

## Teoria Dialógica do Discurso (TDD) e pesquisa com grandes corpora: processo de composição de discursos sobre a divulgação científica

### Dialogical discourse theory and large corpora research: the compositional process of discourses about science communication

Giselle Liana FETTER<sup>1</sup>

**RESUMO:** As pesquisas com extensos *corpora* baseadas na Teoria Dialógica do Discurso (TDD), que segue os pressupostos do Círculo de Bakhtin, não são comumente realizadas, porém, tais estudos são possíveis se o pesquisador tiver um objeto e um foco específico. Assim, este artigo objetiva apresentar o processo de composição de um extenso *corpus* para a análise com base na TDD. O objeto de estudo é o discurso de professores-pesquisadores de universidades brasileiras sobre a divulgação científica, tema ainda escasso de estudos. Para a estruturação do *corpus*, foram coletados, a partir do termo *divulgação científica* 226 artigos científicos sobre a divulgação científica, publicados entre 2013 e 2018, na ferramenta Google Acadêmico. Como critérios de recorte, delimitou-se que a autoria fosse de professores-pesquisadores de pós-graduação *stricto sensu* de universidades brasileiras, vinculados a sete grandes áreas do conhecimento da CAPES, o que resultou em 114 artigos. Observou-se que a aplicação de procedimentos metodológicos fundamentados nos preceitos da Linguística de *Corpus* (LC) e o uso de uma ferramenta de análise de *corpus* – *Wordsmith Tools* – auxiliou na delimitação do objeto de estudo e na organização dos discursos.

**PALAVRAS-CHAVE:** Teoria dialógica do discurso. Linguística de corpus. Corpora. Divulgação científica.

**ABSTRACT:** Research with extensive *corpora* based on Dialogic Discourse Theory (DDT), which follows the precepts of Bakhtin Circle, are not commonly implemented, but these studies are possible if the researcher has a specific object and focus. Therefore, this paper aims to present the composition process of an extensive *corpus* for the analysis based on DDT. The object of study is the discourse of professors from Brazilian universities about science communication, a topic that is still scarcely studied. The *corpus* was structured with 226 scientific papers about science communication, published between 2013 and 2018, in Google Scholar, by the search of the term *divulgação científica*. As for the delimitation criteria, it was determined that the authorship of the papers belonged to professors of postgraduate programs, belonging to seven CAPES' major areas of knowledge, which resulted in 114 papers. It was observed that the application of methodological procedures based on *Corpus Linguistics* (CL) and the use of a *corpus* analysis program – *Wordsmith Tools* – collaborated in the delimitation of the object of study and in the organization of the discourses.

**KEYWORDS:** Dialogical discourse theory. Corpus Linguistics. Corpora. Science communication.

---

<sup>1</sup> Doutora em Letras/Linguística da Escola de Humanidades da Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil. E-mail: gisellerevisora@gmail.com. Orcid: <https://orcid.org/0000-0002-5577-2214>.



## Introdução

A Teoria Dialógica do Discurso (doravante TDD) se alicerça na filosofia da linguagem dos russos Mikhail Bakhtin (1895-1975), Valentin Volóchinov (1895-1936) e Pável Medviédev (1891-1938), integrantes do grupo que veio a ser nomeado, por estudiosos de seus trabalhos, de Círculo de Bakhtin (FARACO, 2009). A TDD constitui-se como um aporte teórico-metodológico em vista de sua proposta de análise sociológica da linguagem, que nos evidencia tanto um caminho para a compreensão do discurso, visto como um processo de interação entre os indivíduos, quanto uma dimensão metodológica que situa o discurso em sua existência concreta.

As orientações metodológicas, propostas por Volóchinov (2017), nos condicionam à análise do objeto – a linguagem – em seu ambiente real, que é a interação social. A linguagem é viva justamente por estar nesse ambiente, no qual e pelo qual os sentidos se desenvolvem, se articulam e se relacionam dialogicamente. Sendo assim, ao iniciarmos uma pesquisa científica, “não se pode construir uma definição, apenas indicações metodológicas: é preciso, antes de mais nada, apalpar o objeto real da pesquisa, destacá-lo da realidade circundante e apontar previamente seus limites” (VOLÓCHINOV, 2017, p. 143).

Nas pesquisas em linguagem, a Linguística de *Corpus* (doravante LC), por ser uma área de estudos que possibilita a análise a partir de textos reais ou, nas palavras de Berber Sardinha (2000, p. 325), “através de evidências empíricas”, tem contribuído para diferentes pesquisas linguísticas. Nesse sentido, especialmente no caso da TDD, que, como mencionamos, mantém seu enfoque na interação social, ou seja, em contextos reais, a LC oferece ferramentas e subsídios para a delimitação dos limites da pesquisa, permitindo a organização e a exploração de conjuntos de textos, de pequenos, médios e grandes *corpora*, visto que estes não são, conforme explica Rastier (2019), um mero agrupamento de textos, mas se originam de critérios pré-estabelecidos pelos pesquisadores.

Ainda que sejam poucas as pesquisas fundamentadas em TDD que utilizem extensos *corpora*, Brait (2004, p. 191) esclarece que as análises sob essa perspectiva linguística “advêm necessariamente de ‘arquivos’, de *corpus*, de conjuntos de textos, e não tem como meta a análise de um texto, de um trecho ou de uma sequência (embora possa fazê-lo)”. Além disso, a autora explica que, ao analisarmos as partes de um texto, estamos observando aspectos constitutivos de um todo; a teoria está

[...] a serviço da reflexão sobre a linguagem, sobre os discursos, sobre o homem e seu estar no mundo, e nunca em função do esquiteamento anatômico de um *corpus*, conforme as normas de um manual de instruções (BRAIT, 2004, p. 191).

Bakhtin (2018) desenvolveu sua tese acerca da polifonia com base no estudo das obras do autor russo Fiódor Dostoiévski (1821-1881). Segundo Bakhtin, a “multiplicidade de vozes e consciências independentes e imiscíveis e a autêntica polifonia de vozes plenevalentes constituem, de fato, a peculiaridade fundamental dos romances de Dostoiévski” (BAKHTIN, 2018, p. 4). Ao afirmar que Dostoiévski “é o criador do romance polifônico”, Bakhtin (2018, p. 5) tem como objeto dessa afirmação um *corpus* composto pelas obras desse autor russo, além de trabalhos do século XX referentes a Dostoiévski e “[...] ao problema de sua *poética*”. Desse modo, e conforme explicitado por Brait (2004), Bakhtin analisa as partes em função do todo. Embora a palavra *corpus* não tenha sido empregada por Bakhtin (2018) em seus estudos, como vemos, o autor parte de um conjunto de textos, de um conjunto de romances delimitados.

Para além dessa contextualização introdutória a respeito dos estudos com *corpora* e da TDD, cabe-nos mencionar que o processo de construção de um extenso *corpus* foi motivado por nossa pesquisa de doutorado, intitulada *A divulgação científica como arena discursiva nas universidades brasileiras: (des)encontro de vozes nos dizeres de professores-pesquisadores* (FETTER, 2022a), por objetivo geral investigar as concepções de divulgação científica no discurso dos professores-pesquisadores das universidades brasileiras, a fim de contribuir para o aprimoramento da divulgação da ciência do Brasil. Assim, julgamos relevante expor, de maneira sucinta, as motivações para a análise dos discursos sobre a divulgação de conhecimentos científicos, objeto de estudo dessa investigação.

A divulgação científica é um tema que instiga pesquisadores de diversos campos de estudos. Especialmente na área da Linguística, há trabalhos de diferentes níveis acadêmicos – graduação, mestrado e doutorado – que investigam a relação entre ciência e sociedade sob o enfoque das mais variadas teorias da linguagem. Contudo, percebemos uma lacuna de estudos dessa área que se concentrem nas percepções, nos pontos de vista, nas valorações atribuídas à divulgação científica. Considerando que a elaboração dessa pesquisa de doutorado se situa no contexto da pandemia de covid-19, percebemos um destaque e um aumento significativo de cientistas e pesquisadores bem como de instituições científicas e universidades brasileiras envolvidas em divulgar e propagar a ciência. Assim, elegemos esse objeto de estudo com vistas a contribuir para o aprimoramento da divulgação científica do Brasil, que a nosso ver, apresenta algumas deficiências e insuficiências.

Para que possamos atingir nosso objetivo de apresentar o processo de composição de um extenso *corpus* para a análise com base na TDD, o presente artigo organiza-se em quatro seções, além da introdução. Na primeira seção, que segue, abordaremos a relação dos estudos baseados em extensos *corpora* com a análise do discurso e a TDD. Em seguida, na segunda seção, exporemos as escolhas metodológicas realizadas para organização do *corpus*. Na terceira seção, apresentaremos o uso da ferramenta *Concord* do programa *Wordsmith Tools* na análise do discurso dos professores-pesquisadores de universidades brasileiras sobre a divulgação científica, indicando algumas possibilidades de sua aplicação e como os preceitos da LC nos auxiliaram na estruturação de nosso *corpus* de 114 artigos. Por fim, na quarta e última seção, traremos algumas considerações acerca da pesquisa realizada a partir de um extenso *corpus*.

## Estudos baseados em *corpora* e sua relação com a análise do discurso e a TDD

Um *corpus* é definido, segundo Rastier (2019, p. 32, grifos do autor), como “*um reagrupamento estruturado de textos integrais, documentados, eventualmente enriquecidos por etiquetas e reunidos*”. Dessa compreensão, mais precisamente, o autor explica que um *corpus* se estrutura a partir de uma pesquisa determinada, ou seja, não se trata de um emaranhado de textos, mas de textos organizados com base em critérios e princípios metodológicos. Além disso, é importante salientar que um *corpus* se constitui apenas de textos reais, de evidências empíricas dos usos de uma língua.

A disponibilidade de grandes *corpora* e a possibilidade de trabalhar com dados linguísticos de vasta dimensão deram origem, segundo Leon (2006, p. 52), nos anos 1990, a um avanço nos estudos em Linguística, suscitando também a “[...] renovação do interesse pelos tratamentos estatísticos e probabilísticos que, mais ou menos diretamente, questionam a Linguística sobre seu objeto, seus métodos e seus fundamentos”. Na opinião da autora, a LC se tornou uma área dominante nos estudos linguísticos. De mesmo modo, Berber Sardinha (2000) explica que a LC influencia fortemente as pesquisas na Linguística.



No Brasil, segundo Novodvorski e Finatto (2014, p. 8-9), os estudos em *corpora* já possuem certa trajetória, inclusive “[...] alcança áreas que poderiam parecer, num primeiro momento, incompatíveis ou inimagináveis”. Dentre essas áreas, podemos incluir a Análise do Discurso, que, conforme Baker (2006), aparentemente não possui muitos pontos em comum com a LC, já que esta costuma ser empregada, de maneira mais expressiva, em estudos quantitativos, e não qualitativos, como é o caso da TDD. Por essa razão, de acordo com o autor, há pesquisadores da área do discurso que se opõem ou não se interessam em unir esses campos linguísticos.

Oliveira (2009) comenta que há certo equívoco por parte de pesquisadores tanto de *corpus* quanto de outras vertentes teóricas, por estes julgarem que a LC se restringe apenas a pesquisas estatísticas e por alguns daqueles realizarem, justamente, pesquisas voltadas apenas ao levantamento estatístico e não considerarem a importância do contexto sócio-histórico que circunda os textos. A autora não desqualifica os dados numéricos proporcionados pelos estudos com *corpora*, mas enfatiza que esses dados devem ser analisados e compreendidos sob a luz de posicionamentos teórico-metodológicos e serem “interpretados pelo pesquisador para que possam servir de apoio para novas descrições linguísticas ou para a proposta de novas perspectivas teóricas” (OLIVEIRA, 2009, p. 51).

De acordo com Novodvorski e Finatto (2014, p. 7), a LC caracteriza-se como uma área que oferece métodos de análise de textos bem como serve de pilar teórico “diferenciado dos Estudos da Linguagem”. Dessa forma, os pesquisadores podem aplicá-la, conforme apresentaremos neste artigo, na realização de procedimentos metodológicos, sobre os quais “[...] nada será cobrado em termos de uma filiação teórica – ou epistemológica” (NOVODVORSKI; FINATTO, 2014, p. 8), ainda que os autores defendam que a LC também é uma maneira de compreender a linguagem.

Para Baker (2006), há três problemas relacionados à LC nos estudos do discurso. Primeiramente, está o desconhecimento ou a rejeição pelo uso de computadores para a análise de *corpus*. Em segundo lugar, encontra-se a percepção de que trabalhar com *corpus* é realizar análises estritamente quantitativas. E terceiro, há o fato de as análises de discurso demandarem um excessivo trabalho e, conseqüentemente, serem mais complexas para seus analistas.

Avaliando cada um desses aspectos contrários à LC, no que tange à TDD – foco deste artigo –, podemos afirmar que a utilização de programas ou ferramentas de tratamento linguístico possam gerar algum receio, principalmente por que muitos pesquisadores precisam dominar o funcionamento e, até mesmo, a programação desses recursos informáticos. Sobre os levantamentos puramente estatísticos, na quarta seção deste artigo, apresentaremos o uso de uma ferramenta computacional que não teve a intenção de realizar uma investigação estatística, mas de facilitar e auxiliar na organização do *corpus* de análise. Quanto a complexidade do trabalho do analista, compreendemos que a TDD estuda a interação social a partir de diversos fatores, que tais ferramentas não podem identificar. Concordamos, como analistas do discurso, que estudar a linguagem a partir dos preceitos da teoria em pauta exige dedicação e tempo de pesquisa, porém entendemos que a LC pode colaborar nesse sentido, embora enfatizemos que esta não deve ser aplicada em detrimento da TDD.

Nesse sentido, para o pesquisador da TDD, é possível realizar uma análise apoiado no uso de um extenso *corpus*, uma vez que há, nessa teoria, fundamentos teóricos e metodológicos para determinar alguns recortes que possibilitem o foco no objeto de estudo. Tal argumento tem embasamento nos estudos do Círculo de Bakhtin.

Conforme comentamos na introdução deste artigo, Bakhtin (2018) elucida a sua tese acerca do romance polifônico a partir das obras de Dostoiévski e destaca trechos que fundamentem as suas proposições. Em seu ensaio, *Lendo Razlúka de Púchkin: a voz do outro*

na poesia lírica, Bakhtin (2021) igualmente evidencia alguns excertos da obra de Aleksandr Púchkin (1799-1837) e afirma que a sua “[...] análise apenas vai destacar aqueles momentos que nós aqui necessitamos, e vai esquecer todo o resto, por vezes mesmo que seja muito significativo [...]”. Desse modo, vemos a aplicabilidade da TDD juntamente aos estudos com extensos *corpora*, já que a teoria permite que o pesquisador faça escolhas metodológicas para a constituição e a apresentação de sua análise.

Muitas pesquisas em análise do discurso, de modo geral, ainda realizam coleta e seleção de *corpus* de maneira manual. E, de acordo com Sargentini (2005, p. 5), esse tipo de pesquisa ocasiona, para o analista, a “[...] dificuldade de se trabalhar com uma quantidade extensa de textos a serem analisados”. Segundo a autora, um grande número dessas investigações, no Brasil, pode ter sofrido certas limitações em relação ao material a ser analisado “[...] devido à falta de banco de dados e à vulnerabilidade em relação ao levantamento e seleção do *corpus*” (SARGENTINI, 2005, p. 5). A autora também defende que os resultados das análises de discursos não devem se concentrar apenas em ferramentas computacionais, mas aplicá-las como uma forma de apoio.

Dessa forma, na seção seguinte, apresentaremos as escolhas metodológicas realizadas para a delimitação do *corpus* de discursos sobre a divulgação científica e, na sequência, na terceira seção, discorreremos sobre alguns procedimentos executados por meio de uma ferramenta de análise de *corpus* – *Wordsmith Tools*.

### **Coleta e escolhas metodológicas para a estruturação do corpus de discursos sobre a divulgação científica**

A proposta de construção de um extenso *corpus* de discursos sobre a divulgação científica, discutidos nesta seção, foram motivados, como mencionamos, por nossa pesquisa de doutorado (FETTER, 2022a). Contudo, para essa pesquisa, foram aplicados 56 artigos, uma vez que realizamos um recorte temporal (2016 a 2018) menos abrangente. Assim, nesta seção, discorreremos sobre o processo de composição do *corpus* que nos possibilitou reunir 114 artigos científicos publicados entre os anos de 2013 e 2018.

Para realizarmos essa investigação, definimos que o *corpus* seria estruturado com artigos científicos de professores-pesquisadores de universidades brasileiras. Resolvemos analisar os discursos de professores-pesquisadores desse nível acadêmico, pois há maior exigência, nos programas de pós-graduação, para publicar artigos científicos (ANDRÉ; PINTO; LUCAS, 2017). Há também a questão orientação acadêmica de alunos de mestrado e doutorado, ou seja, de possíveis futuros professores-pesquisadores, que nos permite compreender que os discursos dos docentes de pós-graduação evidenciam, de maneira mais expressiva, as concepções de divulgação científica que circulam entre os grupos sociais nos quais se inserem.

Além do critério referente ao vínculo dos professores-pesquisadores à pós-graduação, definimos que os artigos deveriam estar enquadrados de acordo com a *Tabela de áreas do conhecimento da CAPES*, na qual constam 9 (nove) grandes áreas. Como critério de recorte dessas áreas, excluímos artigos científicos vinculados à grande área Linguística, Letras e Artes e à área de avaliação Comunicação, pertencente à grande área Ciências Sociais Aplicadas. Tal decisão se deve ao fato de identificarmos, em nossas leituras de trabalhos sobre a divulgação científica, que essas áreas tendem a valorizar a temática em questão, e, portanto, poderiam influenciar os resultados de nossa investigação. Além disso, não foram encontrados artigos da grande área Ciências Agrárias, totalizando 7 (sete), a citar: Ciências Exatas e da Terra;

Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Sociais Aplicadas; Ciências Humanas; e Multidisciplinar.

Desse modo, coletamos, por meio da busca de **divulgação científica**, na ferramenta *Google Acadêmico*, trabalhos publicados entre os anos de 2013 e 2018. Essa ferramenta, conforme explicam López-Cózar, Orduna-Malea e Martín-Martín (2019), permite rastrear todas as publicações acadêmicas disponíveis em *websites* como livros, capítulos de livros, artigos de periódicos e de eventos, teses, dissertações, entre outras, indexando-as independentemente de sua qualidade e da forma de avaliação.

Realizar esse tipo de coleta demanda do pesquisador algum tempo de dedicação, dependendo do tamanho do *corpus* ou do período delineado para o estudo. Conforme Leon (2006), as pesquisas em LC possuem alguns objetivos, e colocar *corpora* à disposição é um deles. Assim, esse *corpus* estruturado com artigos científicos estará disponível em Fetter (2022a). Trata-se de um aspecto relevante para nosso propósito com o presente artigo, pois as investigações em TDD, embora costumem aplicar análises em pequenos *corpora*, podem se beneficiar de material coletado previamente para outros estudos.

Nossa coleta na ferramenta *Google Acadêmico*, previamente à aplicação dos critérios ainda a serem descritos nesta seção, reunimos um *corpus* de 226 artigos científicos, recuperados pela busca de **divulgação científica**<sup>2</sup>. Anteriormente à coleta no *Google Acadêmico*, decidimos ter como foco os artigos que continham as palavras-chave **divulgação científica** e **divulgação da ciência**, pois, nos referenciais teóricos sobre o assunto, observamos que há uma tendência ao uso desse termo de maneira generalizada, ou seja, ao longo das publicações, mesmo que os autores escolham outros termos – popularização da ciência, jornalismo científico, comunicação da ciência, percepção pública da ciência, entre outros –, o termo **divulgação científica** é mencionado em algum trecho da publicação ou empregado como palavra-chave para agregá-la a outros trabalhos da categoria. Frente a essa limitação da pesquisa no *Google Acadêmico*, os resultados recuperados podem apresentar documentos que contenham uma única ocorrência de certo termo, palavra ou expressão, que, muitas vezes, não tem nenhuma relação com os trabalhos acadêmicos que o usuário deseja consultar. Embora compreendamos que existam limitações no uso do *Google Acadêmico*, como relatam López-Cózar, Orduna-Malea e Martín-Martín (2019), julgamos que essa ferramenta era a mais acessível em termos de busca por artigos científicos em vista das vantagens já anteriormente exploradas.

Nesse sentido, outro objetivo da LC, apresentado por Leon (2006), é a exploração de fatos linguísticos nos *corpora*, ou seja, realizar o levantamento de palavras ou sequências de palavras em um *corpus*. Ao percebermos a limitação do *Google Acadêmico* em termos de dimensão da busca por **divulgação científica** como palavra-chave dos trabalhos científicos, decidimos utilizar uma ferramenta computacional.

É importante salientar, primeiramente, que as investigações fundamentadas na TDD não partem da análise isolada de palavras ou sequências de palavras dentro de um gênero discursivo. Como explicamos na seção sobre essa teoria, o conceito de linguagem transcende o código linguístico, pois envolve relações e contextos culturais, sociais e históricos que não podem ser percebidos pela LC tampouco identificados por ferramentas automáticas de tratamento da linguagem.

Ao nos propormos à apresentação das contribuições dos estudos com *corpora* para a TDD, pretendemos pontuar os subsídios daquela área de pesquisa para esta teoria, no sentido

---

<sup>2</sup> A coleta foi realizada durante o mês de janeiro de 2019.



de proporcionar ao analista do discurso sugestões, orientações e, inclusive, maneiras possíveis de abreviar certos procedimentos metodológicos que teriam maior duração caso fossem realizados de forma manual.

Dito isso, ao definirmos que os artigos científicos tivessem como palavra-chave **divulgação científica**, foi necessário, dentre os 226 artigos coletados, realizar a etapa de recorte a partir desse critério. Para esse procedimento, utilizamos a ferramenta *Concord* do programa *Wordsmith Tools*, a ser apresentada na seção seguinte.

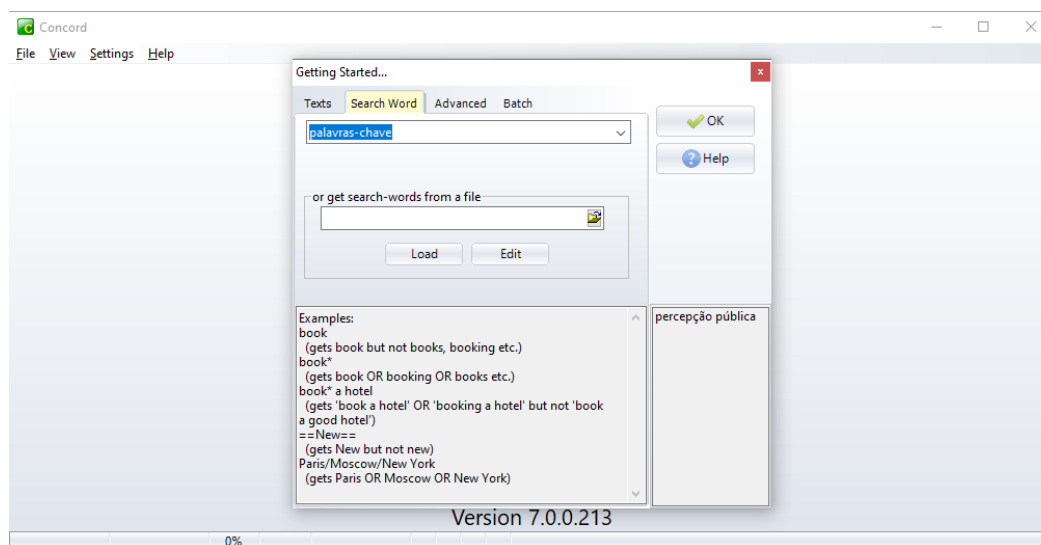
## Organização do corpus com o auxílio da ferramenta *Wordsmith Tools*

O programa *WordSmith Tools* foi idealizado e desenvolvido por Mike Scott nos anos de 1990, e sua primeira versão foi lançada em 1996 (WILKINSON, 2011). O programa já está em sua versão 8.0, lançada em 2020, porém, para esta pesquisa, adotamos a versão 7.0, por ser a mais atual no momento da investigação. O *WordSmith Tools* conta com diversos recursos para a análise de textos. Como explica Berber Sardinha (1999, p. 2), tais recursos são “úteis e poderosos”, pois fornecem subsídios para “a composição lexical, a temática de textos selecionados, e a organização retórica e composicional de gêneros discursivos”. O programa oferece três ferramentas à disposição do analista: *WordList*, *KeyWords* e *Concord*.

Interessa-nos, para esta pesquisa, a ferramenta *Concord*, que serve para a produção de listas de ocorrências desejadas pelo analista. Berber Sardinha (1999) explica que o tipo de concordância mais comum de uso é a *Key Word in Context* (KWIC), que consiste em uma busca em que a palavra desejada aparece centralizada, antecedida e seguida pelas partes do texto do *corpus* de estudo, mais especificamente, podemos visualizar até cinco palavras à direita e à esquerda. No caso desta análise, essa ferramenta possibilitou a busca por palavras-chave bem como a identificação dos termos mais comuns empregados para se referir à relação entre ciência e sociedade, isto é, à divulgação científica, como explicaremos mais adiante.

Devido a essa possibilidade de centralizar uma palavra e, ao mesmo tempo, poder observar as palavras próximas, a ferramenta *Concord* auxiliou na verificação do critério de recorte do *corpus* por palavra-chave. Na captura de tela 1, apresentamos o recurso KWIC:

**Figura 1:** Uso do recurso *Search Words* do KWIC

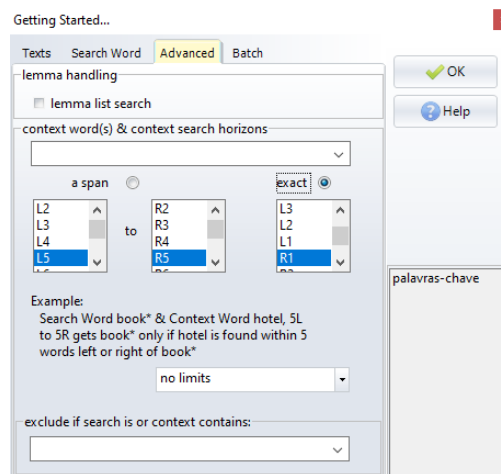


**Fonte:** Wordsmith Tools, versão 7.0.



Podemos observar na captura de tela 1, na aba *Search Word*, o campo destacado na cor azul, que indica a palavra escolhida para busca. Escolhemos procurar por **palavras-chave** – no plural –, pois é dessa forma que os artigos fazem referência aos termos de indexação. Além disso, optamos pela busca pelo termo **palavras-chave**, pois se pesquisássemos pelo termo **divulgação científica**, a ferramenta apresentaria a sua ocorrência em qualquer lugar do artigo, e nosso enfoque era apenas olhar, primeiramente, a indexação. Na aba *Advanced*, mostrada na captura de tela 2, os campos destacados na cor azul correspondem a seleção de palavras que o pesquisador deseja que estejam à direita ou à esquerda da palavra a ser procurada:

**Figura 2:** Uso do recurso *Context Word(s)* do KWIC



Fonte: WordSmith Tools, versão 7.0.

Após essas etapas, o pesquisador clica em OK, e uma nova tela aparece com os resultados da busca. Nesta próxima janela, em captura de tela 3, é possível observar **palavras-chave** centralizado entre as palavras da direita e da esquerda:

**Figura 3:** Resultados de busca na ferramenta *Concord*

N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	H...	H...	Secl	Secl	File	Da
		#		#	Pos.	#	Pos.	#	Pos.	#	Pos.	#		
1	para uma educação CTS nos textos analisados. PALAVRAS-CHAVE: divulgação científica. CTS.	245	5	2	0	244			0	244			2018-CONTRIBUIÇÕES	2021/Jul/29 0
2	palavras associadas. Quadro 01 – Quadro com as palavras-chave e sua respectiva frequência textual	4,315	198	8	01	4			01	4			2018-PLANTAS	2021/Jul/29 0
3	discutir isso com os estudantes em sala de aula. Palavras-chave: Divulgação científica. Livro didático	179	5	2	0	178			0	178			2018-O caráter híbrido	2021/Jul/29 0
4	da escolarização na tomada de decisões. Palavras-chave: Divulgação científica. Produção de	143	8	2	0	142			0	142			2018-O QUE PENSAM	2021/Jul/29 0
5	a interação com diferentes contextos escolares. Palavras-chave: Extensão Universitária;	475	41	2	0	474			0	474			2018-OLHARES SOBRE A	2021/Jul/29 0
6	de Lev Vigotski, Georges Snyders e Paulo Feire. Palavras-chave: Rock e Ciência; Educação	232	8	2	0	231			0	231			2018-RITA DIVULGAÇÃO	2021/Jul/29 0
7	campo da divulgação jornalística-científica. PALAVRAS-CHAVE: João Ribeiro; Visões de	162	5	2	0	161			0	161			2018-A VISÃO DE	2021/Jul/29 0

Fonte: WordSmith Tools, versão 7.0.

Na captura de tela 3, em destaque na cor azul, está **palavras-chave**. Nessa tela, podemos verificar a presença de palavras à direita que correspondem às palavras-chave dos artigos. Para que a visualização dos resultados estivesse disposta desse modo, foi necessário que cada artigo correspondesse a um arquivo, embora o *WordSmith Tools* possibilite e suporte arquivos mais extensos. Organizamos os artigos em arquivos separados para que pudéssemos identificar quais deles continham a palavra-chave **divulgação científica** e quais



deveriam ser desconsiderados do *corpus* por não atenderem a esse critério. Na captura de tela 3, também podemos observar, em destaque na cor vermelha, a aba correspondente a *file* (arquivo, em inglês), na qual consta o nome do arquivo (a ser designado pelo pesquisador). No caso desta pesquisa, escolhemos nomear os arquivos pelo ano de publicação e título do artigo científico.

Esse procedimento metodológico foi viabilizado pela ferramenta *Concord*, pois, ao organizarmos um *corpus* prévio de 226 artigos, a busca manual pelas palavras-chave se torna, indubitavelmente, exaustiva e prolongada. Após finalizadas essas etapas, reunimos um *corpus* de 114 artigos científicos, como discutiremos adiante.

Conforme Baker (2006, p. 13, tradução nossa), as análises de discurso podem ser favorecidas a partir do uso desse tipo de ferramenta, principalmente, como uma etapa inicial de pesquisa: “[...] uma única palavra, frase ou construção gramatical sozinhas podem sugerir a existência de um discurso. Mas ao invés de confiarmos na nossa intuição (e possíveis vieses), às vezes pode ser difícil dizer se tal discurso é típico ou não [...]”<sup>3</sup>. Nossa proposta de expor as contribuições da LC para a TDD não é enquadrar todos os procedimentos metodológicos de uma análise dialógica do discurso em etapas automatizadas. Assim como sugere Baker (2006), o uso das ferramentas computacionais para tratamento da linguagem, como preconizamos neste artigo, favorece a coleta, a preparação, a composição do *corpus* nas fases preliminares ao estudo do discurso realizado pelo analista.

Essa afirmação de Baker (2006) também nos ajuda a esclarecer outro aspecto de nossa análise relacionado à tipicidade dos discursos. Ao definir os critérios de investigação do nosso *corpus*, decidimos por verificar a terminologia empregada pelos professores-pesquisadores em referência à divulgação científica. É importante frisar, para a apresentação desse procedimento, que, embora o termo **divulgação científica** seja o mais usado conforme estudo realizado por Rocha e Massarani (2017), há outros termos para se referir à relação entre ciência e sociedade/público não especializado. Para realizar a busca por esses termos, adotamos a premissa de Caribé (2015) sobre a terminologia mais comum, como por exemplo, **popularização científica**, **cultura científica**, **jornalismo científico**, entre outros.

Na captura de tela 4, que corresponde aos artigos publicados em 2018, ilustramos a busca de um dos termos relacionados à divulgação científica. Como é possível observar, o termo **popularização científica/da ciência** aparece em vários artigos:

**Figura 4:** Resultados de busca pelo termo *popularização científica/da ciência*

N	Concordance	Set	Tag	Word #	Sent #	Para #	H...	H...	Sect #	Pos
1	UFRJ, 2002. MOREIRA, I de C. A Inclusão social e a popularização da ciência e tecnologia no Brasil.			6,850	307	7	03...	9	03...	2018
2	: 18 out 2017." GERMANO, M. G.; KULESZA, W. A. Popularização da Ciência: uma revisão con-			2,621	103	2	02...	0	02...	2018
3	, v. 7, n. 2, p. 25-153, 2001. HAMBURGER, E. W. A popularização da ciência no Brasil. In: CRESTANA,			6,631	261	3	03...	0	03...	2018
4	culturais apresentam-se como possibilidades de popularização da ciência nas escolas. O projeto			708	30	21	0	707	0	707
5	, justifica-se a importância de promover ações de popularização da ciência, divulgação científica,			788	34	48	0	787	0	787
6	e a participação em atividades públicas de popularização da ciência (Feiras e Olimpíadas			1,284	39	47	01...	3	01...	2018
7	et al., 2008)." Entretanto, esta compreensão ou popularização científica requer uma linguagem			1,450	56	7	01...	9	01...	2018

Fonte: Wordsmith Tools, versão 7.0.

<sup>3</sup> No original: “[...] a single word, phrase or grammatical construction on its own may suggest the existence of a discourse. But other than relying on our intuition (and existing biases), it can sometimes be difficult to tell whether such discourse is typical or not [...]”.



Se considerarmos que as ocorrências desses termos aparecem em artigos de 10 a 20 páginas, uma posterior seleção do *corpus* poderia descartá-las da análise caso dependêssemos apenas de nosso olhar de pesquisador. Tais procedimentos metodológicos proporcionam um recorte mais objetivo do *corpus*.

O estudo com *corpus* apresenta algumas vantagens como a coleta de materiais – textos – autênticos; a exploração e a descrição de gêneros do discurso; a análise de padrões discursivos, entre outras. Porém, a organização de um *corpus* depende, essencialmente, dos objetivos de pesquisa, que serão empregados para nortear os procedimentos metodológicos. Conforme Rastier (2019), os *corpora* não se configuram como um mero agrupamento de textos, visto que a estruturação de um *corpus* se origina com base em critérios pré-estabelecidos pelos pesquisadores. O *corpus* utilizado, efetivamente, para a pesquisa de doutorado contém 56 artigos, pois definimos que o recorte temporal seria entre os anos de 2016 e 2018. Desse modo, com base nos pressupostos de Rastier (2019), compomos para a pesquisa de doutorado um *corpus de referência* – 56 artigos científicos – e um *corpus de estudo* – 34 artigos científicos. Esse *corpus* menor corresponde àqueles artigos que contribuíram para o cumprimento de nossos objetivos de pesquisa.

Ao nos colocarmos nesse espaço investigativo, novamente, verificamos a aplicabilidade teórico-metodológica dos pressupostos do Círculo de Bakhtin, que nos permite fazer escolhas para a constituição desse *corpus de estudo*, de maneira que este contribua para a condução da pesquisa à qual nos propormos e que, por sua vez, materialize os sentidos em torno das concepções de divulgação científica.

As etapas que seguiram após esta concentraram-se na análise dialógica dos discursos sobre a divulgação científica. Sendo assim, não foram descritas aqui por estarem fundamentadas exclusivamente na TDD.

## Considerações finais

A proposta deste artigo era apresentar o processo de composição de um extenso *corpus* para a análise com base na TDD a partir da análise do discurso de professores-pesquisadores das universidades brasileiras sobre a divulgação científica. Para discorrermos a respeito das contribuições, apresentamos alguns procedimentos metodológicos adotados na investigação desses discursos com o uso da ferramenta *Concord* do programa *Wordsmith Tools*.

Tendo como enfoque esses procedimentos, a interface entre a LC e a TDD se mostrou produtiva, se levarmos em consideração o tamanho de nosso *corpus*. A análise de médios e grandes *corpora* demanda tempo do analista e pode, por vezes, tornar-se exaustiva quando não são utilizadas ferramentas computacionais. Assim, para organizarmos um *corpus* prévio de 226 artigos científicos, ao qual definimos, primeiramente, como critério, que contivessem a palavra-chave **divulgação científica**, podemos afirmar que o programa *Wordsmith Tools*, em particular, beneficiou o desenvolvimento da estruturação do *corpus* de estudo. Além disso, o uso dessa ferramenta auxiliou na busca por termos relacionados à divulgação científica, complementando a análise que seria realizada a partir de nosso olhar de pesquisadores, que pode, por vezes, não ser igualmente objetivo e exato. Desse modo, apresentamos métodos diferentes a serem empregados pelo analista como possibilidade de investigação com *corpora*.

O objetivo de nossa pesquisa de doutorado foi investigar como os professores-pesquisadores das universidades brasileiras concebem a divulgação científica a fim de contribuir para o seu aprimoramento no Brasil. Desse modo, identificar a terminologia da



divulgação científica nos artigos científicos nos proporcionou uma visão mais abrangente dos enunciados que faziam referência aos termos da área<sup>4</sup>, já que o olhar do pesquisador pode, por vezes, não ser tão acurado quanto às ferramentas de automatização da linguagem e, portanto, alguns enunciados considerados significativos para a análise podem ser, involuntariamente, descartados. De acordo com Baker (2006, p. 13, tradução nossa), a LC e a análise do discurso se complementam, pois as ferramentas auxiliam a observar e reconhecer certos aspectos pontuais a respeito da linguagem enquanto a interpretação do analista se concentraria em “[...] descobrir como a linguagem é empregada, frequentemente de maneira sutil, para revelar discursos subjacentes”<sup>5</sup>.

Cabe-nos ressaltar que o ponto de vista do pesquisador/analista não deve ser substituído ou colocado em segundo plano quando nos fundamentamos na TDD, já que não há ferramentas computacionais que sejam capazes de abranger aspectos da linguagem que ultrapassam o código linguístico. De mesmo modo, enfatizamos que os procedimentos metodológicos a partir da LC se configuram como etapas iniciais para a construção e a organização do *corpus* e, portanto, não determinam os resultados finais da pesquisa, que, novamente, partem da interpretação e visão do analista com base na teoria.

Esperamos que este artigo sirva de inspiração para novas investigações em TDD de maneira a incentivar pesquisadores a desenvolverem análises com extensos *corpora*. Ao refletirmos sobre a multiplicidade de discursos aos quais podemos acessar com o advento da internet, percebemos a possibilidade de que a LC seja aplicada em um maior número de pesquisas, diferentemente do que observamos nos últimos anos.

## Referências

- BAKER, Paul. *Using corpora in discourse analysis*. London: Continuum, 2006.
- BAKHTIN, Mikhail. *Os gêneros do discurso*. Organização, tradução, posfácio e notas de Paulo Bezerra. São Paulo: Editora 34, 2016.
- BAKHTIN, Mikhail. Fragmentos dos anos 1970-1971. In: BAKHTIN, Mikhail. *Notas sobre literatura, cultura e ciências humanas*. Organização, tradução posfácio e notas de Paulo Bezerra. Notas da edição russa de Serguei Botcharov. São Paulo: Editora 34, 2017. p. 21-56.
- BAKHTIN, Mikhail. *Problemas da poética de Dostoiévski*. Tradução, notas e prefácio de Paulo Bezerra. 5. ed. Rio de Janeiro: Forense Universitária, 2018.
- BARROS, Diana. *Contribuições de Bakhtin às teorias do discurso*. In: Brait, Beth. (org.). *Bakhtin, dialogismo e construção do sentido*. São Paulo, Editora da Unicamp, 2007, p. 25-36.
- BERBER SARDINHA, Tony. Linguística de corpus: histórico e problemática. *DELTA*,

---

<sup>4</sup> A análise da terminologia está organizada em Fetter (2022b). Portanto, não será apresentada neste artigo, já que esse não é nosso objetivo.

<sup>5</sup> No original: “[...] to uncover how language is employed, often in [quite subtle ways, to reveal underlying discourses”.



São Paulo, v. 16, n. 2, 2000. Disponível em:  
[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-44502000000200005&lng=pt&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005&lng=pt&tlng=pt). Acesso em: 1 mar. 2016.

BERBER SARDINHA, Tony. Tamanho do corpus. *The Specialist*, v. 23, n. 2, 2002. p. 103-122. Disponível em:  
<https://revistas.pucsp.br/index.php/esp/article/view/9381/6952>. Acesso em: 28 jul. 2021.

BERBER SARDINHA, Tony. *Linguística de Corpus*. Barueri, SP: Manole, 2004.

BRAIT, Beth. Revisitando mitologias pelas lentes dialógicas. *Desenredo*, v. 10, n. 1, p. 9-30, 2014. Disponível em: <http://seer.upf.br/index.php/rd/article/view/4094>. Acesso em: 14 jan. 2020.

FARACO, Carlos Alberto. *Linguagem & diálogo: as ideias linguísticas do Círculo de Bakhtin*. São Paulo: Parábola, 2009.

FETTER, Giselle Liana. *A divulgação científica como arena discursiva nas universidades brasileiras: (des)encontro de vozes nos dizeres de professores-pesquisadores*. 2022. 289f. Tese (Doutorado em Letras/Linguística) – Escola de Humanidades, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2022a. No prelo.

FETTER, Giselle Liana. Variação terminológica nas pesquisas sobre divulgação científica: análise dos termos empregados por professores-pesquisadores das universidades brasileiras. *Trabalhos em Linguística Aplicada*, v. 61, n. 1, 2022b. No prelo.

LEON, Jaqueline. A linguística de corpus: história, problemas, legitimidade. *Filologia e Linguística Portuguesa*, n. 8, p. 51-81, 2006. Disponível em:  
<https://www.revistas.usp.br/flp/article/view/59745>. Acesso em: 20 jun. 2016.

MEDVIÉDEV, P. *O método formal nos estudos literários: introdução crítica a uma poética sociológica*. Tradução de Sheila Camargo Grillo e Ekaterina Vólkova Américo. Prefácio de Sheila Camargo Grillo. São Paulo: Contexto, 2012.

MUSSALIM, Fernanda. Análise do discurso. In: MUSSALIM, Fernanda.; BENTES, Ana Cristina (org.). *Introdução à Linguística: domínios e fronteiras*. São Paulo: Cortez, 2003. v. 2.

NOVODVORSKI, Ariel; FINATTO, Maria José Bocorny. Linguística de Corpus no Brasil: uma aventura mais do que adequada. *Letras & Letras*, v. 30, n. 2, p. 7-16, 18 dez. 2014. Disponível em:  
<http://www.seer.ufu.br/index.php/letraseletras/article/view/28516/15799>. Acesso em: 1 mar. 2016.

OLIVEIRA, Lúcia Pacheco de. Linguística de *corpus*: teoria, interfaces e aplicações. *Matraga*, v. 16, n. 24, p. 48-76, 2009. Disponível em: <https://www.e->



publicacoes.uerj.br/index.php/matraga/article/view/27796/19917. Acesso em: 27 fev. 2022.

RASTIER, François. Da semântica estrutural à semiótica das culturas. *Galaxia*, set./dez. 2019. Disponível em: <https://www.scielo.br/j/gal/a/sQWWSFtprFrxFs8JjN3Ggt/abstract/?lang=pt>. Acesso em: 19 jul. 2021. Edição especial.

SARGENTINI, Vanise. A noção de formação discursiva: uma relação estreita com o corpus na Análise do Discurso. In: Seminário de Estudos em Análise do Discurso, 2, Porto Alegre. *Anais eletrônicos...* Porto Alegre: UFRGS, 2005. p. 1-6. Disponível em: <https://www.ufrgs.br/analisedodiscurso/anaisdosead/2SEAD/SIMPOSIOS/VaniceMariaOliveiraSargentini.pdf>. Acesso em: 16 maio 2016.

VOLÓCHINOV, Valentín. *Marxismo e filosofia da linguagem: problemas fundamentais do método sociológico na ciência da linguagem*. Tradução de Sheila Vieira de Camargo Grillo e Ekaterina Vólkova Américo. 1. ed. São Paulo: Editora 34, 2017.

VOLÓCHINOV, Valentín. *A palavra na vida e a palavra na poesia: ensaios, artigos, resenhas e poemas*. Organização, tradução, ensaio introdutório e notas de Sheila Grillo e Ekaterina Vólkova Américo. São Paulo: Editora 34, 2019.

WILKINSON, M. WordSmith Tools: The best corpus analysis program for translators?. *Translation Journal*, v. 15, n. 3, 2011. Disponível em: <http://translationjournal.net/journal/57corpus.htm>. Acesso em: 20 abr. 2017.

