

Linguística de *corpus* e Garantia literária: contribuições metodológicas para a elaboração de tesouros e vocabulários controlados

Isabel Cristina Pereira dos Santos¹ Rita do Carmo Ferreira Laipelt²

RESUMO

Este trabalho tem por objetivo investigar como a Linguística de *Corpus* (LC), em um encontro com a Garantia Literária (GL), pode contribuir como método de análise de textos técnico-científicos para a elaboração de tesouros e vocabulários controlados no âmbito da Organização do Conhecimento (OC).

Palavras-chave: Linguística de *corpus*; Garantia literária; Texto técnico-científico; Organização do Conhecimento; Terminologia.

INTRODUÇÃO

No processo de elaboração de tesouros e vocabulários controlados, é preciso ter bastante clareza quanto à abordagem teórico-metodológica que será usada para a constituição do *corpus* e aplicada para a seleção de termos, a fim de que a escolha de determinado termo, em detrimento de outro, possa ser justificada de forma convicta e transparente. Além disso, é importante que o percurso metodológico seja descrito didaticamente para que possa ser replicado por outros pesquisadores que desejem seguir a mesma metodologia.

Tendo em vista a constituição de *corpus* para fins de pesquisa terminológica, busca-se, por meio de pesquisa bibliográfica, examinar duas abordagens teórico-metodológicas - a Linguística de *Corpus* (LC) e a Garantia Literária (GL) -, utilizadas

¹ Mestranda em Ciência da Informação; UFRGS; isabel.pereira@ufrgs.br

² Professora do Programa de Pós-Graduação em Ciência da Informação (PPGCIN) da UFRGS.

pela Terminologia e pela Organização do Conhecimento (OC), campos de conhecimento distintos, mas que se inter-relacionam por seus objetivos e aplicações, a fim de investigar de que forma a LC, em um encontro com a GL, pode contribuir como método de análise de textos técnico-científicos para a elaboração de tesouros e vocabulários controlados no âmbito da Organização do Conhecimento.

METODOLOGIA

Para atingir o objetivo proposto, primeiro, fez-se um levantamento, nos *sites* da Brapci e Web of Science, no período 1990-2020, do número de publicações sobre essas temáticas a fim de se conhecer a realidade da LC e da GL dentro da Ciência da Informação (CI). A seguir, realizou-se pesquisa bibliográfica sobre a história da LC e da GL, suas principais características, bem como a forma como elas tratam o *corpus* e realizam a análise terminológica, conforme o entendimento de autores como Barité (2019, 2018, 2017, 2010, 2007) Berber Sardinha (2004, 2000), McEnery e Wilson (2001) e Beghtol (1995, 1986), principais pesquisadores dessas abordagens.

RESULTADOS

Somando-se os resultados dos dois *sites*, foram recuperadas 27 publicações sobre GL e 3 publicações sobre LC para um período de 30 anos no âmbito da CI. Esse resultado se mostrou bastante irrisório para o período. Constata-se que há uma escassez de literatura científica sobre essas metodologias no campo da CI, o que significa que muito ainda pode ser explorado sobre o potencial dessas metodologias para pesquisas na área da OC.

Como metodologia, a LC mostra-se mais eficiente para a coleta e seleção de termos, por estabelecer princípios mais claros de organização do *corpus* e por propor ferramentas computacionais que auxiliam na seleção e análise de termos; já a GL deixa

a desejar, uma vez que não costuma mencionar em sua literatura quais instrumentos utiliza para a extração de termos do *corpus*.

Apesar de terem propósitos diferentes (a LC é voltada para estudos da linguagem e a GL para estudos da OC), essas metodologias podem se complementar como métodos de análise e em questões relacionadas a justificativas para validação de uso dos termos e de conceitos, seja em glossários e dicionários, seja em sistemas de organização da informação. Entende-se que explicitar a aplicação da LC e da GL nas metodologias de pesquisa possa vir a contribuir para uma organização mais consistente, baseada em justificativas mais claras e sólidas, propiciando a reprodutibilidade da pesquisa.

CONSIDERAÇÕES FINAIS

A pesquisa mostrou que o encontro entre a LC e a GL é bastante profícuo para o aperfeiçoamento do trabalho terminológico para a elaboração de tesouros e vocabulários controlados. Dentro da OC, a LC pode contribuir com suas práticas metodológicas e ferramentas de gerenciamento de *corpus*, tornando a pesquisa mais ágil e aumentando a precisão dos dados; enquanto a GL se mantém como uma teoria importante de justificação e delimitação da representatividade de um termo na pesquisa baseada em *corpus*.

REFERÊNCIAS

BARITÉ, M. *et al.* Garantia literária: elementos para uma revisão crítica após um século. *TransInformação*, Campinas, v. 22, n. 2, p. 123-138, maio/ago. 2010.

BARITÉ, M. *et al.* Hacia un tesoro de Organización del Conocimiento. In: JORNADAS DE INVESTIGACIÓN DE LA FACULTAD DE INFORMACIÓN Y COMUNICACIÓN, 2017, Montevideo. *Anales...* Montevideo: Facultad de Información y Comunicación, 2017.

BARITÉ, M. La garantía literaria: vigencia y proyección teórico-metodológica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. *Anais...* Salvador: Ancib, 2007.

_____. Literary warrant. In: HJØRLAND, B.; GNOLI, C. (Ed.). *ISKO Encyclopedia of Knowledge Organization*. 2018. Disponível em: http://www.isko.org/cyclo/literary_warrant. Acesso em: 21 abr. 2020.

_____. Towards a general conception of warrants: first notes. *Knowledge Organization*, v. 46, n. 8, 647-655, 2019.

BASE DE DADOS REFERENCIAL DE ARTIGOS DE PERIÓDICOS EM CIÊNCIA DA INFORMAÇÃO - BRAPCI. Disponível em: <http://www.brapci.inf.br>. Acesso em: 15 out. 2020.

BEGHTOL, C. Domain Analysis, Literary Warrant, and Consensus: the case of fiction studies. *Journal of the American Society for Information Science*, v. 46, n. 1, p. 30-44, 1995.

_____. Semantic validity: concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, v. 30, n. 2, p. 109-125, Apr.-June 1986.

BERBER SARDINHA, A. B. Lingüística de Corpus: histórico e problemática. *Delta*, v. 16, n. 2, p. 323-367, 2000.

_____. *Linguística de corpus*. Barueri, SP: Manole, 2004.

MCENERY, T.; WILSON, A. Early corpus linguistics and the Chomskyan revolution. In: *Corpus linguistics: an introduction*. 2. ed. Edinburgh University Press, 2001. Cap. 1. p. 1-27.

WEB OF SCIENCE. *Web of Science: confident research begins here*. Disponível em: <https://clarivate.com/webofsciencegroup/solutions/web-of-science>. Acesso em: 15 out. 2020.