

Do Corpus Lingüístico À Extração Terminológica: Insumos Para a Criação De Uma Plataforma Web Baseada Na Normalização Terminológica Do Domínio Da Conservação-restauração De Bens Patrimoniais (Short Paper)

Luciana Pissolato¹ Carles Tebé²

¹ Pontificia Universidad Católica de Chile, Av Vicuña Mackenna, 4680, Santiago, Chile

² Pontificia Universidad Católica de Chile, Av Vicuña Mackenna, 4680, Santiago, Chile

Resumo

Neste artigo, abordaremos a constituição e exploração de um corpus linguístico monolíngue em espanhol -variante chilena- no campo da conservação-restauração de bens patrimoniais, a fim de correlacionar os diferentes gêneros textuais que compõem o corpus com a extração de sua terminologia. Discutiremos, mais especificamente, os problemas relacionados com a variação terminológica na área e proporemos uma forma de abordá-los por meio do desenho de recursos tecnológicos que colaborem com sua fixação e difusão entre os principais produtores/consumidores desse conhecimento. As discussões e os resultados preliminares apresentados neste trabalho fazem parte do projeto de pesquisa “*Plataforma para la identificación de las alteraciones a los bienes patrimoniales chilenos*” (ID22110052), concedido em setembro de 2022, e cuja conclusão está prevista para o início de 2024. Esta iniciativa é financiada pela *Agencia Nacional de Investigación y Desarrollo de Chile -ANID-*, no âmbito do concurso Fondef IDEA I+D 2022, cujo objetivo é apoiar o cofinanciamento de projetos de P&D aplicados com forte componente científico para o desenvolvimento de tecnologias que possam se tornar novos produtos, processos ou serviços, com probabilidade de gerar impactos produtivos, econômicos e sociais.

Palavras-chave

Normalização terminológica, conservação-restauração de bens patrimoniais, corpus linguístico, extração terminológica, variação terminológica

1. Introdução

No Chile, um problema crítico enfrentado pelas organizações públicas que produzem e transmitem conhecimento especializado é a falta de terminologia normalizada. O problema particular que abordamos neste artigo está focalizado no campo temático da conservação e restauração de bens patrimoniais e em uma área geolinguística específica –o espanhol do Chile–.

Os problemas relacionados com a falta de uma terminologia normalizada no *Centro Nacional de Conservación y Restauración* (em adiante, CNCR), dependente do *Ministerio de las Culturas, las Artes y el Patrimonio*, são graves e situam-se em dois níveis: problemas de fixação ou representação do conhecimento e problemas de comunicação ou transferência desse conhecimento.

No campo da fixação do conhecimento, o principal problema é a falta de consenso em torno aos limites de muitos dos conceitos que utilizam, e nas denominações (termos especializados) que os designam. Em trabalhos exploratórios preliminares, observam-se fenômenos de sinonímia, polissemia, variação denominativa e variação conceitual, que precisam ser abordados de forma sistemática. No campo da comunicação, os problemas manifestam-se a) na redação dos documentos necessários à

2nd International Conference on “Multilingual Digital Terminology Today. Design, Representation Formats and Management Systems”, (MDTT) 2023, June 28-29, 2023, Lisbon, Portugal

EMAIL: lupissolato@uc.cl (L. Pissolato); ctebe@uc.cl (C. Tebé)

ORCID: 0000-0002-8188-499X (L. Pissolato); 0000-0002-2300-7243 (C. Tebé)



© 2023 Copyright for this paper by its authors.

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

CEUR Workshop Proceedings (CEUR-WS.org)

realização de seus trabalhos de conservação e restauração –diagnósticos, fichas clínicas, relatórios de intervenção –; b) nas ações de formação realizadas por conservadores e restauradores no próprio Centro e em todo o Chile (cursos, seminários, workshops) e c) também se manifestam em intercâmbios com outros especialistas externos ao CNCR (simpósios, congressos, jornadas científicas). A falta de uma terminologia normalizada na área dos bens patrimoniais acarreta uma dificuldade de se fazer um diagnóstico e, assim, pactar as ações necessárias para solucionar o problema que esse bem apresenta; e, conseqüentemente, na impossibilidade de comunicar eficazmente à contraparte interessada o estado de conservação daquele bem.

A nível de país, sendo o CNCR uma instituição de âmbito nacional vinculada ao *Ministério de las Culturas, las Artes y el Patrimonio*, a falta de uma terminologia normalizada impacta em uma dificuldade de elaborar um diagnóstico preciso a nível nacional do patrimônio cultural, elemento base para a execução de planos e programas futuros para sua conservação.

No campo da conservação-restauração de bens patrimoniais, existem muitos recursos externos em formato de dicionários, alguns internacionais e multilíngues, outros monolíngues em espanhol. No entanto, nenhum desses recursos terminológicos está normalizado, de modo que entre as diferentes fontes disponíveis para consulta:

1. existem discrepâncias em todos os níveis: seleção de termos, indicação de sinônimos ou termos preferentes, relações entre conceitos, definições, etc.
2. a maioria dessas obras é iberocêntrica; em outros casos, são traduções de obras originais escritas em outros idiomas e, sobretudo,
3. não incluem os usos do espanhol chileno, que possui terminologia própria devido a sua rica tradição na área,
4. estão em formatos diferentes, e
5. não são facilmente acessíveis ou interativos.

Este estado de coisas dificulta a comunicação inequívoca de certos conceitos fundamentais para o desempenho do trabalho profissional dos restauradores e conservadores no país. Se consideramos o papel central desempenhado pelo CNCR como centro de referência nacional – além de seu prestígio internacional – no domínio da conservação-restauração do patrimônio, abordar conjuntamente a normalização da sua terminologia com uma metodologia rigorosa de trabalho, é uma tarefa de grande relevância para este campo especializado, e que pode servir de modelo para a normalização terminológica de outras áreas das ciências humanas e sociais que apresentem as mesmas necessidades no país.

Dado que não é possível quantificar com precisão o número total de termos utilizados em uma disciplina que não possui uma nomenclatura unificada e validada internacionalmente, como ocorre com outras disciplinas científicas e técnicas, em trabalho exploratório anterior optou-se por delimitar uma subárea concreta para realizar o projeto de normalização da terminologia das alterações de bens patrimoniais, com o objetivo de abordar o trabalho em um número limitado de conceitos, que permita a validação da metodologia e a implementação dos resultados na comunidade de restauradores e conservadores do Chile.

Dentro do processo metodológico de intervenção de um bem patrimonial, a etapa de diagnóstico é uma das mais críticas, já que as decisões de intervenção que serão tomadas dependerão precisamente de um diagnóstico. Um dos seus pilares é a identificação e descrição das alterações apresentadas pelos bens que lhes são entregues para estudo e eventual intervenção posterior. Na documentação que é gerada para cada trabalho patrimonial, antes da sua eventual intervenção, existem dois tipos de documentos essenciais, que são as fichas clínicas e os relatórios de intervenção: a) as fichas clínicas são instrumentos de trabalho utilizados para identificar, descrever e registrar o estado de conservação a proposta de intervenção de conservação e/ou restauração, juntamente com os tratamentos realizados. b) Os relatórios de intervenção são instrumentos metodológicos que reúnem toda a informação resultante do processo realizado, além da análise de cada caso de intervenção de conservação e/ou restauração. Através dos relatórios, são revelados os principais antecedentes do valor histórico, social e cultural do objeto de estudo, bem como os resultados das análises científicas necessárias ao seu conhecimento tecnológico.

Como se pode observar, ambos tipos de documentos constituem um corpus de trabalho de extrema relevância para este projeto, pois correspondem a textos reais produzidos por especialistas na área

temática e são, portanto, uma fonte de evidência sobre o uso que se dá à terminologia referente às alterações dos bens patrimoniais nos diferentes laboratórios do CNCR.

2. Objetivos

Nesse contexto, este artigo abordará a constituição e exploração de um corpus linguístico monolíngue em espanhol -variante chilena-, a fim de correlacionar os diferentes gêneros textuais que compõem o corpus com a extração de sua terminologia.

A primeira fase do projeto visa estabelecer um primeiro vocabulário terminológico de cerca de 300 conceitos-chave no campo das alterações ou danos aos bens patrimoniais. Este é um campo transversal e essencial para o exercício da profissão de conservador-restaurador que abrange as cinco áreas temáticas que compõem o corpus linguístico: escultura e monumentos, papel e livros, pintura, arqueologia e têxtil. Em cada uma dessas áreas temáticas, os textos que compõem o corpus pertencem a diferentes gêneros textuais: fichas clínicas que descrevem o estado do bem patrimonial; relatórios de assessoria, que orientam o proprietário do bem sobre qual tipo de intervenção é necessária; relatórios de intervenção, que descrevem e justificam as ações de conservação e restauração realizadas, e artigos científicos, que relatam, numa perspectiva acadêmica, os avanços das técnicas e métodos de trabalho na área.

O objetivo dessa primeira etapa é coletar toda a variação terminológica existente no corpus, fazer uma primeira descrição dos tipos e causas dessa variação e criar um banco de dados terminológico que sirva de base para a etapa posterior de normalização desse vocabulário. O produto final deste projeto é disponibilizar um recurso tecnológico -em formato de plataforma web-, que colabore com o estabelecimento e disseminação de tal conhecimento.

Neste artigo, nos concentraremos nos primeiros resultados da extração terminológica do corpus, que nos permitem observar fenômenos interessantes de variação terminológica na disciplina. Em alguns casos, diferentes variantes podem ser correlacionadas de acordo com o gênero textual em que estão documentadas.

3. Conclusões

Nesta exposição, apresentamos os problemas relacionados com a normalização da terminologia de conservação-restauração de bens patrimoniais no Chile e o desenvolvimento deste projeto de pesquisa interdisciplinar. Os resultados e produtos derivados deste estudo estarão disponíveis para consulta pública no início de 2024, e serão divulgados, oportunamente, pela equipe de pesquisadores.

4. Referências

1. Bojanoski, S., Michelon, F., Bevilacqua, C. (2017). *Os termos preservação, restauração, conservação e conservação preventiva de bens culturais: uma abordagem terminológica*. Revista Calidoscópico, v. 15, n.3, pp.43-454.
2. Bowker, L., Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.
3. Bowker, L. & Hawkins, S. (2006). Variation in the organisation of medical terms: Exploring some motivations for term choice. *Terminology*, 12(1), 79-110.
4. Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.
5. Cabré T., Estopà R. & Vivaldi J. (2001). Automatic term detection. A review of current systems. En Bourigault D., Jaquemin C., y L'Homme M.-C. (eds.), *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia: John Benjamins, p. 53-87.
6. Calvo, A. (2016) "La normalización terminológica aplicada a la conservación y restauración de patrimonio cultural", en AA.VV. *El lenguaje sobre el patrimonio. Estándares documentales para la descripción y gestión de colecciones*. Madrid: Ministerio de Educación, Cultura y Deporte, 2016, pp. 131-139.
7. Ciapuscio, G. (2003). *Textos especializados y Terminologías*. Barcelona: Publicaciones del IULA, Universidad Pompeu Fabra.

8. Costa, R. (2006). Texte, terme et contexte. Actes des VIIes Journées scientifiques du Réseau Lexicologie, Terminologie et Traduction. "Mots, Termes et contextes". Daniel Blampain / Philippe Thoiron / Marc Van Campenhoudt [ed.] Paris: Editions des archives contemporaines, pp. 79 – 88.
9. Crawfor, W. & Csomay, E. (2016). *Doing Corpus Linguistics*. London: Routledge.
10. Desagulier, G. (2017). *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. Switzerland: Springer.
11. Freixa, J. (2006). Causes of Denominative Variation in Terminology: A Typology Proposal. *Terminology*, 12(1), 51-77.
12. Garside, R., Leech, G. & McEnery, T. (2013). *Corpus Annotation. Linguistic Information from Computer Text Corpora*. London, Routledge.
13. Gómez, P. C. (2013). *Statistical Methods in Language and Linguistics Research*. Sheffield: Equinox.
14. ICOM-CC (2008). Terminology to characterize the conservation of tangible cultural heritage. Resolution adopted by the ICOM-CC membership at the *15th Triennial Conference*, New Delhi, 22-26 September 2008.
15. Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P., & Suchomel, V. (2014). *The sketch engine: Ten years on*. *Lexicography*, 1(1):7–36.
16. Peñas, A., Verdejo, F., & Gonzalo, J. (2001). *Corpus-based terminology extraction applied to information access*.
17. Rogers, M. & Ahmad, K. (2001). Corpus Linguistics and Terminology Extraction. In *Handbook of Terminology Management. Vol. 2.*, pp.725-760; John Benjamins.
18. Sandu, A., Spiridon, P., Sandu, I. (2016). *Current studies and approaches in the field of cultural heritage conservation science. Harmonising the terminology in an interdisciplinary context*. IJCS.