ejemplo: artículos (la/a; el/o); pronombres (nosotros/nós; él/el), preposiciones (a, de).

Como se explicó, el contraste del corpus de estudio con el corpus de referencia, permite identificar las palabras específicas del primer corpus -*Keyness (clavicidad)*. De hecho, utilizar la *stoplist* y el corpus de referencia ya pueden ser considerados criterios de selección de los términos, pues ayudan a filtrarlos.

Para poder establecer un parámetro de comparación entre las dos herramientas utilizadas y posteriormente entre los textos académicos y los de divulgación, se estableció un corte en la *clavicidad* de 50.00, considerado adecuado en función del tamaño pequeño de los corpus. A partir de este corte, los candidatos a términos seleccionados se colocaron en una grilla Excel. Se mezclaron las listas generadas por los dos programas para identificar los términos comunes y los distintos. En una primera etapa de limpieza se retiraron: los nombres propios (Quiroga, Güichen); los topónimos (Montevideo, São Paulo); las siglas (CNPq, ISBN, ISSN). Sobre la lista resultante, se aplicaron los siguientes criterios:

1) criterio de categoría nominal: los términos deberían ser de categoría nominal, lo que permitió excluir verbos, adjetivos y adverbios (*simbolizar, diferentes, conocer, etc.*);

2) criterio de sinonimia/pertinencia pragmática: permitió seleccionar términos que ocurrían con las conjunciones “y/e” e “o/u” o expresiones como “llamado también de”, uso de paréntesis. Por ejemplo: *salvamento y rescate; negligencia o descuido; Descarte llamado también expurgo o selección negativa; amidos (colas)*.

3) criterio de pertinencia temática: se seleccionaron términos que podían formar parte del mapa conceptual del área de conservación y restauración en papel elaborado por Bojanoski⁵. Eso permitió excluir sustantivos como *años, siglos, artes escénicas*. Con estos criterios se obtuvieron:

País	Textos especializados	Textos de divulgación
Uruguay	152	148
Brasil	224	210

Algunos ejemplos de los términos más frecuentes seleccionados en cada corpus fueron:

Cuadro 2 - Ejemplos de términos extraídos de los corpus de estudio.

Términos textos académicos uy	Términos textos académicos pt	Términos textos divulgación uy	Términos textos divulgación pt
Acervo	Acervo	Acervo	Acervo
Acervo bibliográfico	--	--	Acervo bibliográfico
Acervo cultural	Acervo cultural	--	--
Catalogación	--	Catalogación	Catálogo
Documentos	Documentos	Documentos	Documentos
Documentos	--	Documentos	Documentos
históricos	Documentos gráficos	históricos	históricos
--	Restauração	--	--
Preservación		Restauración	Restauração

Fuente: Autora.

⁵El mapa conceptual utilizado es uno de los resultados de la tesis de Silvana Bojanoski cuyo título es *Terminologia em Conservação de Bens Culturais em papel: produção de glossário para profissionais em formação*, a ser presentada en marzo de 2018 en el Programa de Pós-Graduação em Memória Social e Patrimônio Cultural dessa Universidade Federal de Pelotas. La autora es codirectora de la tesis.

3 CONSIDERACIONES FINALES

A partir de las etapas metodológicas y del marco teórico que fundamenta el proyecto fue posible describir los textos del corpus de estudio (objetivo específico a) e identificar los términos (objetivo específico b), así como seleccionar un conjunto de términos (Cuadro 2) que será analizado más detalladamente para explicar las semejanzas y diferencias en su uso en los textos especializados y los de divulgación. Así, la continuidad de la investigación prevé las etapas siguientes: análisis de los términos más frecuentes en cada corpus y su inclusión en el mapa conceptual elaborado por Bojanoski; análisis de la variación terminológica; y establecimiento de la relación de los resultados anteriores con aspectos teóricos de la divulgación.

Se espera, de este modo, poder contribuir para el avance de la interdisciplinariedad tan necesaria entre las áreas de Terminología, Información y Comunicación.

REFERENCIAS BIBLIOGRÁFICAS

- Berber Sardinha, A. (2004). *Linguística de Corpus*. Barueri: Manole.
- Biber, D., Conrad, S. y Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Cabré, M. T. (1999). *Terminología: representación y comunicación: Una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Calsamiglia, H. (1997). Divulgar: Itinerarios discursivos del saber: Una necesidad, un problema, un hecho. *Quark*, (7), 9-18. Recuperado de <http://quark.prbb.org/7/estrella.htm>.
- Ciapuscio, G. (2003). *Textos especializados y terminología*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Jiménez Aleixandre, M. P. (2003). La cultura científica en las clases de ciencia: Comunidad de aprendizaje. *Cultura Científica*, (28-29), 1-7. Recuperado de <http://quark.prbb.org/28-29/028057.htm>. Accedido el: 29 abr 2017.