

# Normalización terminológica en el ámbito de los bienes patrimoniales chilenos: del corpus a la plataforma web (Short Paper)

Ignacia Montero, Carles Tebé y Luciana Pissolato

Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago de Chile

## Abstract

This paper deals with the process of terminological standardization in a monolingual corpus in Chilean Spanish in the field of conservation-restoration of heritage assets carried out within the framework of the FONDEF IDeA I+D project "Platform for the identification of alterations to Chilean heritage assets". Specifically, we will explain the methodological stages that constitute the standardization process: 1) constitution of the corpus, 2) extraction of terminological information from the corpus, 3) preparation of the materials for the standardization sessions, 4) the standardization sessions and 5) elaboration of the output dossier, prior to the transition stage that corresponds to the transfer of the standardization results to the web platform. The objective of this standardization process is to obtain terms related to concepts of alterations to Chilean heritage assets that are well delimited and agreed upon by specialists in the field for the creation of a terminology database accessible through a web platform. The uniqueness of the standardization carried out is that, although the final decisions rely on the heritage specialists, the process to arrive at those decisions is guided by terminologists. This paper will focus on presenting the particularities and challenges that arose in the different stages and the solutions adopted in each of them, and relate them to the characteristics of the disciplinary field, the corpus and the objectives and needs of the project.

## Resumen

En esta ponencia se aborda el proceso de normalización terminológica en un corpus monolingüe en español de Chile del ámbito de la conservación-restauración de bienes patrimoniales realizado en el marco del proyecto FONDEF IDeA I+D "Plataforma para la identificación de las alteraciones a los bienes patrimoniales chilenos". En específico, se explicarán las etapas metodológicas que constituyen el proceso de normalización: 1) constitución del corpus, 2) extracción de información terminológica del corpus, 3) preparación de los materiales para la sesión de normalización, 4) las sesiones de normalización y 5) elaboración del dossier de salida, previas a la etapa de transición que corresponde al traslado de los resultados de la normalización a la plataforma web. El objetivo de este proceso de normalización es obtener términos relativos a conceptos de alteraciones de los bienes patrimoniales chilenos bien delimitados y consensuados por los especialistas del ámbito para la creación de una base de datos terminológica accesible mediante una plataforma web. La singularidad de la normalización llevada a cabo es que, si bien las decisiones finales recaen en los especialistas del ámbito, el proceso para llegar a esas decisiones está guiado por terminólogos. Esta ponencia se centrará en presentar las particularidades y desafíos que surgieron en las diferentes etapas y las soluciones adoptadas en cada una de ellas, y relacionarlas con las características del ámbito disciplinar, del corpus y con los objetivos y necesidades del proyecto.

## Palabras clave

Normalización terminológica, conservación-restauración de bienes patrimoniales, lingüística de corpus, variación terminológica, recursos terminológicos.

## 1. Introducción

Esta ponencia aborda el proceso de normalización terminológica de las alteraciones a los bienes patrimoniales chilenos llevado a cabo en el marco del proyecto FONDEF IDeA I+D

---

3rd International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2024, June 27-28, 2024, Granada, Spain.

✉ ignacia@uc.cl (I. Montero); ctebe@uc.cl (C. Tebé); lupissolato@uc.cl (L. Pissolato)

ORCID 0000-0001-7031-0854 (I. Montero); 0000-0002-2300-7243 (C. Tebé); 0000-0002-8188-499X (L. Pissolato)

© 2024 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

*“Plataforma para la identificación de las alteraciones a los bienes patrimoniales chilenos”*, en específico, sus particularidades y desafíos en relación con el ámbito, el corpus y los objetivos del proyecto en que se enmarca. Este proyecto surge a partir de los problemas de comunicación que enfrentan los especialistas del Centro Nacional de Conservación y Restauración (en adelante, CNCR), debido a la falta de fijación de los conceptos que utilizan y sus denominaciones. El proyecto tiene como objetivo general “diseñar, implementar y validar una plataforma para identificación de alteraciones en los bienes patrimoniales de Chile, que permita a la comunidad de conservadores y restauradores reducir la variación terminológica utilizada en la identificación/descripción de las alteraciones a los bienes patrimoniales, a través de conceptos y denominaciones normalizadas, con la finalidad de abordar formas más eficientes de enfrentarlas”. Para cumplir con ese objetivo general, es fundamental establecer una metodología y protocolos específicos que permitan llevar a cabo un proceso de normalización riguroso de la terminología del ámbito.

## **2. Etapas metodológicas**

A continuación se explican las etapas metodológicas que componen este proceso de normalización y las dificultades, limitaciones y soluciones adoptadas en cada una de ellas:

### **2.1. Constitución del corpus**

La primera etapa de constitución del corpus consistió en la recopilación y selección de los textos, la caracterización general de dichos textos según su tipo y grado de especialidad, las especificidades del proceso de compilación y la presentación de los datos estadísticos. Este corpus fue compilado con el propósito de estudiar el uso real que hacen los especialistas de los términos de alteraciones de los bienes patrimoniales. Por ello, el criterio de selección de textos fue la incorporación de todos los documentos disponibles redactados por especialistas del CNCR y facilitados por la misma institución. Estos documentos se dividen en cuatro subcorpus, correspondientes a los cuatro laboratorios del CNCR: 1) arqueología, 2) escultura y monumentos, 3) pintura y 4) papel y libros. Esta división permite detectar eventuales diferencias en el uso de la terminología según el área de especialidad, como posible causa de variación. Con respecto a los tipos de texto que conforman el corpus, estos se clasifican en: informes de intervención, informes de asesoría y fichas clínicas. En ellos se abordan diferentes etapas del trabajo de conservación-restauración de bienes patrimoniales en Chile, desde el estudio histórico de los bienes, su descripción y caracterización, hasta los diferentes procesos de evaluación y clasificación de las alteraciones que presenta el objeto, y las consiguientes acciones de conservación y restauración que se llevan a cabo. Estos tres tipos de texto presentan características muy diferentes entre sí en cuanto a extensión y grado de especialización en el uso del lenguaje (y, por lo tanto, densidad terminológica específica del ámbito). Entre las limitaciones del proceso de compilación se encuentran aspectos como la falta de regularización del formato y estructura de los documentos, que presentan variación en las secciones que los componen y su nombramiento. Otra de las limitaciones en la compilación es la diferencia en la disponibilidad de textos por parte de los diferentes laboratorios de la institución, que responde a factores como la cantidad de bienes patrimoniales que reciben y las intervenciones que realizan. Esto se traduce en un desbalance en la representatividad de cada laboratorio en el corpus. Si bien estas limitaciones dificultan la obtención de conclusiones específicas sobre la variación de los diferentes laboratorios de forma independiente, y también sobre la relación entre la variación y el género textual, sí es posible obtener conclusiones sobre la variación de forma transversal en el ámbito a través de la recuperación de gran parte de las variantes y relaciones entre los términos, lo que contribuye a delimitar la estructura conceptual.

## 2.2. Extracción de información terminológica del corpus

La segunda etapa consistió en el procesamiento y análisis del corpus que se realizó con la herramienta de gestión de corpus Sketch Engine. Esta herramienta incorpora lematización automática para los textos en español, permite búsquedas de concordancia y colocaciones para obtener información sobre la frecuencia y el comportamiento sintáctico y semántico de las unidades en el corpus, por lo que resulta adecuada para el reconocimiento de terminología. El análisis tuvo como punto de partida la base de datos terminológica elaborada por Silva [1] y Adriaola [2] a partir de diferentes glosarios proporcionados por el CNCR. En específico, se tomó cada término del glosario y se buscaron todas las entradas del término (por laboratorio) con sus variantes. Así, por ejemplo, el término “faltante” presenta las siguientes variantes:

**Arqueología:** faltante estructural, faltante superficial

**Escultura y Monumentos:** pérdida de elementos, falta de material, laguna, faltante, pérdida de material

**Pintura:** faltante estructural, faltante superficial, faltante de capa

**Papel y libros:** faltante

Estas informaciones de la base de datos se contrastan con el corpus extrayendo la frecuencia de uso de cada término y variante, consignando colocaciones frecuentes y detectando posibles nuevas variantes a través de diferentes estrategias (marcadores discursivos, marcas ortotipográficas y búsquedas dirigidas). Finalmente, según los resultados de esta exploración, se conformaron series terminológicas compuestas por los términos candidatos a variantes para un mismo concepto.

## 2.3. Preparación de los materiales para la sesión de normalización

En la tercera etapa, las informaciones obtenidas en el análisis del corpus se unificaron y sintetizaron en un formato que permitiera reflejar los diferentes aspectos involucrados en el uso real que se hace de los términos en los textos del corpus. Para ello, se diseñó una ficha en la que se consigna cada serie terminológica con los datos tanto cuantitativos como cualitativos más relevantes desde el punto de vista lingüístico (frecuencia de uso de los términos por laboratorio, colocaciones frecuentes/expresiones terminológicas). Adicionalmente, se obtuvieron desde Sketch Engine contextos de uso para cada término por laboratorio y se solicitó al CNCR el envío de imágenes de las alteraciones correspondientes a los términos de las series cuando estuvieran disponibles. Todos estos materiales se juntaron en una carpeta para cada serie, a modo de evidencias y argumentos que contribuyeran a la discusión con los especialistas y favorecieran el consenso. Las carpetas de series fueron enviadas previamente a la realización de su respectiva sesión de normalización para preparación de los especialistas.

## 2.4. Sesiones de normalización

La cuarta etapa del proceso de normalización corresponde a las sesiones de normalización, que consistieron en sesiones conjuntas de discusión entre los lingüistas del proyecto y especialistas de los diferentes laboratorios del CNCR. En estas sesiones se presentaron las series terminológicas y la evidencia obtenida a partir del análisis del corpus, reunida en los materiales presentados anteriormente. La discusión fue guiada por los lingüistas través de preguntas dirigidas a los especialistas para determinar, en primer lugar, si los términos correspondían/se clasificaban como alteraciones, y en segundo lugar, si todos los términos de una misma serie hacían referencia al mismo concepto de alteración o si había alguna diferencia conceptual entre ellos. En la discusión, además, se evaluaron diferencias en el uso de los términos entre los laboratorios (ya sea diferencias en el concepto al que refiere un mismo término, o diferencias en el término que se utiliza para referir a un mismo concepto de alteración). El propósito de estas sesiones fue confirmar las series terminológicas, y el eventual establecimiento de variantes

preferentes (una o más), a partir del consenso alcanzado por los especialistas. Los términos rechazados se incluirán igualmente en la plataforma para información de sus usuarios.

## **2.5. Elaboración del dossier de salida**

En la quinta etapa, los resultados de las sesiones de normalización se consignan en un documento denominado dossier de salida. Este dossier tiene una doble función: por una parte, debe documentar las decisiones y conclusiones del proceso de normalización de cada término/serie terminológica, y por otra parte debe sintetizar y estructurar la información en un formato unificado proyectable en/trasladable a una plataforma para ser utilizada como recurso terminológico. En este sentido, el mayor desafío fue determinar tanto el contenido del dossier como su formato. Así, se incluye la frecuencia de uso y el área de especialidad en el que se utiliza (ya sea una o más de una), y las expresiones terminológicas asociadas. Además, se incluyen las observaciones consensuadas alcanzadas en la sesión de normalización, enfatizando aquellos aspectos relevantes como matices, casos particulares y relaciones con otros términos que son pertinentes de reflejar en la plataforma para consultas terminológicas. Finalmente, se incluye la decisión de normalización que sintetiza los términos confirmados para un mismo concepto y se especifica si existe algún término preferente.

Posteriormente a las etapas del proceso de normalización terminológica presentadas se pasa a una etapa de transición entre la normalización terminológica y la elaboración de la plataforma web. Esta etapa no será desarrollada en profundidad en esta ponencia, sin embargo, se presenta a modo general como parte de los objetivos del proyecto que determinan algunas decisiones en las etapas de la normalización. Esencialmente, esta etapa consiste en el traspaso de los resultados obtenidos en el proceso de normalización y unificados en los dossiers a la plataforma web de recursos terminológicos del ámbito, con la ayuda de programadores y especialistas en software.

## **3. Conclusiones**

El uso de terminología normalizada es fundamental para conseguir una comunicación eficaz y sin ambigüedades en ámbitos de conocimiento especializado. La falta de consenso y fijación del lenguaje utilizado en el ámbito de los bienes patrimoniales en Chile acarrea problemas de comunicación tanto a nivel interno, entre los mismos especialistas de la disciplina, como en su proyección externa en instancias de divulgación e intercambio nacional o internacional. El objetivo general del proyecto que enmarca este trabajo es la creación de una plataforma como recurso terminológico que contribuya a mejorar la comunicación y el trabajo en torno a la conservación-restauración.

Para cumplir con este objetivo es necesario llevar a cabo un proceso de normalización riguroso constituido por diferentes etapas que van desde la compilación y análisis del corpus lingüístico hasta la creación de dossiers de normalización. El objetivo de este proceso de normalización es obtener términos relativos a conceptos de alteraciones de los bienes patrimoniales chilenos bien delimitados y consensuados por los especialistas del ámbito para la creación de una base de datos que se difundirá en una plataforma terminológica.

En esta ponencia se presentan las particularidades y desafíos que surgieron en las diferentes etapas del proceso de normalización guiado por terminólogos, derivadas de aspectos específicos del ámbito disciplinar, del corpus y de los objetivos del proyecto, y las decisiones y criterios adoptados frente a ellos.

## **Referencias**

- [1] M. Silva, Creación de una base de datos terminológica en el área de conservación y restauración de obras de arte, Tesis de magíster, Pontificia Universidad Católica de Chile (PUC), Santiago, Chile, 2017.

- [2] A. Adriazola, Terminología de las alteraciones patrimoniales en el área de la conservación y restauración de obras de arte, Tesis de magíster, Pontificia Universidad Católica de Chile (PUC), Santiago, Chile, 2019.
- [3] S. Bojanoski, F. Michelon, C. Bevilacqua, Os termos preservação, restauração, conservação e conservação preventiva de bens culturais: uma abordagem terminológica, *Revista Calidoscópico* 3 (2017) 43-454.
- [4] L. Bowker, J. Pearson, J. *Working with Specialized Language: A Practical Guide to Using Corpora*. Routledge, London, 2002.
- [5] L. Bowker, S. Hawkins, Variation in the organisation of medical terms: Exploring some motivations for term choice, *Terminology* 1 (2006) 79-110.
- [6] V. Brezina. *Statistics in Corpus Linguistics: A Practical Guide*. University Press, Cambridge, 2018.
- [7] T. Cabré, R. Estopà R, J. Vivaldi, Automatic term detection. A review of current systems, in Bourigault D., Jaquemin C., y L'Homme M.-C. (Eds.), *Recent Advances in Computational Terminology*, John Benjamins, Amsterdam, Philadelphia, 2001, pp. 53-87.
- [8] A. Calvo, La normalización terminológica aplicada a la conservación y restauración de patrimonio cultural in AA.VV, *El lenguaje sobre el patrimonio. Estándares documentales para la descripción y gestión de colecciones*, Ministerio de Educación, Cultura y Deporte, Madrid, 2016, pp. 131-139.
- [9] G. Ciapuscio, *Textos especializados y Terminologías*, Publicaciones del IULA, Universidad Pompeu Fabra Barcelona, 2003.
- [10] R. Costa, Texte, terme et contexte in: D. Blampain, P. Thoiron, Marc Van Campenhoudt (Eds.), *Actes des VIIes Journées scientifiques du Réseau Lexicologie, Terminologie et Traduction. "Mots, Termes et contextes"*, Editions des archives contemporaines, Paris, 2006, pp. 79 – 88.
- [11] W. Crawford, E. Csomay, *Doing Corpus Linguistics*. Routledge, London, 2016.
- [12] G. Desagulier, *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*, Springer, Switzerland, 2017.
- [13] J. Freixa. Causes of Denominative Variation in Terminology: A Typology Proposal, *Terminology*, 1 (2006), 51-77.
- [14] R. Garside, G. Leech, T. McEnery, *Corpus Annotation. Linguistic Information from Computer Text Corpora*, Routledge, London, 2013.
- [15] P. C. Gómez, *Statistical Methods in Language and Linguistics Research*, Equinox, Sheffield, 2013.
- [16] ICOM-CC, Terminology to characterize the conservation of tangible cultural heritage. Resolution adopted by the ICOM-CC membership at the 15th Triennial Conference, New Delhi, 2008.
- [17] A. Kilgarriff, V. Baisa, J. Busta, M. Jakubicek, et al., The sketch engine: Ten years on, *Lexicography*, 1(2014), 7-36.
- [18] M. Rogers, K. Ahmad, *Corpus Linguistics and Terminology Extraction in: Handbook of Terminology Management. Vol. 2*, John Benjamins, 2001, pp.725-760.