

Estudos exploratórios

em Linguística de Corpus

Ariel Novodvorski
Joel Victor Reis Lisboa
Raphael Marco Oliveira Carneiro
(organizadores)

2

Estudos exploratórios em Linguística de *Corpus* 2

Ariel Novodvorski
Joel Victor Reis Lisboa
Raphael Marco Oliveira Carneiro
(organizadores)

Estudos exploratórios em Linguística de *Corpus* 2

Araraquara
Letraria
2022

FICHA CATALOGRÁFICA

**Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)**

Estudos exploratórios em linguística de corpus 2 [livro eletrônico] / organização Ariel Novodvorski, Joel Victor Reis Lisboa, Raphael Marco Oliveira Carneiro. - 1. ed. - Araraquara, SP : Letraria, 2022.

PDF.

Bibliografia.

ISBN 978-85-69395-99-7

1. Línguas e linguagem I. Novodvorski, Ariel. II. Lisboa, Joel Victor Reis. III. Carneiro, Raphael Marco Oliveira.

22-111672

CDD-418.4

Índices para catálogo sistemático:

1. Linguística de corpus: Terminologia: Linguística aplicada
410.285

Aline Grazielle Benitez - Bibliotecária - CRB-1/3129

Conselho editorial

Silvelena Cosmo Dias (UFMS/CPTL)

Neubiana S. V. Beilke (UFU)

Vera Lúcia Dias dos Santos Augusto (Faculdade de Caldas Novas)

| Sumário

Estudos Descritivos e Linguística de <i>Corpus</i> : abrindo caminhos para a descrição linguística	8
Ariel Novodvorski, Joel Victor Reis Lisboa e Raphael Marco Oliveira Carneiro	
O suposto antagonismo de <i>vida e morte em corpora</i> : otimismo em Hay e pessimismo em Cioran à luz da Linguística de <i>Corpus</i>	14
Lucas Amâncio Mateus	
Uma nova leitura de <i>Rayuela</i> : análise lexical baseada em <i>corpus</i>	27
Lidiane Carlos Ramos	
O vocabulário distópico em <i>Fahrenheit 451</i> : uma análise pautada na Linguística de <i>Corpus</i>	40
Terezinha de Assis Oliveira	
O nome "Barack" em <i>Minha História</i> (2018) sob a perspectiva sistêmico-funcional	49
Sarah Cristina de Oliveira Sebba	
Análise de orações existenciais em <i>corpus</i> literário bilíngue	60
Victor Mariotto Palma	
Como se traduz "língua de acolhimento"? Análise exploratória em <i>corpus</i> paralelo bilíngue	68
Joel Victor Reis Lisboa	
AVALIATIVIDADE em <i>corpus</i> de comentários: um olhar sobre o feminino	79
Mayra Natanne Alves Marra	
O uso do vocábulo <i>bitch</i> em letras de <i>rap</i> : uma análise em <i>corpus</i>	95
Mariana Souza Santos	
Contribuições da Linguística de <i>Corpus</i> para a sala de aula: prosódia semântica do item lexical <i>mulher</i>	106
Cássia Beatriz de Moraes Silva	
A Língua da Tabatinga: exploração inicial à luz da Linguística de <i>Corpus</i>	121
Roberta Gê-Acaiaba	

A identificação de metáforas em <i>corpus</i> jornalístico comparável bilíngue de opinião e política Wagner da Cunha Nunes	137
Elos coesivos na tese e indicação de impessoalidade: estudo exploratório em <i>corpus</i> de redações estilo ENEM Rosena Caixeta Silva Rodrigues de Sousa	148
Coletando <i>corpus</i> da internet com auxílio de <i>scripts</i> de programação Heitor Carvalho de Almeida Neto	159
Exploração do significado do lexema <i>medo</i> : uma análise pautada na Análise do Discurso e na Linguística de <i>Corpus</i> Thaís dos Santos Souza	171
Linguística de <i>Corpus</i> e C-ORAL-BRASIL: análise da fala espontânea em Belo Horizonte-MG Maria de Oliveira Rodrigues	185
Sobre os autores e organizadores	198

Estudos Descritivos e Linguística de Corpus: abrindo caminhos para a descrição linguística

Ariel Novodvorski¹

Joel Victor Reis Lisboa²

Raphael Marco Oliveira Carneiro³

1 Doutor em Estudos Linguísticos (UFMG) com pós-doutorado pela UFRGS. Professor Associado do Instituto de Letras e Linguística da Universidade Federal de Uberlândia. Docente do Programa de Pós-Graduação em Estudos Linguísticos, vinculado à Linha 1: Teoria, descrição e análise linguística. Lattes: <http://lattes.cnpq.br/2882362453894798>. E-mail: arivorski@ufu.br.

2 Doutorando e Mestre em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/7657573383244824>. E-mail: joelvictorlisboa@gmail.com.

3 Doutorando e Mestre em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/2014869652199855>. E-mail: raphael.olic@gmail.com.

É com imenso prazer que apresentamos o segundo número de *Estudos Exploratórios em Linguística de Corpus*. Assim como o primeiro (NOVODVORSKI; LISBOA, 2021), esta publicação reúne estudos exploratórios realizados no âmbito da disciplina *Estudos Descritivos e Linguística de Corpus*, ofertada pelo Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU).

Assim como na publicação anterior, a oferta da disciplina buscou despertar a percepção e a sensibilidade dos pós-graduandos para fatos e fenômenos linguísticos, por meio da utilização empírica e prática de ferramentas e recursos próprios da Linguística de *Corpus* (LC), assumida tanto por sua relevância enquanto abordagem como por seu potencial teorizador, somados ao conjunto de procedimentos metodológicos que a caracterizam. Para além de um eixo inicial norteador quanto à perspectiva, história e alcances da LC, um segundo segmento no plano da disciplina buscou nuclear as etapas e princípios envolvidos na compilação de *corpora*, levantamento e sistematização de dados, entre outros passos, em que buscamos consolidar percursos possíveis para pesquisas exploratórias guiadas/baseadas em *corpus*. O terceiro eixo do programa da disciplina foi pensado no intuito de tentar cobrir tanto quanto possível a diversidade de trabalhos práticos em que foi testada a aplicabilidade da LC em pesquisas de base empírica. Desse modo, cada uma das leituras realizadas durante a disciplina foi um convite para a testagem de ferramentas, para a configuração e ajuste de instrumentos, para a exploração de dados refazendo ou (re)criando caminhos de pesquisa.

Os trabalhos que compõem este volume são resultantes da atividade final da disciplina, a saber: o desenvolvimento de uma breve pesquisa num *corpus* a definir, em que fossem empregados princípios, procedimentos e ferramentas da LC, culminando na escrita de um artigo, com critérios definidos. Para além desse, que foi o trabalho final, e da participação nas discussões teórico-metodológicas durante as aulas, também foi desenvolvido coletivamente um Vocabulário de termos e conceitos-chave da disciplina, com auxílio dos recursos do *Moodle* institucional da UFU. Outra atividade da disciplina foi a realização de um Seminário, para apresentação individual das pesquisas que dariam origem aos trabalhos finais. Tudo precisou ser realizado em formato remoto, em decorrência da continuidade do estado de pandemia pela COVID-19 durante o ano de 2021.

Organizamos esta publicação objetivando demonstrar a produtividade da LC para estudos linguísticos de caráter descritivo, bem como para evidenciar sua relevância na formação (continuada) de pesquisadores da Linguística e Linguística Aplicada. As pesquisas exploratórias que compõem este livro vinculam-se a diferentes áreas, como Dialetologia, Lexicologia, Sociolinguística, Ensino de Línguas, Terminologia, Tradução, Análise do Discurso, Linguística Textual, Linguística Computacional, dentre outras. Os *corpora* analisados são, também, de

natureza diversa, incluindo *corpora* monolíngues e bilíngues, orais transcritos, de entrevistas, de comentários, de letras de música, acadêmicos, literários, jornalísticos, dentre outros. A seguir, sintetizamos os estudos exploratórios apresentados em cada capítulo desta publicação.

No capítulo intitulado “O suposto antagonismo de *vida e morte* em *corpora*: otimismo em Hay e pessimismo em Cioran à luz da Linguística de *Corpus*”, Lucas Amâncio Mateus descreve os usos de uma amostra de unidades lexicais relacionadas aos temas das obras do filósofo Emil Cioran e da autora motivacional Louise Hay. Por meio das ferramentas do *WordSmith Tools* 6.0 (SCOTT, 2012), o estudo ressalta as visões de mundo dos diferentes autores e como essas visões estão lexicalmente constituídas em *corpora* de suas obras mais conhecidas.

Na sequência, Lidiane Carlos Ramos, em “Uma nova leitura de *Rayuela*: análise lexical baseada em *corpus*”, analisa o romance de Julio Cortázar por meio do programa *WordSmith Tools* 4.0 (SCOTT, 2004) e suas três ferramentas, *Concord*, *KeyWords* e *WordList*. Com base na LC e na Teoria de Metáforas Conceptuais, o estudo evidencia a relevância de somatismos na composição das obras cortazarianas, bem como a proficiência da utilização de ferramentas computacionais na análise de campos lexicais em *corpora* literários.

Em “O vocabulário distópico em *Fahrenheit 451*: uma análise pautada na Linguística de *Corpus*”, Terezinha de Assis Oliveira apresenta um recorte da análise dos cem substantivos comuns mais frequentes em um *corpus* literário composto pela obra *Fahrenheit 451* (BRADBURY, 1953). Com o auxílio das ferramentas *WordList* e *Concord* do *WordSmith Tools* 6.0 (SCOTT, 2012), a autora analisa o *corpus* em busca de compreender como esses substantivos se relacionam e o que revelam sobre a obra, focalizando posteriormente suas análises no substantivo *casa*.

No capítulo seguinte, intitulado “O nome ‘Barack’ em *Minha História* (2018) sob a perspectiva sistêmico-funcional”, Sarah Cristina de Oliveira Sebba realiza uma análise sistêmico-funcional em torno do nome próprio *Barack*, especificamente pelo sistema da TRANSITIVIDADE, com alcance nos Processos, Participantes e Circunstâncias. A autora se serve dos recursos do programa *LancsBox*[®] (BREZINA; WEILL-TESSIER; McENERY, 2020) para o tratamento e análise do *corpus*. Com base na análise das orações selecionadas para o capítulo, a autora se questiona acerca da humanização de Barack no texto de Michelle Obama.

Victor Mariotto Palma, em “Análise de orações existenciais em *corpus* literário bilíngue”, apresenta os resultados parciais de uma análise de orações existenciais prototípicas em língua inglesa e de suas traduções em um *corpus* literário paralelo bilíngue (inglês-português) unidirecional. O *corpus* é composto pelos contos wildianos da coletânea *A house of pomegranates*, em língua inglesa, e de duas traduções para o português brasileiro. As análises foram realizadas por meio do *WordSmith Tools* 4.0 (SCOTT, 2004) e do *Wordfast Anywhere* (WORDFAST, 2021).

Em “Como se traduz ‘língua de acolhimento’? Análise exploratória em *corpus* paralelo bilíngue”, por meio da análise de um *corpus* de resumos e palavras-chave de dissertações e teses nas línguas inglesa e portuguesa, Joel Victor Reis Lisboa descreve como uma amostra de unidades fraseológicas especializadas da área de Português como Língua de Acolhimento foram traduzidas. Embasada por estudos em Terminologia e LC, a investigação evidencia a proficuidade dessa combinação teórico-metodológica para a exploração de equivalentes terminológicos.

Na sequência, Mayra Natanne Alves Marra, em “AVALIATIVIDADE em *corpus* de comentários: um olhar sobre o feminino”, realiza uma análise em um *corpus* de 75 comentários do vídeo “Xerecou: empoderamento feminino na linguagem” (XERECOU, 2021), publicado no YouTube. A análise, realizada por meio das ferramentas *WordList*, *KeyWords* e *Concord* do *WordSmith Tools* 6.0 (SCOTT, 2012) e amparada pela Linguística Sistêmico-Funcional, focaliza a criatividade lexical em torno do substantivo *xereca*, bem como descreve escolhas lexicogramaticais atitudinais utilizadas para expressar avaliações.

No capítulo intitulado “O uso do vocábulo *bitch* em letras de *rap*: uma análise em *corpus*”, Mariana Souza Santos descreve as nuances semânticas de *bitch* a partir de dois *corpora* de letras de músicas interpretadas por *rappers* americanos sob a variável independente *sexo*. Com a base teórico-metodológica da Estilística de *Corpus* e da Sociolinguística, a pesquisa revela como a variável *sexo* pode influenciar as escolhas linguísticas e as conotações adquiridas pelo vocábulo em análise.

Em “Contribuições da Linguística de *Corpus* para a sala de aula: prosódia semântica do item lexical *mulher*”, por meio da análise de um *corpus* composto por letras de canções sertanejas brasileiras, Cássia Beatriz de Moraes Silva analisa usos do item lexical *mulher* a fim de identificar sua prosódia semântica e propor uma sequência didática. A metodologia inclui o uso das ferramentas *WordList* e *Concord* do programa *WordSmith Tools* 4.0 (SCOTT, 2004), com o embasamento teórico da Lexicologia e da Lexicultura.

Roberta Gê-Acaiaba, em “A Língua da Tabatinga: exploração inicial à luz da Linguística de *Corpus*”, analisa um *corpus* oral sincrônico transcrito, oriundo de 10 entrevistas realizadas com cidadãos de Bom Despacho-MG, objetivando analisar a presença do léxico Tabatinga nas produções linguísticas dos entrevistados. As análises apresentadas, realizadas por meio das ferramentas *WordList* e *Concord* do *WordSmith Tools* 4.0 (SCOTT, 2004) e sustentadas pela Sociolinguística Variacionista, focalizam os substantivos mais frequentes no *corpus*, *cuete* e *ocaia*, bem como a polissemia apresentada nas utilizações do verbo *tipurar*.

Em “A identificação de metáforas em *corpus* jornalístico comparável bilíngue de opinião e política”, por meio de análises impressionísticas e hipertextuais, Wagner da Cunha Nunes realiza a identificação, o mapeamento dos domínios e a análise de duas expressões metafóricas retiradas de um *corpus* comparável bilíngue (português-espanhol) composto por textos das seções de opinião e política de dois jornais brasileiros (*Estadão* e *Folha de São Paulo*) e dois argentinos (*Clarín* e *Perfil*). O estudo exploratório apresentado está embasado pela Teoria da Metáfora Conceptual, de Lakoff e Johnson (1980), e foi realizado por meio do *Sketch Engine* (KILGARRIFF *et al.*, 2003).

Na sequência, em “Elos coesivos na tese e indicação de impessoalidade: estudo exploratório em *corpus* de redações estilo ENEM”, Rosena Caixeta Silva Rodrigues de Sousa explora, com o auxílio das ferramentas *WordList* e *Concord* do *WordSmith Tools* 6.0 (SCOTT, 2012), um *corpus* composto por sete redações que alcançaram nota máxima no Exame Nacional do Ensino Médio (ENEM). A autora analisa as relações entre elos coesivos e impessoalidade na construção das teses dessas redações, focalizando, posteriormente, suas discussões na utilização do conectivo *porém*.

Heitor Carvalho de Almeida Neto, em “Coletando *corpus* da internet com auxílio de *scripts* de programação”, propõe uma metodologia de coleta, conversão e armazenamento de textos digitais que inclui o uso da linguagem de programação *Python*, exemplificada com um *corpus* de obras de Machado de Assis. O estudo mostra como técnicas computacionais podem facilitar os procedimentos metodológicos da pesquisa com *corpus*, permitindo que o pesquisador otimize o processo computacional e dedique-se mais aos aspectos analíticos de sua investigação.

No capítulo intitulado “Exploração do significado do lexema *medo*: uma análise pautada na Análise do Discurso e na Linguística de *Corpus*”, Thaís dos Santos Souza apresenta um estudo exploratório localizado na interface entre Linguística Aplicada e Criminologia. Por meio das ferramentas *WordList* e *Concord* do *WordSmith Tools* 4.0 (SCOTT, 2004), a autora analisa um *corpus* de transcrições de 21 entrevistas, objetivando explorar o campo semântico do lexema *medo* e compreender os significados e as relações físicas e sociais atreladas a este lexema nos discursos dos sujeitos entrevistados.

Finalizando o compilado de estudos exploratórios, Maria de Oliveira Rodrigues, em “Linguística de *Corpus* e C-ORAL BRASIL: análise da fala espontânea em Belo Horizonte-MG”, descreve usos do português brasileiro por meio de um *corpus* oral da fala belo-horizontina. Ao comparar as diferentes visões da gramática tradicional e da gramática descritiva, o estudo mostra a presença marcante de variações na fala em relação ao aspecto da concordância verbal.

Sem mais delongas, esperamos que a leitura deste novo volume seja enriquecedora e motivadora para a contínua exploração do mundo fascinante da linguagem verbal.

| Referências

BRADBURY, R. **Fahrenheit 451**. Nova York: Ballantine Books, 1953.

BREZINA, V.; WEILL-TESSIER, P.; McENERY, A. **#LancsBox version 5**. Lancaster: Lancaster University, 2020. Disponível em: <http://corpora.lancs.ac.uk/lancsbox>. Acesso em: 14 jan. 2022.

KILGARRIFF, A. *et al.* **Sketch Engine**. East Sussex: Lexical Computing Limited, 2003. Disponível em: <https://www.sketchengine.eu/>. Acesso em: 10 jan. 2022.

LAKOFF, G.; JOHNSON, M. **Metaphors we live by**. Chicago: Chicago University Press, 1980.

NOVODVORSKI, A.; LISBOA, J. V. R. (org.). **Estudos exploratórios em Linguística de Corpus**. Araraquara: Letraria, 2021. Disponível em: <https://www.lettraria.net/estudos-exploratorios-em-linguistica-de-corpus/>. Acesso em: 18 jan. 2022.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 12 jan. 2022.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://lexically.net/wordsmith/version6/>. Acesso em: 12 jan. 2022.

WORDFAST. **Wordfast Anywhere**. Version 6.2.5. New York: Wordfast, 2021. Disponível em: <https://www.freetm.com/>. Acesso em: 14 jan. 2022.

XERECOU: Empoderamento feminino na linguagem | Karen Jonz | Mini Saia | Saia Justa. **Canal GNT**, 12 ago. 2021. (11min12s). YouTube. Disponível em: <https://youtu.be/6JwED9VRdBY>. Acesso em: 14 jan. 2022.

O suposto antagonismo de *vida e morte* em *corpora*: otimismo em Hay e pessimismo em Cioran à luz da Linguística de *Corpus*¹

Lucas Amâncio Mateus²

¹ Apoio e financiamento: Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG.

² Doutorando em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/0101102792314230>. E-mail: lucas.mateus@ufu.br.

1 Introdução

Ao observarmos algumas literaturas com abordagens antagônicas – a do pessimismo e a do otimismo – percebemos que nelas há concepções de mundo e conceitos dualistas como vida e morte, que merecem ser investigados e analisados com maior profundidade, a fim de se perceber como as noções são constituídas e quais suas definições subjacentes, que podem ser expressas até mesmo por meio de metáforas, em alguns casos. Para isso, os princípios e as ferramentas da Linguística, mais especificamente da Linguística de *Corpus* (LC), podem auxiliar com eficácia a detecção de quais palavras possuem maior peso em cada um desses posicionamentos, além de permitir a identificação de suas frequências, seus ambientes textuais, usos, concordâncias, definições diretas (marcadas por contextos definitórios) e definições indiretas (presentes em contextos explicativos).

Isso posto, o objetivo deste capítulo é contrastar algumas noções aparentemente opostas, aquelas que foram mais frequentes em *corpora* compostos de obras de destaque dos autores Emil Cioran (1911-1995) e Louise Hay (1926-2017). A partir de nossas análises, foi possível perceber as construções conceituais pessimistas e otimistas dos autores à luz da LC.

Emil Cioran foi reconhecido pelo *site* de crítica literária do Los Angeles³ como tendo sido o maior niilista do ocidente, desde Nietzsche. Cioran é um filósofo romeno, famoso por seu pessimismo, niilismo⁴, antinatalismo e ceticismo. Seus interesses de estudo e publicações repousam sobre conceitos como morte, angústia, absurdo, infinito, insignificância, caos, agonia, suicídio e loucura.

Louise Hay foi uma autora motivacional⁵, estadunidense. Ela é considerada uma das fundadoras do gênero autoajuda e foi precursora em publicações dessa categoria. Suas obras giram em torno dos temas do autodesenvolvimento e do poder dos pensamentos otimistas, e alguns de seus livros são consideradas *best-sellers* devido ao grande alcance quanto ao número de leitores. Seus livros abordam concepções de saúde emocional, amor, vida, cura e afirmações positivas.

No bojo desse suposto antagonismo, esta pesquisa procura investigar se há uma definição específica para os termos *vida* e *morte* nas obras de Cioran e de Hay. Nesse caso, a hipótese subjacente é que *corpora* escritos de textos autênticos desses autores oferecem a ambiência necessária para levantamento de tais conceitos, expressos por meio de seus itens lexicais

3 Para informações mais detalhadas sobre o autor e as críticas tecidas a partir de seu trabalho, consultar Bradatan (2016).

4 De acordo com o dicionário *Michaelis Online*, o niilismo é o “pensamento que considera as crenças e os valores tradicionais da sociedade como infundados e inúteis”, enfocando assim noções de não existência. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/niilismo/>. Acesso em: 22 out. 2021.

5 Para mais detalhes sobre Louise Hay, consultar informações disponíveis em: <https://refornari.com/louise-hay-saiba-mais> e/ou em: <https://www.louisehay.com/about/>. Acesso em: 22 out. 2021.

que se constituem como termos dentro das obras. Outrossim, os contextos contidos nos *corpora* permitirão perceber que a especificidade das definições de *vida* e *morte* para esses autores está diretamente relacionada a suas visões de mundo, que entre si são contrastantes: a concepção cética de Cioran e a otimista de Hay.

De antemão, temos a hipótese prévia de que a análise do *corpus* de Hay estará cingida de uma temática espiritualista, apregoando a existência divina e de valores sociais de paz e esperança, já que os títulos de suas obras indicam esse estilo. Em contraposição, a natureza das obras de Cioran permite-nos hipotetizar que, na tratativa de seu *corpus* de essência pessimista, tornar-se-á evidente a ausência de valores espirituais e a ênfase em tendências negativas e derrotistas. Tais premissas, no entanto, poderão ser validadas ou refutadas no decorrer do trabalho, realizado sob o critério da LC.

2 Fundamentação teórica

A partir do que ficou estabelecido por Tagnin (2005, p. 21), considera-se que a LC é um ramo específico do saber, que possibilita a investigação e oferece uma metodologia “que veio facilitar muito a identificação das unidades convencionais da língua”, posição consoante com a de Berber Sardinha (2004, p. 37) de que a LC é “uma maneira de se chegar à linguagem”. Stubbs (2001), por sua vez, concebe a LC como um instrumento que permite que enxerguemos melhor as amostras linguísticas e tratemos um volume maior de dados de maneira mais ágil, eficaz e precisa. Ela é também definida como uma “metodologia para a investigação das línguas e da linguagem, a qual permite levar a cabo investigações empíricas em contextos autênticos” (PARODI, 2010, p. 15 *apud* BEILKE, 2016, p. 70).

De acordo com Assunção e Araújo (2019), há autores, no campo de ação de diversas áreas da Linguística, que consideram a LC como uma abordagem e uma metodologia. Neste estudo, a LC é tomada como uma abordagem, por admitir que “cada *corpus* que tive a chance de examinar, mesmo pequeno, ensinou-me fatos que não poderia imaginar encontrar de nenhum outro modo” (FILLMORE, 1992, p. 35 *apud* BEILKE, 2016, p. 74). Portanto, compreendemos a LC não só como uma metodologia empírica eficaz para os estudos descritivos, mas também como uma abordagem com características e princípios próprios que contribui para diversos tipos de investigações da linguagem, ainda que em alguns momentos seja necessário focar mais em seu aspecto metodológico, a fim de permitir a eficiência dos procedimentos e garantir o alcance dos objetivos de pesquisa, conforme se dá na circunscrição restrita do presente texto.

Tendo em vista nossa busca pelas definições de vida e morte em Cioran e Hay, nos apoiamos na conceitualização de contextos definitórios e explicativos formulada por Aubert (2001). Alguns contextos definicionais foram identificados sob a forma de metáforas, portanto, como base de

nossas análises, partimos dos postulados de Berber Sardinha (2007, 2008), que disserta sobre dois tipos de metáforas. Dentre os demais tipos existentes, ele aborda brevemente as metáforas linguísticas, que são, conforme o autor, aquelas constituídas de expressões metafóricas e que formam uma unidade de sentido. Na escrita, elas formam uma oração, já na fala, formam um enunciado. Para o autor, trata-se de expressões que contêm palavras usadas metaforicamente, como, por exemplo, “ele subiu na vida”. Nesse caso, as metáforas linguísticas são do tipo que podem ou não terem sido entendidas na mente do falante/ouvinte como metáfora.

Em seguida, Berber Sardinha (2007, 2008) enfoca seu objeto de interesse que são as metáforas conceituais, o que também nos interessa no presente estudo. Portanto, adotamos a definição desse linguista, que as entende como expressões linguísticas que são a manifestação de uma metáfora conceptual, ou seja, advêm de uma conceituação metafórica. Segundo o exemplo do autor, “nosso casamento está indo muito bem” (BERBER SARDINHA, 2007, p. 107) é uma expressão que advém da metáfora conceptual AMOR É UMA VIAGEM, pois ela contém um domínio (fonte e alvo), mapeamentos (as relações feitas entre os domínios) e os desdobramentos (as inferências que podemos fazer a partir de uma metáfora conceptual).

Lembramos também que “existem metáforas conceituais que subjazem às metáforas linguísticas identificadas na concordância” (BERBER SARDINHA, 2007, p. 108). O autor também descreve o que e como são as metáforas conceituais: “As metáforas conceituais não são verbalizadas como tais. [...] As metáforas conceituais são representações mentais dos conceitos” (BERBER SARDINHA, 2008, p. 98).

Portanto, em suma, o autor esclarece que as metáforas conceituais expressam conceitos que são metafóricos. Com base nisso, quando buscamos as noções de *vida* e *morte* em ambos os *corpora*, levamos em conta essas considerações e observamos se elas apareceram metaforizadas ou até mesmo expressas nos contextos definitórios e explicativos por meio de metáforas conceituais. Adicionalmente, conforme explica Berber Sardinha (2008), os *softwares* utilizados na LC por si só não podem encontrar ocorrências de metáforas, o que o analista pode fazer é buscar por palavras que normalmente possuem usos metafóricos.

3 Metodologia

O primeiro procedimento foi compilar dois *corpora*, um de cada autor escolhido, para obtermos amostras legítimas a respeito da nossa temática. Sendo assim, compilamos cinco obras⁶ de Cioran, a saber: *Nos Cumes do Desespero*, *O Livro das Ilusões*, *Breviário de*

⁶ Os títulos originais em romeno e francês, com suas primeiras datas de publicação, são: *Pe culmile disperării* (1934), *Cartea amăgirilor* (1936), *Précis de décomposition* (1949), *Syllogismes de l'amertume* (1952). Já a entrevista *A Filosofia Irritada* (idioma original desconhecido) fora publicada em forma de livro em 1995, concedida ao escritor alemão Heinz-Norbert Jocks, publicada no nº 5 da revista *Kulturchronik*, editada em Bonn pela InterNations.

Decomposição, Silogismos da Amargura e A Filosofia Irritada – entrevista publicada em forma de livro. Igualmente, de Hay, reunimos cinco obras⁷, conforme se segue: *Você pode curar sua vida* (1984), *Cure seu corpo* (1984), *O poder das afirmações positivas* (2005), *Está tudo bem* (2013) e *A vida ama você – 7 passos para curar sua vida* (2014).

Após a compilação e organização dos dois *corpora*, realizamos a análise dos dados por meio do *WordSmith Tools* versão 6.0 (SCOTT, 2012) – doravante WST, uma suíte de ferramentas para realização de análises lexicais, tais como geradores de lista de palavras, de linhas de concordância, de lista de palavras-chave, dentre outras.

O primeiro passo foi configurar o *software* para a língua portuguesa e, em seguida, gerar duas listas de palavras, uma de cada *corpus*. Após salvar a lista do *Corpus Cioran* e do *Corpus Hay*, prosseguiu-se para a aplicação de *stoplist*⁸, a fim de ausentar da listagem as palavras gramaticais, que embora sejam também relevantes, não são o foco deste capítulo, que focaliza as palavras lexicais. Na Figura 1, a seguir, são apresentadas ambas as listas agrupadas após esse tratamento.

Figura 1: Recorte das listas de palavras do *Corpus Cioran* e do *Corpus Hay* após indexação da *stoplist*.

N	Word	Freq.	%	Texts	% Lemmas	Set
1	É	2.090	1,21	5	100,00	
2	VIDA	995	0,58	5	100,00	
3	MUNDO	710	0,41	5	100,00	
4	SER	706	0,41	5	100,00	
5	QUANDO	509	0,30	5	100,00	
6	MORTE	505	0,29	5	100,00	
7	HOMEM	471	0,27	5	100,00	
8	PODE	391	0,23	5	100,00	
9	TEMPO	385	0,22	5	100,00	
10	SÃO	369	0,21	5	100,00	
11	AMOR	302	0,18	4	80,00	
12	EXISTÊNCIA	280	0,16	5	100,00	
13	HÁ	279	0,16	5	100,00	
14	HOMENS	255	0,15	4	80,00	
15	TEM	248	0,14	5	100,00	
16	ESPÍRITO	242	0,14	4	80,00	
17	DEUS	222	0,13	5	100,00	
18	TER	205	0,12	5	100,00	
19	ALMA	199	0,12	4	80,00	
20	ESTÁ	195	0,11	5	100,00	
21	CONSCIÊNCIA	184	0,11	5	100,00	
22	MEDO	180	0,10	5	100,00	
23	SEJA	177	0,10	4	80,00	
24	SOFRIMENTO	177	0,10	4	80,00	
25	HISTÓRIA	176	0,10	5	100,00	

N	Word	Freq.	%	Texts	% Lemmas	Set
1	É	1.680	1,03	5	100,00	
2	VIDA	1.544	0,95	5	100,00	
3	OU	837	0,51	5	100,00	
4	ESTÁ	625	0,38	5	100,00	
5	SER	616	0,38	5	100,00	
6	AMOR	614	0,38	5	100,00	
7	ESTOU	592	0,36	5	100,00	
8	PROBLEMAS	546	0,34	4	80,00	
9	SOU	534	0,33	5	100,00	
10	QUANDO	509	0,31	5	100,00	
11	MEDO	474	0,29	5	100,00	
12	ALEGRIA	385	0,24	5	100,00	
13	OF	379	0,23	3	60,00	
14	CORPO	360	0,22	5	100,00	
15	SÃO	338	0,21	5	100,00	
16	PENSAMENTOS	321	0,20	5	100,00	
17	VER	315	0,19	5	100,00	
18	SAÚDE	311	0,19	5	100,00	
19	PODE	301	0,19	5	100,00	
20	SEGURO	301	0,19	4	80,00	
21	PESSOAS	297	0,18	5	100,00	
22	RAIVA	292	0,18	4	80,00	
23	AND	263	0,16	1	20,00	
24	TER	263	0,16	5	100,00	
25	AFIRMAÇÕES	256	0,16	5	100,00	

Fonte: *WordList*.

⁷ Seguem os títulos originais em inglês, com suas primeiras datas de publicação: *You can heal your life* (1984), *Heal your body: the mental causes for physical illness and the metaphysical way to overcome them* (1984), *I can do it: how to use affirmations to change your life* (2004), *All is well: heal your body* (2013) e *Life loves you: 7 spiritual experiments to heal your life* (2014).

⁸ *Stoplist*: “lista de palavras que é anexada ao *corpus* e faz uma limpeza no texto ao excluir palavras que nela se encontram” (CARVALHO, 2015, p. 80).

Como é possível observar na Figura 1, o item lexical mais frequente em ambos os *corpora* é o verbo *ser* conjugado como “é”, totalizando, respectivamente, 2.090 ocorrências no *Corpus Cioran* e 1.680 ocorrências no *Corpus Hay*, seguido pelo item *vida*, o segundo mais frequente em ambos os *corpora*.

A etapa seguinte foi a seleção e preparação de um *corpus* de referência. De acordo com Berber Sardinha (2004), o *corpus* de referência deve ser em torno de cinco vezes maior que o *corpus* de estudo, a fim de funcionar como um parâmetro para a produção de palavras-chave em contraste com a lista de palavras do *corpus* de estudo. Para isso, utilizamos textos das áreas de humanas, principalmente de religiões e filosofia, disponíveis no Lácio-Web⁹.

Logo após, procedemos à geração das palavras-chave, passamos a gerar linhas de concordâncias a partir das palavras mais frequentes e das palavras-chave dos *corpora*. Realizamos várias buscas e fizemos observações comparativas, além de salvar algumas telas com os resultados obtidos em forma de linhas de concordância de buscas como “vida”, “vida é”, “vidas” e “vidas são”. Para as buscas de “vidas são”, “morte” e “morte é” só foram encontradas concordâncias resultantes no *Corpus Cioran*, por isso, realizamos o agrupamento de todas as palavras do campo lexical da *morte* no *Corpus Hay* (*morre*, *morrendo*, *morrera*, *morreram*, *morrerem*, *morreríamos*, *morri*, *mortais*, *mortalidade*, *mortes* e *morto*), já que a palavra *morte* no singular possui apenas 19 ocorrências. Então geramos linhas de concordâncias a partir delas, a fim de verificarmos rapidamente como essas noções estariam conceituadas nas obras da autora.

4 Resultados e discussão

Como resultado da aplicação de nossos procedimentos metodológicos, obtivemos alguns dados, os quais passamos a descrever de agora em diante.

O processamento dos dados resultou nas estatísticas de 172.341 *tokens* (itens) e 17.708 *types* (formas) do *Corpus Cioran* e de 160.618 *tokens* e 14.109 *types* do *Corpus Hay*. A Figura 2 comprova essas informações.

⁹ Disponível em: <http://143.107.183.175:22180/lacioweb/index.htm>. Acesso em: 1 nov. 2021.

Figura 2: Resultados estatísticos do *Corpus* Cioran e do *Corpus* Hay.

N	text file	tokens (running words) in	tokens used for word list	types (distinct words)	type/token ratio (TTR)	standardise TTR	STTR std.dev.
1	Overall	172.458	172.341	17.708	10,27	47,75	48,61
2	Emil Cioran - Silogismos da amargura-E	16.163	16.132	4.624	28,66	53,31	42,80
3	Emil-Cioran-Breviário-de-decomposiçã	57.293	57.259	10.269	17,93	51,76	46,91
4	Emil-Cioran-Nos-Cumes-do-Desespero	32.307	32.299	5.308	16,43	46,85	50,19
5	Emil-Cioran-O-livro-das-ilusões-Rocco-I	65.006	64.968	8.102	12,47	43,28	55,54
6	Emil-Cioran_A-Filosofia-Irritada-by-_z-lil	1.689	1.683	713	42,36	48,70	

N	text file	tokens (running words) in	tokens used for word list	types (distinct words)	type/token ratio (TTR)	standardised TTR	STTR std.dev.
1	Overall	162.603	160.618	14.109	8,78	44,85	54,29
2	Louise Hay_ Está-tudo-bem-_z-lib.org_	78.186	76.593	9.463	12,35	44,94	52,25
3	Louise Hay_ Você-pode-curar-sua-vida	50.200	50.091	6.549	13,07	45,21	53,00
4	Louise Hay_ Cure-Seu-Corpo-z-lib.org_	15.185	15.062	2.820	18,72	43,67	50,79
5	Louise Hay_ Poder das Afirmações Po	11.196	11.141	2.309	20,73	44,80	49,48
6	Louise Hay_ A vida ama você - 7 pass	7.836	7.731	1.932	24,99	43,81	47,70

Fonte: Captura de tela dos *corpora* no WST 6.0.

Os dois *corpora* estão equilibrados pela quantidade de livros selecionados, tendo em vista que não é possível ter controle sobre a quantidade de palavras deles, já que utilizamos suas obras completas e mais famosas. Acredita-se que a variação na quantidade de itens entre os *corpora* se deve ao fato de que o gênero filosófico discorre comumente de forma mais difusa e subjetiva sobre um tema, enquanto a modalidade autoajuda costuma ser mais direta e objetiva, residindo aí já um primeiro contraste.

De modo geral, a observação das listas e das linhas de concordância demonstrara que as noções que investigamos – *vida e morte* – estão presentes nas obras, principalmente *vida*, que é muito frequente, visto que há uma satisfatória quantidade de contextos definitórios (marcados pela presença do verbo *ser* conjugado como *é*) e, também, de contextos explicativos, que descrevem as noções em questão. Sendo assim, consideramos os *corpora* representativos para o escopo desta investigação.

Geramos duas listas de palavras-chave, uma de cada um dos *corpora*, cujos recortes estão apresentados na Figura 3.

Figura 3: Recorte das listas de palavras-chave do *Corpus Cioran* e do *Corpus Hay*.

CIORAN CORPUS_KWL_files.kws								HAY CORPUS_KWL_files.kws							
N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness
1	→ VIDA	995	0,58	5	2.192	0,06	2.428,00	1	→ VIDA	1.544	0,95	5	2.192	0,06	4.899,00
2	→ MORTE	505	0,29	5	720	0,02	1.543,00	2	ESTÁ	625	0,38	5	4.860	0,13	483,00
3	MUNDO	710	0,41	5	2.270	0,06	1.347,00	3	→ AMOR	614	0,38	5	263		2.836,00
4	HOMEM	471	0,27	5	1.056	0,03	1.136,00	4	ESTOU	592	0,36	5	307		2.617,00
5	AMOR	302	0,18	4	263		1.119,00	5	PROBLEMAS	546	0,34	4	1.557	0,04	1.180,00
6	ETERNIDADE	175	0,10	4	14		988,60	6	SOU	534	0,33	5	323		2.270,00
7	ESPÍRITO	242	0,14	4	301		784,60	7	QUANDO	509	0,31	5	5.020	0,14	255,80
8	TRISTEZA	154	0,09	4	37		772,20	8	→ MEDO	474	0,29	5	243		2.101,00
9	EXISTÊNCIA	280	0,16	5	511	0,01	758,20	9	ALEGRIA	385	0,24	5	109		1.923,00
10	ALMA	199	0,12	4	161		755,90	10	OF	379	0,23	3	2.516	0,07	367,60
11	SOFRIMENTO	177	0,10	4	104		738,70	11	CORPO	360	0,22	5	658	0,02	1.011,00
12	SOLIDÃO	134	0,08	5	24		700,00	12	PENSAMENTOS	321	0,20	5	45		1.761,00
13	→ DEUS	222	0,13	5	328		667,30	13	VER	315	0,19	5	870	0,02	695,30
14	IDEIA	104	0,06	5	1		634,80	14	SAÚDE	311	0,19	5	1.118	0,03	566,50
15	INFINITO	134	0,08	4	50		621,70	15	SEGURO	301	0,19	4	142		1.360,00
16	ÊXTASE	119	0,07	4	26		605,20	16	PESSOAS	297	0,18	5	2.359	0,06	221,70
17	VIVER	170	0,10	4	188		577,90	17	RAIVA	292	0,18	4	42		1.598,00
18	→ MEDO	180	0,10	5	243		563,40	18	AND	263	0,16	1	1.698	0,05	264,50
19	SANTIDADE	93	0,05	3	2		558,40	19	TER	263	0,16	5	2.655	0,07	125,60
20	DESESPERO	112	0,06	5	32		546,10	20	AFIRMAÇÕES	256	0,16	5	42		1.380,00
21	HOMENS	255	0,15	4	694	0,02	543,20	21	PENSAMENTO	250	0,15	5	469	0,01	692,80
22	MELANCOLIA	102	0,06	4	16		541,40	22	FAZER	249	0,15	5	2.469	0,07	123,60
23	ABSOLUTO	126	0,07	5	74		525,90	23	EMOCIONAL	239	0,15	4	41		1.282,00
24	IDEIAS	86	0,05	4	5		495,90	24	PROBLEMA	235	0,14	5	1.816	0,05	183,10
25	LÁGRIMAS	95	0,06	5	18		492,60	25	PAZ	232	0,14	4	248		824,00

Fonte: KeyWords.

Constatou-se a presença dos seguintes itens lexicais e suas frequências dentre as 25 palavras-chave em ambos os *corpora*: *Corpus Cioran* - vida (995), morte (505), amor (302), Deus (222) e medo (180); *Corpus Hay* - vida (1.544), amor (614) e medo (474); como nesse último as palavras *morte* e *Deus* não apareceram dentre as palavras-chave, buscamos por ordem alfabética em todo o *corpus* a fim de verificar se elas estariam em alguma medida presentes, conforme demonstra a Figura 4, abaixo.

Figura 4: Presença de palavras do grupo lexical de *morte* e da palavra *Deus* no *Corpus Hay*.

N	Word	Freq.	%	Texts	%
11.392	MORRE	1		1	20,00
11.393	MORRENDO	1		1	20,00
11.394	MORRERA	1		1	20,00
11.395	MORRERAM	1		1	20,00
11.396	MORREREM	1		1	20,00
11.397	MORRERIAM	1		1	20,00
11.398	MORRERÍAMOS	1		1	20,00
11.399	MORRI	1		1	20,00
11.400	MORTAIS	1		1	20,00
11.401	MORTALIDADE	1		1	20,00
11.402	MORTES	1		1	20,00
11.403	MORTO	1		1	20,00
11.404	MOSTRAM-NOS	1		1	20,00

HAY CORPUS_SL_files.lst					
N	Word	Freq.	%	Texts	%
4.128	DETESTA	6		3	60,00
4.129	DETESTAM	1		1	20,00
4.130	DETESTAMOS	1		1	20,00
4.131	DETESTOU	1		1	20,00
4.132	DETRIMENTO	4		1	20,00
4.133	DEU	17	0,01	5	100,00
4.134	DEU-LHE	1		1	20,00
4.135	DEU-NOS	1		1	20,00
4.136	→ DEUS	26	0,02	3	60,00
4.137	DEUSES	1		1	20,00
4.138	DEVA	1		1	20,00
4.139	DEVASTADORAS	1		1	20,00

Fonte: WordList.

Também foi realizada uma busca para saber se haveria a presença do plural *vidas* nos *corpora*, talvez pressupondo a crença na existência de mais de uma vida e na ideia de reencarnação, e foram encontradas em ambos: 9 ocorrências para *vidas* no *Corpus Cioran* e 33 no *Corpus Hay*¹⁰. Para Louise Hay, a citação de *vidas* encaixa-se num contexto geral e amplo, ao ressaltar que podemos modificar e melhorar nossas vidas, a partir de mudanças na forma de pensar. A ideia de reencarnação vem apenas com uma citação, em que a autora preconiza a regressão a vidas passadas – aludindo então à possibilidade de transmigração da alma – como uma técnica benéfica para se trabalhar com a mente. Já no *Corpus Cioran*, *vidas* é um termo tratado como um simples plural de *vida*, no entanto, é atrelado a seu estilo pessimista, quando o autor assevera que as vidas neste mundo são oprimidas, asfixiadas e vestidas de luto. Há tão-somente uma menção em que *vidas* promove a possibilidade reencarnatória, quando em um trecho o autor diz que no passado vivera várias vidas.

No *Corpus Cioran*, o item *morte* ocorre 505 vezes, sendo o segundo com maior chavidade. Como palavra-chave, ele não ocorre no *Corpus Hay*, apenas a palavra *mortalidade*, com 11 ocorrências; nenhuma outra palavra desse campo lexical foi identificada nesse *corpus*.

Ao avaliar qualitativamente os dados presentes nos contextos, por meio da ferramenta *Concord* do WST, que gera linhas de concordâncias, listamos o modo como *vida* e *morte* aparecem e são conceituadas em ambos os *corpora*. Também identificamos por meio das linhas de concordância tanto contextos definitórios quanto explicativos em que há a presença de algumas metáforas.

Foram identificados 95 contextos definitórios para *vida* (*vida* é) no *Corpus Hay* e 1.449 contextos explicativos. Devido à extensão das ocorrências, optamos em analisar somente os contextos definitórios. Neles, a *vida* aparece na concepção de Hay como “eterna e cheia de alegria”, “mudança”, “muito simples”, “solitária”, “fácil e alegre”, “**uma droga**¹¹”, “um processo”, “minha aliada”, “dirigida pela atenção emocional”, “boa, segura e alegre”, “**uma fraude**”, “doce”, “favorável a mim”, “eterna e plena de alegria”, “feita de relacionamentos”, “um fluxo”, “movimento”, “**cheia de perigos**”, “**sua inimiga**”, “experiências”, “para ser vivida plenamente”, “profundamente afirmativa”, “divinamente guiada”, “repleta”, “sempre nova”, “guiada tanto pela graça divina quanto pela escolha”, “uma alegria”, “pura alegria”, “perfeita”, “excitante”, “espelho dos pensamentos”, “uma conselheira”.

No *Corpus Cioran*, foram encontrados 43 contextos definitórios e 952 explicativos para *vida*. Nos detivemos então a observar esses contextos, e neles, a *vida* aparece como: “ocupação de

¹⁰ O autor Cioran não faz alusão à reencarnação por meio do termo *vidas*. Louise Hay sim.

¹¹ Para a autora, as definições de *vida*, em negrito no parágrafo, representam as crenças limitantes do ser humano. Ou seja, para ela, a *vida* é algo sempre bom, e o que está marcado como ruim seria o tipo de pensamento que devemos evitar e combater.

um inseto”, “milagre que a amargura destrói”, “lugar de minhas paixões”, “ponto de interrogação”, “romance da matéria”, “a piedade da duração”, “um élan (ímpeto)”, “frágil”, “problemática”, “bem perdido”, “doença crônica”, “lembrança suprema”, “realidade suprema”, “remoída”, “cativa da morte”, “mediocre”, “fragmentária”, “falta de gosto que nem a morte”, “batismos de sombras”, “limitada”, “pretexto”, “in-eternidade”, “passo na morte”, “originária”, “eterna para o espírito e efêmera diante da morte”, “medo atravessado de lampejos”, “não ter talento para a poesia”, “uma evidência”, “torpor no claro-escuro”, “inércia entre luzes e sombras”, “o que se decompõe a todo momento”, “é ilusória”, “um estrondo”, “o cume mais alto”.

Já o termo *morte* alcançou 21 contextos definitórios (morte é) no *Corpus* Cioran onde ela aparece como “única surpresa da solidão”, “igual ao mendigo”, “a única obsessão que não pode se tornar voluptuosa”, “motivo de vergonha” e “condição de sobrevivência” e 484 explicativos nos quais a *morte* aparece como “algo que chega a seu tempo”, “limite da vida”, “um apagar-se”, “algo que não merecemos, assim como a vida”, “não absoluta”, “subjativa”, “repugnante”, “dissolução imaterial no infinito”.

Para *morte*, não encontramos nenhum contexto definitório no *Corpus* Hay, apenas 19 contextos explicativos, neles, a *morte* aparece como “a saída do filme da vida” ou “abandonar o filme da vida”, e ainda a morte de alguém (o cônjuge) como possibilidade de mudança da própria vida. Pelo fato de *morte* não ser palavra-chave no *Corpus* Hay e não ter muita frequência na lista de palavras, geramos uma linha de concordância para cada uma das ocorrências *hápax legomenon*¹² para *morre*, *morrendo*, *morrera*, *morreram*, *morrerem*, *morreríamos*, *morri*, *mortais*, *mortalidade*, *mortes* e *morto*. Nelas não há contextos definitórios nem explicativos, exceto por um contexto explicativo onde *morrer* aparece como algo de que se tem medo.

Há sim algumas posturas duais no ideário dos autores Hay e Cioran, pois enquanto para Hay a vida é simples, eterna e perfeita, para Cioran ela é problemática, efêmera/in-eternidade e mediocre. Enquanto para Hay a morte é colocada de forma negativa quando o morrer é algo a se temer, porém é metaforizada como “saída” ou “abandono do filme da vida”, já para Cioran a morte é “condição de sobrevivência” e “algo que chega a seu tempo”.

A vida também aparece como milagre, romance, lugar de paixões, lembrança suprema em Cioran, o que contraria a visão totalmente pessimista e leva suas noções para a zona da ambivalência, inclusive quando o autor define a morte como “algo que não merecemos, assim como a vida” e define a vida como “eterna para o espírito e efêmera diante da morte”. Há também noções do campo semântico negativo em Hay quando ela coloca a vida como uma droga, cheia de perigos, inimiga e solitária, porém, se referindo às crenças limitantes que existem, de modo geral, em suas obras.

12 Termo que vem do grego e designa palavras que são usadas (ou no contexto da LC, que ocorrem) uma única vez.

Percebemos em Hay a metáfora conceitual a VIDA É UMA VIAGEM quando ela a coloca como fluxo, movimento, guiada, dirigida, já em Cioran quando ele cita que a vida é efêmera e um passo; identificamos em Cioran a metáfora conceitual a VIDA É UMA DOENÇA quando ele se refere a ela como uma doença crônica e a VIDA É UMA ESCALADA, quando o autor a define como “o cume mais alto”.

5 Considerações finais

A partir dos resultados apresentados neste capítulo, considera-se que o olhar para as obras de ambos os autores, Cioran e Hay, por meio das ferramentas de LC, nos permitiu perceber fatos surpreendentes, tendo em vista que Cioran é considerado um autor pessimista e ateu e, no entanto, *Deus* e *amor* são duas das palavras mais frequentes no *corpus* de suas principais obras. Além disso, inferíamos de antemão que *morte* fosse mais frequente, porém, *vida* possui mais ocorrências no *corpus* do reconhecido autor cético. Em contrapartida, o *corpus* de Hay, composto pelas principais obras da autora, que vem originalmente de um círculo religioso, não possui o item *Deus* dentre os mais frequentes e nem mesmo tantas palavras do campo semântico ou lexical da espiritualidade, conforme esperávamos.

Esses fatos reforçam, a nosso ver, o quanto que o olhar para os dados por meio da LC é diferenciado, o quanto a percepção sobre os fatos de linguagem, quando estudados por meio dessa abordagem e postos à prova por meio dessa perspectiva, pode contrariar nossas expectativas prévias, derrubar preconceitos, refutar hipóteses antes concebidas ou rearranjá-las, além de trazer à tona novas hipóteses e redirecionar nossas análises.

Reconhecemos que, nesse curto espaço, não foi possível aprofundar o estudo qualitativo dos resultados alcançados e nem mesmo aprofundar a análise sobre a questão das metáforas, que definem conceitualmente algumas ideias e noções altamente frequentes nos *corpora* como *vida e morte*, o que se revelou um campo rico, que merece estudos posteriores, inclusive para a análise das palavras *Deus* e *amor*.

Em suma, a dualidade está presente nas obras dos autores em comparação entre si, porém, percebemos que há também ambivalências entre as noções de morte e vida dentro das mesmas obras, quando enfocamos a abordagem de cada autor de forma isolada. Por fim, consideramos que a LC permitiu aplicar com eficácia seus métodos, a fim de testarmos o breve estudo desse tema aqui proposto à luz de seus recursos e princípios.

| Referências

ASSUNÇÃO, C.; ARAÚJO, C. Linguística de *Corpus*: teoria, perspectivas metodológicas e ensino das línguas. **Filologia e Linguística Portuguesa**, São Paulo, v. 21, n. 2, p. 271-288, 2019. DOI: <https://doi.org/10.11606/issn.2176-9419.v21i2p271-288>.

AUBERT, F. H. **Introdução à metodologia da pesquisa terminológica bilíngüe**. 2. ed. São Paulo: FFLCH/CITRAT, 2001.

BEILKE, N. S. V. **Pommersche Korpora**: uma proposta metodológica para compilação de *corpora* dialetais. 2016. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2016. DOI: <http://doi.org/10.14393/ufu.di.2016.426>.

BERBER SARDINHA, T. Lula e a metáfora da conquista. **Linguagem em (Dis)curso**, Tubarão, v. 8, n. 1, p. 93-120, 2008. DOI: <https://doi.org/10.1590/S1518-76322008000100005>.

BERBER SARDINHA, T. Análise de metáfora em corpora. **Ilha do Desterro**, Florianópolis, n. 52, p. 167-199, 2007. Disponível em: <https://periodicos.ufsc.br/index.php/desterro/article/view/11715>. Acesso em: 8 nov. 2021.

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

BERBER SARDINHA, T. **Metáfora**. São Paulo: Parábola, 2007.

BRADATAN, C. The philosopher of failure: Emil Cioran's heights of despair. **Los Angeles Review of Books**, Los Angeles, 28 nov. 2016. Essays. Disponível em: <https://lareviewofbooks.org/article/philosopher-failure-emil-ciorans-heights-despair/>. Acesso em: 28 jan. 2022.

CARVALHO, D. A. **VocTur**: proposta de vocabulário bilíngüe bidirecional do Turismo. 2015. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2015. DOI: <https://doi.org/10.14393/ufu.di.2015.320>. Acesso em: 28 jan. 2022.

FILLMORE, C. J. "Corpus linguistics" or "Computer-aided armchair linguistics". In: SVARTVIK, J. (ed.). **Directions in Corpus Linguistics**: proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991. Berlin, New York: Mouton de Gruyter, 1992. p. 35-60.

PARODI, G. **Linguística de Corpus**: de la teoría a la empiria. Madrid: Iberoamericana, 2010.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 6 out. 2021.

STUBBS, M. Texts, corpora, and problems of interpretation: a response to Widdowson. **Applied Linguistics**, Oxford, v. 22, n. 2, p. 149-172, 2001. DOI: <https://doi.org/10.1093/applin/22.2.149>.

TAGNIN, S. E. O. **O jeito que a gente diz**: expressões convencionais e idiomáticas. São Paulo: Disal, 2005.

Uma nova leitura de *Rayuela*: análise lexical baseada em corpus

Lidiane Carlos Ramos¹

¹ Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/3825836697652282>. E-mail: lidiane.ramos@ufu.br.

1 Introdução

A Linguística de *Corpus*, segundo Berber Sardinha e Almeida (2008, p. 18), “é uma área interdisciplinar que vem tendo um grande desenvolvimento desde a década de oitenta na Europa e, mais tarde, em outras partes do mundo, incluindo o Brasil”. Segundo Gonçalves (2008, p. 389), “vem sendo cada vez mais utilizada como um poderoso instrumento de pesquisa e observação do texto” e “possibilita um estudo mais completo do fenômeno literário”. A partir de e de acordo com essas considerações, faremos uma breve análise do romance *Rayuela* de Cortázar (1973 [1963]), em sua versão original em espanhol, partindo de uma lista de palavras (*WordList*) e, posteriormente, da análise de palavras-chave (*KeyWords*) a fim de encontrar resultados que possam caracterizar essa obra em comparação com outras obras de Cortázar por meio de um estudo do léxico.

A comparação quantitativa do léxico, que nos trouxe informações sobre a linguagem do autor e sobre as temáticas principais presentes em *Rayuela*, somente foi possível por meio das ferramentas que nos oferecem a Linguística de *Corpus* e, em especial, o programa *WordSmith Tools 4.0* (SCOTT, 2004), doravante WST, uma vez que se trata de muitos textos extensos em número de palavras, o que inviabilizaria uma análise manual. A análise de cada palavra com significativa frequência foi feita levando em conta seus contextos linguísticos, apresentados pelo concordanciador (*Concord*).

Para a realização deste estudo, empregamos o arcabouço teórico-metodológico da Linguística de *Corpus* (BERBER SARDINHA, 2004, 2009), da Metáfora Conceptual (LAKOFF; JOHNSON, 2002 [1980]), à obra *Rayuela* de Cortázar (1973 [1963]) e alguns artigos, capítulos e livros relacionados à Linguística de *Corpus* e à análise da obra de Julio Cortázar. Para apresentar a tradução dos trechos extraídos para análise, utilizamos Cortázar (2019), edição traduzida para o português por Eric Nepomuceno.

2 Sobre o autor e a obra

Julio Cortázar foi um escritor, tradutor e professor argentino, um dos grandes percursores do Realismo Fantástico na América Latina. Nascido em Bruxelas, na Bélgica, em 1914, viveu grande parte de sua vida entre a Argentina e a França. Podemos afirmar que, com grandes obras como *Rayuela*, Cortázar é um dos escritores argentinos mais reconhecidos no Brasil e no mundo.

Segundo Fioruci (2007, p. 2), o autor “unia o atrevimento literário e a aventura estética com o compromisso político em favor dos povos do terceiro mundo, num desejo de subverter não só a linguagem, mas sim de nos fazer descobrir universos que o homem é incapaz de ver”. Para Dill (2013, p. 4-5),

[...] a linguagem de Cortázar é encantatória e lúdica, feita de improvisos e imprevistos, usando construções linguísticas absolutamente inventivas. Mas, em nenhum momento, Cortázar quer mostrar-se inatingível. Pelo contrário, ele vai ao encontro do leitor e revela a sua humanidade. [...] O léxico esbarra ora para um modo coloquial, ora culto, por vezes conotativo, por vezes denotativo.

Essa linguagem de Cortázar, tão bem descrita por Dill (2013), está presente em todas as suas criações literárias, inclusive em *Rayuela*, nosso objeto de estudo. Obra consagrada do autor, *Rayuela*, romance traduzido ao português como “O jogo da amarelinha”, foi um marco da literatura do século XX pela ousadia na forma e na linguagem e por considerar o leitor um participante ativo na construção de sua leitura e interpretação. Segundo o próprio Cortázar, “*Rayuela* é um livro cuja feitura não correspondeu a nenhum plano” (PREGO, 1991, p. 99). Segundo o autor, somente quando teve “todas as páginas de *Rayuela* em cima de uma mesa, ou seja, aquela enorme quantidade de capítulos e fragmentos” é que ele sentiu “a necessidade de pôr um pouco de ordem naquilo tudo” (PREGO, 1991, p. 99). Ou seja, *Rayuela* é a união de pequenas narrativas que não foram escritas como uma unidade e, devido a essa particularidade, são possíveis as diversas leituras com diferentes desfechos, o que muitas vezes desperta no leitor um interesse especial.

Nosso propósito, neste trabalho, diferentemente de outros que buscam analisar aspectos da narrativa, é desvendar a linguagem do autor, deixando-nos guiar pelo *corpus*, utilizando a Linguística de *Corpus* e as ferramentas que ela nos oferece.

3 Metodologia

Tomamos como *corpus* de estudo a edição de 1973 de *Rayuela*, em sua versão completa e no idioma original (espanhol). Esta escolha deve-se ao fato de tratar-se de um texto longo, de um enredo complexo e representativo da obra cortazariana. Nosso *corpus* de estudo totaliza 172.581 *tokens* e 20.622 *types*.

Após obter os textos em formato .pdf, utilizamos o conversor de arquivos *on-line Convertio*² (gratuito para conversão de um número limitado de arquivos por dia) para converter os arquivos em .pdf para .txt, formato de texto plano que possibilita a leitura das obras pelo WST. Salvamos os arquivos com a codificação ANSI. Realizamos, em seguida, a limpeza dos arquivos em .txt eliminando a parte inicial de cada texto nas quais apareciam os índices (subtítulos com as respectivas páginas), os grandes espaçamentos em branco entre os parágrafos, o início de cada capítulo onde se repetia o título da obra e, no final de cada página correspondente ao arquivo original, apagamos os números das páginas que se colocavam entre as linhas de texto.

² Programa disponível em: <https://convertio.co/pt/>. Acesso em: 20 nov. 2020.

Utilizamos o recurso de edição “Localizar/Substituir” do Bloco de notas (editor de texto simples do Microsoft Windows) para eliminar alguns dados que se repetiam e que não nos interessavam para nossa pesquisa, como nome do autor, títulos e algumas referências que se repetiam a cada página em alguns dos textos. Após esses procedimentos, o *corpus* estava pronto para o processamento.

Ao abrir o programa WST, configuramos a língua para *Spanish* em *Settings > Languages*, tendo em vista proceder ao processamento do arquivo. Ao ler o texto da narrativa em .txt, salvo com codificação ANSI, utilizando a ferramenta *WordList*, desconsideramos a grande repetição de palavras de uso muito frequente como preposições, conjunções, pronomes, artigos e nomes próprios (dos personagens). Então, destacou-se, entre outras classes de palavras que não selecionamos para este estudo, a grande quantidade de substantivos que nomeiam partes do corpo (doravante, somatismos).

Após a observação do campo lexical “somatismos”, recorrente na obra, e a análise de algumas unidades fraseológicas (UFs) formadas a partir da palavra “mano” e seus derivados, buscamos responder a um questionamento que surgiu neste momento de desenvolvimento da pesquisa: ao comparar *Rayuela* a um *corpus* de referência que compilamos, composto por obras do mesmo autor, os somatismos se destacarão como característicos desta obra em particular?

Para iniciar nossa análise lexical, criamos uma primeira lista de palavras que denominamos *Rayuela – WordList 1*. Eliminamos da lista as preposições, as conjunções, os pronomes e os artigos que apareciam em grande número e não nos interessavam para esta pesquisa, também a cerquilha (#) que, segundo Berber Sardinha (2009, p. 82), “é empregada pelo WordSmith Tools para substituir algarismos”. Criamos assim a “*Rayuela – WordList 2*” com 20.622 *types* (palavras distintas).

Em uma primeira leitura do *corpus* de estudo pela *WordList*, nos chamou a atenção, especialmente, a grande recorrência de palavras de uso tipicamente argentino. É o caso do pronome pessoal de segunda pessoa do singular, “vos” (freq. 372), dos verbos conjugados concordando com esse pronome, do pronome pessoal de uso formal “usted” (freq. 115) e da interjeição “che” (freq. 117). Embora relevantes, já esperávamos esses resultados devido à nacionalidade de Cortázar, escritor argentino. Então, decidimos observar atentamente os substantivos para encontrar novos achados, que poderiam resultar em uma análise mais direcionada e menos abrangente da obra.

Observando os substantivos, identificamos a frequente repetição de palavras que denominam partes do corpo humano, como exemplificado na Tabela 1.

Tabela 1: Somatismos e suas ocorrências no *corpus* de estudo.

Palavra	Frequência
<i>mano</i>	207
<i>cara</i>	145
<i>ojos</i>	145
<i>boca</i>	118
<i>cabeza</i>	109
<i>manos</i>	98
<i>pelo</i>	77

Fonte: Elaborada pela autora.

Especialmente, notamos o uso recorrente da palavra “mano” e seus derivados em toda a narrativa que compõe nosso *corpus* de estudo.

Após definirmos como objeto de nossa análise lexical a palavra “mano”, fizemos a junção das palavras pertencentes à mesma família lexical por meio do recurso de lematização que, segundo Berber Sardinha (2009, p. 144), “é a redução das diversas formas derivadas de uma palavra à sua forma canônica ou dicionarizada, chamada de lema”. A lematização em nossa *WordList* foi feita manualmente, arrastando cada uma das palavras relacionadas semanticamente à palavra “mano”, unindo-as em uma mesma linha, como mostra a Figura 1.

Figura 1: Lematização da palavra *mano*.

N	Word	Freq.	%	Texts	%
41	YA	404	0,23	1	100,00
42	SIN	402	0,23	1	100,00
43	VOS	372	0,22	1	100,00
44	TALITA	371	0,21	1	100,00
45	VEZ	348	0,20	1	100,00
46	ESTABA	337	0,19	1	100,00
47	CUANDO	331	0,19	1	100,00
48	ESTÁ	323	0,19	1	100,00
49	MANO	321	0,12	[1] manotazo[1] manotazos[2] manoteaba[1] manoteando[5] manoteó[1] manotón[2] manotones[2]	
50	TAN	318	0,18	1	100,00
51	DOS	317	0,18	1	100,00
52	PORQUE	315	0,18	1	100,00
53	POCO	311	0,18	1	100,00
54	HAY	308	0,18	1	100,00
55	BIEN	306	0,18	1	100,00
56	ESA	296	0,17	1	100,00
57	NADA	295	0,17	1	100,00
58	SUS	292	0,17	1	100,00
59	HASTA	291	0,17	1	100,00
60	SER	285	0,16	1	100,00
61	ENTRE	284	0,16	1	100,00
62	ALGO	281	0,16	1	100,00
63	ASÍ	280	0,16	1	100,00

Fonte: *WordList*.

Após juntarmos as palavras *mano*, *manos*, *manotazo*, *manotazos*, *manotón*, *manotones*, *manualmente* e as formas do verbo *manotear*, obtivemos um total de 321 ocorrências (Figura

1). O passo seguinte foi utilizar a ferramenta *Concord* para analisar os diversos contextos em que esses vocábulos se inserem.

A ferramenta *Concord* “realiza concordâncias, ou listagens de uma palavra específica (o ‘nóculo’, *node word* ou *search word*) juntamente com parte do texto onde ocorreu” (BERBER SARDINHA, 2009, p. 8). O uso dessa ferramenta foi fundamental para localizarmos de uma maneira prática e rápida em que contexto(s) Cortázar utilizou cada uma dessas palavras.

Como não conseguimos gerar as linhas de concordância diretamente a partir da palavra “*mano*” lematizada na *WordList*, abrimos a ferramenta *Concord* e buscamos por “*man**”. Esse procedimento nos possibilitou, sem muita dificuldade, com um resultado de 763 ocorrências, fazer uma leitura atenta e eliminar todas as palavras que não fazem parte do campo lexical de “*mano*”, como “*mancha*”, “*manera*” ou flexões do verbo “*mandar*”, chegando a um resultado de 326 linhas de concordância relacionadas ao lema “*mano*” e que, portanto, nos interessam nesta pesquisa.

4 Análise e resultados

Observamos que a palavra *mano(s)*, em grande parte dos contextos analisados por meio das linhas de concordância, foi utilizada pelo autor em seu sentido literal, como parte do corpo humano, como exemplificamos nos trechos apresentados abaixo:

*[...] una mujer que me pasaba su **mano** fina y transparente por los muslos...³ (linha de concord. nº 57)*

*La Maga se quedaba triste, juntaba una hojita al borde de la vereda y hablaba con ella un rato, se la paseaba por la palma de la **mano**, la acostaba de espaldas o boca abajo...⁴ (linha de concord. nº 253)*

*En el quinto piso los esperaban Ronald y Babs, cada uno con una vela en la **mano** y oliendo a vodka barato⁵ (linha de concord. nº 275)*

*Ronald había cerrado los ojos, las **manos** apoyadas en las rodillas marcaban apenas el ritmo⁶ (linha de concord. nº 147).*

As linhas de concordância nos mostram partes do texto em que ocorreu determinada palavra e, observando cada uma delas, decidimos, para esse breve trabalho de análise das

3 “[...] uma mulher que passava a mão fina e transparente pelas minhas coxas...” (CORTÁZAR, 2019, p. 22).

4 “A Maga ficava triste, apanhava uma folhinha na beira da calçada e falava com ela um pouco, passeava a folhinha pela palma da mão, a deitava de costas ou de bruços...” (CORTÁZAR, 2019, p. 34).

5 “No quinto andar, Ronald e Babs esperavam por eles, cada um com uma vela na mão e cheirando a vodka barata” (CORTÁZAR, 2019, p. 45)

6 “Ronald havia fechado os olhos, as mãos apoiadas nos joelhos marcavam levemente o ritmo” (CORTÁZAR, 2019, p. 55).

UFs somáticas na obra de Cortázar, iniciar uma busca por ocorrências que indicassem usos metafóricos do somatismo *mano* ou de seus derivados.

Segundo Berber Sardinha (2009, p. 40), metáfora “não mais significa apenas um recurso linguístico para ornamentar o discurso literário, mas fundamentalmente um tipo de processamento mental. Esse processamento é o que nos permite entender conceitos abstratos, como amor, tempo e vida, entre outros”. A Teoria da Metáfora Conceptual, que usamos como base para a análise de algumas UFs metafóricas, foi proposta por George Lakoff e Mark Johnson no livro *Metaphors We Live By* (Metáforas da Vida Cotidiana), publicado em 1980. Segundo os autores, nossa linguagem do dia a dia está repleta de metáforas, das quais nem sempre nos damos conta e “a essência das metáforas é compreender e experienciar uma coisa em termos de outra” (LAKOFF; JOHNSON, 2002, p. 47-48).

A metáfora, segundo Lakoff e Johnson (2002, p. 45), diferentemente do que pensa a maioria das pessoas, não é exclusivamente um recurso da imaginação poética e um ornamento retórico, uma questão de linguagem extraordinária e não de linguagem ordinária. Segundo os autores, “a metáfora está infiltrada na vida cotidiana, não somente na linguagem, mas também no pensamento e na ação”.

Na expressão *en la mano del recuerdo*, percebemos o uso da palavra *mano* atribuída a *recuerdo* com um sentido metafórico.

*Con ese par de zapatos en la mano del recuerdo, el resto venía solo: la cara de doña Manuela, por ejemplo, o el poeta Ernesto Morroni. Pero los rechazaba porque el juego consistía en recobrar tan sólo lo insignificante, lo inostentoso, lo perecido.*⁷

Neste trecho em que Oliveira se lembra da Maga com saudades (trata-se do casal protagonista do romance *Rayuela*), na expressão *en la mano del recuerdo*, percebemos o uso da palavra *mano* atribuída al *recuerdo* com um sentido metafórico. É como se a “lembrança” fosse um ser que possui uma “mão (humana)” para sustentar um par de sapatos. Claramente se entende o *par de zapatos* como um objeto que traz à mente do personagem as outras recordações mencionadas. Podemos comparar a “mão da lembrança” ou, como Nepomuceno preferiu traduzir (CORTÁZAR, 2019), as “mãos da memória” a um “registro de memória”, que também é um conceito abstrato. Outra possível interpretação é que *ese par de zapatos* vem pela *mano del recuerdo*, como se se tratasse de um caminho de mão dupla por onde se transita. Essa última interpretação nos leva à metáfora conceptual **RECUERDO ES UNA PISTA DE UNA AVENIDA**.

⁷ “Com esse par de sapatos nas mãos da memória, o resto vinha sozinho: o rosto de dona Manuela, por exemplo, ou o poeta Ernesto Morroni. Mas eu rejeitava isso, porque a brincadeira consistia em recuperar apenas o que fosse insignificante, o inostentoso, o perecido” (CORTÁZAR, 2019, p. 16).

A ação de *manotear*, nos trechos a seguir, também identificados pelas linhas de concordância geradas para “man*”, assim como o substantivo concreto *mano* no exemplo anterior, se torna metafórica nas expressões criadas por Cortázar. Vejamos alguns exemplos:

[...] dijo Oliveira **manoteando** mentalmente algo que decir...⁸

Que a cada sucesiva derrota hay un acercamiento a la mutación final, y que el hombre no es sino que busca ser, proyecta ser, **manoteando** entre palabras y conducta y alegría salpicada de sangre y otras retóricas como esta.⁹

Nestes dois exemplos, o verbo “*manotear*” está relacionado com a procura por uma linguagem propícia à organização do pensamento para dizer alguma coisa. *Manotear*, que pode ser traduzido como fazer movimentos bruscos com as mãos ou gesticular, está associado, neste contexto, à ideia do abstrato, invisível, como pensar, refletir etc. Ou seja, a mente é considerada uma entidade viva que possui mãos, capaz de tocar e selecionar palavras que, por sua vez, também se tornam concretas e palpáveis. Portanto, “*manoteando*”, que aparece no gerúndio nos dois exemplos, pode ser interpretado como uma ação feita de um determinado modo, *mentalmente*: MENTE ES UN SER CON MANOS.

Para exemplificar o uso literal do verbo *manotear* por Cortázar, observemos a seguinte passagem em que uma das personagens está literalmente “tocando suas saias com as mãos” para procurar uma garrafa: “[...] dijo Emmanuèle **manoteando** entre sus innúmeras faldas para encontrar la otra botella”¹⁰. No entanto, a análise das linhas de concordância anteriores nos sugere que Julio Cortázar metaforizou as palavras *mano* e *manoteando*, em muitos casos, ao relacioná-las com atividades mentais e conceitos abstratos.

Para concluir e complementar este breve estudo, trazemos um novo exemplo:

Una **mano** de humo lo llevaba de la **mano**, lo iniciaba en un descenso, si era un descenso...¹¹

A pequena oração que inicia o trecho veiculado se constitui numa metáfora, na qual a palavra *mano* aparece duas vezes. Segundo López e Jorgensen (2009), “*llevar (alguien) de la mano*” significa guiar, orientar alguém. Ou seja, Oliveira, personagem principal do romance de Cortázar, era guiado em suas ações pela “mão de fumaça”, pela sua ação de fumar, que sugere uma personificação da expressão que é capaz de segurar a mão de uma pessoa e guiá-la em um caminho: LA MANO DE HUMO ES UN GUÍA.

8 “[...] disse Oliveira, tateando mentalmente o que dizer...” (CORTÁZAR, 2019, p. 116).

9 “De que a cada derrota sucessiva há uma aproximação à mutação final, e de que o homem não é senão o que procura ser, planeja ser, agitando as mãos entre palavras e comportamento e alegria salpicada de sangue e outras retóricas do tipo” (CORTÁZAR, 2019, p. 337).

10 “[...] disse Emmanuèle manobrando entre suas inúmeras saias até encontrar a outra garrafa” (CORTÁZAR, 2019, p. 201).

11 “Uma mão de fumaça o levava pela mão, o iniciava numa descida, se é que era uma descida...” (CORTÁZAR, 2019, p. 56).

4.1 As palavras-chave em *Rayuela*: criando um corpus de referência

Para comprovar a importância do léxico encontrado nas primeiras linhas da lista de palavras e perceber o que há de especial na composição lexical de *Rayuela*, usamos um *corpus* de referência composto por obras do mesmo autor. Como nos interessa analisar e perceber o que nos oferece *Rayuela* de diferencial em termos de léxico com relação a outras narrativas, para compor nosso *corpus* de referência, optamos pelos romances e contos de Cortázar.

Portanto, selecionamos 17 obras do mesmo autor. Dentre as obras compiladas para o *corpus* de referência, se encontram cinco obras anteriores a *Rayuela*, nove posteriores à sua edição de 1963 e três obras póstumas que foram publicadas a partir de anotações que Cortázar deixou, que foram editadas e publicadas por Carles Garriga e Aurora Bernárdez, sua ex-mulher.

A Tabela 2, a seguir, apresenta as obras que compõem nosso *corpus* de referência, juntamente com seus anos de publicação e número de *tokens* e *types*.

Tabela 2: Obras de Cortázar: *corpus* de referência.

CRONOLOGIA DE OBRAS – JULIO CORTÁZAR				
	Obra	Ano	Nº Tokens	Nº Types
01	<i>Bestiario</i>	1951	32.109	6.348
02	<i>Final del juego</i>	1956	48.677	7.725
03	<i>Las armas secretas</i>	1959	51.918	7.178
04	<i>Los premios</i>	1960	129.718	15.604
05	<i>Historias de cronopios y famas</i>	1962	22.479	5.636
06	<i>Todos los fuegos el fuego</i>	1966	50.607	7.713
07	<i>62, modelo de amar</i>	1968	87.701	11.671
08	<i>Prosa del observatorio</i>	1972	6.433	2.095
09	<i>Libro de Manuel</i>	1973	106.124	15.064
10	<i>Octaedro</i>	1974	34.926	6.066
11	<i>Alguien que anda por ahí</i>	1977	46.089	7.452
12	<i>Un tal Lucas</i>	1979	29.463	7.142
13	<i>Queremos tanto a Glenda</i>	1980	4.015	1.372
14	<i>Deshoras</i>	1982	42.485	6.616
15	<i>Divertimento (póstuma)</i>	1986	28.581	6.082
16	<i>El examen (póstuma)</i>	1986	70.856	11.047
17	<i>Papeles inesperados (póstuma)</i>	2009	133.168	18.035

Fonte: Elaborada pela autora.

Como afirma Berber Sardinha (2009, p. 225), os *corpora* de referência devem ser maiores do que o *corpus* de estudo. Neste caso, temos um *corpus* de referência aproximadamente cinco vezes maior do que o *corpus* de estudo se considerarmos o número total de *tokens*. No *corpus* de referência, que denominamos “Cortázar”, na tela *Statistics*, após criar as listas de palavras, constatamos a presença de 925.349 *tokens* e 50.473 *types*. Já no *corpus* de estudo, como mencionado anteriormente, temos um total de 172.581 *tokens* e 20.622 *types*.

Nesta etapa, os mesmos procedimentos de limpeza e conversão de formato e de extensão realizados com o *corpus* de estudo foram empreendidos na preparação do *corpus* de referência para processamento.

Para gerar a lista de palavras-chave na ferramenta *KeyWords*, optamos por restringir a frequência mínima de palavras para 5, desconsiderando as que se repetem menos de 5 vezes, e definimos o valor de *p* (poder de significância estatística) como 0,000001, que reduz a possibilidade de os resultados serem fruto do acaso. Essas configurações estão disponíveis na ferramenta *KeyWords* na aba *Settings* em *Specific to KeyWords*.

Segundo Berber Sardinha (2009, p. 8), a ferramenta *KeyWords* “[...] extrai palavras de uma lista cujas frequências são estatisticamente diferentes (maiores ou menores) do que as frequências das mesmas palavras num outro *corpus* (de referência). Calcula também palavras-chave chave, que são chave em vários textos”.

Os procedimentos descritos resultaram em uma lista de palavras-chave composta por 139 itens. Como é possível perceber na Figura 2, que apresenta um recorte da lista de palavras-chave, as palavras que haviam se destacado pela frequência na lista de palavras do *corpus* de estudo não constam na lista de palavras-chave, nem positivas e nem negativas. As palavras-chave negativas, destacadas em vermelho na Figura 2, são aquelas cuja “frequência é significativamente mais alta no *corpus* de referência do que no de estudo” (BERBER SARDINHA, 2009, p. 210). Como podemos notar na figura em questão, a lista de palavras-chave foi ordenada em ordem alfabética para confirmar a ausência da palavra *mano*. Em uma análise mais minuciosa, notamos a ausência de todo o campo lexical “partes do corpo humano”.

Figura 2: Recorte da lista de palavras-chave em ordem alfabética.

N	Key word	Freq.	%	RC. Freq.	RC. %	Keyness	P	Lemmas	Set
64	KLEE	12		2		33,61	0,0000000038		
65	LENGUAJE	41	0,02	77		25,64	0,0000004080		
66	LÉONIE	10		0		37,01	0,0000000000		
67	LLUVIA	36	0,02	50		33,40	0,0000000046		
68	LUEGO	12		276	0,03	-39,05	0,0000000000		
69	MADRID	9		1		27,15	0,0000001857		
70	MAMÁ	10		459	0,05	-97,28	0,0000000000		
71	MANDADOS	12		2		33,61	0,0000000038		
72	MATE	69	0,04	91		67,71	0,0000000000		
73	MON	26	0,02	16		45,87	0,0000000000		
74	MONDRIAN	20	0,01	1		66,32	0,0000000000		
75	MONJES	8		0		29,61	0,0000000500		
76	MONOD	7		0		25,90	0,0000003558		
77	MONSIEUR	6		191	0,02	-33,82	0,0000000031		
78	MONTACARGAS	16		0		59,21	0,0000000000		
79	MONTEVIDEO	34	0,02	12		77,13	0,0000000000		
80	MORELLIANA	9		0		33,31	0,0000000049		
81	MUNDO	150	0,09	456	0,05	32,84	0,0000000071		
82	NACIDOS	11		2		30,23	0,0000000355		
83	PALANGANAS	16		3		43,66	0,0000000000		
84	PALOMA	20	0,01	12		35,78	0,0000000001		
85	PARÍS	90	0,05	236	0,03	29,62	0,0000000496		

Fonte: KeyWords.

Como afirma Berber Sardinha (2009, p. 8) a ferramenta *KeyWords* “calcula também palavras-chave chave, que são chave em vários textos” e essa foi nossa conclusão nesta análise: a palavra *mano*, assim como todas as outras que denominam partes do corpo, são **palavras-chave chave** em Cortázar, por estarem presentes também nas 17 obras do *corpus* de referência, sem exceção, como comprovamos na Figura 3:

Figura 3: Recorte da lista de palavras do *corpus* de referência, evidenciando a frequência de *mano* e o número de arquivos em que ocorreu.

N	Word	Freq.	%	Texts	% Lemmas	Set
73	NOCHE	1.160	0,13	17	100,00	
74	MUY	1.129	0,12	16	94,12	
75	COSA	1.129	0,12	17	100,00	
76	ELLA	1.119	0,12	17	100,00	
77	MISMO	1.118	0,12	17	100,00	
78	COSAS	1.087	0,12	17	100,00	
79	SER	1.082	0,12	17	100,00	
80	TAMBIÉN	1.076	0,12	17	100,00	
81	ENTONCES	1.045	0,11	17	100,00	
82	MEJOR	1.042	0,11	17	100,00	
83	MANO	1.034	0,11	17	100,00	
84	CADA	1.030	0,11	17	100,00	
85	LADO	980	0,11	17	100,00	
86	VOS	978	0,11	15	88,24	
87	ANTES	978	0,11	17	100,00	
88	TODOS	965	0,10	17	100,00	
89	AHÍ	965	0,10	17	100,00	
90	TENÍA	957	0,10	16	94,12	
91	UNO	953	0,10	17	100,00	
92	NI	946	0,10	17	100,00	
93	MOMENTO	915	0,10	17	100,00	
94	CARA	903	0,10	17	100,00	
95	HUBIERA	872	0,09	17	100,00	

Fonte: WordList.

Esta breve análise que apresentamos em torno do somatismo *mano* foi realizada principalmente no intuito de explorar os *corpora* de estudo e de referência de Cortázar com as ferramentas do WST. O uso dessas ferramentas nos possibilitou testar as hipóteses formuladas durante nossa análise qualitativa dos contextos apresentados. A chavicidade nos indicou uma proporção de ocorrência maior do que a frequência simples, ou seja, esse critério nos possibilitou constatar que o somatismo *mano* não é só frequente na obra analisada, *Rayuela*, mas se destaca em todo o conjunto de obras selecionadas do mesmo autor.

5 Considerações finais

Neste capítulo, apresentamos uma breve análise da obra *Rayuela* de Julio Cortázar viabilizada pelas ferramentas da Linguística de *Corpus*. Nossa análise foi guiada pelos resultados obtidos através do uso das ferramentas *WordList*, *Concord* e *KeyWords* do WST.

Apesar de já haver lido o romance de maneira convencional há algum tempo e saber que a linguagem de Cortázar não é de fácil interpretação por fazer uso, muitas vezes, de recursos poéticos e linguagem metafórica, sem a Linguística de *Corpus* como aliada para a análise do texto, seria altamente laborioso e demorado chegar aos resultados que obtivemos. Especialmente, as listas de palavras e de palavras-chave tornaram possível evidenciar um campo lexical recorrente não só em *Rayuela*, mas em todas as 18 obras de Cortázar que compuseram nossos *corpora*: somatismos. Essa interessante recorrência passaria despercebida em outros tipos de leitura, pois a obra *Rayuela* é extensa e, como já mencionado, apresenta um vocabulário e enredo complexos que prendem a atenção do leitor em outros aspectos da obra, como na trama narrativa.

Como era nosso objetivo inicial perceber o que o nosso *corpus* de estudo nos mostraria nesta primeira análise com as ferramentas da Linguística de *Corpus* sem nos estendermos muito, fizemos um pequeno recorte dentre as diversas possibilidades de análise lexical. Deixamos abertas, portanto, com este trabalho, algumas possibilidades de novas descobertas e novos caminhos que poderão desvendar outros segredos guardados no “jogo da amarelinha” de Cortázar.

| Referências

BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Manole, 2004.

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

BERBER SARDINHA, T.; ALMEIDA, G. M. B. A Linguística de *Corpus* no Brasil. In: TAGNIN, S. E. O.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 17-40.

CORTÁZAR, J. **O jogo da amarelinha**. Tradução Eric Nepomuceno. São Paulo: Companhia das Letras, 2019.

CORTÁZAR, J. **Rayuela**. 15. ed. Buenos Aires: Editorial Sudamericana, 1973 [1963].

DILL, D. E. **O hábito e o estranhamento na obra de Julio Cortázar**. 2013. Disponível em: <https://home.unicruz.edu.br/mercosul/pagina/anais/2013/LINGUAGEM%20E%20DESENVOLVIMENTO%20SOCIOCULTURAL/ARTIGOS/>. Acesso em: 23 nov. 2020.

FIORUCI, W. R. Aproximações a Julio Cortazar. **Urutágua**, Maringá, n. 11, p. 1-6, 2007. Disponível em: <http://www.urutagua.uem.br/011/11fioruci.pdf>. Acesso em: 23 nov. 2020.

GONÇALVES, L. B. Linguística de *corpus* e análise literária: o que revelam as palavras-chave. In: TAGNIN, S. E. O.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 387-405.

LAKOFF, G.; JOHNSON, M. **Metáforas da vida cotidiana**. Campinas: EDUC/Mercado de Letras, 2002 [1980].

LÓPEZ, J. A. M.; JORGENSEN, A. M. **Diccionario de expresiones y locuciones del español**. Madrid: Ediciones de la Torre, 2009.

PREGO, O. **O fascínio das palavras**: entrevistas com Julio Cortázar. Tradução Eric Nepomuceno. Rio de Janeiro: José Olympio, 1991.

SCOTT, M. **WordSmith Tools version 4**. Liverpool: Lexical Analysis Software, 2004. Disponível em: <https://lexically.net/wordsmith/version4/index.htm>. Acesso em: 05 jan. 2021.

O vocabulário distópico em *Fahrenheit 451*: uma análise pautada na Linguística de *Corpus*

Terezinha de Assis Oliveira¹

¹ Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/3140346526525800>. E-mail: terezinha63@ufcat.edu.br.

1 Introdução

As inovações tecnológicas e os avanços científicos que aconteceram desde o fim do século XIX trouxeram entusiasmo e esperança para a humanidade. Com base nas concepções de Silva (2006), podemos afirmar que, como aplicação prática dos conhecimentos, as novas descobertas e técnicas traziam a possibilidade de um mundo melhor, mais correto, humano e com muita confiança no advento do novo milênio. Nesse contexto se iniciou o século XX, com a expectativa de melhores condições de vida, imbuído também com a ideia de mais tempo disponível para o lazer e a família. Ainda não eram pauta de discussão as implicações éticas entre técnica e ciência, o que efetivamente importava era a instrumentalidade e a eficácia desses novos aparatos. Todas essas mudanças impactaram sobremaneira a vida das pessoas, com uma sociedade cada vez mais voltada para o consumismo e imediatismo.

Diante desse cenário, alguns escritores deixaram registradas suas preocupações, como um aviso, por meio de suas obras. Nessas narrativas, a realidade não era apenas assumida como ela era de fato, mas as suas práticas e tendências negativas eram em muito ampliadas, o que favorecia a construção de um mundo grotesco e, na maioria dos aspectos, triste. Essas obras são distopias literárias que têm o seu foco em um futuro hipotético, no qual o controle exercido sobre os indivíduos é total e irrestrito, e tenta, inclusive, alcançar a memória dos sujeitos, como forma de controlar seu passado e, também, seu futuro.

A partir do exposto, apresentamos neste capítulo um recorte com os resultados de uma análise dos substantivos comuns presentes em *Fahrenheit 451*, de Ray Bradbury, com o objetivo de identificar como eles se apresentam e o que nos revelam sobre a obra. Nesse sentido, este texto se apoia na hipótese de nossa pesquisa de doutorado, em andamento, em que um romance distópico, com a característica de apresentar várias críticas sociais, revela também aspectos de como a linguagem é percebida e compreendida pelas pessoas, pois aborda temas como controle social, liberdade de escolha e diferenças. Sendo assim, buscamos analisar as escolhas lexicais feitas pelo autor.

Para a análise do *corpus*, utilizamos o programa *WordSmith Tools 6.0* (doravante WST) (SCOTT, 2012) para verificar as ocorrências presentes na obra, que a constituem como um alerta, visto que o autor tem como referência a observação de fatos e acontecimentos de sua época.

2 Fundamentação teórica

A Linguística de *Corpus* (doravante LC) é uma área do conhecimento que, com auxílio computacional, se ocupa do estudo da linguagem, utilizando, para esse fim, grandes quantidades de dados que se referem ao uso efetivo da língua. De acordo com Berber Sardinha (2004, p. 3),

[...] ocupa-se da coleta e exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador.

Em concordância com o autor supracitado, um *corpus* não significa apenas uma coletânea de textos, já que seu propósito é retratar a língua, ou ao menos uma parte dela. Ainda segundo Berber Sardinha (2004), alguns pontos merecem ser considerados na elaboração de um *corpus*, tais como: que os dados sejam autênticos, que o objeto de estudo seja linguístico, que os textos sejam selecionados de maneira criteriosa, sejam legíveis por computador e que o *corpus* seja vasto para que tenha representatividade. O autor apresenta uma definição mais completa que compreende essas características mencionadas, designando *corpus* como

[...] um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise (SÁNCHEZ, 1995 *apud* BERBER SARDINHA, 2004, p. 18).

Uma característica da LC é a possibilidade de observação de dados empíricos de uma ou mais línguas, ou suas variantes, que são armazenados em bancos de dados. Além disso, a LC conta com ferramentas eletrônicas que foram concebidas com o objetivo de auxiliar o pesquisador na análise, contribuindo com o trabalho de verificação dos fenômenos linguísticos.

Com a extensa possibilidade de pesquisa dentro da LC, o universo literário se apresenta como um rico e frutífero meio para investigações linguísticas, principalmente no campo lexical, possibilitando “um estudo mais completo do fenômeno literário” (GONÇALVES, 2008, p. 389). O caráter hermenêutico de estudos de textos dessa categoria exige do pesquisador uma quantidade significativa de tarefas que demandariam muito tempo, caso fossem feitas manualmente. Análises muito específicas e, em determinadas situações, repetitivas podem ser realizadas de uma maneira mais rápida e automática por meio de ferramentas computacionais que propiciam resultados com muito mais precisão e economia de tempo. Gonçalves (2004, p. 311) destaca que

[...] assim, a Linguística de *Corpus* se evidencia imediatamente como uma metodologia extremamente facilitadora do trabalho de pesquisadores e críticos literários, pois vai além de apenas listar palavras, fornecendo entre outras possibilidades palavras-chave dos textos estudados e linhas de concordâncias de diferentes tamanhos, dependendo do objetivo da pesquisa.

Dessa forma, a capacidade de processar grandes quantidades de dados de modo extremamente ágil coloca o computador como um grande aliado do especialista em estudos literários, podendo este se dedicar a focalizar mais no seu trabalho de investigação da área interpretativa, que é o cerne de atuação do crítico literário.

A opção por traçar este estudo na obra de Bradbury justifica-se, primeiramente, por compor o *corpus* da pesquisa do meu doutoramento e, também, porque queremos responder às seguintes perguntas, já que o autor apresenta como destaque a queima dos livros: se a distopia se caracteriza pelo pessimismo em relação à sociedade, com destaque aos sistemas totalitários que controlam a vida das pessoas, como se apresentam as escolhas lexicais do autor, nomeadamente os substantivos comuns? O que nos mostram esses vocábulos? São encontrados lexemas com significados de subversão, que causariam um impacto linguístico? Pensamos que o vocabulário utilizado revela aspectos como a manipulação da sociedade, que caracteriza o gênero distopia.

Em *Fahrenheit 451*, a ação da trama acontece nos Estados Unidos, em um tempo futuro, mas não são feitas referências a nenhuma cidade em especial. Apesar de o lugar não ser especificado, é apresentado como um espaço tecnologicamente sofisticado e desenvolvido, com todas as comodidades e conveniências da vida moderna, que são percebidas conforme os personagens descrevem situações cotidianas como, por exemplo, quando Montag “na porta de sua casa, enfiou a mão no orifício em forma de luva e seu toque foi identificado. A porta deslizou, abrindo-se” (BRADBURY, 2012, p. 25).

As pessoas viviam completamente alienadas e estavam entregues a supostos prazeres hedonistas², em um mundo cercado por diversões banais no qual as televisões, ou telões, eram a principal atração. Aos poucos, os hábitos de leitura e os livros foram abandonados e as pessoas chegaram a um ponto de dependência desses aparelhos que as visitas que, porventura, recebiam, eram para assistirem juntas aos mesmos programas televisivos, interagindo com os apresentadores e tornando esta uma das principais atividades a serem feitas no seu cotidiano.

Nesse mundo futurista, muitas situações são diferenciadas, sendo que uma delas é a característica das casas serem à prova de fogo. Assim, os bombeiros não apagam incêndios, mas a função deles passa a ser outra, que é oposta à que usualmente desempenhariam: devem atear fogo e destruir todo e qualquer livro que porventura encontrarem. Isso explica o próprio título do livro, *Fahrenheit 451*, que é a temperatura na qual o papel entra em combustão.

Em sua narrativa, Ray Bradbury destaca o papel do livro enquanto elemento de formação de opinião, pois ele é o responsável por existirem cidadãos conscientes e atuantes no mundo.

² Característica da pessoa adepta do hedonismo, que é a busca incessante pelo prazer como bem supremo, uma excessiva busca pelo prazer como modo de vida.

O conformismo e a alienação que são apresentados funcionam como crítica, já que esses tipos de comportamento não geram questionamentos, incertezas, mudanças e crescimento, mas tornam as pessoas alvos fáceis e passíveis de manipulação.

3 Metodologia

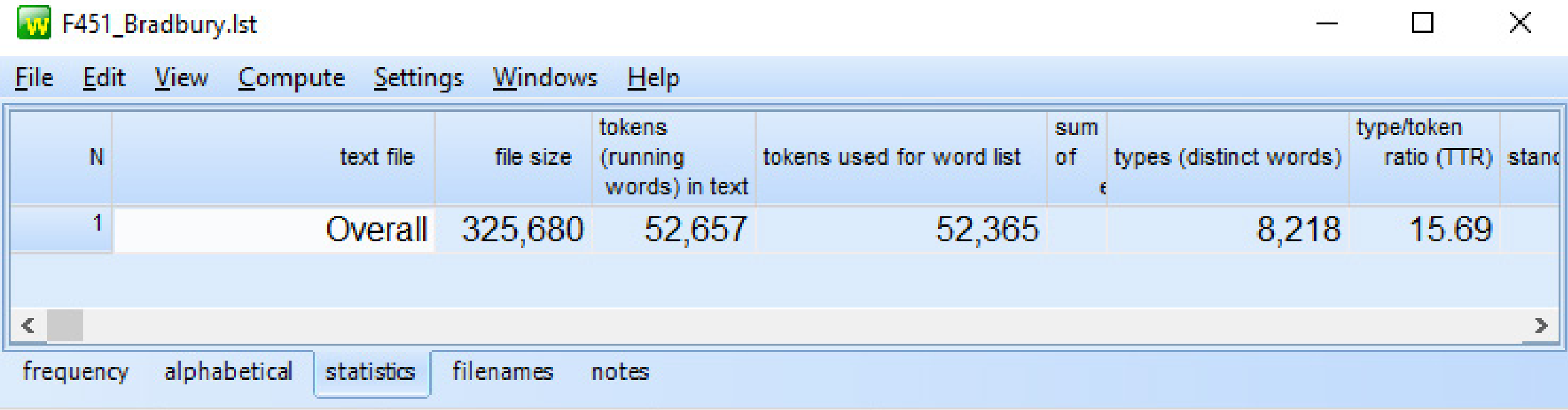
Como exposto anteriormente, este capítulo focaliza um recorte dos resultados de uma análise da obra *Fahrenheit 451*, que, juntamente com *1984* de George Orwell, *Admirável Mundo Novo* de Aldous Huxley e *O Presidente Negro* de Monteiro Lobato, compõe o *corpus* de estudo de minha pesquisa de doutoramento.

A pesquisa foi desenvolvida em várias etapas:

1. Seleção de *Fahrenheit 451*, entre as outras obras que fazem parte do *corpus*.
2. Conversão do texto em formato .pdf para .docx e, em seguida, para .txt.
3. Revisão e limpeza do arquivo em .txt.
4. Processamento do arquivo no programa WST 6 (SCOTT, 2012).
5. Geração da lista de palavras por meio da ferramenta *WordList*.
6. Geração de linhas de concordância por meio da ferramenta *Concord*.
7. Análise dos substantivos comuns no contexto expandido.

A extensão do *corpus* de estudo é a seguinte: 52.365 *tokens* (itens lexicogramaticais ou palavras totais, incluindo repetições) e 8.218 *types* (formas ou palavras individuais). A razão forma/item (*type/token ratio*) é de 15,69%. Esses dados estão indicados na Figura 1, que apresenta um recorte da tela do recurso *Statistics*, da ferramenta *WordList*.

Figura : Extensão do *corpus Fahrenheit 451*.



N	text file	file size	tokens (running words) in text	tokens used for word list	sum of types (distinct words)	type/token ratio (TTR)	stand
1	Overall	325,680	52,657	52,365	8,218	15.69	

Fonte: *WordList*.

A Figura 2 ilustra o resultado da lista de palavras ordenada por frequência (*frequency*) e por ordem alfabética (*alphabetical*), num recorte dos primeiros 20 itens.

Figura : Lista de palavras do *corpus* em ordem de frequência e alfabética.

N	Word	Freq.	%	Texts
1	DE	1,802	3.42	1
2	E	1,695	3.22	1
3	O	1,639	3.11	1
4	A	1,561	2.96	1
5	QUE	1,312	2.49	1
6	SE	852	1.62	1
7	NÃO	732	1.39	1
8	PARA	722	1.37	1
9	UM	716	1.36	1
10	UMA	568	1.08	1
11	EM	567	1.08	1
12	MONTAG	541	1.03	1
13	OS	496	0.94	1
14	ELE	481	0.91	1
15	COM	453	0.86	1
16	AS	426	0.81	1
17	DO	411	0.78	1
18	VOCÊ	401	0.76	1
19	É	381	0.72	1
20	DA	376	0.71	1

N	Word	Freq.	%	Texts
1	A	1,561	2.96	1
2	À	156	0.30	1
3	ABAFAR	1		1
4	ABAIXAR	1		1
5	ABAIXO	8	0.02	1
6	ABAIXOU	1		1
7	ABALADO	2		1
8	ABALANDO	1		1
9	ABALARA	1		1
10	ABALOU	2		1
11	ABANDONADA	1		1
12	ABANDONADO	1		1
13	ABANDONADOS	2		1
14	ABANDONAR	2		1
15	ABANDONO	1		1
16	ABARROTADO	1		1
17	ABAS	1		1
18	ABASTECER	1		1
19	ABATIDOS	1		1
20	ABELHA	2		1

Fonte: *WordList*

Ao analisar a lista de palavras, nos chamou a atenção o fato de entre essas primeiras ocorrências não estar presente nenhum substantivo comum. De fato, o único substantivo que ocorre é o nome do protagonista, Montag. Procedemos à análise dos cem (100) primeiros *types* e selecionamos todos os substantivos comuns. Para este capítulo, focalizaremos no substantivo *casa*, o segundo mais frequente no *corpus*. Passemos à análise e discussão dos dados.

4 Análise e discussão

Ao analisar as cem (100) primeiras ocorrências, a *WordList* já nos apresenta os seguintes substantivos comuns utilizados pelo autor, conforme mostra a Tabela 1.

Tabela 1: Lista de substantivos comuns por ordem de ocorrência.

Substantivo	Frequência absoluta
livros/livro	138/95
casa	129
noite	121
tempo	108
coisa/coisas	90/69
olhos	89
mundo	84
homem	78
fogo	73
pessoas	72

Fonte: Elaborada pela autora.

Essas ocorrências demonstram a opção do autor por substantivos comuns que se referem a situações do cotidiano. Essa seleção de palavras evidencia, para além da frequência, que em muitos excertos do texto elas apareceriam com proximidade.

O próximo passo da pesquisa concentrou-se na geração das linhas de concordância, que, de acordo com Berber Sardinha (2004, p. 105), “são listagens das ocorrências de um item específico (chamado de busca ou nódulo, que pode ser formado por uma ou mais palavras) acompanhado do texto ao seu redor (co-texto)”. Na Figura 3, apresentamos uma tela com as concordâncias da palavra *casa*, por ordem de ocorrência.

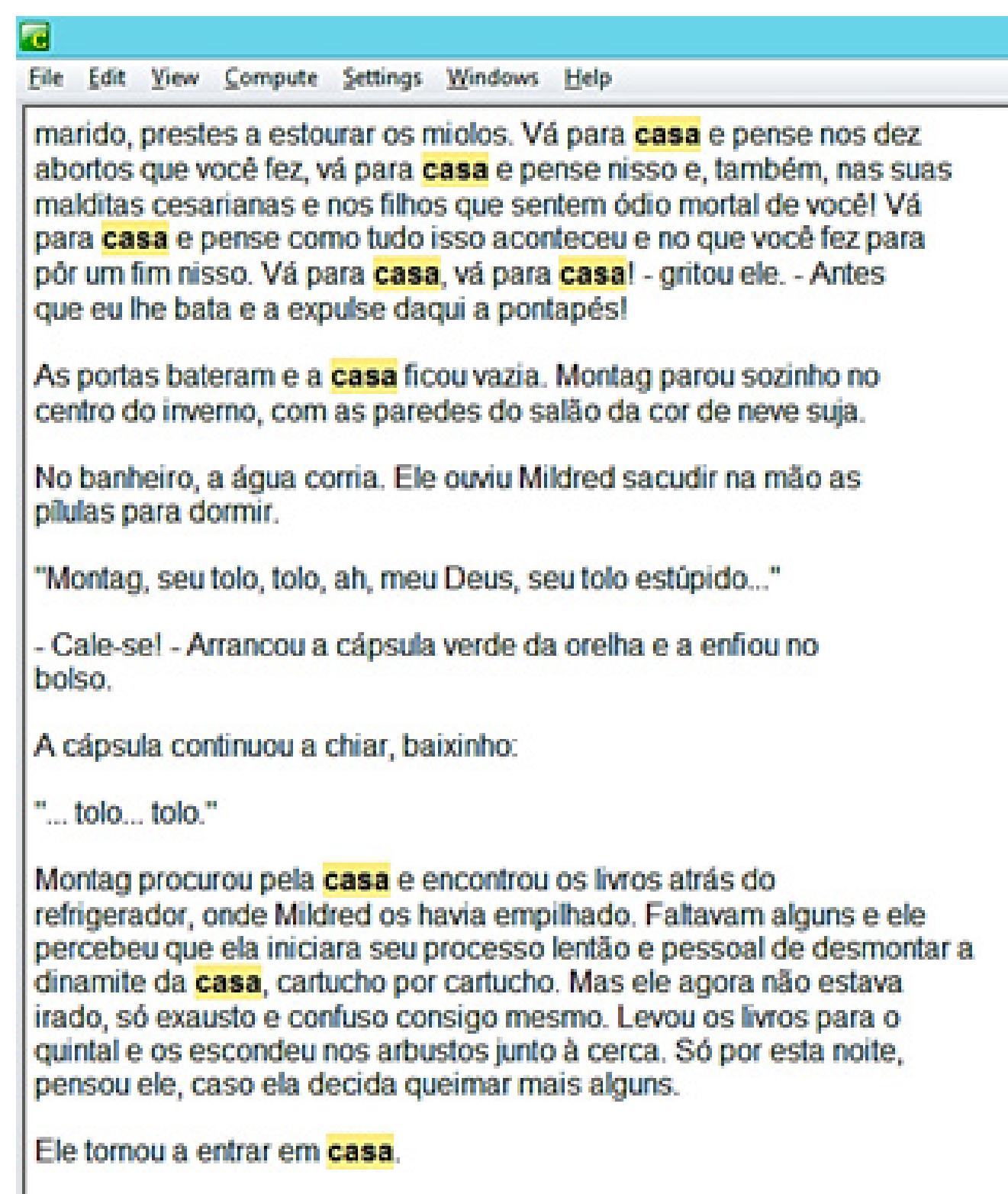
Figura 3: Recorte das linhas de concordância da palavra *casa*.



Fonte: Concord.

Na sequência passamos para a expansão das linhas de concordância para então selecionar os trechos que seriam analisados, como mostra a Figura 4.

Figura 4: Palavra *casa* em contexto expandido.



Fonte: *Concord*.

Ao analisar a palavra *casa* nesse contexto expandido da Figura 4, percebemos que, mesmo que não estejam todos os outros substantivos comuns presentes na Tabela 1, o sentido deles se faz perceber. Além de *casa* estão neste trecho *noite* e *livros*. Os vocábulos *pessoas* e *fogo* não ocorrem, mas se fazem perceber pelo contexto, pois Montag discute com uma amiga de Mildred (*pessoas*) e depois decide esconder os livros “[...] caso ela decida queimar mais alguns.” (BRADBURY, 2012, p. 121), em referência ao *fogo*.

5 Considerações finais

Ao discorrer sobre a narrativa distópica de *Fahrenheit 451*, percebemos como Ray Bradbury apresenta sua preocupação com o futuro da humanidade diante de situações desencadeadas pelo desenvolvimento cada vez maior da tecnologia e ciência. Ele expõe circunstâncias que retratam o cotidiano das pessoas, pois, enquanto escritor, sabia o quão importante eram os livros na formação, pois o foco da sua crítica recai também na morosidade e passividade das pessoas.

Retomando, com base em nossa hipótese da pesquisa de base – segundo a qual um romance distópico, com a característica de apresentar várias críticas sociais, revela também aspectos de como a linguagem é percebida e compreendida pelas pessoas –, acreditamos que a obra

de Bradbury cumpre o seu papel social de alerta, já que a língua é indiscutível na construção social e é um instrumento essencial nas relações humanas e no fomento da cultura.

A análise das escolhas lexicais revela que os substantivos comuns que se apresentam nessa obra distópica são palavras da rotina das pessoas, ratificando que o autor se baseava na observação de fatos e acontecimentos de seu tempo, para deixar registrado o seu aviso. Como resposta à última pergunta, até o momento não foram encontrados vocábulos com significados de subversão ou de difícil compreensão. Como o próprio Bradbury pontuou: “não escrevo para prever as coisas, mas para evitá-las” (SILVA, 2006, p. 319). À guisa de conclusão, consideramos que *Fahrenheit 451* traz à baila discussões e questionamentos que são tão importantes hoje quanto o foram na época em que foi escrita.

Referências

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

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

BRADBURY, R. **Fahrenheit 451**: a temperatura na qual o papel do livro pega fogo e queima. Tradução Cid Knipel. Rio de Janeiro: Biblioteca Azul, 2012.

GONÇALVES, L. B. Linguística de *Corpus* e análise literária: o que revelam as palavras-chave. In: TAGNIN, S. E. O.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008, p. 387-405.

GONÇALVES, L. B. Linguística de *Corpus* e análise literária: uma aplicação a *Dubliners* de Joyce. **Crop**, São Paulo, v. 10, p. 309-328, 2004. Disponível em: https://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/lourdes_joyce.pdf. Acesso em: 29 out. 2021.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://lexically.net/wordsmith/version6/>. Acesso em: 29 out. 2021.

SILVA, A. M. **Literatura inglesa para brasileiros**: curso completo de cultura e literatura inglesa para brasileiros. Rio de Janeiro: Ciência Moderna, 2006.

O nome “Barack” em *Minha História* (2018) sob a perspectiva sistêmico-funcional

Sarah Cristina de Oliveira Sebba¹

¹ Doutoranda em Estudos da Linguagem pela Universidade Federal de Catalão. Lattes: <http://lattes.cnpq.br/5523573670912856>. E-mail: sarah_sebba@hotmail.com.

1 Introdução

A escolha da obra *Minha História* (OBAMA, 2018), autobiografia de Michelle Obama, se deve ao fato de esta obra ter se tornado um sucesso, um recorde em números de vendas e acessos em diversas plataformas digitais. Foi também considerada inspiradora em vários aspectos, tanto na esfera social, das lutas de classe, como no âmbito cultural e político, no que tange à representatividade da mulher contemporânea.

Diante disso, focalizamos a análise do nome “Barack” no discurso da autora, que não apenas ocupa o papel social de mulher, mas também de esposa, ícone e influenciadora social. Vale ressaltar que a análise linguística holística e semiótica centrada no uso, ou seja, sistêmica e funcional, visa entender a produção dos significados em situações de comunicação reais, o que torna pertinente o uso de *corpora* em nossa investigação.

Nesse íterim, a Linguística de *Corpus* (LC) propõe que as palavras não são dispostas de modo aleatório nos textos. Sob uma visão estatística, Halliday (1994) enfatiza que a linguagem compreende um sistema de probabilidades, cuja face mais notável é a frequência de uso das palavras. Logo, a LC promove a possibilidade de indagar *corpus*, com o objetivo de desvendar as relações linguístico-culturais, por meio de pesquisas que valorizam a representatividade das escolhas lexicais em atos de fala e de escrita reais (TEUBERT *et al.*, 2004).

De acordo com Halliday e Matthiessen (2004), Halliday e Hasan (1989), Eggins (1994), Martin e Rose (2003), Thompson (2004) e Martin e White (2005), a Linguística Sistêmico-Funcional (LSF) considera os conceitos de linguagem, língua, cultura e identidade por meio do uso e de seus contextos de produção. Também com base nos estudos da Tradução (BAKER, 1993, 1999), pretendemos associar as leituras sobre normas de Toury (1995), voltadas aos contextos socioculturais de produção de Textos Meta (TMs), aos princípios da LSF e à compreensão de como os tradutores de Michelle Obama lidaram com os contextos de situação, no tocante ao uso do nome “Barack”.

As seções subsequentes apresentam a fundamentação teórica, os procedimentos metodológicos do estudo em questão, os resultados obtidos com uma breve discussão, bem como as considerações finais.

2 Princípios da linguística sistêmico-funcional

Segundo Almurashi (2016), a LSF foi desenvolvida por Michael Halliday a partir dos anos 1960 no Reino Unido e concentra-se em analisar os discursos tanto orais quanto escritos, com foco na função da linguagem. Os termos “Sistêmico” e “Funcional” são explicados por Fuzer e

Cabral (2014), quando identificam que sistemas linguísticos e estruturas de linguagem podem cooperar para o significado em um texto e, em seguida, explicar o efeito ou impacto que tais combinações podem causar ao utilizarmos da linguagem. Assim:

[...] todo e qualquer uso que fazemos do sistema linguístico é funcional relativamente às nossas necessidades de convivência em sociedade. Ao usarmos a linguagem fazemos, portanto, uma série de escolhas dentre as possibilidades que o sistema linguístico disponibiliza. Em vista disso, precisamos desenvolver nossa consciência sobre os significados que as palavras e suas combinações em textos geram para alcançarmos efetivamente nossos propósitos em contextos específicos (FUZER; CABRAL, 2014, p. 19).

Dessa forma, uma visão sistêmico-funcionalista propõe explorar a estruturação de uma língua em variados contextos, focando a linguagem pelas relações sociais, que motivam, conscientemente, as escolhas lexicogramaticais. Sendo assim, a funcionalidade recai no fato de a gramática interpretar as formas linguísticas, bem como a descrição semântica, uma vez que a linguagem, considerada prática social, sempre terá um objetivo, uma finalidade específica.

O interesse da LSF baseia-se em verificar como se realiza a comunicação, compreender a relação entre texto e contexto se faz necessária. O “texto”, para Halliday e Matthiessen (2004, p. 4-5), é “[...] qualquer instância da linguagem, em qualquer meio, que faz sentido a alguém que conhece a linguagem”. Uma das grandes contribuições da LSF é, certamente, lançar um olhar para a língua que ultrapassa sua estrutura interna, e que englobe toda a esfera social exercendo a função comunicativa, uma visão global, do todo e não mais de partes estanques. Assim, podemos perceber os textos para além dos seus aspectos linguísticos, mas também considerando suas características extralinguísticas, que aqui são consideradas como contexto de cultura e de situação.

A LSF é uma perspectiva teórico-metodológica que não vê a língua dissociada do contexto em que ela se insere. Nesse sentido, Fuzer e Cabral (2014, p. 26) afirmam que “[...] o contexto em que o texto se desenvolve está encapsulado no texto através de uma relação sistemática entre o meio social e a organização funcional da linguagem”. Pode-se depreender que, segundo Halliday (1994, 2004), a língua é percebida a partir de dois conceitos-chave relativos ao contexto, ao gênero e ao registro, que podem estar relacionados tanto à cultura, quanto à situação, pois:

[...] o gênero pode ser também denominado contexto de cultura, e o registro, contexto de situação. O gênero é entendido como um nível mais abstrato, mais geral do contexto; enquanto o registro diz respeito a um conceito semântico que corresponde a um nível intermediário entre o contexto de uso e a linguagem, isto é, o lugar onde as variáveis linguísticas são realizadas. Desta forma, o registro descreve o contexto situacional no qual o texto é produzido (ALMEIDA, 2010, p. 19).

Visto isso, considera-se, aqui, que a obra de Michelle Obama realiza o contexto situacional, que pode ser descrito por meio do Registro. Lembrando que os termos registro e gênero para a LSF referem-se respectivamente ao contexto de situação (microcontexto – o que está mais próximo ao texto) e ao contexto de cultura (macrocontexto – valores, crenças já estabelecidas). Estes conceitos apresentados serão explorados ao longo da pesquisa e das análises, uma vez que, para a LSF, o “texto” está intrinsecamente ligado aos contextos em que a língua é instanciada.

2 Material e método

Um *corpus* paralelo bilíngue unidirecional (inglês-português), constituído pela obra *Becoming*, de Michelle Obama (2018), e a respectiva tradução para o português, sob o título *Minha História*, elaborada por Débora Landsberg, Denise Bottman e Renato Marques e distribuída no mesmo ano em nosso país.

3 Metodologia

Esta pesquisa se insere no paradigma quanti-qualitativo. Pretendemos produzir resultados analisados, não apenas por meio de procedimentos estatísticos ou outras técnicas de quantificação, mas também pelos métodos de interpretação qualitativa.

Além de o estudo ser de cunho bibliográfico, consideramos que a sua natureza é descritiva, porque categorizamos as escolhas lexicais e semântico-discursivas usadas por Michelle Obama no Texto Fonte (TF) e seus tradutores, respectivamente. Também pode ser caracterizada como analítica, por apurarmos e interpretarmos as representações discursivas selecionadas que compõem o *corpus*.

Para a organização do *corpus*, utilizamos o *software LancsBox*[®], versão 5.0 (BREZINA; WEILL-TESSIER; McENERY, 2020), que classifica e nomeia os dados em cinco categorias, a saber: KWIC, GRAPHCOLL, WHELK, WORDS, NGRAMS, TEXT e WIZARD. De maneira sucinta, cada uma está ligada a determinada funcionalidade do referido programa e de suas ferramentas. Especificamente em nossa investigação, utilizamos:

- GRAPHCOLL, que apresenta as colocações em formato de gráfico;
- WHELK, na qual são organizadas as informações de comportamento do item lexicogramatical ou como ele está distribuído no *corpus* de pesquisa;
- WORDS, que produz uma lista de palavras, as identifica e as visualiza em determinado texto;
- NGRAMS, com a qual se encontra um grupo de palavras que aparecem mais vezes juntas – o N se refere ao número de palavras a serem identificadas pelo usuário;
- TEXT, que apresenta o contexto relacionado ao item previamente escolhido para a pesquisa.

4 Análise dos dados e discussão

Com base no uso inicial da ferramenta *LancsBox*[®], levantamos o número de itens lexicais (total de palavras) e a quantidade de palavras diferentes no texto. Além disso, vale ressaltar aqui que o *software* tem a função de lematizar o *corpus*, por isso verificamos os lemas, ou seja, uma representação canônica das entradas de um dicionário (BIDERMAN, 1999), conforme a Tabela 1:

Tabela 1: Número de itens, formas e lemas.

CORPUS	ITENS	FORMAS	LEMAS
<i>"Becoming"</i>	163.508	14.543	12.890
<i>"Minha História"</i>	162.693	17.733	18.892

Fonte: Elaboração da autora.

Notamos que o TM possui 2.815 palavras a menos que o TF, embora haja 3.190 formas a mais; logo, os tradutores obtiveram uma densidade vocabular maior. No que diz respeito a WORD, observamos que a palavra mais frequente, nos dois *subcorpora* do *corpus* paralelo, é o nome "Barack", relacionado ao então ex-presidente dos Estados Unidos e marido de Michelle; desse modo, corroboramos a pertinência da proposta inicial do estudo proposto.

A seguir, apresentamos as Tabelas 2 e 3 por ordem de frequência, que indicam as palavras mais frequentes com base substantival (nomes) e a frequência absoluta no *corpus*:

Tabela 2: Palavras de maior frequência no TF.

SUBSTANTIVO (NOME)	FREQUÊNCIA
<i>Barack</i>	667
<i>Time</i>	372
<i>People</i>	350
<i>House</i>	303
<i>School</i>	301
<i>Day</i>	296
<i>Work</i>	248
<i>Years</i>	248
<i>Home</i>	241
<i>Life</i>	214

Fonte: Elaboração da autora.

Tabela 3: Palavras de maior frequência no TM.

SUBSTANTIVO (NOME)	FREQUÊNCIA
Barack	877
Casa	499
Anos	348
Dia	324
Vida	307
Tempo	305
Pessoas	267
Trabalho	240
Família	199
Escola	177

Fonte: Elaboração da autora.

Nesse prisma, depreendemos que a tradução para a língua portuguesa utilizou um conjunto maior de palavras para construir significados semelhantes, com variação na frequência dos substantivos mais utilizados (em comparação à tabela anterior), principalmente no que concerne ao nome “Barack”, abordado 210 vezes a mais no TM do que no TF. Assim, podemos pressupor que a avaliatividade relativa ao nome próprio mais frequente utilizado pela autora também deve ser percebida ao substituí-lo por pronomes que se refiram a ele.

Convém salientar que no TM e nos excertos analisados a seguir, respectivamente, há a tradução literal, a substituição de palavras (nomes por pronomes, por exemplo), o acréscimo e a supressão de vocábulos e, novamente, a adição de palavras. A autora só se refere a “Barack” ao final da primeira parte do livro (composto de três partes), quando acontece o primeiro contato profissional entre ambos.

A Gramática Sistêmico-Funcional (GSF) denomina Processos, no plano da Semântica do discurso, aquilo que a Gramática Tradicional trata como verbos. Cada Processo é ligado a, pelo menos, uma das três metafunções da linguagem que, de acordo com Fuzer e Cabral (2014, p. 32), dizem respeito às manifestações, “[...] no sistema linguístico, dos propósitos que estão subjacentes a todos os usos da língua: compreender o meio (ideacional), relacionar-se com os outros (interpessoal) e organizar a informação (textual)”.

Interessa-nos evidenciar, no discurso de Michelle Obama, as representações dos participantes envolvidos nos processos das orações. Nesse sentido, a metafunção ideacional e sua função experiencial utilizam a oração pelo sistema de transitividade e demonstram como ocorre a representação do indivíduo ao expressar suas experiências, sejam elas materiais ou a nível de consciência (FUZER; CABRAL, 2014).

Em linhas gerais, a GSF, por meio da TRANSITIVIDADE, considera três componentes da oração: o Processo (grupos verbais), os Participantes (grupos nominais) e a Circunstância (grupos adverbiais). Diante dos dados apresentados nas Tabelas 4 e 5, selecionamos 67 excertos que compõem a primeira parte do livro, para eles serem analisados sob o viés da LSF, principalmente no TM:

Tabela 4: Tipos de processos e suas ocorrências no TM em orações com “Barack”.

TIPOS DE PROCESSO	QUANTIDADE	PORCENTAGEM
Material	24	35,82
Mental	14	20,90
Relacional	16	23,88
Comportamental	5	7,46
Verbal	7	10,45
Existencial	1	1,49
TOTAL	67	100

Fonte: Elaboração da autora.

Tabela 5: “Barack” como participante dos processos e suas ocorrências no TM.

PARTICIPANTES NO PROCESSO MATERIAL	QUANTIDADE
Ator	16
Meta	2
Escopo processo	2
Beneficiário recebedor	0
Beneficiário cliente	4
Atributo	0
PARTICIPANTES NO PROCESSO MENTAL	QUANTIDADE
Experenciador	6
Fenômeno	8
PARTICIPANTES NO PROCESSO RELACIONAL	QUANTIDADE
Portador	9
Atributo	0
Identificado	4
Identificador	2
Possuído	0
Possuidor	1

PARTICIPANTES NO PROCESSO COMPORTAMENTAL	QUANTIDADE
Comportante	4
Comportamento	1
PARTICIPANTES NO PROCESSO VERBAL	QUANTIDADE
Dizente	6
Verbiagem	0
Receptor	1
Alvo	0
PARTICIPANTES NO PROCESSO EXISTENCIAL	QUANTIDADE
Existente	0
Circunstância	1
TOTAL	67

Fonte: Elaboração da autora.

De acordo com Halliday e Matthiessen (2014), três processos principais elencam as experiências do ser humano – Materiais (referem-se ao fazer), Mentais (relativos à consciência) e Relacionais (que estabelecem relações) – e outros três que estariam à margem dos primeiros: Comportamentais (comportamento humano), Existenciais (representação do que existe ou acontece) e Verbais (atividades linguísticas e processos do dizer). Nesse caso, cada um trata os participantes e as circunstâncias sob óticas diversas e recebe outras denominações.

Coincidentemente (ou não), os três principais processos pensados *a priori* por Halliday e Matthiessen (2014) foram os mais recorrentes na parte da obra analisada. Destacamos o Processo Material, que apareceu 35,82% das vezes, isto é, 24 ocorrências de 67 excertos, seguido pelo Relacional, com 23,88% (16 ocorrências), e o Mental, com 20,90% (14 ocorrências).

O Processo Material conta com vários tipos de participantes, a saber: Ator, Meta, Escopo-entidade, Escopo-processo, Beneficiário recebedor, Beneficiário cliente, Atributo (uma característica dos participantes) e várias Circunstâncias, como Extensão, Localização, Modo, Causa, Contingência, Acompanhamento, Papel, Assunto e Ângulo.

As Tabelas 4 e 5 explicitam as ocorrências do *corpus* analisado e comprovam a densidade dos números, que foram posteriormente analisados em *Microsoft Office Word*[®]. Em suma, as escolhas linguísticas da autora demonstram que ela decide representar o então marido e ex-presidente dos Estados Unidos por meio de processos que estabelecem mudanças no decorrer dos eventos, em que ele é o participante envolvido, na maioria das vezes, por esse fazer/acontecer. Isso lhe confere uma posição de positividade frente à realidade apresentada e ao contexto em questão.

No entanto, para este estudo, visamos demonstrar que, pela teoria da LSF, mais precisamente da metafunção experiencial, as escolhas linguísticas da autora indicam as representações e as experiências de Barack no mundo – isso reforça que a presente análise não é simplista, tampouco especulativa, mas sim baseada em fatos, ao invés das hipóteses. Seguem alguns exemplos pertinentes para reafirmar que a classificação de transitividade, mesmo que em uma análise inicial, não se encerra neste momento. De fato, existem outros desdobramentos e podemos ir além da identificação de participantes, processos e circunstâncias; porém, os elementos abordados no texto são suficientes para cumprir o propósito deste trabalho. Aqui priorizamos o participante Barack e o(s) processo(s) envolvido(s) nele e focalizamos na obra traduzida:

Barack Obama	chegou atrasado no primeiro dia
Ator	Oração Material

Quando chegou o outro dia de <i>happy hour</i> , deixei	Barack
Oração Material	Beneficiário Cliente

Barack	sabia exatamente como eu me sentia
Experienciador	Oração Mental Cognitiva

Barack	era sério sem se levar muito a sério
Portador	Oração Relacional

Barack	tinha um sorriso que parecia se esticar por toda a extensão do rosto
Possuidor	Oração Relacional

A seguir, apresentamos algumas considerações, uma vez feitas as análises ao longo desta seção.

5 Considerações finais

Interessante ressaltar que o mais difícil, talvez, é salientar que escolhas diferentes podem representar um mesmo construto (MATTHIESSEN, 2001). Não se espera que os textos de partida e de chegada sejam exatamente iguais, dado que a retextualização do tradutor precisa passar por novas decisões de textualização e salvaguardar os contextos e as experiências de vida da autora. Conforme os excertos analisados, há importância de a atividade tradutória em se alinhar aos pensamentos de Halliday (1985, 1994) ou de Halliday e Matthiessen (2004). Eles

veem a linguagem como uma prática social comunicativa, na qual as escolhas dos usuários (autor e tradutor) se (inter)relacionam em uma estrutura sociocultural mais abrangente, o que ocorre nas relações interpessoais, na construção da experiência humana e na organização da mensagem.

Até o momento, a positividade presente na autobiografia analisada nos faz questionar acerca da voz autoral de Michelle quanto a esse gênero. A perfeição de seu marido nunca é tangida quanto a um caráter duvidoso ou que apresente angústias, debilidades, entre outros, uma vez que ela recorre a recursos negativos somente em termos voltados à moda e à vestimenta de seu companheiro (deficiente noção de estilo). Isso nos possibilita questionar o fato de tal recurso ser empregado para humanizar Barack em seu texto, algo que ainda está em investigação, principalmente quanto ao modo de se comunicar e transmitir confiança, seriedade etc. Vale ressaltar que estas considerações se baseiam na análise das orações elencadas na primeira parte do livro.

Destarte, em sua autobiografia, a autora, ao mencionar o então marido, realiza escolhas linguísticas avaliativas de apreciação por meio de adjetivos e predicativos do sujeito que o colocam frequentemente em uma posição de excelência, o que nos leva a considerar os pontos que a aproximam do gênero autobiográfico e os que a distanciam de fato. Tais questionamentos serão realizados em uma análise mais aprofundada e, por enquanto, ficam a título de reflexão.

Referências

ALMEIDA, F. A. S. D. P. **A avaliação na linguagem**: os elementos de atitude no discurso do professor. São Carlos: Pedro & João, 2010.

ALMURASHI, W. A. The effective use of YouTube videos for teaching English language in classrooms as supplementary material at Taibah University in Alula. **IJELLR**, v. 4, n. 3, p. 32-47, 2016. Disponível em: <https://www.eajournals.org/wp-content/uploads/The-Effective-Use-of-Youtube-Videos-for-Teaching-English-Language-in-Classrooms-as-Supplementary-Material-at-Taibah-University-in-Alula.pdf>. Acesso em: 7 dez. 2021.

BAKER, M. Corpus linguistics and translation studies: implications and applications. *In*: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (ed.). **Text and technology**. Amsterdam/Philadelphia: Benjamins, 1993. p. 233-250.

BAKER, M. The role of corpora in investigating the linguistic behaviour of professional translators. **International Journal of Corpus Linguistics**, v. 4, n. 2, p. 281-298, 1999. DOI: <https://doi.org/10.1075/ijcl.4.2.05bak>.

- BIDERMAN, M. T. C. Conceito linguístico de palavra. **Palavra**, Rio de Janeiro, v. 1, p. 81-97, 1999.
- BREZINA, V.; WEILL-TESSIER, P.; McENERY, A. **#LancsBox version 5**. Lancaster: Lancaster University, 2020. Disponível em: <http://corpora.lancs.ac.uk/lancsbox>. Acesso em: 14 jan. 2022.
- EGGINS, S. **An introduction to Systemic Functional Linguistics**. London: Pinter, 1994.
- FUZER, C.; CABRAL, S. R. S. **Introdução à Gramática Sistêmico-Funcional em língua portuguesa**. Campinas: Mercado de Letras, 2014.
- HALLIDAY, M. A. K. **An introduction to Functional Grammar**. London: Arnold, 1985.
- HALLIDAY, M. A. K. **An introduction to Functional Grammar**. 2. ed. London: Edward Arnold, 1994.
- HALLIDAY, M. A. K.; HASAN, R. **Language, context, and text: aspects of language in a social-semiotic perspective**. 2. ed. Oxford: Oxford University Press, 1989.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. I. M. **An introduction to functional grammar**. 3. ed. Oxford: Oxford University Press, 2004.
- MARTIN, J. R.; ROSE, D. **Working with discourse: meaning beyond the clause**. London: Continuum, 2003.
- MARTIN, J. R.; WHITE, P. R. R. **The language of evaluation: appraisal in English**. London: Palgrave MacMillan, 2005.
- MATTHIESSEN, C. I. M. The environments of translation. *In*: STEINER, E.; YALLOP, C. (ed.). **Beyond content: exploring translation and multilingual text**. Berlin: de Gruyter, 2001.
- OBAMA, M. **Becoming**. New York: Crown, 2018.
- OBAMA, M. **Minha história**. Tradução Débora Landsberg, Denise Bottman, Renato Marques. Rio de Janeiro: Objetiva, 2018.
- TEUBERT, W. *et al.* **Lexicology and Corpus Linguistics: an introduction**. London: MGP Books, 2004.
- THOMPSON, G. **Introducing Functional Grammar**. 2. ed. London: Arnold, 2004.
- TOURY, G. **Descriptive translation studies and beyond**. Amsterdam: J. Benjamins, 1995.

Análise de orações existenciais em *corpus* literário bilíngue

Victor Mariotto Palma¹

¹ Mestrando em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/6432124077851888>. E-mail: victorm@ufu.br.

1 Introdução

O presente trabalho é parte de uma pesquisa de mestrado em andamento, cujo objetivo é identificar e analisar a ocorrência de Processos Existenciais em *corpus* paralelo bilíngue constituído pelos contos wildianos de *A house of pomegranates* e duas traduções para o português brasileiro, a primeira de 1961 e a segunda de 2012. Por meio desta pesquisa, buscamos testar a hipótese da retradução (BERMAN, 2017) e contribuir para a descrição sistêmico-funcional da realização de significados existenciais no português brasileiro, pelo viés dos estudos da tradução.

As orações existenciais (OE) importam ao discurso, pois realizam gramaticalmente a existência e o acontecimento. A identificação e a quantificação dos Processos Existenciais nos três *subcorpora* de pesquisa foram realizadas com o auxílio de *software* e segundo os pressupostos metodológicos da Linguística de *Corpus*. Os resultados foram, então, analisados e cotejados à luz dos preceitos da Linguística Sistêmico-Funcional.

Apresenta-se aqui resultados parciais da análise das ocorrências de orações existenciais realizadas por Processos Existenciais prototípicos em inglês e em português em *corpus* literário.

2 Orações existenciais

Conforme apontam Halliday e Matthiessen (2014), a linguagem constrói a experiência humana por meio do sistema de TRANSITIVIDADE subsumido ao componente experiencial da metafunção ideacional, que organiza o fluxo de eventos na forma de *figuras*, compostas de Processo, Participante e Circunstância, dos quais o primeiro é o elemento central e é responsável por criar a figura, cuja experiência é realizada na oração. Para o sistema de TRANSITIVIDADE da língua inglesa, os autores arrolam seis tipos de Processos, a saber, Material, Mental, Relacional, Comportamental, Verbal e Existencial.

O Processo Existencial, foco de nossa pesquisa, situa-se entre os Processos Material de acontecimento e Relacional de atribuição, ou seja, “os processos existenciais estão localizados no limite do mundo das relações abstratas e do mundo físico” (FERREGUETTI; PAGANO; FIGUEREDO, 2012, p. 283).

Por meio do Processo Existencial, constata-se a existência de todo tipo de fenômeno, indica-se que algo existe ou acontece. Embora pouco frequente no discurso, esse Processo pode prestar-se a introduzir participantes no estágio inicial de uma narrativa, apresentar fenômenos ao longo do curso da narrativa ou indicar lugares de interesse (HALLIDAY; MATTHIESSEN, 2014; LIMA; BARROS, 2018).

A figura de *existir* compõe-se obrigatoriamente do Processo Existencial e de um participante, o Existente. Eventualmente, Circunstâncias de tempo ou de lugar podem ligar-se à figura.

3 Metodologia

O material de análise desta pesquisa compõe-se de três *subcorpora*: *corpus* Wilde, em inglês, constituído da versão digitalizada dos quatro contos que integram a coletânea wildiana *A house of pomegranates* – publicada pela primeira vez em 1891 e, hoje, disponibilizada em Internet Archive; e os *corpora* Mendes e Salgado, em português brasileiro, contendo a versão digitalizada das traduções de 1961 (*Uma casa de romãs*, de Oscar Mendes) e 2012 (*A casa das romãs*, de Luciana Salgado) da coletânea wildiana. O *corpus* tem as seguintes características (Tabela 1):

Tabela 1: Características do *corpus* de estudo.

CARACTERÍSTICA	WILDE	MENDES	SALGADO
tamanho	173.952	172.212	173.430
<i>tokens</i> no texto	33.466	30.460	30.575
<i>token</i> para lista de palavras	33.466	30.459	30.575
types	3.780	5.546	5.407
TTR	25,67	18	18
TTR normalizada	41,10	46,21	46,29
TTR normal. desvio padrão (base 1.000)	46,87	48,20	48,08
orações	1.558	1.610	1.587
parágrafos	540	812	521

Fonte: Dados da pesquisa.

Para mapear os Processos Existenciais utilizados nos contos, recorreu-se às ferramentas *WordList* e *Concord* da suíte *WordSmith Tools 4* (SCOTT, 2004): com o primeiro, extraíram-se, inicialmente, os dados gerais do *corpus* de estudo e uma lista de palavras, por meio da qual se localizou a forma *there*; e com o segundo, apuraram-se as 64 linhas de concordâncias de interesse deste estudo – aquelas que constituem Processos Existenciais – com base nas 100 ocorrências totais da forma *there*.

A fim de se localizarem as traduções correspondentes às linhas de concordância de interesse, os *subcorpora* foram alinhados com o *Wordfast Anywhere* (WORDFAST, 2021), gerando-se uma planilha eletrônica com cada *corpus* em uma coluna distinta.

4 Resultados e discussão

Apresentam-se, a seguir, os resultados parciais da análise das orações existenciais no *corpus* de estudo.

4.1 OE em inglês

As OE em inglês são prototipicamente realizadas por *there be*, em que *there*, originalmente um locativo, não tem outra função que não a de sujeito do verbo em nível interpessoal, e nada realiza no sistema de TRANSITIVIDADE. Ademais, outros verbos lexicais podem realizar OE com o uso de *there*, que ocupa, então, a posição temática, de modo a indicar que se vai apresentar um Existente, geralmente um ente não específico (PAGANO; FIGUEREDO; FERREGUETTI, 2012).

Davidse (1999) distingue OE cardinais e OE enumerativas que quantificam e enumeram, respectivamente, as instanciações do hiperônimo a que corresponde o Existente. Nas cardinais, o Existente pode ou não ser específico e *there* é opcional, ao passo que, nas enumerativas, o Existente é específico (genérico ou definido) e a presença de *there* é obrigatória.

No *corpus* de estudo, foram identificadas 64 ocorrências de OE prototipicamente realizadas por *there be* (Tabela 2):

Tabela 2: Ocorrências dos Processos Existenciais prototípicos no *corpus* de estudo.

PROCESSOS VERBAIS	CONTO 1	CONTO 2	CONTO 3	CONTO 4	TOTAL
There is	5	1	16	5	27
There are	0	0	2	1	3
There was	0	7	5	7	19
There were	0	5	5	0	10
There be	0	0	2	0	2
There being	1	0	0	0	1
There will be	0	0	1	0	1
There had been	1	0	0	0	1
Total	7	13	31	13	64

Fonte: Dados da pesquisa.

A maior ocorrência de OE em *The fisherman and his soul* (O pescador e sua alma, conto 3) deve-se, em parte, à maior extensão do conto e ao fato de ser, em grande medida, um relato que a personagem Alma faz de sua peregrinação pelo mundo a um jovem pescador.

Em sua maioria, os Existentes compreendem entes não específicos (45 ocorrências), como em “under the shadow of a rock **there was a figure that had not been there before**” (“sob a sombra de um rochedo, havia uma figura que não estava ali antes”). Também são maioria as OE cardinais. Foram identificadas apenas cinco instâncias de OE enumerativas, como “**There were opals and sapphires**, the former in cups of crystal, and the latter in cups of jade” (“Havia opalas e safiras, as primeiras em taças de cristal e as segundas em taças de jade”).

4.2 OE em português

Diferentemente do inglês, o verbo *haver*, realizador prototípico de OE no português, não é acompanhado de elemento que exerça a função interpessoal de sujeito – a figura de existir compõe-se apenas do Processo e do Existente e eventualmente de Circunstâncias de tempo e lugar.

Pelo viés da semântica discursiva, Franchi, Negrão e Viotti (1998) arrolam verbos lexicais que podem realizar OE em português brasileiro, notadamente os verbos *ter*, *existir* e *haver* (aqui apresentados em ordem decrescente de frequência no *corpus* dos autores). Afirmam ainda os autores que “as construções existenciais com *ter* constituem uma singularidade do PB [...] em relação às construções existenciais com *haver*” (FRANCHI; NEGRÃO; VIOTTI, 1998, p. 106, grifos dos autores).

Analisando a ocorrência de OE em dois *corpora*, um monolíngue e outro paralelo bilíngue português-italiano, Ferreguetti, Pagano e Figueredo (2012) identificaram que nas traduções havia menos OE e que, em alguns casos, ou foram substituídas por outros Processos (notadamente relacionais, materiais e mentais) ou não foram realizadas.

No *corpus* de estudo, verificou-se uma tendência a se traduzir as OE prototípicas inglesas pelo verbo *haver* em português, que responde por 68,75% das ocorrências no *subcorpus* Mendes e 65,63% no *subcorpus* Salgado; o verbo *existir*, por sua vez, responde por 7,81% em Mendes e 21,87% em Salgado.

Tabela 3: Ocorrências dos Processos Existenciais prototípicos nas traduções.

PROCESSOS VERBAIS	MENDES	SALGADO	TOTAL
Haver	44	42	86
Existir	5	14	19
Total	49	56	105

Fonte: Dados da pesquisa.

Não se verificou nenhuma ocorrência do grupo verbal *ter* na realização de Processo Existencial na tradução das OE prototípicas inglesas do *corpus*. Nesse sentido, os dados divergem dos resultados de Franchi, Negrão e Viotti (1998), ressalvadas as diferenças metodológicas deste e daquele estudo.

Além disso, a análise dos textos alinhados permitiu perceber mudanças de tipo de Processo nos textos traduzidos como resultado das escolhas tradutórias para os verbos lexicais que no texto fonte realizam Processos Existenciais, como indica o Quadro 1:

Quadro 1: Traduções das OE prototípicas inglesas.

WILDE	MENDES	PROCESSO	SALGADO	PROCESSO
The Lizards were extremely philosophical by nature, and often sat thinking for hours and hours together, when there was nothing else to do, or when the weather was too rainy for them to go out.	As Lagartixas são extremamente filosóficas por natureza e muitas vezes passam horas e horas ininterruptas a pensar, quando não há outra coisa a fazer, ou quando o tempo está demasiado chuvoso para que elas possam sair.	Existencial	Lagartixas são extremamente filosóficas por natureza, e sempre se sentam juntas por horas e horas, quando não têm nada mais a fazer, ou quando o tempo está chuvoso demais para poderem sair.	Relacional
There was no other sound save the sound of a wave fretting the smooth pebbles below.	Não se ouvia ruído senão o das ondas agitando-se sobre os seixos polidos.	Mental	Não havia nenhum outro som além do barulho das ondas roçando os seixos macios lá embaixo.	Existencial
There are nine gates to this city	Tem nove portas essa cidade	Relacional	há nove portões naquela cidade	Existencial

Fonte: Dados da pesquisa.

Outros Processos que traduzem as OE prototípicas inglesas constantes do *corpus* de estudo são *escutar*, *estar*, *obter*, *repartir* e *ser*. Ademais, quatro instâncias em Mendes e duas em Salgado não realizaram nenhum Processo, como: “[there is a little city hard by] in which **there is** a garden of tulip-trees”, por “[Muito perto daqui há uma pequena cidade] com um jardim de tulipas” no *subcorpus* Mendes.

5 Considerações finais

Apresentou-se, aqui, uma pequena parte dos dados de uma pesquisa de mestrado em andamento. Percebe-se, entretanto, que a metodologia e o tipo de análise se mostram úteis para a investigação de OE em *corpus* paralelo bilíngue unidirecional inglês-português.

Assim como em Ferregueti, Pagano e Figueredo (2012), verificou-se também menos OE nas traduções, o que pode ser explicado pelas escolhas tradutórias que produziram outros Processos, quais sejam, Mental e Relacional, ou simplesmente não realizaram nenhum Processo. Ainda assim, verificou-se que no *subcorpus* Salgado há mais Processos Existenciais traduzindo as OE prototípicas inglesas do que no *subcorpus* Mendes, o que pode corroborar a hipótese da retradução de Berman (2017). Conforme essa hipótese, a primeira tradução de um texto tende a se aproximar mais da língua e da cultura do público da tradução, ao passo que a retradução tende a se aproximar mais do texto-fonte.

Ademais, dada a limitada extensão dos dados analisados, os resultados devem ser tratados apenas como indicativos de tradução e não podem ser generalizados.

Referências

BERMAN, A. A retradução como espaço da tradução. **Cadernos de Tradução**, Florianópolis, v. 37, n. 2, p. 261-268, 2017. DOI: <https://doi.org/10.5007/2175-7968.2017v37n2p261>.

DAVIDSE, K. The semantics of cardinal versus enumerative existential constructions. **Cognitive Linguistics**, Birmingham, v. 10, n. 3, p. 203-250, 1999. DOI: <https://doi.org/10.1515/cogl.2000.001>.

FERREGUETTI, K.; PAGANO, A. S.; FIGUEREDO, G. P. Significados existenciais no português brasileiro: um estudo contrastivo em textos traduzidos e não traduzidos. In: ENCONTRO DE LINGUÍSTICA DE CORPUS, 10., 2012, Belo Horizonte. **Anais [...]**. Belo Horizonte: Faculdade de Letras da UFMG, 2012. p. 280-293. Disponível em: <http://www.lettras.ufmg.br/site/e-livros/Anais%20do%20X%20Encontro%20de%20Lingu%C3%ADstica%20de%20Corpus.pdf>. Acesso em: 6 nov. 2021.

FRANCHI, C.; NEGRÃO, E. V.; VIOTTI, E. Sobre a gramática das orações impessoais com ter/haver. **DELTA**, São Paulo, v. 14, n. esp., p. 105-131, 1998. DOI: <https://doi.org/10.1590/S0102-44501998000300009>.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **Halliday's introduction to functional grammar**. 4. ed. Milton Park Abingdon Oxon: Routledge, 2014.

LIMA, L. R.; BARROS, N. C. Contribuição de orações existenciais às fases de uma narrativa. In: CABRAL, S. R. S.; BARBARA, L. (org.). **Estudos de transitividade em linguística sistêmico-funcional**. Santa Maria: UFSM; PPGL, 2018. Disponível em: <https://repositorio.ufsm.br/bitstream/handle/1/18870/Livro%20SF1.pdf?sequence=1&isAllowed=y>. Acesso em: 6 nov. 2021.

PAGANO, A. S.; FIGUEREDO, G. P.; FERREGUETTI, K. Mapeamento das orações existenciais no português brasileiro. *In: ENCONTRO DE LINGUÍSTICA DE CORPUS*, 10., 2012, Belo Horizonte. **Anais** [...]. Belo Horizonte: Faculdade de Letras da UFMG, 2012. p. 240-251.

Disponível em: <http://www.lettras.ufmg.br/site/e-livros/Anais%20do%20X%20Encontro%20de%20Lingu%C3%ADstica%20de%20Corpus.pdf>. Acesso em: 6 nov. 2021.

SCOTT, M. **Wordsmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 6 nov. 2021.

WORDFAST. **Wordfast Anywhere**. Version 6.2.5. New York: Wordfast, 2021. Disponível em: <https://www.freetm.com/>. Acesso em: 6 nov. 2021.

Como se traduz “língua de acolhimento”? **Análise exploratória em corpus paralelo bilíngue¹**

Joel Victor Reis Lisboa²

1 Este trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior – Brasil – CAPES – Código do financiamento 001.

2 Doutorando em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/7657573383244824>. E-mail: joelvictorlisboa@gmail.com.

1 Introdução

Em pesquisas anteriores (LISBOA, 2020, 2021a, 2021b), constatamos, a partir de *corpora* especializados monolíngues (português), que há uma alta variação terminológica formal e conceitual na área de Português como Língua Não Materna (PLNM), à qual se vincula a área de Português como Língua de Acolhimento (PLAc). Neste estudo, objetivamos analisar e descrever traduções de **língua de acolhimento, Português como Língua de Acolhimento e PLAc**, bem como verificar se há variações formais nessas traduções. Nossa hipótese é que se não há padronização em português, o mesmo poderá ser verificado em língua inglesa, haja vista que a alta variação terminológica é um traço expressivo e característico da área de PLNM.

Este estudo vincula-se à Terminologia e utiliza a Linguística de *Corpus* como metodologia. Para realizar as análises, partimos de um *corpus* paralelo bilíngue (português-inglês) unidirecional, composto por resumos e palavras-chave de dissertações e teses brasileiras vinculadas à área de PLAc. Esse *corpus* compreende dois *subcorpora*, um em português, oriundo de um estudo bibliométrico realizado anteriormente (LISBOA *et al.*, 2021), e um em inglês, compilado especificamente para este estudo exploratório. As análises foram realizadas por meio do utilitário *Aligner* do *WordSmith Tools* 6.0 (SCOTT, 2012), doravante WST.

Na seção seguinte, introduzimos alguns pontos teóricos que subjazem a este estudo. Em seguida, apresentamos a Linguística de *Corpus*, o *corpus* e os procedimentos metodológicos realizados. Por fim, discutimos os dados obtidos, tecemos nossas considerações finais e apresentamos alguns possíveis encaminhamentos futuros.

2 Fundamentação teórica

PLNM é uma área acadêmico-profissional vinculada à Linguística Aplicada que, grosso modo, se ocupa da pesquisa e do ensino-aprendizagem de português para/por pessoas não lusófonas. Essa área se desdobra em diferentes subáreas, dentre elas, o PLAc (LISBOA, 2021b). Sumariamente, PLAc é uma área voltada à pesquisa e ao ensino-aprendizagem da língua portuguesa para/por (i)migrantes de crise, isto é, por pessoas não lusófonas, recém imigradas em países de língua majoritária portuguesa, que se encontram em vulnerabilidade socioeconômica e emocional em virtude do deslocamento, em geral forçado, a que foram submetidas (LOPEZ; DINIZ, 2018; OLIVEIRA, 2019; SILVA; COSTA, 2020).

No escopo de PLAc, o conceito de língua de acolhimento é central, inclusive funcionando como delimitador da própria área e de suas esferas de atuação. Esse conceito, que começou a ser popularizado no início do século XXI em Portugal (GROSSO, 2010; OLIVEIRA, 2019), refere-se a uma língua que não é materna para o aprendente. Seu uso é majoritário na sociedade de

imigração e “vinculado à urgente necessidade de participação social dos recém-chegados e à vivência de uma nova cultura para que, conseqüentemente, possam atuar efetivamente na sociedade e em suas organizações” (SOARES; TIRLONI, 2019, p. 83). Em vista disso, este é um conceito muito próximo daquele de segunda língua, mas ambos possuem traços conceituais que nos permitem diferenciá-los. Uma discussão mais aprofundada desses conceitos e de suas fronteiras conceituais foi desenvolvida por Lisboa (2021b).

Sob o prisma da Terminologia, **língua de acolhimento** e **Português como Língua de Acolhimento** são concebidos como unidades fraseológicas especializadas (UFEs). Para compreender o que entendemos por UFEs, vale retomar o que são unidades terminológicas (UTs). Em síntese, UTs são signos que representam nódulos no sistema conceitual de uma área de especialidade, veiculando conceitos, processos, propriedades, técnicas etc., basilares para a área da qual fazem parte (REY, 1995). Quando assumem a forma de signos linguísticos, as UTs podem ser compostas por um só item lexical ou podem ser agrupamentos lexicais sintagmáticos de valor terminológico, o que designamos como UFEs³.

Como mencionado na seção introdutória, em pesquisas anteriores, a partir de *corpora* especializados monolíngues (português), voltamo-nos a análises de UFEs designativas de área, subáreas e de conceitos de língua do PLNM, e constatamos que a variação formal e conceitual é característica marcante da terminologia analisada. Todavia, como essas pesquisas foram realizadas em *corpora* monolíngues, alguns questionamentos surgiram: como essas UFEs são traduzidas? Há padronização nessas traduções? Haveria mais de um equivalente tradutório? Se sim, haveria alguma forma mais recorrente e estável?

Tendo em vista esses questionamentos, focalizamos neste estudo as UFEs **língua de acolhimento** e **Português como Língua de Acolhimento** e, por conseguinte, o acrônimo **PLAc**, partindo da hipótese de que, assim como em português, não haveria padronização terminológica nas traduções para o inglês, haja vista que a alta variação é um traço característico das utilizações terminológicas no PLNM e, portanto, no PLAc. A metodologia utilizada para realizar as análises está descrita na seção seguinte.

³ É relevante evidenciar que as designações do que chamamos de UTs e de UFEs variam na literatura da área. As UTs são também designadas como termos, lexias especializadas, signos terminológicos, unidades lexicais terminológicas/especializadas etc. Por sua vez, o que chamamos de UFEs também pode ser encontrado sob as seguintes designações: termos sintagmáticos, sintagmas terminológicos, unidades terminológicas/especializadas poliléxicas, fraseologismos terminológicos/especializados, colocações especializadas, dentre outras. Em alguns casos, as diferentes designações pressupõem diferentes conceitualizações, isto é, a variação formal acontece em função da (ou para marcar a) variação conceitual, mas, em outros casos, as diferentes designações não necessariamente representam diferentes conceitualizações. Devido à limitação de espaço, optamos por apenas assinalar essa variação, mas não nos ocuparemos dessa discussão neste capítulo.

3 Metodologia

A Linguística de *Corpus* (doravante LC) foi utilizada neste estudo como metodologia, principalmente por facilitar a organização dos dados e, por conseguinte, agilizar a observação simultânea das UFEs-alvo e de suas respectivas traduções em seus contextos de ocorrência. Em síntese, a LC se constitui como uma metodologia/abordagem⁴ de análise e descrição linguística, realizadas com auxílio computacional e com base em *corpora*. As pesquisas em LC privilegiam a empiricidade, a replicabilidade, a utilização de ferramentas computacionais e a análise linguística em abrangentes quantidades de textos autênticos (BERBER SARDINHA, 2004).

Em LC, *corpus* é concebido como “uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo de pesquisa a que se destina” (TAGNIN, 2013, p. 29). Como estão intimamente vinculados aos objetivos da pesquisa, há diferentes tipos de *corpora*, dentre eles, os *corpora* paralelos. Grosso modo, um *corpus* paralelo (unidirecional) compreende dois *subcorpora*, um composto por textos originais e outro composto por traduções desses mesmos textos, sendo que sua maior vantagem “reside justamente no aproveitamento da sua disposição em paralelo”, que possibilita “pesquisar originais e tradução em simultâneo” (FRANKENBERG-GARCIA, 2008, p. 118).

Sendo assim, nosso *corpus* de análise também é composto por dois *subcorpora*, um de resumos e palavras-chave de dissertações e teses em português e outro composto pelas respectivas traduções em língua inglesa. O *subcorpus* em português é oriundo de um estudo bibliométrico realizado anteriormente (LISBOA *et al.*, 2021), em que analisamos dissertações e teses da área de PLAc disponibilizadas na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), obtidas por meio da expressão de busca “língua de acolhimento”. No estudo mencionado, com base nos metadados, resumos e palavras-chave, apresentamos um panorama quali-quantitativo da área de PLAc com base nas pesquisas obtidas. Todavia, nossas análises não se restringiram à terminologia e aos possíveis equivalentes tradutórios utilizados nesses arquivos.

A partir das tabelas⁵ de análise construídas no estudo supracitado, selecionamos os resumos e palavras-chave para compor o *subcorpus* em português. Para a construção do *subcorpus* em inglês, retornamos aos documentos obtidos na pesquisa anterior e compilamos as traduções dos resumos e palavras-chave presentes nessas dissertações e teses.

4 Discussões sobre o estatuto da LC estão presentes desde os primeiros manuais publicados. Seria ela uma área, uma teoria, uma abordagem ou uma metodologia? Devido às limitações de espaço e ao escopo deste capítulo, não entramos nessa discussão neste texto, apenas sinalizamos que, assim como nós, grande parte dos pesquisadores a consideram como uma metodologia (quando utilizada somente como um instrumental para análise, obtenção ou testagem de dados) ou abordagem (quando utilizada como perspectiva de estudo da língua que parte de princípios específicos, como os de probabilidade, padronização e não aleatoriedade da variação linguística). Para um breve panorama da discussão sobre o estatuto da LC, ver Shepherd (2009) e Parodi (2010).

5 Essa tabela auxiliou não somente na compilação dos *subcorpora*, mas também na criação dos códigos de nomeação dos arquivos, pois todos os metadados das dissertações e teses estavam minuciosamente detalhados.

Sumariamente, este é um *corpus* paralelo bilíngue (português-inglês) unidirecional, composto por resumos e palavras-chave de 12 dissertações e 6 teses da área de PLAc disponibilizadas na BDTD, defendidas em 8 instituições de ensino superior entre os anos de 2016 e 2020. O *subcorpus* em português possui 6.837 *tokens* e 1.715 *types*, ao passo que a dimensão do *subcorpus* em inglês é de 6.638 *tokens* e 1.479 *types*.

Em síntese, os procedimentos metodológicos realizados foram os seguintes:

- Formulação dos códigos⁶ de nomeação, tendo em vista facilitar a identificação e padronizar a nomeação dos arquivos de ambos os *subcorpora*;
- Organização⁷ e atribuição de códigos aos arquivos do *subcorpus* em português;
- Compilação e atribuição de códigos aos arquivos do *subcorpus* em inglês;
- Conversão da codificação dos arquivos por meio do utilitário *Text Converter* do WST⁸;
- Alinhamento paralelo (por períodos) de cada arquivo no utilitário *Aligner*⁹ do WST;
- Análise e tabelamento dos dados, gerando uma tabela por arquivo.

4 Resultados e discussão

Antes de passarmos à análise das traduções, evidenciamos que em um dos arquivos não houve ocorrências das UFEs ou acrônimo-alvo no resumo ou nas palavras-chave. Além disso, é importante fazer alguns apontamentos sobre a variação formal identificada em ambos os *subcorpora*. Em relação à UFE **língua de acolhimento**, identificamos no *subcorpus* em português variação entre letras iniciais maiúsculas e minúsculas, prevalecendo o uso de minúsculas. No que tange à UFE **Português como Língua de Acolhimento**, além da variação entre iniciais maiúsculas e minúsculas, identificamos a alternância na utilização do advérbio “como”, mas observamos que a tendência é a utilização de letras iniciais maiúsculas e do advérbio. Essas observações também foram verificadas por Lisboa (2021b).

6 As informações contidas nesses códigos são, respectivamente, língua (PT/EN), ano de defesa, sigla/abreviação da instituição de ensino, especificação do documento (dissertação ou tese) e último sobrenome do autor. O ordenamento das informações foi pensado visando a possibilitar, em pesquisas futuras, a identificação de possíveis padrões tradutórios vinculados principalmente aos anos de defesa ou a determinadas instituições de ensino.

7 Nas etapas de organização do *subcorpus* em português e de compilação do *subcorpus* em inglês, cada resumo e grupo de palavras-chave foi copiado e colado no Bloco de Notas do Windows para que fosse salvo em .txt, formato mais profícuo para processamento pela suíte de análise lexical escolhida.

8 O *Text Converter* é um utilitário que permite a conversão em massa de arquivos em diversas codificações para a codificação Unicode. Esse procedimento é necessário porque o WST geralmente apresenta erros de leitura quando os arquivos são processados em outra codificação que não a Unicode.

9 O utilitário *Aligner* foi utilizado por fazer o alinhamento paralelo por períodos de forma semiautomática, viabilizando a análise simultânea de originais e traduções e, por conseguinte, facilitando a identificação das UFEs e acrônimo-alvo e de suas respectivas traduções.

Nos equivalentes tradutórios no *subcorpus* em inglês, também observamos a variação formal quanto à utilização de iniciais maiúsculas e minúsculas, bem como no uso da conjunção “as” e do artigo “a” (ex. **Portuguese (as (a)) Host Language**), casos similares às variações identificadas em português. Além disso, especificamente no equivalente **host language**, observamos que, por vezes, ele é grafado entre aspas, mas essas utilizações não são tão representativas em nosso *corpus* de estudo (4 ocorrências em apenas 2 arquivos).

Visando a simplificar a apresentação e discussão dos resultados, optamos por apenas assinalar as variações formais referentes à utilização de iniciais maiúsculas e minúsculas e de aspas, mas desconsiderá-las nas tabelas a seguir. Dito isso, iniciemos pelas traduções de **Português (como) Língua de Acolhimento**. Em nosso *corpus* de estudo, identificamos três possíveis traduções, sendo que duas delas são mais frequentes e uma ocorre em apenas um arquivo. Também identificamos um caso em que a UFE em questão não foi traduzida. Esses dados estão apresentados na Tabela 1.

Tabela 1: Traduções de Português (como) Língua de Acolhimento.

UNIDADE-FONTE	TRADUÇÕES	FREQUÊNCIA	ARQUIVOS
Português (como) Língua de Acolhimento	Portuguese (as a) Welcoming Language	17	8
	Portuguese (as (a)) Host(ing) Language	17	7
	Portuguese as a Shelter Language	2	1
	-	1	1

Fonte: Elaborada pelo autor.

Portuguese (as a) Welcoming Language ocorreu como tradução da UFE em questão 17 vezes em 8 arquivos distintos. Percebemos a preferência pela grafia utilizando “as a”, haja vista que **Portuguese Welcoming Language** ocorreu duas vezes, mas em apenas um arquivo. Por sua vez, **Portuguese (as (a)) Host(ing) Language** ocorreu também 17 vezes, mas em 7 arquivos distintos. Quanto às preferências de grafia, observamos que **Portuguese as a Host Language** foi a mais recorrente (10 vezes em 4 arquivos), seguida de **Portuguese Host Language** (3 vezes em 2 arquivos). As demais variações (**Portuguese as Host Language** e **Portuguese as a Hosting Language**) ocorreram em apenas um arquivo cada¹⁰. Em nosso *corpus*, identificamos também **Portuguese as a Shelter Language**, ocorrendo duas vezes em um único arquivo.

Em relação às traduções de **língua de acolhimento**, identificamos quatro possibilidades tradutórias, sendo que **host(ing) language** foi a mais recorrente (7 vezes em 4 arquivos). Além disso, observamos que o segundo caso mais recorrente foi a não tradução da UFE-fonte (2 vezes em 2 arquivos). Esses dados estão apresentados na Tabela 2, a seguir.

¹⁰ O total de arquivos apresentados nessa discussão difere do total de arquivos exibido na Tabela 1, pois em um mesmo arquivo identificamos duas variantes sendo utilizadas (**Portuguese as a Host Language** e **Portuguese Host Language**).

Tabela 2: Traduções de língua de acolhimento.

UNIDADE-FONTE	TRADUÇÕES	FREQUÊNCIA	ARQUIVOS
língua de acolhimento	host(ing) language	7	4
	-	2	2
	PWLg	2	1
	welcoming language	1	1
	Portuguese as a Shelter Language	1	1

Fonte: Elaborada pelo autor.

No que se refere a **host(ing) language**, a tendência é a grafia como **host language**, visto que **hosting language** ocorre 2 vezes, mas em apenas um arquivo. As demais traduções, **PWLg** (em referência à **Portuguese Welcoming Language**), **welcoming language** e **Portuguese as a Shelter Language**, ocorreram como tradução da UFE em questão em apenas um arquivo cada.

Comparando as informações apresentadas nas Tabelas 1 e 2, é possível perceber que ao traduzir **Português (como) Língua de Acolhimento**, geralmente opta-se por UFEs com os constituintes **welcoming** e **host**, mas no caso das traduções de **língua de acolhimento**, o constituinte **welcoming** não é comum, pois a tendência é a utilização de **host**. Portanto, duas UFEs que possuem constituintes idênticos em português (língua de acolhimento) são traduzidas com diferentes constituintes a depender da UFE-fonte.

Nosso *corpus* também nos permitiu analisar as traduções de **PLAc**, utilizado em português como acrônimo de **Português como Língua de Acolhimento**. Os dados referentes a essas traduções estão apresentados na Tabela 3.

Tabela 3: Traduções de PLAc.

UNIDADE-FONTE	TRADUÇÕES	FREQUÊNCIA	ARQUIVOS
PLAc	-	12	9
	PLAc	5	4
	Portuguese as a Welcoming Language	9	3
	PHL	10	2
	PWLg	4	1
	Portuguese as a Shelter Language	4	1
	host language	3	1
	PWL	2	1
	Portuguese	1	1
	Portuguese teaching for immigrants and refugees	1	1
	Portuguese for forced immigrants	1	1

Fonte: Elaborada pelo autor.

Como é possível observar na tabela apresentada, o mais comum é a não tradução de **PLAc**, observada por meio da supressão do acrônimo (12 vezes em 9 arquivos) ou pela utilização do mesmo acrônimo-fonte na tradução (5 vezes em 4 arquivos). Os dois outros casos mais comuns são a não utilização de um acrônimo/sigla/abreviatura na tradução, optando-se por apresentar a UFE-tradução por extenso (**Portuguese as a Welcoming Language**), e a tradução por **PHL** (em referência a **Portuguese as (a) Host Language**). Os demais casos ocorrem em apenas um arquivo cada, apesar de alguns ocorrerem mais de uma vez em um mesmo arquivo. Os dois últimos casos apresentados na Tabela 3 nos chamaram a atenção por não serem especificamente UFEs, mas por se constituírem como explicitações do conceito ao qual **PLAc** faz referência (cf. Seção 2).

Por fim, observamos alguns casos singulares. Em um dos arquivos, há a ocorrência de uma UFE que se constitui como uma extensão de **Português como Língua de Acolhimento**, a saber, **Português como Língua de Acolhimento para Fins Acadêmicos**, sendo traduzida por **Portuguese as a Welcoming Language for Academic Purposes**, e cujo acrônimo utilizado tanto em inglês como em português é **PLAc-FA**. Em outro arquivo, identificamos duas palavras-chave (língua portuguesa e língua de acolhimento) que foram traduzidas por uma só UFE (**Portuguese as "host language"**). Além desses casos, identificamos em um arquivo que **Português (como) Língua de Acolhimento** foi abreviado como **PLA**, tanto em inglês como em português, diferenciando-se da forma abreviada mais frequente (**PLAc**).

5 Considerações finais

Neste estudo exploratório, voltamo-nos à análise de traduções das UFEs **língua de acolhimento, Português como Língua de Acolhimento** e, por conseguinte, do acrônimo **PLAc** em um *corpus* paralelo bilíngue (português-inglês) unidirecional, composto por resumos e palavras-chave de 12 dissertações e 6 teses brasileiras vinculadas à área de **PLAc**. Nossa hipótese inicial foi confirmada. Identificamos que, assim como nas UFEs em português, há variações formais nas traduções, especificamente em relação à utilização de letras iniciais maiúsculas/minúsculas e de aspas, ao uso de conjunção e artigo (*as a*) e à alternância dos constituintes das UFEs. Não obstante, salvo algumas exceções, a variação formal referente à alternância de unidades constituintes não foi identificada em um mesmo arquivo, mas na comparação entre arquivos distintos. Portanto, especificamente no caso de alternância de constituintes, as traduções em um mesmo arquivo são geralmente padronizadas.

A tendência observada neste *corpus* de estudo é que **Português (como) Língua de Acolhimento** seja traduzido por **Portuguese as a Welcoming Language** (equivalente mais recorrente) ou por **Portuguese as a Host Language**. Por sua vez, a UFE **língua de acolhimento**

tende a ser traduzida por **host language** (caso mais recorrente) ou não é traduzida em inglês. Por fim, **PLAc** tende a não ser traduzido (suprime-se o acrônimo ou utiliza-se o mesmo acrônimo utilizado em português); em outros casos menos frequentes, opta-se por apresentar por extenso a UFE a que **PLAc** se refere ou explicitar o conceito ao qual esse acrônimo está vinculado, sem necessariamente fazer uso de uma UFE específica no texto traduzido.

Alguns dos possíveis encaminhamentos futuros deste estudo são os seguintes: (i) analisar a presença de padrões tradutórios vinculados aos anos de defesa ou a determinadas instituições de ensino¹¹; (ii) compilar e analisar um *corpus* em inglês, composto por textos especializados em língua de acolhimento, para verificar se essas mesmas variações são encontradas e se há algum tipo de padronização; (iii) analisar a terminologia em língua inglesa utilizada por organizações ou agências oficiais voltadas ao tema de migração e refúgio (como a ACNUR¹²), tendo em vista verificar se as traduções aqui identificadas são utilizadas por esses órgãos ou se há um outro tipo de padrão terminológico que poderia ser utilizado como equivalente das UFEs analisadas neste estudo.

À guisa de conclusão, evidenciamos a proficuidade da LC como metodologia para estudos de descrição linguística, principalmente por facilitar a organização e observação de dados, agilizando o processo de análise, inclusive em estudos de caráter mais manual como o aqui apresentado. Além disso, este estudo demonstra que é possível reutilizar/reconstruir, ou “reciclar” (FROMM; YAMAMOTO, 2021), *corpora* originalmente compilados em pesquisas anteriores, desde que o (re)desenho atenda aos objetivos da nova pesquisa.

Referências

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

FRANKENBERG-GARCIA, A. Compilação e uso de *corpora* paralelos. In: TAGNIN, S. E. O.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 117-136.

FROMM, G.; YAMAMOTO, M. I. Compilação, reciclagem e padronização de um *Corpus* Colaborativo de Linguística: percursos metodológicos. **Rev. Estud. Ling.**, Belo Horizonte, v. 29, n. 3, p. 2041-2078, 2021. DOI: <http://dx.doi.org/10.17851/2237-2083.29.3.2041-2078>.

¹¹ Em uma breve análise em nossa tabela de dados, percebemos que as pesquisas vinculadas à UFMG tendem a utilizar **Portuguese as a Welcoming Language** nas traduções, ao passo que as vinculadas à UnB tendem a utilizar **Portuguese (as (a)) Host Language**. Seria possível identificar outros padrões a partir dos metadados destas pesquisas? Será que a partir de determinado ano as utilizações terminológicas foram sendo padronizadas? Para isso, como o número de dissertações e teses brasileiras vinculadas à área de PLAc é baixo, poder-se-ia expandir ou redesenhar o *corpus* abrangendo outros tipos de textos ou até mesmo pesquisas não necessariamente brasileiras.

¹² O Alto Comissariado das Nações Unidas para Refugiados é uma agência da Organização das Nações Unidas voltada a questões políticas e ao asseguramento de direitos de pessoas refugiadas e apátridas.

GROSSO, M. J. R. Língua de acolhimento, língua de integração. **Horizontes de Linguística Aplicada**, Brasília, v. 9, n. 2, p. 61-77, 2010. DOI: <https://doi.org/10.26512/rhla.v9i2.886>.

LISBOA, J. V. R. *et al.* Breve panorama de dissertações e teses da área de Português como Língua de Acolhimento. In: SOUZA, V. V. S.; NASCIMENTO, M. R.; MARQUES, M. M. D. (org.). **Mapeando a Linguística Aplicada em sua multiplicidade: (re)descobertas**. São Carlos: Pedro & João Editores, 2021. p. 111-132. Disponível em: <https://pedroejoaeditores.com.br/site/mapeando-a-linguistica-aplicada-em-sua-multiplicidade-redescobertas/>. Acesso em: 4 out. 2021.

LISBOA, J. V. R. Português para Falantes de Outras Línguas: uma proposta inicial de árvore de domínio à luz da Linguística de *Corpus*. In: NOVODVORSKI, A.; LISBOA, J. V. R. (org.). **Estudos exploratórios em Linguística de Corpus**. Araraquara: Letraria, 2021a. p. 102-111. Disponível em: <https://www.lettraria.net/estudos-exploratorios-em-linguistica-de-corpus/>. Acesso em: 4 out. 2021.

LISBOA, J. V. R. **Proposta de harmonização da terminologia designativa de área e subáreas do Português como Língua Não Materna baseada em corpus**. 2021b. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2021b. DOI: <http://doi.org/10.14393/ufu.di.2021.161>.

LISBOA, J. V. R. Reflexões iniciais sobre a terminologia designativa de área e subáreas do Português para Falantes de Outras Línguas: um estudo exploratório. **Revista GTLex**, Uberlândia, v. 5, n. 2, p. 283-311. 2020. DOI: <https://doi.org/10.14393/Lex9-v5n2a2020-5>.

LOPEZ, A. P. A.; DINIZ, L. R. A. Iniciativas jurídicas e acadêmicas brasileiras para o acolhimento de imigrantes deslocados forçados. **Revista SIPLE**, Brasília, v. 6, n. 1, p. 31-56, 2018.

OLIVEIRA, M. R. Implantando PLAc em Caicó/RN: o ensino de língua portuguesa para refugiados venezuelanos. **RECEI**, Mossoró, v. 5, n. 15, p. 576-585, 2019. Disponível em: <http://natal.uern.br/periodicos/index.php/RECEI/article/view/1699>. Acesso em: 26 out. 2021.

PARODI, G. **Lingüística de Corpus: de la teoría a la empiria**. Madrid: Iberoamericana, 2010.

REY, A. **Essays on Terminology**. Tradução e edição de Juan Carlos Sager. Amsterdam: John Benjamins Publishing Company, 1995.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 9 set. 2021.

SHEPHERD, T. M. G. O estatuto da Linguística de Corpus: metodologia ou área da Linguística? **Matraga**, Rio de Janeiro, v. 16, n. 24, p. 150-172. 2009. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/matraga/article/view/27801>. Acesso em: 14 out. 2021.

SILVA, F. C.; COSTA, E. J. O ensino de Português como Língua de Acolhimento (PLAc) na linha do tempo dos estudos sobre o Português Língua Estrangeira (PLE) no Brasil. **Horizontes de Linguística Aplicada**, Brasília, v. 19, n. 1, p. 125-143, 2020. DOI: <https://doi.org/10.26512/rhla.v19i1.24117>.

SOARES, L.; TIRLONI, L. P. Literatura popular em projeto de ensino de Português como Língua de Acolhimento para imigrantes haitianos. In: FERREIRA, L. C. *et al.* (org.). **Língua de acolhimento: experiências no Brasil e no mundo**. Belo Horizonte: Mosaico Produção Editorial, 2019. p. 83-101.

TAGNIN, S. E. O. **O jeito que a gente diz**: combinações consagradas em inglês e português. Barueri: Disal, 2013.

AVALIATIVIDADE em *corpus* de comentários: um olhar sobre o feminino

Mayra Natanne Alves Marra¹

¹ Professora do ensino básico, técnico e tecnológico do Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro - *Campus* Ituiutaba. Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/4955128268478795>. E-mail: mayra.marra@ufu.br.

1 Introdução

Este estudo exploratório é parte de uma pesquisa de doutorado em andamento. Neste capítulo, objetivamos, a partir da identificação de palavras-chave, descrever as escolhas lexicogramaticais atitudinais utilizadas para expressar avaliações em comentários escritos de espectadores do episódio “Xerecou: empoderamento feminino na linguagem”, publicado em agosto de 2021, pelo canal GNT, no YouTube.

O tema movimentou a imprensa, acompanhando uma discussão na mídia nacional sobre a utilização da palavra *xerecou* pela atleta Karen Jonz, skatista tetracampeã e comentarista de *skate* das Olimpíadas de Tóquio, durante a transmissão de uma prova, ao vivo, em um canal de TV. Ao narrar a participação de uma competidora e relatar a queda da atleta após uma manobra, a comentarista utilizou a palavra *xerecou* para descrever um choque contra o corrimão durante um movimento executado de forma inadequada pela skatista olímpica. Ao narrar a situação, Karen disse: “Xerecou no campeonato!”.

A palavra *xerecou*, então, tornou-se uma das mais citadas, apareceu nos *Trending Topics* do *Twitter* e gerou muitos memes² e comentários diversos. De acordo com Barton e Lee (2015), falar sobre a língua e linguagem se tornou uma prática comum em nosso mundo social, cada vez mais mediado por textos.

As novas mídias também proporcionam novas oportunidades e espaços para que os atores sociais reflitam sobre questões linguísticas, especialmente na forma escrita. [...] os participantes da *web* costumam falar sobre a língua e seu aprendizado; as pessoas avaliam o conhecimento de sua própria língua e da dos outros; [...] eles não precisam ser usuários perfeitos de determinada língua para avaliá-la e para saber como ela deve ou não ser utilizada (BARTON; LEE, 2015, p. 162).

Karen Jonz também foi convidada a participar de programas televisivos e *podcasts* e o fato virou notícia³ em diversos veículos de comunicação, que destacaram a utilização da palavra *xerecou* e de outras expressões usadas por ela.

A utilização desse neologismo⁴ também suscitou debates sobre o uso de determinadas palavras por mulheres, sobre a adequação da linguagem ao contexto etc. Os comentários de Karen, além de terem sido considerados pela imprensa, de maneira geral, como engraçados e espontâneos, também promoveram uma reflexão sobre o “constrangimento” em relação ao uso de palavras relacionadas ao órgão genital feminino⁵ e à utilização de palavras tabus.

2 Disponível em: <https://twitter.com/search?f=top&q=%22xerecou%22&src=typed>. Acesso em: 14 jan. 2022.

3 Disponível em: <https://www.uol.com.br/esporte/ultimas-noticias/2021/09/26/karen-jonz-xerecou-altas-horas.htm>. Acesso em: 14 jan. 2022.

4 Entendemos por neologismos palavras que não estejam dicionarizadas e que ainda não integram o vocabulário da língua portuguesa – VOLP.

5 Disponível em: <https://bit.ly/3Oc6MPT>. Acesso em: 14 jan. 2022.

O vídeo⁶ escolhido para este trabalho é um produto do programa *Saia Justa*⁷, do canal de televisão por assinatura GNT⁸. O programa original é transmitido integralmente na televisão fechada e, depois, é disponibilizado na internet, editado e com tamanho reduzido, sendo considerado pela emissora uma seleção dos melhores momentos do programa. Esse produto é intitulado *Mini Saia Justa*. Nele, são apresentados recortes das falas das apresentadoras e da pessoa entrevistada. Tal programa está hospedado na plataforma de vídeos YouTube, no canal da emissora. Importante destacar que o canal da emissora no YouTube existe desde julho de 2010, possui 1,92 milhões de inscritos e seus vídeos contam com mais de 500 milhões de visualizações.

Cabe mencionar que o YouTube é um *site* que permite que os seus usuários compartilhem vídeos, disponibilizando-os na internet. Nesse ambiente virtual, também é possível interagir com os autores/divulgadores por meio de comentários escritos, publicados na página do vídeo. Como ressaltam Barton e Lee (2015, p. 60), “o YouTube é visto como *site* de rede social, devido às relações únicas desenvolvidas entre quem sobe o vídeo e seus espectadores”. Portanto, segundo os autores, o YouTube é “rico em espaços de escrita, pois além das legendas e anotações no vídeo, a seção de comentário é o principal espaço de escrita interativa do *site* [...] os comentários podem ser avaliados pelos usuários (votar a favor ou contra)”.

Barton e Lee (2015, p. 62) citam, ainda, a popularidade desse espaço virtual, o qual abrange diferentes modos de construção de significados e diferentes dinâmicas de interação. Destacam que o “YouTube se baseia no aspecto visual e inclui bastante interação com estranhos”. Ademais, os pesquisadores pontuam que o YouTube tem um perfil etário amplo entre os usuários da rede e possui funções mais abrangentes que as redes sociais.

Por isso, devido às suas características particulares, um estudo em *corpus* de comentários de um vídeo do YouTube é produtivo, pois possibilita a descrição e o estudo da inter-relação entre as práticas de produção de texto e os espaços de produção escrita em novas mídias digitais, em que as pessoas empregam a linguagem *on-line* quando participam de *sites* como o YouTube, e através dela expressam suas opiniões, avaliações e atitudes sobre temas variados.

Os usuários dessa plataforma precisam articular suas opiniões, sentimentos e/ou atitudes em relação a algo ou alguém para interagir a partir do vídeo publicado. Charaudeau (2006, p. 176) explica que o gênero “comentário” está para a argumentação, uma vez que ele “problematiza os acontecimentos, constrói hipóteses, desenvolve teses, traz provas, impõem conclusão”. O

6 Disponível em: <https://youtu.be/6JwED9VRdBY>. Acesso em: 14 jan. 2022.

7 O programa *Saia Justa* é apresentado e mediado por Astrid Fontenelle e vai ao ar toda semana. Justamente com Fontenelle, as apresentadoras Mônica Martelli, Pitty e Gaby Amarantos promovem conversas, debates e discussões sobre temas diversos com pessoas convidadas, sempre de uma perspectiva que destaca diferentes experiências e vozes femininas.

8 Endereço eletrônico da GNT: <https://gnt.globo.com/>. Acesso em: 14 jan. 2022.

comentário, segundo o pesquisador, “põe o leitor em questão: exige uma atividade intelectual, um trabalho de raciocínio, uma tomada de posição contra ou a favor, e desta atividade não há ninguém, no fim da troca, que saia incólume (o comentário é histórico)”.

Outro aspecto a ser destacado sobre o espaço de escrita “comentários” no YouTube, segundo Barton e Lee (2015, p. 119), é o fato de que esse espaço de escrita licencia a produção conjunta de posições sobre vários aspectos do vídeo publicado. “Alguns podem se alinhar [...], enquanto outros podem assumir uma postura de oposição. Às vezes, esses comentários podem se referir diretamente ao conteúdo do vídeo enviado; em outros momentos, os comentários podem mudar temas e iniciar novos tópicos de discussão entre si”.

Assim, coadunando com Barton e Lee (2015, p. 65), os “usuários comuns da internet têm um poder sem precedentes de escolha e criatividade”. Por isso, conforme as palavras dos pesquisadores, “parece razoável dar uma olhada nas atividades e práticas sociais reais que cercam essa escrita”.

Outras possibilidades a serem investigadas através do *corpus* aqui proposto, para além da AVALIATIVIDADE e de neologismos, são a utilização de *emoticon*; as regras de netiqueta utilizadas pelos comentaristas dos vídeos, como, por exemplo, o uso de letras maiúsculas; a linguagem ofensiva e agressiva em mensagens insultuosas e hostis, que geralmente violam a polidez linguística; o uso de abreviações, de repetições, dentre outras características da linguagem escrita *on-line*.

Portanto, investigar a linguagem específica da internet através de comentários escritos é viável, uma vez que “usuários da língua *on-line* também desenvolveram seu próprio repertório e convenções linguísticas para a língua que empregam *on-line*”, conforme destacam Barton e Lee (2015, p. 149).

Acreditamos que descrever a utilização e refletir sobre a criação de novas palavras e sobre seus contextos de utilização é relevante. Como destaca Fairclough (2001, p. 246), “novas palavras geram novos itens lexicais” (HALLIDAY, 1966). Assim, parafraseando Fairclough, a criação de itens lexicais permite conceber as perspectivas particulares dos domínios da experiência segundo uma visão teórica, científica, cultural ou ideológica mais abrangente. Portanto, investigar a utilização do termo *xerecou* em um *corpus* de comentários do YouTube permite refletir sobre a utilização do português brasileiro na atualidade.

Este é um trabalho de base descritiva e que leva em consideração os usos que se faz da linguagem, especificamente a linguagem avaliativa, filiando-se à Linguística Sistêmico-Funcional (LSF), à Avaliatividade e à Linguística de *Corpus* (LC). Utilizamos o programa *WordSmith Tools* 6.0 (SCOTT, 2012), doravante WST, para o processamento e a análise dos dados.

Considerando o objetivo aqui proposto, fizemos um recorte para trabalharmos com a metafunção interpessoal da linguagem, especificamente com o sistema de AVALIATIVIDADE, proposto por Martin e White (2005). Este estudo é pautado no sistema de ATITUDE, a partir da perspectiva *corpus driven*, isto é, guiada por *corpus*.

2 Os estudos descritivos e a linguística sistêmico-funcional

Este trabalho seguiu o pressuposto apresentado por Perini (2006, p. 36), de que o objetivo do linguista não se resume à “apresentação de dados”; é preciso observar atentamente “a descrição da estrutura da língua, o conjunto de regras, elementos, classes e princípios que governam as associações dos diversos elementos da língua e seus significados”. Coadunando com Perini (2008, p. 57), um dos principais objetivos da linguística é descrever as línguas naturais, “prever a aceitabilidade e a inaceitabilidade de sequências formais e de suas associações e representações semânticas”. Assim, estudos como este auxiliam na compreensão de fenômenos linguísticos no português brasileiro em espaços de escrita *on-line*.

Aqui, nos interessa relatar o potencial criativo linguístico e descrever a forma como o item lexical *xereca*, o qual classifica-se, segundo a gramática normativa, como um substantivo, passou a ser utilizado como um verbo de primeira conjugação, ou seja, terminado em AR – *xerecar*, mais especificamente, flexionado no pretérito perfeito, na terceira pessoa do singular *xerecou*, criando um neologismo.

Diante de tal fato linguístico, a utilização do item lexical *xereca* nos interessa, uma vez que tem sido sistêmica a criação de novos itens baseados na mudança de classe gramatical, passando de um nome a um verbo. Esse tipo de movimento foi descrito no trabalho de Bispo (2019). Segundo a pesquisa, o item *sextou*, o qual deriva do substantivo *sexta-feira*, é empregado em redes sociais. Gama (2017) explica o neologismo *bombar*, derivado da palavra *bomba*, que também é de base nominal e dá origem a outras formas como *bombou* e *bombando* e foram utilizados em um jornal popular.

A LSF preocupa-se em compreender as funções da linguagem, por isso é uma teoria de base funcionalista que considera o funcionamento e o uso da língua. Assim, nos amparamos na gramática Sistêmico-Funcional (GSF), proposta por Halliday (1994) e Halliday e Matthysien (2004), para realizar a análise apresentada.

Logo, o estudo aqui proposto, embasado na LSF, concorda que o texto e o contexto devem ser levados em consideração e que as análises linguísticas devem ser concebidas através de uma interpretação descritiva, baseada no uso da língua.

Halliday defende a ideia de que a linguagem e seu desenvolvimento, enquanto sistema semiótico, devem ser analisados em contextos de uso, por isso, propõe uma análise “sociofuncional” da linguagem, apontando três Metafunções que aprofundam as relações entre sistema linguístico e as necessidades de comunicação dos indivíduos. São elas: a Ideacional, a Interpessoal e a Textual. Essa organização possibilita a compreensão das funções básicas da linguagem e a multiplicidade de usos que o ser humano faz dela.

Este estudo centrou-se na Metafunção Interpessoal, especificamente, no subsistema de Atitude, a fim de identificar, descrever e analisar os usos lexicogramaticais presentes em comentários escritos, a partir da lista de palavras-chave, à luz do sistema de AVALIATIVIDADE, sugerido por Martin e White (2005).

2.1 A AVALIATIVIDADE e o Subsistema de Atitude

De acordo com Vian Jr. (2010, p. 22), o sistema de AVALIATIVIDADE pode ser caracterizado como interpessoal no nível da semântica do discurso, que está articulado, simultaneamente, a outros dois sistemas: Negação e Envolvimento.

O sistema de AVALIATIVIDADE é formado por três grandes subsistemas: a Atitude – foco deste trabalho, a Gradação e o Engajamento. Assim, esse sistema está relacionado a todo o potencial que a língua oferece para realizarmos significados avaliativos. Logo, estudar a AVALIATIVIDADE em *corpus* de comentários escritos é relevante, pois esse gênero possibilita o uso de expressões, pontos de vista sobre determinado conteúdo e/ou pessoa que está sendo exposta ou produto publicizado, por exemplo.

O Subsistema de Atitude é um componente que possibilita que o falante/escritor faça avaliações sobre entidades, estado de coisas e acontecimentos, negativa ou positivamente. Esse subsistema subdivide-se em três campos semânticos: o Afeto, responsável por expressar emoções; o Julgamento, responsável por representar as avaliações éticas no âmbito da moralidade; e a Apreciação, responsável por expressar as avaliações sobre as coisas e os objetos no âmbito da estética.

3 Procedimentos metodológicos

Este estudo exploratório foi guiado pelo *corpus*, isto é, pelos resultados encontrados através da *chavidade* e da identificação das temáticas do *corpus* de comentários (cf. NOVODVORSKI, 2013; BERBER SARDINHA, 2009), neste caso, focalizamos nossas análises no uso da palavra *xerecou*.

Ao investigar a realização das escolhas linguísticas de avaliação presentes no *corpus* de comentários, coadunando com Castilho (2010, p. 59), contribuímos para uma importante reflexão sobre a língua, já que, para isso, é preciso refletir e tomar conhecimento sobre os processos linguísticos, e, antes de tudo, é necessário descrever a língua.

O *corpus* desta investigação é composto por 75 comentários, apresentando uma variedade lexical satisfatória aos propósitos deste trabalho. São 1.226 *tokens* (itens) e 567 *types* (formas). Através do *corpus* de comentários, foi possível encontrar inúmeras avaliações, pois esse tipo de escrita possibilita a interação entre diferentes espectadores que dialogam na rede e demonstram, por meio de suas “falas”, aspectos culturais e linguísticos diversos e, ainda, estabelecem uma relação de troca com o programa, apresentadoras e emissora responsável pelo canal, reafirmando a função do espaço de escrita “comentários”.

A escolha do episódio para subsidiar a coleta dos comentários ocorreu por tratar-se de uma discussão relacionada à linguagem, especificamente sobre o uso da palavra *xerecou*. A utilização de tal vocábulo reverberou, nos comentários, avaliações sobre a situação atual das mulheres e aspectos ligados ao feminino na sociedade.

Tendo em vista os atuais embates sociais em relação a temas que envolvem a mulher, a sexualidade feminina e a noção de feminino, não foi possível encontrar, nos comentários, material suficiente para traçar padrões sexistas na utilização da linguagem. Entretanto, demonstrou-se o apoio da audiência sobre a postura ética da entrevistada, destacando sua relevância como exemplo a ser seguido.

3.1 A Linguística de Corpus e o programa WordSmith Tools

A Linguística de *Corpus* (LC) surgiu nos anos 60 e é uma abordagem teórico-metodológica. Nas palavras de Berber Sardinha (2008), a LC possui um carácter interdisciplinar, possibilitando o diálogo entre áreas diversas. Ela permite a observação de manifestações linguísticas de diversos segmentos, em vários textos. Assim, a LC dedica-se à coleta e à exploração de *corpora* ou de conjuntos de dados linguísticos (cf. BERBER SARDINHA, 2000).

De acordo com Novodvorski e Finatto (2014, p. 8), a LC compreende “a língua como um sistema probabilístico de combinatórias, no qual uma unidade se define pelas associações que mantêm com outras unidades”. Por isso, observar dados sistematizados em contextos autênticos de uso é de grande importância, pois permite a identificação de características peculiares aos *corpora*. Assim, descobrir e reconhecer padrões de utilização da língua é fundamental, pois toda teorização parte de uma observação atenta.

Para que este estudo exploratório fosse possível, utilizamos o programa WST. Ele é um pacote integrado de ferramentas, desenvolvido com o propósito de auxiliar na descrição e análise linguísticas em *corpus*.

O WST oferece diversas ferramentas, dentre elas, *Wordlist*, *KeyWords* e *Concord*, que foram as ferramentas utilizadas neste estudo. A primeira ferramenta, talvez a mais importante, gera a lista de palavras e contabiliza os *types* e o *tokens* do *corpus*. Ela também permite ordenar a lista alfabeticamente ou por número de frequência e organizar agrupamentos de formas em um mesmo item, por exemplo, as formas *xerecou* e *xerecaram* foram agrupadas sob o lema *xereca*.

Já a ferramenta *Keywords* é responsável por gerar lista de palavras que possibilitam a identificação de palavras-chave do texto em um *corpus*, através de uma comparação entre um *corpus* de estudo e um de referência, que deve ser entre 3 e 5 vezes maior que o *corpus* de estudo.

Por último, há a ferramenta *Concord*. Com ela, é possível visualizar as linhas de concordância em que as ocorrências de determinada palavra, ou várias, estão listadas. A ferramenta também permite acessar os contextos de uso destas palavras.

3.2 Coleta, organização e categorização dos dados

A metodologia utilizada nesse trabalho seguiu a proposta da LC, por isso, o primeiro passo foi a compilação do *corpus* de estudo. Os primeiros procedimentos realizados para a compilação do *corpus* foram:

- I. Escolha do *corpus* de estudo;
- II. Extração manual dos comentários por meio de seleção e transferência dos segmentos para arquivo de texto plano (.txt), legível pela ferramenta WST.
- III. Armazenamento dos arquivos em .txt, limpeza e organização do *corpus* de comentários.

Em seguida, depois do trabalho manual realizado para possibilitar a leitura dos comentários pelo programa WST, partimos para a etapa em que foi gerada a *WordList*, tanto para o *corpus* de estudo quanto para o de referência⁹. Sendo assim, demos continuidade às seguintes etapas:

- IV. Geração das listas de palavras tanto do *corpus* de estudo quanto do de referência, por meio da ferramenta *WordList*;
- V. Geração da lista de palavras-chave.

⁹ Utilizei como *corpus* de referência entrevistas que foram parte do *corpus* da minha dissertação de mestrado (MARRA, 2017).

Após a criação da lista de frequência de palavras, foi possível identificar algumas características do *corpus* através da ferramenta *Statistics*. Com as listas, também foi possível criar a lista de palavras-chave. Em seguida, identificamos e analisamos a única palavra-chave encontrada, “Karen”, nome da entrevistada do programa.

3.3 O *corpus* de estudo e o *corpus* de referência

O *corpus* de estudo é composto por 75 comentários¹⁰, que se caracterizam por apresentar uma linguagem informal, apresentando, muitas vezes, abreviações, tabuísmos e gírias. Foi criada uma etiqueta para identificação do início de cada comentário (C -), assim, sempre que se inicia um novo comentário, foi acrescentada essa marcação. Também optamos por retirar os nomes dos comentaristas.

O tamanho do *corpus* é pequeno, possui 1.226 tokens e 567 types e a riqueza lexical (*type/token ratio*) é de 46,25. As informações sobre o *design* do *corpus* de estudo foram sintetizadas e estão apresentadas no Quadro 1.

Quadro 1: O *corpus* de estudo.

Língua	Português brasileiro
Modo	Escrito
Tipologia textual	Expositivo/Argumentativo – comentários
Data da publicação	Agosto de 2021
Conteúdo	Comentários de usuários da plataforma YouTube
Tamanho	Pequeno (menos de 3.000 de palavras)
Nível de codificação	Com etiqueta

Fonte: Quadro adaptado de Marra (2017).

O *corpus* de referência é composto por 3 entrevistas da *Revista da Cultura*, da cidade de São Paulo, publicadas em 2014. Os assuntos abordados nas entrevistas versam sobre literatura, música, teatro, cinema etc. As entrevistas apresentam introdução sobre o entrevistado antes das perguntas estilo pingue-pongue e mantêm linguagem informal, uma vez que aparecem tabuísmos e gírias. O *corpus* de referência possui tamanho 5 vezes maior que o *corpus* de estudo, apresentando 7.511 *tokens* e 2.105 *types*.

¹⁰ Inicialmente, o *corpus* apresentava 105 comentários, porém, com a limpeza, foram excluídos comentários duplicados e desconexos com o vídeo ou tema ali tratado. É relevante destacar que optamos por manter os *emoticons* presentes nos comentários, pois eles ajudam a recuperar o contexto e a interpretar o comentário.

Após a organização dos *corpora* de estudo e referência, foi extraída a lista de palavras-chave. O *corpus* nos guiou para a palavra: “Karen”, o nome da convidada do programa. A partir deste resultado, iniciamos a análise das linhas de concordância com a palavra-chave, a fim de identificar avaliações e, assim, aplicar os conceitos do sistema de AVALIATIVIDADE. Foram identificadas 12 linhas de concordância, como mostra a Figura 1.

Figura 1: Linhas de concordância do nóculo “Karen”.

N	Concordance	Set	Word #	Sen	Pará	Hea	Sec	File	Date	%
1	desse assunto não ser tabu! A Karen é incrível... C - Definir		190	15	27	0	18	Comentários	2021/nov/23	14%
2	muito essa entrevista C - A Karen é incrível...amo demais C		404	37	50	0	37	Comentários	2021/nov/23	31%
3	CAUSA DO FEMINISMO C - A Karen é incrível...amo demais C		1,052	72	64	0	97	Comentários	2021/nov/23	79%
4	Sempre me acabo de rir com a Karen, menina leve, do bem... C		105	9	73	0	10	Comentários	2021/nov/23	8%
5	do bem... C - Coisa mais linda a Karen bombando, ficando cada		116	10	57	0	11	Comentários	2021/nov/23	9%
6	C - FOLLOW ME C - Karen C - Pitty aqui no Rio		720	54	35	0	66	Comentários	2021/nov/23	56%
7	Eles xerecam C - Karen maravilhosa! C -		1,083	74	10	0	10	Comentários	2021/nov/23	82%
8	kkkkk C - KAREN JONZ EU TE		499	44	36	0	46	Comentários	2021/nov/23	39%
9	é incrível...amo demais C - Karen é foda C - Karen		410	37	73	0	38	Comentários	2021/nov/23	32%
10	C - Com certeza kkkkk C - Karen é incrível, que mulher.		39	0	91	0	4%	Comentários	2021/nov/23	3%
11	demais C - Karen é foda C - Karen maravilhosa! C - Pitty		416	37	96	0	38	Comentários	2021/nov/23	32%
12	AMEI os looks e o cabelo da KAREN. C - Definitivamente		761	57	10	0	70	Comentários	2021/nov/23	59%

Fonte: Concord.

Observamos que, das 12 linhas de concordância com a palavra “Karen”, todas contêm avaliações sobre a skatista, seja sobre sua pessoa ou sobre sua aparência, fato que pode ser explicado devido ao aspecto visual do YouTube.

Outro item lexical que foi investigado foi a palavra *xerecou*, por ser um dos temas centrais do episódio e também por entendermos que esta palavra exemplifica a criatividade lexical, demonstrando uma nova manifestação linguística do ponto de vista gramatical, uma vez que *xerecou* não integra os dicionários e vocabulários do português brasileiro. É um item criado a partir do substantivo *xereca* e foi utilizado como verbo, sendo flexionado no pretérito perfeito e obedecendo às regras de formação de palavras do português. Sendo *xerecou* uma unidade lexical criada a partir de um processo de derivação, esse neologismo¹¹ demonstra a potência criativa da língua, isto é, uma palavra fruto da necessidade de denominar uma nova realidade, corroborando Correia e Almeida (2012, p. 23) que afirmam que “os neologismos podem constituir palavras formalmente novas, preexistentes que adquirem um novo significado, ou, ainda, palavras que passam a ocorrer em registros linguísticos nos quais não costumavam ocorrer”.

Almeida (2020), em estudo sobre o substantivo *xereca*, mostra que ele é um africanismo que se refere ao órgão genital ou sexual feminino. Ele é substantivo compreendido como tabuísmo, “uma palavra de uso restrito, censurada por pudor ou por crença, sendo ainda vista como

11 Neste trabalho, seguimos a proposta de Alain Rey (1976 *apud* CORREA; ALMEIDA, 2012, p. 23) e compreendemos que neologismo é uma unidade lexical cuja forma significante ou cuja relação significado-significante, caracterizada por um funcionamento efetivo num determinado modelo de comunicação, não se tinha realizado no estágio anterior do código da língua.

popular e jocosa”. Zavaglia (2018, p. 5), em trabalho sobre tabuísmos, afirma que a palavra *xereca* é “considerada grosseira, ofensiva e chula”, significados esses que podem contribuir para o entendimento da repercussão da palavra *xerecou*. A Figura 2 demonstra as ocorrências encontradas no *corpus* de comentários, que corroboram o valor semântico averiguado pelos trabalhos citados.

Figura 2: Linhas de concordância do nóculo *xereca*.

N	Concordance	Set	Word #	Sen	Sen	Para	Para	Hea	Hea	Sec	Sec	File	Date	%
1	quando a competidora bateu a xereca. C - Eu uso buceta C -	XE	1,202	83	10	0	92			0	92	Comentários	2021/nov/23	92%
2	C- Karen maravilhosa! ♥ C - Xerecar? arma que vcs, mulheres	XE	1,088	75	10	0	84			0	84	Comentários	2021/nov/23	82%
3	pra obter BENEFICIOS C - "Xereca" s. f. genitália feminina.	XE	1,097	76	71	0	84			0	84	Comentários	2021/nov/23	83%
4	☺ C - Eu xereco Tu xerecas Ela xereca Nós xerecamos Vós	XE	1,074	74	50	0	82			0	82	Comentários	2021/nov/23	81%
5	xerecamos Vós xerecais Eles xerecam ♥♥♥♥ C- Karen	XE	1,080	74	10	0	10			0	10	Comentários	2021/nov/23	81%
6	C - Pitty pontuo bem. ☺ C - Eu xereco Tu xerecas Ela xereca	XE	1,070	74	10	0	10			0	10	Comentários	2021/nov/23	81%
7	uma parte intima no ferro falar "xerecou". C- Minha avó materna	XE	466	43	10	0	10			0	10	Comentários	2021/nov/23	36%
8	tempo por causa da música xerecard é só pq foi em rede	XE	525	44	10	0	10			0	10	Comentários	2021/nov/23	41%
9	Tu xerecas Ela xereca Nós xerecamos Vós xerecais Eles	XE	1,076	74	10	0	10			0	10	Comentários	2021/nov/23	81%
10	C - É claro que verdade também xerecar é claro que eu acho que	XE	1,290	91	10	0	10			0	10	Comentários	2021/nov/23	99%
11	pontuo bem. ☺ C - Eu xereco Tu xerecas Ela xereca Nós	XE	1,072	74	10	0	10			0	10	Comentários	2021/nov/23	81%
12	Ela xereca Nós xerecamos Vós xerecais Eles xerecam ♥♥♥♥	XE	1,078	74	10	0	10			0	10	Comentários	2021/nov/23	81%

Fonte: Concord.

No *corpus* deste trabalho, encontramos 12 ocorrências que foram agrupadas sob o lema¹² *xereca*. Através dos exemplos, percebe-se que a utilização dessas lexias está relacionada à necessidade de criação de palavras que “verbalizem” os substantivos, transformando seus significados em ações que refletem sentidos literais ou metafóricos, demonstrando, assim, que a criação de palavras pelos falantes se dá em contextos diversos de comunicação, com o propósito de veicular ações, ideias, conceitos etc.

4 Resultados e discussão

Buscando observar como o *corpus* de estudo representa a palavra Karen, analisamos as linhas de concordância. Encontramos a recorrência de um certo tipo de comentário, os elogiosos. Foram identificados os adjetivos “incrível” (4 ocorrências) e “maravilhosa” (2 ocorrências). Por outro lado, encontramos ocorrências que nos chamam a atenção para o uso coloquial e metafórico da linguagem, como os itens “foda” e “bombando”; expressões como “menina leve”, “que mulher!”, “do bem” e *emoticons* de corações e rostos apaixonados.

Assim, seguindo o Sistema de AVALIATIVIDADE e com base no que o *corpus* de comentários nos guiou, destacamos a ocorrência de algumas dessas avaliações. Considerando que o Sistema de AVALIATIVIDADE subdivide-se em Afeto (emoção), Julgamento (ética) e Apreciação

¹² No WST, a coluna ‘lemmá’ é destinada aos ‘lemas’, que, de acordo com Berber Sardinha (2009, p. 154), “são formas canônicas de palavras, como, por exemplo, ‘casa’, que encapa ‘casa’, ‘casinha’, ‘casas’, ‘casinhas’, ‘casarão’, etc”.

(estética), foi possível observar que há mais avaliações do tipo Julgamento, do tipo Capacidade Positiva, como é possível observar nos exemplos a seguir:

C 1 – Karén é **incrível!**

C 2 – Karén é **maravilhosa!**

C 3 – Karén é **foda!**

É possível perceber que os adjetivos aplicados a Karen são positivos e estão relacionados à sua capacidade diante da sociedade, o que é capaz de causar admiração, como uma atitude positiva. Já com relação às avaliações do tipo Afeto, elas se realizam por meio do processo mental “amar”, conforme mostram os exemplos:

C 4 – KAREN JONZ EU TE **AMOOO**

C 5 – Karen é incrível...**amo** demais!

Veja que as realizações de avaliações do tipo Emoção demonstram um sentimento positivo, de felicidade. Já com relação às realizações de Apreciação, encontramos:

C 6 – **AMEI** os *looks* e o cabelo da Karen.

C 7 – Coisa mais linda a Karen **bombando**, ficando cada vez mais famosa.

Nesses exemplos, é possível dizer que ocorrem avaliações em relação à estética de Karen, relacionadas à sua composição (aparência). Além disso, encontramos um exemplo de apreciação da situação na qual a Karen está envolvida (evidência), sendo avaliada de forma positiva por estar sendo bastante citada, conseqüentemente, fazendo sucesso. A seguir, apresentamos o Quadro 2 para sintetizar as ocorrências encontradas.

Quadro 2: Resultados AVALIATIVIDADE.

Classificação das ocorrências do tipo Atitude			
	Afeto	Julgamento	Apreciação
Número de ocorrências	4	7	2

Fonte: Elaborado pela autora.

No que se refere às ocorrências relacionadas ao lema *xerec*, encontramos o substantivo *xereca*, para se referir ao órgão genital feminino, corroborando os estudos de Almeida (2020) e Zavaglia (2018), conforme os exemplos:

C 8 – Quando a competidora bateu a **xereca**.

C 9 – “**Xereca**” s. f. genitália feminina.

Já o neologismo *xerecard*, resultado do cruzamento vocabular das palavras *xereca* e *card* (cartão, em inglês), foi utilizado como um substantivo e referia-se ao nome de uma música do MC Brisola¹³. A noção de *xereca* como instrumento também apareceu em outro exemplo com o verbo no infinitivo *xerecar*, em que se avalia negativamente a ação de *xerecar*, sendo esta entendida como um ato pejorativo e sexualizado, diferente do sentido utilizado pela skatista e comentarista Karen Jonz, conforme os exemplos:

C 10 – Por causa da música **xerecard** é só pq foi em rede aberto

C 11 – **Xerecar**? arma que vcs mulheres usam para obter BENEFICIOS

As formas verbais conjugadas no presente do indicativo, *xereco*, *xerecas*, *xereca*, *xerecamos*, *xerecais* e *xerecam*, ocorreram em comentário que demonstrava a possibilidade de conjugação verbal desse neologismo, exemplificando mais uma vez o potencial de criatividade linguística. Também destacamos a utilização da palavra flexionada no pretérito perfeito para a terceira pessoa do singular, conforme os exemplos a seguir:

C 12 – Eu **xereco** Tu **xerecas** Ela **xereca** Nós **xerecamos** Vós **xerecais** Eles **xerecam**

C 13 – Parte íntima no ferro falar “**xerecou**”

Após a observação dos itens lexicais aqui descritos, fica evidente que o recurso linguístico de criação de palavras a partir de uma base substantiva é um fato linguístico sistêmico e que ocorre no português brasileiro, como é o caso de *xerecou* e tantos outros usos, como os já citados *sextou* e *bombou* e outros que ainda serão descritos, como “Adrenalizou”, título da canção de Vitor Kley¹⁴.

Porém, o que difere no fato linguístico descrito neste trabalho com o item *xerecou* é que este foi um neologismo criado a partir de um vocábulo tabu que carrega, em nossa sociedade, questões culturais diversas que perpassam as noções sobre o órgão genital feminino, estereótipos de práticas discursivas femininas, como a forma de falar e se expressar, o cuidado vocabular etc.

5 Considerações finais

O objetivo deste estudo exploratório foi descrever, a partir da identificação de palavras-chave, as escolhas lexicogramaticais de Atitude, utilizadas para expressar avaliações em comentários de espectadores do episódio “Xerecou: Empoderamento feminino na linguagem”, programa publicado em agosto de 2021.

13 Endereço eletrônico da letra da música *Xerecard*: <https://www.lettras.mus.br/mc-brisola/xerecard/>. Acesso em: 14 jan. 2022.

14 Endereço eletrônico da letra da canção *Adrenalizou*: <https://www.lettras.mus.br/vitor-kley/adrenalizou/>. Acesso em: 14 jan. 2022.

Ainda que de maneira inicial, esse estudo demonstrou a aplicabilidade da LC e da AVALIATIVIDADE para descrever avaliações realizadas em situações reais de uso da língua. Logo, também evidenciou a possibilidade de trabalho com *corpus* de comentários escritos em ambientes *on-line*, destacando a criatividade lexical e a presença de avaliações nesse tipo de texto.

Os excertos escolhidos para análise evidenciam um olhar positivo da audiência sobre a entrevistada, destacando o seu perfil ético, através de julgamentos do tipo comportamento humano e de comentários favoráveis. Esse aspecto observado vai ao encontro do tema do episódio, o qual discute o empoderamento feminino através da utilização de um vocábulo tabu por mulheres na sociedade atual.

Assim, através desta análise, fica evidente a receptividade da audiência com relação ao assunto abordado no episódio e à pessoa entrevistada. Além disso, pôde-se perceber os significados que o item lexical *xerecou* carrega, segundo o seu contexto de uso, resultando em diferentes ações, as duas literais, porém uma é involuntária, de caráter acidental, e a segunda voluntária, de caráter sexual, portanto, veiculando ideias/conceitos distintos, com finalidades distintas.

Além disso, também é possível refletir sobre o estigma existente frente à utilização de tabuísmos por mulheres, na sociedade atual, corroborando o que Fairclough (2001, p. 264) mencionou como “práticas de masculinidade – por exemplo, a pressupostos de que ser homem implica práticas discursivas agressivas e obscenas”, demonstrando, assim, que o léxico e sua utilização refletem os valores da sociedade que o utiliza.

Referências

ALMEIDA, A. A. D. A categorização em perspectiva sociocognitiva de um africanismo do português do Brasil: o item léxico *xereca*. **Diadorim**, Rio de Janeiro, v. 22, n. 2, p. 386-408, 2020. DOI: <https://doi.org/10.35520/diadorim.2020.v22n2a38221>.

BARTON, D.; LEE, C. **Linguagem online**: textos e práticas digitais. Tradução Milton Camargo Mota. São Paulo: Parábola Editorial, 2015.

BERBER SARDINHA, T. A linguística de corpus no Brasil. In: TAGNIN, S. E. O.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 17-40.

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

BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. **DELTA**, São Paulo, v. 16, n. 2, p. 323-367, 2000. DOI: <https://doi.org/10.1590/S0102-44502000000200005>.

- BERBER SARDINHA, T. **Pesquisa em Lingüística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.
- BISPO, M. P. **Traços linguísticos-tradutórios em “As Três Marias” de Rachel de Queiroz, à luz dos Estudos da Tradução Baseados em Corpus**: ênfase na análise do léxico mais recorrente. 2019. Dissertação (Mestrado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara, 2019. Disponível em: <http://hdl.handle.net/11449/182361>. Acesso em: 14 jan. 2022.
- CASTILHO, A. T. **Nova Gramática do Português Brasileiro**. São Paulo: Contexto, 2010.
- CHARAUDEAU, P. **Discurso das mídias**. São Paulo: Contexto, 2006.
- CORREIA, M.; ALMEIDA, G. M. B. **Neologia em português**. São Paulo: Parábola, 2012.
- FAIRCLOUGH, N. **Discurso e mudança social**. Brasília: EdUnB, 2001.
- GAMA, V. O. S. **Neologismo em foco**: inovações lexicais no jornal *Massa!* 2017. Dissertação (Mestrado em Estudos Linguísticos) – Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, Feira de Santana, 2017. Disponível em: <http://tede2.uefs.br:8080/handle/tede/597>. Acesso em: 14 jan. 2022.
- HALLIDAY, M. A. K. **Introduction to functional grammar**. London: Edward Arnold, 1994.
- HALLIDAY, M. A. K.; MATTHISSEN, C. M. I. M. **Introduction to functional grammar**. London: Arnold, 2004.
- MARRA, M. N. A. **A avaliatividade em corpus de entrevistas de artistas**: um estudo das escolhas lexicogramaticais de atitude. 2017. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2017. DOI: <http://doi.org/10.14393/ufu.di.2017.450>.
- MARTIN, J. R.; WHITE, P. **The language of evaluation**: appraisal in English. New York: Palgrave, 2005.
- NOVODVORSKI, A. **Estilo das traduções de Sergio Molina de obras de Ernesto Sabato**: um estudo de *corpora* paralelos espanhol/português. 2013. Tese (Doutorado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2013. Disponível em: <http://hdl.handle.net/1843/LETR-96LMU8>. Acesso em: 14 jan. 2022.

NOVODVORSKI, A.; FINATTO, M. J. B. Linguística de *Corpus* no Brasil: uma aventura mais do que adequada. **Letras & Letras**, Uberlândia, v. 30, n. 2, p. 7-16, 2014. DOI: <https://doi.org/10.14393/LL60-v30n2a2014-1>.

PERINI, M. A. **Estudos de gramática descritiva**: as valências verbais. São Paulo: Parábola Editorial, 2008.

PERINI, M. A. **Princípios de linguística descritiva**: introdução ao pensamento gramatical. São Paulo: Parábola Editorial, 2006.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://lexically.net/wordsmith/version6/>. Acesso em: 14 jan. 2022.

VIAN JR., O. O Sistema de Avaliatividade e a linguagem da avaliação. *In*: VIAN JR, O.; SOUZA, A. A.; ALMEIDA, F. A. S. D. P. (org.). **A linguagem da avaliação em língua portuguesa**: estudos sistêmico-funcionais com base no sistema de avaliatividade. São Carlos: Pedro & João Editores, 2010. p. 19-29.

XERECOU: Empoderamento feminino na linguagem | Karen Jonz | Mini Saia | Saia Justa. **Canal GNT**, 12 ago. 2021. (11min12s). YouTube. Disponível em: <https://youtu.be/6JwED9VRdBY>. Acesso em: 14 jan. 2022.

ZAVAGLIA, C. O léxico tabu em dicionários infantis: o Caldas Aulete. *In*: ISQUERDO, A. N.; DAL CORNO, G. O. M. (org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. Campo Grande: Ed. UFMS, 2018. v. VIII. p. 215-230.

O uso do vocábulo *bitch* em letras de *rap*: uma análise em corpus

Mariana Souza Santos¹

¹ Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/6102463770404857>. E-mail: marianasouza2021@yahoo.com.

1 Introdução

De acordo com o *Oxford English Dictionary*, o vocábulo *bitch*, cuja tradução para o português brasileiro equivale a *cadela*, é utilizado de modo pejorativo desde cerca dos anos 1000. Hodgson (2008) afirma que a palavra começou a ser utilizada para se referir de maneira pejorativa às mulheres por volta de 1400, sendo considerado, portanto, um vocábulo vulgar e obsceno na língua inglesa desde então. O uso do vocábulo, que carrega difamação sexual, é considerado um insulto, cujo significado equivale ao mesmo de outras palavras da língua inglesa como *whore*, *slut* e *hoe*². Hoje em dia, o vocábulo continua sendo utilizado com conotação pejorativa para se referir às mulheres. No entanto, o tabu em torno da palavra mudou. A palavra adquiriu novos usos que se expandiram e que, hoje em dia, também fazem parte do uso da língua inglesa por falantes de língua materna e de segunda língua.

O vocábulo em questão é muito utilizado em letras de músicas de *rappers* americanos por ambos os sexos, feminino e masculino. Assumimos a hipótese de que o vocábulo *bitch* é usado por *rappers* do sexo masculino em sua maioria com conotação pejorativa, especialmente para se referirem às mulheres, tratando-as como objeto e expressando domínio sobre elas. Por outro lado, assumimos que o uso do mesmo vocábulo por *rappers* do sexo feminino, em sua maioria, é feito para se referir a outra mulher, mas não de modo depreciativo e sim de maneira afetiva.

A partir dessas hipóteses, temos como objetivo principal analisar o uso do mesmo vocábulo em canções de *rap* interpretadas por *rappers* do sexo masculino e feminino, a fim de verificar os significados atribuídos a ele, partindo da premissa de que as palavras são as nossas ferramentas de pensamento, refletindo a realidade social da comunidade onde estão sendo usadas. Para tal, iniciaremos o processo com a compilação de letras de músicas a partir dos princípios metodológicos da Linguística de *Corpus* (LC). Ademais, utilizaremos o *software WordSmith Tools 6.0* (SCOTT, 2012) para uma análise quali-quantitativa de nosso *corpus* de estudo.

Este trabalho está organizado em cinco seções, que obedecem à seguinte ordem: introdução, fundamentação teórica, metodologia, resultados e discussão de dados e, por fim, considerações finais.

2 Fundamentação teórica

De acordo com Dalzell e Victor (2008), a palavra *bitch* tem sido utilizada de maneiras distintas desde a primeira vez que se soube a seu respeito. Em 1400, o termo era utilizado unicamente para se referir a uma mulher de comportamento sexual desprezível e era considerado um dos termos mais ofensivos da língua inglesa. Hidayat (2018) afirma que, em 1923, foi encontrado o

² As palavras *whore*, *hoe* e *slut* são traduzidas para o português brasileiro como *puta*.

que seria o primeiro registro do vocábulo para designar um homem com papel passivo em uma relação homossexual, colocando em dúvida sua masculinidade. Em 1943, o uso do vocábulo se expandiu, passando a designar uma coisa ou pessoa notável. Nos Estados Unidos, o vocábulo também é empregado como sinônimo da palavra *mulher*. Ainda segundo Hidayat (2018), embora o uso da palavra *bitch* tenha adquirido novos significados ao longo dos anos, a raiz do seu significado de origem, que é basicamente uma mulher de comportamento sexual desprezível, é ainda mais dominante do que outros significados. Atualmente, a palavra *bitch* possui uso recorrente no cenário *pop* e é usada desde canções a manchetes e, até mesmo, em diálogos verbais na televisão em horário nobre.

Os cantores, através de suas músicas, refletem o comportamento linguístico dos indivíduos de uma comunidade de fala e, assim como eles, empregam o vocábulo em questão de diferentes maneiras. Caretta (2011) afirma que os elementos históricos são indispensáveis para compreensão de uma canção, uma vez que determinam o estilo de uma época. Summers (2004) argumenta que um álbum define as sensibilidades musicais de um artista em um determinado momento da história. Uma vez que estudamos o comportamento linguístico de indivíduos em sua comunidade através de letras de músicas, este trabalho está inserido no âmbito da Sociolinguística, cuja função principal é estudar a relação entre língua e sociedade. Weinreich, Labov e Herzog (2006) afirmam que é impossível estudar as variações e mudanças de uma língua sem associá-las ao contexto social em que está inserida. A Sociolinguística nos permitirá, portanto, analisar como o uso linguístico do mesmo vocábulo feito por *rappers* do sexo masculino e feminino reflète o comportamento dos indivíduos na sociedade. Para a análise quantitativa dos dados, utilizaremos a LC, que é uma metodologia de base empírica, voltada para a análise e descrição de dados linguísticos através da análise de *corpus*.

Dentro da LC, trabalhamos com a Estilística de *Corpus*, que combina a metodologia de *corpus* com abordagens mais intuitivas. Para Mahlberg (2007), a Estilística de *Corpus* visa descrever e analisar o estilo, ou seja, o que é distinto na forma como a língua é utilizada por um autor específico, ou numa obra específica, ou num período específico.

Haja vista que a Sociolinguística e a Estilística de *Corpus* permitem a identificação de tendências, relações intertextuais e reflexões dos contextos sociais e culturais, consideramos essas abordagens apropriadas para a análise e descrição deste estudo.

3 Metodologia

Foram selecionadas quarenta letras de músicas do gênero *rap*. Em seguida, as letras foram compiladas e salvas em formato txt e codificação ANSI, a fim de que os arquivos fossem processados pelo *software* de análise linguística WST 6.0 (SCOTT, 2012). As canções foram

divididas em dois *corpora*: *male rappers' songs* e *female rappers' songs*. Foram selecionadas dez canções de cada um dos seguintes *rappers*: Da Babby, Roddy Ricch, Doja cat e Megan thee Stalion. A escolha dos *rappers* foi feita de acordo com os seguintes critérios: 1) todos os *rappers* estão na lista de top 10 artistas do ano de 2020 da renomeada revista de música americana *Billboard*; 2) todos fazem parte do segmento *rap*. As letras das quarenta canções foram retiradas do *website Genius*³, que foi escolhido por ter as letras de músicas verificadas pelos próprios artistas e, também, por exibir o número de visualizações que cada música possui. Foram compiladas as letras das dez canções mais acessadas pelos usuários da plataforma de música de cada um dos artistas acima mencionados.

Após a compilação das quarenta canções, utilizamos a ferramenta *Wordlist* do programa WST para gerar uma lista de palavras. Essa ferramenta permite a criação de uma lista de palavras presente no *corpus*, elencadas em conjuntos com suas frequências absolutas e percentuais (BERBER SARDINHA, 2006). A fim de verificar o tamanho dos dois *corpora* deste estudo, foram confeccionadas duas listas de palavras: uma para o *corpus male rappers' songs* e outra para o *corpus female rappers' songs*. A tabela 1 mostra a extensão do *corpus*, cujo número total de *tokens*⁴ é de 26,314.

Tabela 1: Número de *tokens* do *corpus* de estudo.

NOME DO CORPUS	Nº DE CANÇÕES COMPILADAS	NÚMERO DE TOKENS
<i>Male rappers' songs</i>	20	13,435
<i>Female rappers' songs</i>	20	12,879

Fonte: Elaborada pelo autor.

De acordo com Berber Sardinha (2004), devido ao número total de *tokens*, esse *corpus* é caracterizado como pequeno. No entanto, como nosso objetivo é analisar o uso do vocábulo *bitch* em letras de *rap*, o tamanho do *corpus* se mostra adequado e representativo para esse estudo. Após a elaboração das listas de palavras nos dois *corpora*, observamos que a palavra de conteúdo com maior número de ocorrências em ambas as listas foi o vocábulo *bitch*; guiados por nossa intuição, optamos por trabalhar com esse item lexical. Os itens cujas frequências foram maiores são palavras funcionais, tais como, pronomes, preposições e artigos, que não apresentam traços distintivos para nosso estudo.

Após a confecção das duas listas de palavras obtivemos a seguinte lista de frequência para o item lexical *bitch*.

³ Disponível em: <https://genius.com>. Acesso em: 10 set. 2021.

⁴ *Tokens* se refere ao número total de palavras existentes em um dado texto ou *corpus*.

Tabela 2: Frequência do item lexical *bitch* nos dois corpora.

FEMALE RAPPERS' SONGS	MALE RAPPERS' SONGS
168	83

Fonte: Elaborada pela autora.

Buscando observar como o vocábulo ***bitch*** é utilizado por *rappers* americanos a partir da variável independente sexo, analisamos as linhas de concordância desse item lexical em ambos os corpora. As linhas de concordância exibem todas as ocorrências de uma palavra ou grupo de palavras precedidas e seguidas de seu contexto no texto (RAYSON, 2015). Devido à alta frequência do vocábulo ***bitch*** em ambos os corpora, limitamos o estudo apenas ao seguinte ambiente sintático: adjetivos + substantivo ***bitch***. Após a limpeza das linhas de concordância que não seguiam o ambiente sintático a ser observado, obtivemos o seguinte recorte para as linhas de concordância.

Figura 1: Linhas de concordância da palavra *bitch* no corpus *female rappers' songs* após a limpeza.

N	Concordance	Set	Tag	Word #	Sent.	Para.	File	Date	%
					Sent. Pos. Para	Para. Pos. Hear Hear Sect			
1	think this shit get better with another bitch, go get her (Go and get her)			521	1 66% 0	73%	bitch.txt	2021/out/01	73%
2	see you proud of me I'm just a real ass bitch, give a fuck about a trick On			472	12 14% 0	51%	hot girl summ	2021/set/29	51%
3	some cameras in here I'm a bad bitch, she's a savage, no comparison			626	17 15% 0	74%	savage.txt	2021/set/29	73%
4	Shoulda known you had the baddest bitch Left on read and can't give head			181	0 60% 0	43%	aint shit.txt	2021/set/30	44%
5	cow Bitch, I'm a cow, bitch, I'm a cow Bitch, I'm a cow, bitch, I'm a cow (Moo			396	5 43% 0	56%	moo.txt	2021/set/30	56%
6	Bitch, I'm a cow, bitch, I'm a cow Bitch, I'm a cow, bitch, I'm a cow			301	5 26% 0	43%	moo.txt	2021/set/30	43%
7	wish a bitch would, and I'm a genie Bitch so hot, gotta stay in bikinis He			243	1 63% 0	28%	thot shit.txt	2021/set/29	28%
8	Thee Stallion] Your honor, I'm a freak bitch, handcuffs, leashes Switch my			518	0 85% 0	68%	wap.txt	2021/set/29	67%
9	the ho that you fuckin' with if I find her Bitch, you ain't that busy, I don't give a			209	1 13% 0	29%	bitch.txt	2021/out/01	29%
10	to say jealous All y'all bitches is jealous Bitch, bitch [Chorus] Said play with my			246	0 53% 0	38%	rules.txt	2021/set/30	39%
11	V? (Brr) And who gon' tell him that my bitch is getting her degree? And when			405	9 71% 0	44%	hot girl summ	2021/set/29	44%
12	follow me? 'Cause even in your new bitch, I can see a lot of me And			436	12 2% 0	47%	hot girl summ	2021/set/29	48%
13	like I'm alien, bitch, I'm fucking reptilian Bitch, bitch, aw All y'all bitches was			209	0 45% 0	33%	rules.txt	2021/set/30	33%
14	beat [Verse 1: Megan Thee Stallion] Bitch, I'm a star, got these niggas			42	0 14% 0	5%	cash shit.txt	2021/set/29	5%
15	check the throwback pics, I been that bitch (Uh) [Chorus] I'm a hot girl, I do			523	0 86% 0	86%	captain hook.	2021/set/29	86%

Fonte: WordSmith Tools 6.0.

Figura 2: Linhas de concordância da palavra *bitch* no corpus *male rappers' songs* após a limpeza.

N	Concordance	Set	Tag	Word #	Sent.	Para.	File	Date	%
					Sent. Pos. Para	Para. Pos. Hear Hear Sect			
1	a sniff I'm pure like cocaine in the '80s Bitch [Outro] That ain't the baby, that's			564	0 98% 0	98%	going baby.tx	2021/out/03	98%
2	a hundred plays in a day Put an Arabic bitch inside of the face And I got the			230	1 51% 0	37%	tip toe.txt	2021/set/30	36%
3	(Woo) This ain't no mid from Arizona Bitch, I'm serious (Real), period			463	8 65% 0	59%	said sum.txt	2021/set/29	59%
4	[Verse 3: Yung Miami] You broke ass bitch, you ain't sayin' nothin' When I			382	8 21% 0	49%	said sum.txt	2021/set/29	48%
5	diamonds offset like Cardi Got a bad bitch with me, she a Barbie, huh You			61	0 10% 0	10%	start with me.	2021/set/30	10%
6	dies I make the chrome-chrome fly Bitch, I'm a wave, Roll Tide Bullets at			246	1 49% 0	65%	war baby.txt	2021/out/01	59%
7	out the price on a boat (Okay) My lil' bitch act like Megan Thee Stallion			473	6 16% 0	64%	bop.txt	2021/set/29	64%
8	whip had a cop in it (Woo, okay) My bitch got good pussy, fly her 'cross the			311	3 76% 0	42%	bop.txt	2021/set/29	42%
9	Don't think that she lied to you, nigga (Bitch) Get caught with your ho when			113	0 17% 0	15%	suge.txt	2021/set/29	15%
10	I've been trapping got to get a Patek Bitch I'm rich I can't be livin' in debt			671	1 79% 0	80%	every season	2021/set/30	80%
11	me Clip got thirty-three, Scottie Pippen Bitch nigga ain't heard of me, I got a			31	0 5% 0	5%	start with me.	2021/set/30	5%
12	'Member I used to cheat off a pretty bitch test All the teachers, they thought			631	0 95% 0	83%	suge.txt	2021/set/29	83%
13	Meech, and Chapo That's my top three Bitch I was serving the sidewalk Then I			188	1 22% 0	22%	every season	2021/set/30	23%
14	need a reason, loyalty over treason Bitch nigga, come and see me Put			265	3 36% 0	43%	under the sun	2021/set/29	42%
15	Aquafina, quick trip to Catalina White bitch bad as Selena, fuck her, then a			404	0 68% 0	68%	down below.t	2021/set/30	67%
16	like cocaine in the '80s [Verse 2] Your bitch on my dick, she a junkie (Uhh) I			322	0 56% 0	56%	going baby.tx	2021/out/03	57%
17	'em up bet I'ma shoot 'em (Boom) Your bitch is a dog, she got rabies (Brr,			378	0 66% 0	66%	going baby.tx	2021/out/03	66%

Fonte: WordSmith Tools 6.0.

A partir das ocorrências do vocábulo *bitch* nos dois *corpora* de estudo, optamos por criar campos semânticos, a fim de analisar e descrever os adjetivos que precedem ao vocábulo levando em consideração a variável independente sexo. De acordo com Fischer-Starcke (2009), ao identificar campos semânticos em um *corpus*, os significados dominantes dos dados podem ser revelados.

4 Resultados e discussão

A partir das linhas de concordância das letras de *rap* interpretadas por mulheres, observamos que os adjetivos mais frequentemente empregados com o vocábulo *bitch* possuem conotações positivas referentes ao comportamento da mulher na sociedade, atribuindo a elas o papel de mulheres empoderadas e seguras de si, como é possível observar nas seguintes linhas.

- *I am just a real ass bitch.*
*Eu sou uma mulher foda*⁵.
- *Shoulda known you had the baddest bitch.*
Você deveria saber que você tinha a melhor de todas.

Também foi observado o uso do vocábulo com conotação negativa para se referir a si mesma ou a outra mulher. Atestamos o uso do vocábulo para descrever alguém cuja personalidade é condenada, ora por esta mulher se comportar de maneira inadequada ou por haver algum tipo de rivalidade entre duas mulheres, onde uma deseja possuir o que é da outra. Atestamos tal nas seguintes linhas de concordância.

- *Your honor, I'm a freak bitch.*
Sua excelência, eu sou uma vadia louca.
- *All y'all bitches is jealous bitches.*
Todas vocês são vadias invejosas.

O vocábulo também foi utilizado nas letras de músicas cantadas por mulheres, para se referirem de maneira afetiva a outras mulheres com as quais assumimos que se tenha um relacionamento de amizade. Não havendo, portanto, qualquer significado depreciativo.

- *Who gon' tell him that my bitch is getting her degree?*
Quem vai dizer a ele que minha amiga vai se formar?

Por fim, notamos o uso do vocábulo *bitch* para se referir a outra mulher, cujo comportamento é desagradável e com a qual se tem algum tipo de rivalidade.

⁵ Apesar de haver traduções das canções disponíveis na internet, foram necessários alguns reajustes a fim de que a tradução se adequasse ao contexto retratado na canção.

- *Think this shit get better with another bitch, go get her.*
Você acha que isso vai melhorar com outra vadia, vá buscá-la!

Identificamos, portanto, quatro campos semânticos para o item lexical *bitch* no *corpus female rappers' songs*.

Quadro 1: Campos semânticos do vocábulo *bitch* no *corpus female rappers' songs*.

Campo semântico	Adjetivo + <i>bitch</i>	Significado
Relação de poder	<i>Real ass bitch</i> <i>Bad bitch / the baddest bitch</i> <i>Genie bitch</i>	Mulher de atitude, confiante, inteligente e independente, que lida de maneira direta com seus problemas
Insulto	<i>Freak bitch</i>	Mulher louca e sem pudores.
	<i>Jealous bitch</i>	Mulher invejosa e que fala mal de outra mulher.
	<i>Cow bitch</i>	Uma mulher que te irrita e que faz coisas erradas.
Afeto	<i>My bitch</i> <i>Her bitch</i>	Uma mulher com a qual se tem um relacionamento de amizade.
Competitividade	<i>Another bitch</i> <i>New bitch</i>	Mulher que não te agrada.

Fonte: Elaborado pelo autor.

Diferentemente do que foi observado na análise do uso do vocábulo por mulheres, constatamos que o uso do vocábulo nas letras de *rap* cantadas por homens apresenta teor negativo. As linhas abaixo revelam o tratamento que a figura da mulher recebe, por parte dos *rappers* no *corpus* em questão. Constatamos que a figura feminina é retratada como um objeto sexual cuja única função é a de satisfazer os desejos dos homens.

- *Your bitch on my dick.*
Sua vadia no meu ^{***6}.
- *My bitch got good pussy*
Minha vadia tem a ^{*****} *boa.*
- *My little bitch.*
Minha vadiazinha.

Observamos que o vocábulo foi precedido em algumas linhas de concordância por pronomes adjetivos, que parecem demonstrar a relação de poder e de posse que a figura masculina (aqui representada pelos *rappers*) admite ter sobre a figura feminina. Diferentemente do que foi identificado anteriormente no *corpus* dos *rappers* do sexo masculino, atestamos o uso do

⁶ Optamos por não exibir a tradução de algumas palavras por considerá-las inadequadas.

termo para se referir à mulher de maneira não depreciativa, mesmo que isso tenha ocorrido uma única vez. Foi identificada em uma das linhas de concordância o uso do adjetivo *bad* para descrever o item lexical *bitch*. A expressão *bad bitch* é utilizada por ambos, homens e mulheres, mas com maior frequência atestada na fala das mulheres para se referir a mulheres que são autoconfiantes e seguras de si.

- *Got a bad bitch with me.*
Tenho uma mulher foda comigo.

Também foi atestado o uso do item para se referir a outros homens negros com conotação negativa, colocando em dúvida sua masculinidade.

- *Don't you think that she lied to you, nigga bitch.*
Não acha que ela mentiu para você, seu veado.

Em algumas linhas, percebemos que o vocábulo foi utilizado para descrever uma pessoa detestável, com a qual não se tem relação amigável independente do gênero, podendo se referir a um homem ou a uma mulher.

- *You broke ass bitch.*
Seu(sua) otário(a).

Por fim, notamos que o vocábulo também foi utilizado para se referir a pessoas em geral sem cunho depreciativo. Percebemos que nesse uso o vocábulo *bitch* seria equivalente a outras palavras da língua inglesa como *guys* e *dude*⁷.

- *Top three bitch.*
Os caras do meu top 3.
- *Arizona bitch.*
Um cara do Arizona.

A partir das análises feitas acima, concluímos que, nas letras do *corpus male rappers' songs*, o vocábulo *bitch* foi utilizado em quatro campos semânticos diferentes, como podemos ver no quadro 2.

⁷ Tradução nossa: galera e parceiro

Quadro 2: Campos semânticos do vocábulo *bitch* no *corpus male rappers' songs*.

Campo semântico	Adjetivo + <i>bitch</i>	Significado
Posse/poder	My bitch Your bitch	Mulher que é vista como propriedade do homem e que dele depende.
Empoderamento	<i>Bad bitch</i>	Mulher independente e segura de si.
Ofensa/xingamento	<i>Broke ass bitch</i> <i>Little bitch</i>	Uma pessoa fraca e desprezível e de comportamento desagradável.
Termo de referência	<i>Arizona bitch</i>	Qualquer pessoa.

Fonte: Elaborado pelo autor.

Uma vez apresentadas as análises deste trabalho, passamos à seção de considerações.

5 Considerações finais

O vocábulo *bitch*, através de um processo de *broadening*⁸, adquiriu novos significados ao longo dos anos. Através de procedimentos metodológicos da LC e com o auxílio do *software Wordmith Tools 6.0*, realizamos uma análise do item lexical *bitch* nas letras de canções de *rappers* norte-americanos, onde atestamos que ambos os *rappers*, do sexo masculino e feminino, usam o vocábulo *bitch* de maneira recorrente em suas canções.

A partir das premissas da Estilística de *Corpus* e da Sociolinguística, confirmamos as hipóteses inicialmente apresentadas e comprovamos que há variação de uso, ou seja, de significados do vocábulo *bitch* nas letras de canções de *rap*. Após a análise dos dados, que foi feita a partir da variante independente sexo, podemos afirmar que o significado atribuído ao vocábulo *bitch* depende de três fatores: contexto, pessoa e propósito. A variação estilística das letras de *rap* está relacionada diretamente à pessoa que as interpreta ou as escreve e isso pode ser atestado a partir das análises das escolhas linguísticas do autor ou cantor. Essas escolhas linguísticas parecem mostrar o comportamento desses indivíduos na sociedade, refletindo assim o seu propósito dentro dela. Após análise quali-quantitativa dos dados, atestamos que o termo *bitch* foi utilizado majoritariamente no *corpus male rappers' songs* com conotação negativa, referindo-se principalmente à mulher de modo pejorativo. Através da análise das linhas de concordância, notamos que, nas letras desse *corpus*, a mulher é colocada em uma posição na qual o homem exerce algum tipo de poder sobre ela, resumindo-a apenas a um objeto sexual. Tais resultados parecem refletir o tratamento machista recebido pelas mulheres na sociedade atual, onde alguns homens as enxergam apenas como provedoras de suas vontades sexuais e também como seres dependentes deles tanto financeira como emocionalmente.

⁸ De acordo com Hollmann (2009), *broadening* é um processo de mudança semântica caracterizado pela expansão de significados de uma palavra ao longo dos anos.

Diferentemente do que esperávamos, também atestamos, com apenas uma ocorrência no *corpus* de estudo, o uso do item lexical para se referir às mulheres com o intuito de exaltá-las. Além disso, o vocábulo também foi empregado com outras conotações, tais como, xingamento e como termo geral para se referir a outra pessoa independente do gênero. Em relação ao uso do vocábulo no *corpus female rappers' songs*, atestamos que vários significados foram atribuídos ao item lexical *bitch*. O vocábulo foi utilizado como termo de referência ou de afeto, assim como forma de xingamento a outra pessoa, seja ela homem ou mulher. Comprovamos a hipótese de que o vocábulo é majoritariamente usado pelas *rappers* com conotação positiva. Os resultados desse *corpus* apontaram que o vocábulo *bitch* foi usado em sua maioria com o intuito de descrever as mulheres como figuras empoderadas, confiantes e autossuficientes.

O resultado da análise desse *corpus* parece refletir o papel que a mulher possui na sociedade atual. Ao longo dos anos, atribuiu-se à mulher o papel de um ser dependente da figura masculina, assim como de um ser inferior a ele. No entanto, graças aos movimentos feministas, uma grande mudança tem sido percebida no que diz respeito à figura feminina. A sociedade em geral, e principalmente as mulheres, está mais consciente do seu poder e da sua importância. A língua, que é um produto social, reflete, através das escolhas linguísticas dos seus falantes, o comportamento dos indivíduos de uma sociedade. As mudanças e variações linguísticas de uma língua também são atestadas na música, refletindo o contexto social onde aquele indivíduo está inserido, como observamos através da análise dos dados dos *corpora* deste trabalho.

Referências

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

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. São Paulo: Mercado de Letras, 2006.

CARETTA, A. A. **A canção e a cidade**: estudo dialógico-discursivo da canção popular brasileira e seu papel na constituição do imaginário da cidade de São Paulo na primeira metade do século XX. 2011. Tese (Doutorado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011. DOI: <https://doi.org/10.11606/T.8.2011.tde-28092011-104649>.

DALZELL, T; VICTOR, T. **The concise new partridge dictionary of slang and unconventional English**. London: Routledge, 2008.

FISHER-STARCKE, B. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: a corpus-stylistic analysis. **International Journal of Corpus Linguistics**, v. 14, n. 4, p. 492-523, 2009. DOI: <https://doi.org/10.1075/ijcl.14.4.03fis>.

HIDAYAT, T. Z. **Slang in American and British Hip-Hop/Rap Song Lyrics**. *LEXICON*, v. 5, n. 1, p. 84-94, 2018. DOI: <https://doi.org/10.22146/lexicon.v5i1.41284>.

HODGSON, C. Bitch. **Podictionary**. 8 fev. 2008.

HOLLMANN, W. B. Semantic change. *In*: CULPEPER, J. *et al.* (ed.). **English language: description, variation and context**. Basingstoke: Palgrave, 2009. p. 301-313.

MAHLBERG, M. Corpus stylistics: bridging the gap between linguistic and literary studies. *In*: HOEY, M. *et al.* (ed.). **Text, Discourse and Corpora: theory and analysis**. London: Continuum, 2007.

RAYSON, P. Computational tools and methods for corpus compilation and analysis. *In*: BIBER, D.; REPPEN, R. (ed.). **The Cambridge Handbook of English Corpus Linguistics**. Cambridge: Cambridge University Press, 2015.

SUMMERS, J. **Making and marketing music: the musician's guide to financing, distributing, and promoting albums**. New York: Allworth Press, 2004.

URBAN DICTIONARY. Disponível em: www.urbandictionary.com. Acesso em: 10 set. 2021.

WEINREICH, U.; LABOV, W.; HERZOG, M. **Fundamentos empíricos para uma teoria da Mudança Linguística**. Tradução Marcos Bagno. São Paulo: Parábola, 2006.

Contribuições da Linguística de *Corpus* para a sala de aula: prosódia semântica do item lexical *mulher*

Cássia Beatriz de Moraes Silva¹

¹ Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/3556885127366452>.
E-mail: cassia.de@ufu.br.

1 Introdução

A música faz parte da vida das pessoas desde os primeiros tempos da humanidade, como uma forma de expressão artística que propõe a harmonia entre sons, ritmo, melodia e voz. Por isso, a música tem o poder de emocionar, resgatar memórias, proporcionar lazer, diversão e até mesmo auxiliar em tratamentos médicos, como é o caso da musicoterapia, uma terapia baseada na música e que “está ganhando poder de remédio para silenciar males tão distintos quanto dor e depressão” (BIERNATH, 2019, *on-line*).

O fácil acesso e a versatilidade proporcionados pelos vários estilos fazem com que a música seja uma importante aliada no ensino de línguas, materna e estrangeira, devido a sua riqueza lexical. Por estar inserida no imaginário coletivo e transitar livremente em todas as camadas sociais, a canção é portadora de palavras e expressões com cargas culturais compartilhadas entre a coletividade (BARBOSA, 2005).

Perante o exposto, o objetivo geral do nosso estudo é identificar a conotação positiva, negativa ou neutra do item lexical *mulher* e de itens relacionados a este em letras de canções do estilo sertanejo (*country*). Os objetivos específicos são: 1) analisar o campo lexical das letras de canções por meio do *software WordSmith Tools 4.0* (SCOTT, 2004); 2) identificar os itens lexicais que se associam com maior frequência à palavra *mulher* e 3) desenvolver e apresentar uma sequência didática para discutir em sala de aula o campo lexical relacionado à palavra *mulher* nas letras de músicas, bem como para abordar suas conotações.

Na atual conjuntura, em que tanto se fala sobre igualdade de gênero, sobre as conquistas e os desafios das lutas feministas, e tendo em vista os números crescentes de violência contra a mulher, especialmente no atual contexto pandêmico, acreditamos que a sala de aula seja o espaço para problematizar todas essas questões sob o viés da Lexicultura (BARBOSA, 2009), propondo discussões sobre o item lexical *mulher*, sua carga cultural compartilhada e, por conseguinte, debates acerca da equidade entre os gêneros.

2 Fundamentação teórica

Nosso estudo está baseado nos Estudos do Léxico, no sentido de compreender as dimensões alcançadas pela palavra como detentora de poder (BIDERMAN, 1998), na indissociabilidade entre léxico e cultura (Lexicultura) (BARBOSA, 2009) e na Linguística de *Corpus* (BERBER SARDINHA, 2009).

Com relação ao *country* brasileiro, estilo musical analisado neste estudo, é importante ressaltar que estamos adotando o conceito definido pelo *Rodeo West*, site especializado em marcas constitutivas do universo *country*. Ainda de acordo com o site, o *country* surgiu no

Brasil no final da década de 1970. Ao longo dos anos, o país desenvolveu seu próprio estilo musical *country*, hoje chamado de sertanejo (RODEO WEST, 2017).

Ao recorrermos aos estudos da Lexicologia e da Lexicultura, embora sejam áreas muito semelhantes, reforçamos aqui o nosso entendimento de Lexicologia enquanto área da Linguística que se preocupa apenas com o léxico e que, apesar de considerar a dimensão cultural da palavra, é na Lexicultura que encontramos maior aporte para abordar o léxico marcado culturalmente e voltado para práticas de ensino por meio da abordagem da carga cultural compartilhada pelas palavras.

2.1 Dimensões da palavra e do léxico

A palavra, segundo Biderman (1998), apresenta as dimensões mágica, religiosa, cognitiva e linguística, a partir das quais podemos compreender seu poder na constituição das sociedades e na manutenção de costumes, hábitos e crenças de um povo. Aqui, vamos discorrer apenas sobre as dimensões mágica e religiosa, por entendermos que essas dimensões se relacionam intrinsecamente com as escolhas lexicais que fazemos no processo comunicativo.

Neste sentido, a autora diz que o homem primitivo acredita na relação entre a escolha de um nome e o ser (pessoa, objeto, coisa) e por isso a escolha dos nomes não se dá de maneira arbitrária, mas sim, é uma escolha que envolve crenças, mitos. É aí que a palavra adquire o poder mágico e influenciador no destino de uma pessoa, por exemplo.

Nas numerosas tradições culturais dos homens a linguagem surge com a palavra instituidora que abre ao ser o espaço para ele se manifestar. Todas as culturas nascem de uma *palavra* criadora, dita em tempos imemoriais por um poder divino (BIDERMAN, 1998, p. 84, grifo da autora).

A *palavra* assume assim nos mitos de cada cultura uma força transcendental; nela deitam raízes os entes e os acontecimentos. Por ser mágica, cabalística, sagrada, a palavra tende a constituir uma realidade dotada de poder. Os mitos falam dos segredos e das essências escondidas na palavra instituidora do universo (BIDERMAN, 1998, p. 81, grifo da autora).

Barbosa (2009) enfatiza que o léxico nos auxilia a compreender e a explicar a sociedade da qual fazemos parte ou pretendemos conhecer/integrar. Isso é possível por conta da carga cultural compartilhada das palavras, ou seja, trata-se do valor acrescentado ao sentido referencial da palavra, que é conhecido e compartilhado entre os membros pertencentes a uma cultura e constitui fator de aproximação e de reconhecimento mútuo.

O léxico possui um papel importante para a emissão e para a compreensão de significados, pois está diretamente ligado aos aspectos cognitivos, sociais e culturais de uma língua. De acordo com os princípios de E. Sapir, as palavras são uma forma privilegiada de acesso a uma cultura, uma vez que elas são portadoras de concepções ou de visões de mundo. Desse modo, as características específicas de cada língua podem ser consideradas o reflexo da identidade cultural da sociedade (BARBOSA, 2009, p. 31).

Apresentados esses primeiros pressupostos teóricos, passemos às considerações sobre a sequência didática a ser desenvolvida com o auxílio da Linguística de *Corpus*.

2.2 Sequência didática: contribuição da Linguística de *Corpus* para a sala de aula

Para que possamos refletir sobre qualquer tema em sala de aula e para que essa discussão seja significativa para os estudantes, faz-se necessária a elaboração de sequências didáticas amparadas pelos documentos oficiais que regem a educação nos entes federados. No estado de Goiás, esse documento é o *Currículo Referência do Estado de Goiás*, lançado em 2012, pela Secretaria Estadual da Educação. O documento em questão tem como objetivo contribuir com as unidades educacionais por meio da apresentação de propostas de bimestralização dos conteúdos, para a melhor compreensão dos componentes do currículo e sua utilização na sala de aula (GOIÁS, 2021).

Em 2021, o *Currículo Referência* passou por alterações e ganhou a versão adaptada em virtude do contexto pandêmico da Covid-19, buscando garantir qualidade no ensino da rede pública estadual, apesar dos impactos da pandemia para os estudantes. Embora tenha passado por adaptação, o documento continua prevendo práticas reflexivas, levando em conta aspectos sociais e culturais, entre eles, as discussões relacionadas à mulher. Sendo assim, apresentamos ao final das discussões deste estudo uma proposta de sequência didática desenvolvida com base nas orientações contidas no *Currículo Referência* (GOIÁS, 2021), no intuito de promover momentos de discussão e reflexão em sala de aula da temática trabalhada neste estudo.

2.3 Linguística de *Corpus* e *WordSmith Tools*

Neste estudo, como evidenciado anteriormente, lançamos mão da Linguística de *Corpus* (doravante LC), que, segundo Berber Sardinha (2004, p. 3),

[...] ocupa-se da coleta e exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador.

É importante ressaltar que a LC vem revolucionando cada vez mais a maneira como se investiga a linguagem nos seus mais diversos níveis (BERBER SARDINHA, 2009). Muitas são as possibilidades de aparato tecnológico para que o linguista faça suas análises. Nesse sentido, destacamos a importância da utilização do *software WordSmith Tools* (doravante WST) para o processamento de dados linguísticos. Segundo Berber Sardinha (2009, p. 6),

O programa WordSmith Tools é um conjunto de programas integrados ('suíte') destinado à análise lingüística. Mais especificamente, esse *software* permite fazer análises baseadas na frequência e na co-ocorrência de palavras em *corpora*. Além disso, ele permite pré-processar os arquivos do *corpus* (retirar partes indesejadas de cada texto, organizar o conjunto de arquivos, inserir e remover etiquetas etc.), antes da análise propriamente dita.

Neste estudo, utilizamos a LC pela produtividade na organização e seleção dos dados a serem analisados. O WST, por sua vez, foi utilizado devido ao fato de oferecer ferramentas que permitem análises em diferentes níveis do *corpus*.

3 Procedimentos metodológicos

O primeiro passo para o desenvolvimento deste estudo foi uma pesquisa por conteúdos disponibilizados pelo *site* de músicas Vagalume² e verificamos que o mesmo conta com cerca de quarenta *estilos*, ou gêneros musicais e mais de mil artistas. Feito isto, o critério que adotamos para a escolha do estilo musical foi sua popularidade e levando em consideração que o estilo sertanejo é predominante no estado de Goiás, onde o estudo foi realizado, optamos pela seleção de todas as letras de música do estilo *country*, composto por 29 artistas, com músicas sertanejas variadas, algumas atuais, outras nem tanto, com nomes como Paula Fernandes, Nayara Azevedo, Chitãozinho e Xororó e Edson e Hudson, entre tantos outros. Em seguida, iniciamos o processo de compilação das letras. O *corpus* é composto por 532 arquivos, que foram processados por meio do WST. A análise foi realizada utilizando as ferramentas *WordList* (lista de palavras) e o *Concord* (concordanciador).

4 Análise e discussão

4.1 *WordList*

Na primeira etapa da análise de dados, por meio da ferramenta *WordList*, mensuramos a extensão do *corpus*, que totaliza 87.501 *tokens* (itens) e 6.755 *types* (formas). Desse total, com base na análise da lista de palavras, observamos que a palavra "que" apresentou a maior

² Disponível em: <https://www.vagalume.com.br/>. Acesso em: 25 jan. 2022.

frequência no *corpus*, com 3.101 (3,54%) ocorrências. Aqui, é importante ressaltar que a palavra “que” exerce várias funções em um enunciado, podendo ser conjunção, pronome, substantivo, advérbio, preposição, interjeição ou partícula de realce (RIGONATTO, 2022). Desse modo, a definição da função da palavra “que” depende do enunciado no qual ela está inserida.

A segunda palavra mais frequente foi o pronome pessoal reto “eu”, com 2.814 (3,22%) ocorrências, seguida de “o”, que pode ser artigo masculino ou pronome, com 2.394 (2,74%) ocorrências. O advérbio de negação “não” ocorre 2.165 (2,74%) vezes, enquanto a preposição “de” ocorre 2.057 (2,35%) vezes. O vocábulo “a”, que pode ser artigo feminino, pronome ou preposição, ocorre 2.038 (2,33%) vezes, ao passo que a conjunção “e” possui 2.035 (2,33%) ocorrências. O pronome “você” ocorre 1.738 (1,99%) vezes, o também pronome “me” ocorre 1.404 (1,60%) vezes, e, por fim, a palavra “é”, verbo ser no presente do indicativo, aparece 1.245 (1,42%) vezes no *corpus*.

Aqui, cabe ressaltar que apresentamos as 10 palavras mais frequentes da *WordList*, classificadas como palavras instrumentais: artigos, preposições, pronomes, conjunções etc. (BIDERMAN, 1998). Como podemos observar, com exceção do verbo “ser” no presente do modo indicativo (“é”), as demais palavras não são plenas e dependem de outras para que seu significado seja completo.

Figura 1: Recorte da lista de palavras organizada por ordem de frequência.

N	Word	Freq.	%	Texts	%	Dispersion	Lemmas	Set
1	QUE	3.101	3,54%	493	92,67%	0,97		
2	EU	2.814	3,22%	436	81,95%	0,90		
3	O	2.394	2,74%	478	89,85%	0,95		
4	NÃO	2.165	2,47%	448	84,21%	0,94		
5	DE	2.057	2,35%	465	87,41%	0,98		
6	A	2.038	2,33%	443	83,27%	0,91		
7	E	2.035	2,33%	443	83,27%	0,94		
8	VOCÊ	1.738	1,99%	332	62,41%	0,82		
9	ME	1.404	1,60%	364	68,42%	0,90		
10	É	1.245	1,42%	364	68,42%	0,94		
11	PRA	1.111	1,27%	355	66,73%	0,91		
12	SE	999	1,14%	339	63,72%	0,90		
13	TE	961	1,10%	287	53,95%	0,88		
14	MEU	951	1,09%	313	58,83%	0,94		
15	UM	872	1,00%	309	58,08%	0,91		
16	DO	823	0,94%	281	52,82%	0,95		
17	AMOR	726	0,83%	260	48,87%	0,87		
18	VOU	647	0,74%	200	37,59%	0,87		
19	MAIS	622	0,71%	244	45,86%	0,89		
20	NO	615	0,70%	245	46,05%	0,89		

Fonte: *WordList*.

Ainda seguindo as definições de Biderman (1998), as palavras plenas são aquelas que possuem significação completa, entre elas os substantivos, adjetivos e verbos. A esse respeito, chama nossa atenção o fato de a palavra plena com maior frequência no *corpus* de estudo ser *amor*, com 726 ocorrências, correspondendo a 0,83% do total de *tokens* do *corpus*. Já a segunda palavra plena mais frequente foi *coração*, com 437 ocorrências, representando 0,50% do total de *tokens*. Esses dados podem ser observados nas Figuras 2 e 3, a seguir.

Figura 2: Recorte da lista de palavras com o item *amor* em destaque.

8	VOCE	1.744	1,99	335	62,97
9	ME	1.404	1,60	364	68,42
10	PRA	1.126	1,28	359	67,48
11	SE	1.002	1,14	340	63,91
12	TE	961	1,10	287	53,95
13	MEU	951	1,08	313	58,83
14	DO	878	1,00	289	54,32
15	AMOR	726	0,83	260	48,87
16	MAIS	623	0,71	244	45,86
17	NO	620	0,71	246	46,24

Fonte: *WordList*.

Figura 3: Recorte da lista de palavras com o item *coração* em destaque.

Country_wordlist.lst

N	Word	Freq.	%	Texts	%	emmas	Set
24	NA	541	0,62	210	39,47		
25	MAS	460	0,52	205	38,53		
26	MIM	459	0,52	193	36,28		
27	POR	440	0,50	221	41,54		
28	CORACAO	438	0,50	228	42,86		
29	MINHA	407	0,46	177	33,27		
30	UMA	402	0,46	193	36,28		
31	SEI	395	0,45	162	30,45		
32	QUERO	374	0,43	145	27,26		

Fonte: *WordList*.

Desse modo, guiados pelos dados apresentados até o momento, acreditamos que a frequência estatisticamente significativa dessas palavras comprova que a temática amorosa é predominante nas letras de músicas que compõem nosso *corpus* de estudo.

Na Figura 4, destacamos o objeto central de nosso estudo, a palavra *mulher*. Não obstante, é possível perceber nesse mesmo recorte da lista de palavras algumas variações, como *muié* e *muiezada*, um fenômeno linguístico que tanto pode ocorrer em função do nível de escolaridade do falante quanto em função de situações informais e até mesmo como estilo adotado pelo falante quando ele busca demarcar seu espaço social através das palavras que usa para se

comunicar, conforme aponta Eckert (2003). Nesse caso, essas variáveis são comuns no universo sertanejo, portanto, estando também presentes em letras de canções sertanejas.

Figura 4: Recorte da lista de palavras com o item *mulher* destacado.

N	Word	Freq.	%	Texts	%	Dispersion	Lemmas	Set
4.175	MUIÉ	14	0,02%	10	1,88%	0,52		
4.176	MUIEZADA	1		1	0,19%	0,00		
4.177	MUITA	14	0,02%	13	2,44%	0,63		
4.178	MUITAS	7		7	1,32%	0,45		
4.179	MUITO	92	0,11%	62	11,65%	0,78		
4.180	MUITOS	5		5	0,94%	0,40		
4.181	MULATA	2		2	0,38%	0,35		
4.182	MULATAS	1		1	0,19%	0,00		
4.183	MULHER	77	0,09%	39	7,33%	0,76		
4.184	MULHERADA	26	0,03%	11	2,07%	0,60		
4.185	MULHERÃO	19	0,02%	2	0,38%	0,11		
4.186	MULHERENGO	2		1	0,19%	0,00		
4.187	MULHERES	7		5	0,94%	0,45		

Fonte: *WordList*.

Conforme observa-se na Figura 4, o item lexical *mulher* ocorre 77 vezes (0,09%) no *corpus*. Por sua vez, a palavra *mulata*, filha de pai branco e mãe negra ou vice-versa (HOUAISS, 2009), variante da palavra *mulher*, aparece apenas duas vezes e, flexionada no plural, *mulatas*, aparece somente uma vez.

Apesar da baixa frequência desses dois vocábulos, se levarmos em consideração outras palavras que também são usadas para se referir à mulher no *corpus*, como *muié*, *muiezada*, *mulherada*, é possível verificar a recorrente presença feminina nas letras que compõem esse *corpus*. Além disso, a análise das linhas de concordância permite observar o tratamento que é dado à mulher no *corpus* em questão, bem como examinar a prosódia semântica do item *mulher*.

Após gerar as linhas de concordância para o item *mulher* e organizá-las em ordem alfabética a partir da primeira palavra à esquerda do item em questão, constatamos que as palavras que antecedem o item lexical *mulher* pertencem às seguintes classes gramaticais: artigo feminino, preposição, pronome, conjunção, verbo, adjetivos, numerais e alguns substantivos. Entre os pronomes, o possessivo *minha* possui uma frequência alta, ocorrendo 407 vezes (0,47%).

Em uma das linhas de concordância, *minha* é seguido do substantivo *mulher* em: “hoje ao lado dos meus filhos e da *minha mulher...*”, na letra da canção *Jesus pintou com as mãos*, interpretada por Marco Brasil. Em outros momentos, o possessivo *minha* é seguido dos substantivos *vida*, *menina*, *paixão*, *pequena*, *amada*, *querida*, entre outros vocábulos que sugerem a idealização da mulher, por quem o eu lírico demonstra dependência afetiva. Algumas dessas linhas de concordância estão apresentadas nas Figuras 5 a 8.

Figura 5: Recorte das linhas de concordância com *minha vida* em destaque.

N	Concordance
348	Minha linda, meu amor, meu sabor, minha vida Flor do campo, te desejo Deixa eu te
349	machucada Procurando salvação Dei a ela minha vida Fui afeto, fui saída Fui amante, fui
350	que errei... quero te reconquistar outra vez Amor da minha vida ... ha vida...
351	Era minha escolha preferida A vaga de amor na minha vida Você mas perdeu Quando inesperado
352	meu "véio" se ele souber que eu vou levar Nessa minha vida largada uns 8 anos pra formar
353	um ano que você se foi Refazendo um resumo da minha vida Vejo que tudo que construí não me

Fonte: *Concord*.

Figura 6: Recorte das linhas de concordância com *minha pequena* em destaque.

N	Concordance
275	inferno, A tua calma Eu sou teu tudo, Sou teu nada Minha pequena és minha amada Eu sou teu
276	inferno, A tua calma Eu sou teu tudo, Sou teu nada Minha pequena és minha amada Eu sou teu
277	morrer Eu sei que parte do que eu passei Foi culpa minha porque eu deixei Que outro alguém me
278	morrer Eu sei que parte do que eu passei Foi culpa minha porque eu deixei Que outro alguém me
279	falta de você E se acaso eu não me conter e perder minha postura Não se acanhe, só me beije E me
280	que sou galinha, que ando fora da linha Que a minha praia é a gandaia Que eu tenho cara de
281	que sou galinha, que ando fora da linha Que a minha praia é a gandaia Que eu tenho cara de

Fonte: *Concord*.

Figura 7: Recorte das linhas de concordância com *minha moça* em destaque.

N	Concordance
229	da cidade To chegando de viagem, vim buscar Minha moça , meu amor A mais linda desse
230	Te abençoar, fortalecer Mais que um dever é a minha motivação Irmão amado eu reconheço teu
231	cedinho A viola, eu e deus Tudo que dá na tv minha muié qué fazê não mede as consequências
232	palito As moças de hoje Eu não facilito... Eu mais a minha muié Fizemo combinação Eu vou no
233	mostrar o meu desejo eu senti um beliscão, era minha muié feia torcendo a minha orelha me
234	SÁBADO QUE VEM ELA FICA EU VOU EU MAIS A MINHA "MUIÉ" JÁ VI QUE NÓS "NUM" COMBINA

Fonte: *Concord*.

Figura 8: Recorte das linhas de concordância com *minha menina* em destaque.

N	Concordance
207	não apareceu Eu só quero você perto de mim Vem minha menina estou só Eu só quero um beijo seu
208	não apareceu Eu só quero você perto de mim Vem minha menina estou só Eu só quero um beijo seu
209	menina estou só Eu só quero um beijo seu Vem minha menina estou só Eu só quero um beijo seu
210	não apareceu Eu só quero você perto de mim Vem minha menina estou só Eu só quero um beijo seu
211	menina estou só Eu só quero um beijo seu Vem minha menina estou só Eu só quero um beijo seu
212	menina estou só Eu só quero um beijo seu Vem minha menina estou só Eu só quero um beijo seu
213	não apareceu Eu só quero você perto de mim Vem minha menina estou só Eu só quero um beijo seu
214	menina estou só Eu só quero um beijo seu Vem minha menina estou só Eu só quero um beijo seu

Fonte: *Concord*.

Retomando a análise das linhas de concordância, identificamos o vocábulo *minissaia*, definido pelo Houaiss (2009) como sendo uma saia muito curta que deixa parte da coxa à mostra. Apesar de esse vocábulo ter ocorrido uma única vez no *corpus* (Figura 9), ao ser associado com

a palavra *mulher*, apresenta prosódia semântica negativa, pois precede a expressão “mulher toma sua linha”, sugerindo, no contexto da canção, que ela mude suas vestes, demonstrando uma tentativa de controle do homem sobre o corpo da mulher. Aqui, é importante lembrar que as marcas da oralidade estão presentes também na escrita, suprimindo sinais de pontuação que separam itens lexicogramaticais, como a vírgula no caso dos vocativos. Assim, numa transcrição conforme a norma-padrão, a construção seria: “Minissaia, mulher! Tome sua linha”.

Figura 9: Recorte das linhas de concordância do item *mulher*.

46	do João de Barro Não tem goteira... "Mini-Saia" Mulher toma sua linha Eu sou seu marido Tô			
47	sou o dono da moda Como mais do que soda É só mulher do meu lado 2 litros de whisky 20 caixas			

Fonte: *Concord*.

A partir da reorganização das linhas de concordância em ordem alfabética com base no primeiro item à direita do nóculo *mulher*, identificamos o adjetivo *namoradeira* (*mulher namoradeira*). De acordo com o *Minidicionário Escolar da Língua Portuguesa*, o adjetivo *namoradeira* é definido da seguinte forma: moça jovem que gosta de namorar, que aceita facilmente a corte que se lhe faz (BUENO, 2018). Não discutiremos aqui os processos de formação da palavra *namoradeira*, o que poderá ser feito em estudos posteriores, mas, se levarmos em consideração valores socioculturais que reprovam o comportamento descrito na definição do dicionário, pode-se inferir que a prosódia semântica dessa associação seja negativa.

Ao mencionarmos valores socioculturais em torno da palavra *namoradeira*, temos a convicção de que a discussão não se esgota aqui, pois são necessárias muitas outras ponderações, com testes de percepção acerca das palavras numa abordagem sociolinguística, o que faremos em estudos posteriores. Por enquanto, consideraremos a definição do dicionário sugerindo a prosódia negativa.

Figura 10: Recorte das linhas de concordância do item *mulher*.

47	E o burro impacador Eu corto ele na espora E a mulher namoradeira Eu passo um rôdo E mando	42	0	40	0	40	0	40
48	E o burro impacador Eu corto ele na espora E a mulher namoradeira Eu passo um rôdo E mando	42	0	40	0	40	0	40
49	da medida Mas ela foi traída Mulher não trai, mulher não trai Mulher se vinga Elas se vingam de	112	0	110	0	110	0	110
50	medida Ela foi traída, ela tá ferida Mulher não trai, mulher não trai Mulher se vinga i Mulher se	233	0	231	0	231	0	231
51	E ainda falam por ai que mulher não manda em homem Manda sim!	7	0	5	0	5	0	5
52	Não tem o que fazer E ainda falam por ai que mulher Não manda em homem Manda sim!	108	0	106	0	106	0	106
53	, ela não passava da medida Mas ela foi traída Mulher não trai, mulher não trai Mulher se vinga	109	0	107	0	107	0	107
54	não passava da medida Ela foi traída, ela tá ferida Mulher não trai, mulher não trai Mulher se vinga i	230	0	228	0	228	0	228

Fonte: *Concord*.

Continuando a análise dos itens à direita do nóculo *mulher*, identificamos item lexical *verdadeira* na seguinte construção: “se perdeu a resposta, perdeu a mulher verdadeira”. Se considerarmos a definição de *verdadeiro* como algo legítimo (BUENO, 2018), nesse caso, não se atribui à palavra uma carga positiva ou negativa, e sim neutra, pois atribui-se à mulher o

fato de ser a legítima (casada, legalmente falando). Contudo, se olharmos para o restante da construção: “se a safada foi embora”, então *verdadeira*, que está em contraste com *safada*, se confirma como uma característica positiva, enquanto *safada*, subentendendo “mulher safada”, sugere prosódia semântica negativa para o item lexical *mulher*.

Figura 11: Recorte das linhas de concordância com *mulher verdadeira* em destaque.

73	o coração Dessa mulher Pra conquistar uma mulher Tem que fazer o que ela quer Tem que	103	0	101	0	101
74	do João de Barro Não tem goteira... "Mini-Saia" Mulher toma sua linha Eu sou seu marido Tô	235	0	233	0	233
75	atitudes bonitas Se perdeu a 'responsa' perdeu a mulher verdadeira Se a safada foi embora perdeu	138	0	136	0	136
76	se voce quer aprender Como conquistar uma mulher Você precisa entender de carinho e de	10	0	8	0	8

Fonte: *Concord*.

Os procedimentos analíticos e os dados apresentados nessa seção evidenciam a produtividade da exploração do campo lexical relacionado à palavra *mulher*. Os pontos discutidos podem fornecer bases para levar reflexões e discussões referentes às prosódias semânticas do campo lexical referente à palavra *mulher* para a sala de aula, além de poderem basear discussões sobre igualdade de gênero e propostas de projetos colaborativos. A partir da experiência neste estudo, apresentamos no apêndice desse capítulo uma sequência didática que comporta os tópicos abordados.

5 Considerações finais

A partir da Lexicultura e por meio dos recursos da LC, esse estudo buscou analisar a prosódia semântica da palavra *mulher* e de itens que se associam a essa palavra. Nossas análises e reflexões culminaram na proposta de uma sequência didática com letras de músicas, que segue as orientações presentes no *Currículo Referência da Rede Estadual de Educação de Goiás* (GOIÁS, 2021) e que visa aprimorar o conhecimento metalinguístico dos alunos, bem como levantar discussões sobre a carga cultural presente no léxico e sobre equidade de gênero.

Chegamos ao final deste trabalho com algumas constatações e reflexões importantes. A primeira é com relação à alta frequência de palavras não plenas (BIDERMAN, 1998) no *corpus* de estudo, em que as palavras mais frequentes da *WordList* são as classificadas como palavras instrumentais, como artigos, preposições, pronomes, conjunções etc. Sabemos que o estudo das classes gramaticais ainda gera dúvidas para muitos estudantes, muitas vezes, devido ao fato de eles não perceberem que o tempo todo estão falando, lendo, ouvindo palavras e que essas palavras pertencem a grupos (classes gramaticais). Neste sentido, uma sequência didática mostrando a proximidade entre o conteúdo de língua portuguesa e a rotina do aluno faz com que ele assimile melhor questões morfológicas, já que este conhecimento é essencial para os estudos morfosintáticos, tão exigidos em avaliações ao longo da vida estudantil. Partindo do entendimento de que *minissaia* é uma roupa curta, e da presença da expressão: “Mulher!

Toma sua linha”, concluímos que esta é uma palavra marcada culturalmente (BARBOSA, 2009), e que reforça a ideia de que a mulher não deve usar roupa curta, por isso, a tentativa do homem de controlar a roupa que a mulher deve usar. Ainda que estejamos falando de uma expressão ocorrida uma única vez no *corpus*, acreditamos que sua presença seja a cristalização de uma cultura em que a mulher ainda precisa se submeter ao controle do homem.

Considerando os pontos analisados sobre a prosódia semântica relacionada ao item *mulher* nas letras de músicas que compuseram nosso *corpus*, consideramos produtivo levar essa análise de letras de música do gênero sertanejo para a sala de aula, tendo em vista abordar aspectos gramaticais e discutir questões sociais e culturais que envolvem os debates sobre gênero, buscando a conscientização de alunos e alunas para que possamos vislumbrar um futuro em que as mulheres tenham seus direitos respeitados.

Referências

BARBOSA, L. M. A. O conceito de lexicultura e suas implicações para o ensino-aprendizagem de português língua estrangeira. **Filologia e Linguística Portuguesa**, São Paulo, n. 10-11, p. 31-41, 2009. DOI: <https://doi.org/10.11606/issn.2176-9419.v0i10-11p31-41>.

BARBOSA, L. M. A. **Opacité et transparence lexico-culturelle dans l'apprentissage du portugais langue étrangère au Brésil**: les paroles de chansons, instruments de médiation linguistique et culturelle. 2005. Tese (Doutorado em Estudos Portugueses, Brasileiros e da África Lusofônica) – Université Paris VIII, Paris, 2005.

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

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

BIDERMAN, M. T. C. Dimensões da palavra. **Filologia e Linguística Portuguesa**, São Paulo, n. 2, p. 81-118, 1998. DOI: <https://doi.org/10.11606/issn.2176-9419.v0i2p81-118>.

BIERNATH, A. O que é a musicoterapia e qual o seu potencial? **VEJA SAÚDE**, São Paulo, 18 out. 2019. Bem-Estar. Disponível em: <https://saude.abril.com.br/bem-estar/o-que-e-a-musicoterapia/>. Acesso em: 25 jan. 2022.

BUENO, F. S. **Minidicionário escolar da língua portuguesa**. São Paulo: DCL, 2018.

ECKERT, P. The meaning of style. *In*: ANNUAL SYMPOSIUM ABOUT LANGUAGE AND SOCIETY, 11., 2003, Austin. **Proceedings** [...]. Austin: University of Texas, 2003. p. 41-53. Disponível em: <http://salsa.ling.utexas.edu/proceedings/2003/eckert.pdf>. Acesso em: 25 jan. 2022.

GOIÁS. **Currículo referência da rede estadual de educação de Goiás**: currículo adaptado. Goiânia: Secretaria de Educação, 2021.

HOUAISS, A. **Dicionário eletrônico Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009.

RIGONATTO, M. Funções sintáticas da palavra “que”. **Brasil Escola**, [S.l.], 2022. Gramática. Disponível em: <https://brasilecola.uol.com.br/gramatica/funcoes-sintaticas-palavra-que.htm>. Acesso em: 9 dez. 2021.

RODEO WEST. Conheça o surgimento e crescimento do estilo *country* no mundo. **Rodeo West**, Patos de Minas, 31 ago. 2017. Curiosidades. Disponível em: <https://blog.rodeowest.com.br/curiosidades-rodeio/conheca-surgimento-crescimento-do-estilo-country-no-mundo/>. Acesso em: 25 jan. 2022.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 25 jan. 2022.

| Apêndice - proposta de sequência didática

Objeto de conhecimento: classes gramaticais e prosódia semântica.

Expectativa de aprendizagem:

- 1) analisar o campo lexical das letras de músicas do estilo *country*;
- 2) identificar os itens lexicais que se associam à palavra *mulher*;
- 3) discutir em sala de aula o campo lexical ocupado pela palavra *mulher* nas letras de músicas e suas conotações positivas, negativas ou neutras.

1º Momento: propor aos alunos uma atividade de pesquisa sobre o estilo musical *country* com a história do surgimento do estilo no Brasil e com a seleção de três letras de música desse estilo. A atividade será desenvolvida em sala de aula ou em casa, observando as possibilidades de acesso à internet dos alunos.

2º Momento: apresentação oral dos resultados da pesquisa e da seleção das letras feita pelos alunos, solicitando que informem o nome, o artista e o motivo da escolha.

3º Momento: identificação dos itens lexicais associados à mulher. Para esta atividade, será solicitado que os alunos encontrem a palavra e destaquem a primeira palavra que antecede *mulher* e a primeira depois de *mulher*. Caso a letra escolhida não tenha a palavra *mulher*, o aluno pode procurar sinônimos e se ainda assim ele não encontrar, ele deverá aguardar o próximo momento da aula.

4º Momento: com o auxílio do professor, o aluno fará a classificação morfológica das palavras encontradas antes e depois de *mulher*. Neste momento, será solicitado ao aluno que anote a percepção dele quanto palavra (positiva, negativa ou neutra) e que anote se algum fator cultural contribui para essa percepção.

5º Momento: com a mediação do professor, realizar um debate sobre o tema: igualdade de gênero.

6º Momento: após o debate solicitando que os alunos se reúnam em grupos formados por meninos e meninas em igual quantidade, se for possível, conversem sobre como a comunidade em que eles vivem lida com questões de igualdade de gênero. Em seguida, elaborem juntos um documento que apresente estratégias para promover a igualdade de gênero na comunidade em que eles vivem.

7º Momento: a sequência didática será finalizada com a divulgação do documento para a comunidade escolar e conseqüentemente para a comunidade em que os alunos vivem. Os próprios alunos farão a escolha da mídia a ser utilizada para essa divulgação, se por meio de cartazes, panfletos ou digital, em redes sociais.

Observações: a sequência didática poderá ser adaptada às condições do momento de sua aplicação devido ao contexto pandêmico, atendendo aos protocolos de segurança contra a Covid-19.

A Língua da Tabatinga: exploração inicial à luz da Linguística de *Corpus*

Roberta Gê-Acaiaba¹

¹ Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/0937562879066273>. E-mail: robertage@ufu.br.

1 Introdução

A sociedade brasileira é formada por um encontro étnico. Um encontro a partir do qual surgiram diferentes formas de representar a realidade em que vivemos. Assim, a identidade de nosso país, que se tece hoje frente aos nossos olhos, é pluriétnica, o que se escancara não apenas na formação do nosso povo, mas também, e muitas vezes de forma mais marcante, nas práticas sociais que realizamos em nossas interações diárias.

Não há como negar que, embebido nas tradições asiáticas, europeias, africanas, árabes e indígenas, nosso país reflete uma heterogeneidade de natureza sócio-histórica que fez florescer algo singular no mundo: uma identidade multifacetada que comporta em sua formação elementos distintos e que os relaciona e organiza para a produção de uma cultura plural e única.

Frente a essas práticas sociais heterogêneas que se formaram e ainda se formam em nossa sociedade, a “Língua da Tabatinga” presentifica-se na cidade de Bom Despacho – MG. As origens dessa língua remontam ao passado escravocrata de nosso país, que levou aos mais diversos cantos da nação a barbárie dos homens brancos contra pessoas negras.

De acordo com Queiroz (1998), a Língua da Tabatinga (doravante LT) tem provável origem sudanesa, já que a maioria dos então escravizados que se estabeleceram com seus senhores em Bom Despacho eram provenientes da Vila do Ouro – Pitangui, antiga sede administrativa de Bom Despacho. Os sudaneses possuíam a tradição de trabalharem com mineração e foram explorados em larga escala durante o ciclo do ouro em Pitangui.

Acrescidas às contribuições linguístico-culturais africanas sudanesas, somaram-se as contribuições da língua portuguesa brasileira local, resultando em uma língua verdadeiramente “mestiça”, a LT. Durante o estudo feito por Queiroz (1998), há 23 anos, foi precipitadamente previsto o desaparecimento desta língua. Contudo, em 2021, é possível notar que a LT não desapareceu. Ela se encontra cada vez mais viva e presente nas práticas sociais de linguagem que se realizam na comunidade bom-despachense.

A importância da investigação apresentada neste capítulo reside na noção elementar, conforme Labov (2008), de que não há comunidades linguisticamente homogêneas, e que a heterogeneidade na formação das línguas é algo natural. Logo, o que se observa hoje em Bom Despacho faz parte da dinâmica das línguas, organismos vivos que se moldam ao mesmo tempo em que moldam a sociedade na qual se desenvolvem. Cabe, portanto, evidenciar que a chave do estudo proposto está na possibilidade de compreender cada vez mais sobre o processo de formação das línguas, avaliando as transformações no momento exato em que elas ocorrem, vivenciando concreta e diacronicamente o fenômeno da constituição de uma nova variedade linguística.

Como membro constituinte da sociedade bom-despachense, a partir de um mínimo conhecimento prévio da LT e tendo o português brasileiro como língua materna, a hipótese que surge e que impulsiona o presente estudo é a de que a LT concede à língua portuguesa brasileira da cidade de Bom Despacho um alargamento lexical a partir da inserção de palavras próprias e específicas do léxico “Tabatinga”. Dessa forma, por influência histórica e social, a variedade linguística do português brasileiro existente na cidade de Bom Despacho agrega elementos típicos da variedade linguística “Tabatinga”, o que a torna única em nosso país.

Além disso, há também a hipótese de que a inserção da LT nas práticas de linguagem da sociedade bom-despachense funciona como uma política de autoafirmação e resistência, garantindo a redução do preconceito linguístico e, também, a redução do preconceito social com a comunidade afrodescendente, a partir da qual a LT se originou. Isso demarcaria os motivos pelos quais, apesar de previsto no estudo de Queiroz (1998), tal língua não desapareceu, mas continua cada vez mais presente em Bom Despacho. Assim, a LT despertou nosso interesse, não pelo seu caráter “diferente ou alegórico”, mas pelo seu valor linguístico-cultural na formação da sociedade bom-despachense.

Diante do exposto, este capítulo apresenta uma exploração inicial da LT à luz da Linguística de *Corpus* e amparada pela Sociolinguística Variacionista. Ademais, este estudo é de caráter *corpus-driven*, isto é, guiado por *corpus*. Reconhecemos que analisar e descrever a LT é pensar e produzir saberes sobre a formação linguística de uma variedade dialetal que representa o caráter multifacetado da sociedade brasileira e de sua língua.

2 Fundamentação teórica

A análise e descrição propostas têm como fundamentação teórico-metodológica a Linguística de *Corpus* e a Sociolinguística Variacionista. Tendo como objeto de estudo a LT, a Sociolinguística Variacionista concedeu bases para compreensão da relação “língua x sociedade”. Por sua vez, a Linguística de *Corpus* (doravante LC) nos auxiliou, a partir de uma visão probabilística da linguagem, na observação e análise extensiva do objeto de estudo, amparadas por dados empíricos da língua em uso.

2.1 Linguística de *Corpus*

Berber Sardinha (2004) nos apresenta a LC como uma abordagem empirista da linguagem, que tem como ponto principal encará-la como um sistema probabilístico, tomando por base Halliday (1991). Partindo desse pressuposto, ela se contrapõe aos modelos racionalistas da linguagem, principalmente à linguística chomskiniana, visto que, na LC, o foco do estudo está no desempenho (uso) e não na competência. Dessa forma, a LC geralmente promove nos

estudos da linguagem um deslocamento, que deixa a teorização para um segundo momento e, como ponto de partida para a investigação linguística, utiliza-se de dados obtidos a partir da observação empírica da linguagem autêntica.

Concebemos a LC não apenas como um método, mas também como uma corrente teórica de abordagem da língua. A esse respeito, Berber Sardinha (2004, p. 37) afirma que a LC “é uma perspectiva, uma maneira de chegar à linguagem”, o que vem a corroborar nossa visão ampla a respeito dela.

Para a análise da linguagem, a LC faz uso de ferramentas computacionais que auxiliam os linguistas a analisar e descrever o funcionamento linguístico. Citando Berber Sardinha (2004, p. 3), “a Linguística de *Corpus* ocupa-se da coleta e da exploração de *corpora* [...]. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador”. Para além da análise estatística, a LC permite um tratamento qualitativo dos dados, visto que traz consigo a possibilidade de aliar ao exame do *corpus* a introspecção do pesquisador, conforme evidenciado por Novodvorski (2020).

Isto posto, na análise da LT, a LC colabora efetivamente para a identificação de traços característicos e inerentes a essa variedade linguística, padrões de comportamento linguísticos e variações que só poderiam ser observados através das lentes da LC e que funcionam como evidências autênticas e factuais do funcionamento linguístico do objeto de estudo analisado. Colabora, ainda, para as análises sociais e culturais advindas do estudo da LT.

2.2 Concepções de língua/linguagem

A linguagem humana, segundo Travaglia (2002), apresenta três possibilidades distintas de ser concebida. A primeira afirma ser a linguagem a expressão do pensamento. Segundo essa concepção, a expressão da linguagem se constrói internamente na mente humana, sendo sua exteriorização apenas a tradução do pensamento. É nesta concepção que se encontra a errônea suposição de que as pessoas não se expressam bem porque não pensam, compreendendo, assim, a linguagem como um ato individual independente das interações sociais e do contexto de produção. As leis que regem as criações linguísticas seriam, então, estritamente leis da psicologia individual.

Ainda de acordo com Travaglia (2002), nessa primeira concepção, é presumido que haja regras para serem seguidas a fim de organizar o pensamento e, por conseguinte, a própria linguagem. Entre essas regras estariam as normas gramaticais, que se ligam intrinsecamente ao estudo da Gramática Tradicional. Sendo assim, as interações entre locutor e interlocutor não influenciariam a formação da linguagem, já que ela é vista como algo isolado das interações sociais, através da qual o homem representa, “para si mesmo”, o mundo.

A segunda concepção de linguagem concebe-a como um instrumento de comunicação, sendo sua principal função a transmissão de informações, já que a língua é considerada um código utilizado a fim de transmitir mensagens de um emissor a um receptor. Dessa forma, só há comunicação efetiva se emissor e receptor dominarem o código da linguagem. Nessa visão, o que ocorre entre os indivíduos durante a utilização de uma língua é apenas um processo de decodificação.

A terceira concepção, finalmente, é aquela que encara a linguagem como forma ou processo de interação. Nesse prisma, o locutor age sobre o interlocutor, exigindo dele reações. Além disso, ambos interagem sobre o meio. Logo, o papel concedido à linguagem, nessa concepção, não é só exteriorizar pensamentos ou transmitir informações, mas sim interagir.

A linguagem é, pois, um lugar de interação humana (BAKHTIN, 1999), no qual os usuários da língua criam um domínio cooperativo de interações, ocupando lugares sociais preestabelecidos pela sociedade. Desta forma, Bakhtin (1999, p. 123) afirma que

[...] a verdadeira substância da língua não é constituída por um sistema abstrato de formas linguísticas, nem pela enunciação monológica isolada, nem pelo ato psicofisiológico de sua produção, mas pelo fenômeno social da interação verbal realizada através da enunciação ou das enunciações. A interação verbal constitui assim a realidade fundamental da língua.

Assim, é perceptível que o ponto chave dessa concepção é, sem dúvida, o caráter dialógico da linguagem.

Ainda neste sentido, é preciso considerar as contribuições da Linguística Sistêmico-Funcional (doravante LSF), baseadas nas concepções de Halliday (1978). Por caracterizar a linguagem a partir de uma perspectiva social, a LSF rompe com os paradigmas do formalismo linguístico e estuda a linguagem como uma atividade social, propondo um enfoque centrado nos usos da língua.

Tal enfoque vai ao encontro dos pressupostos da LC, que, ao abordar a língua a partir de uma perspectiva empirista, centrada no uso, considera-a como um sistema probabilístico. Nesse sentido, Berber Sardinha (2004, p. 34-35) afirma que a linguística hallidayana possui uma “visão de linguagem que se encaixa perfeitamente nos preceitos da Linguística de *Corpus* e serve como arcabouço teórico maior no qual ela pode se incluir”. Quando se diz que a linguagem é vista como sistema probabilístico (BERBER SARDINHA, 2004), significa que a língua é vista mais como uma questão de probabilidade do que de possibilidade, ou seja, embora os traços linguísticos sejam teoricamente possíveis, eles não ocorrem com a mesma frequência. E o fato de essas diferenças de frequência não serem aleatórias é o mais importante. Para essa abordagem da língua(gem), há uma correlação entre as características linguísticas e contextuais.

Logo, o que se pode observar é que, ao adotarmos uma concepção linguística empirista, visto que fazemos uso da LC, aproximamo-nos de Halliday, estabelecendo uma relação entre características linguísticas e situacionais, e distanciamos-nos da linguística chomskyana. Segundo Berber Sardinha (2004, p. 30), a linguística chomskyana centra seus estudos na competência, nos universais linguísticos e em uma visão racionalista, que “ênfatiza a determinação de quais agrupamentos sintáticos são possíveis (permissíveis) dado o conhecimento que um falante nativo possui de sua língua”, evidenciando a língua como um sistema autônomo, mental, resultado de uma herança linguístico-genética. Enquanto isso, a LSF ressalta o caráter social da língua.

As pesquisas desenvolvidas através da LSF centram-se na indissociabilidade entre o significado e a forma linguística, entre o léxico e a gramática, entre o contexto e a frase, apoiando-se sempre em dados linguísticos reais, o que se mostra adequado ao propósito deste projeto de pesquisa.

2.3 Diversidade linguística e sociedade

O português brasileiro² descende da língua portuguesa europeia, mas aqui no Brasil tomou sua forma na complexa interação com as diferentes línguas com as quais manteve contato.

A língua autóctone era aqui representada por um grande número de línguas indígenas brasileiras, às quais se somaram numerosas línguas africanas que abarcaram no Brasil no período da escravidão, e também as línguas que emigraram para o Brasil da Europa e da Ásia.

O português brasileiro herdou das línguas indígenas palavras ligadas à fauna e à flora (abacaxi, mandioca, tatu, piranha), bem como nomes próprios e geográficos. Já a influência africana verificou-se em termos ligados à religião, à comida e à cultura africana (candomblé, iansã, feijoada, caçula, moleque, samba). Coube à cultura europeia o papel de introduzir no Brasil seus traços nas variedades linguísticas nordestinas (fato ocasionado pela invasão holandesa em Pernambuco no ano de 1630), além de possuir grande importância na constituição do português brasileiro, já que a língua oficial do Brasil descende, diretamente, da língua portuguesa de Portugal. Resta ainda ressaltar o considerável número de palavras asiáticas introduzidas no círculo linguístico através da tradição oriental nas lutas marciais, que possui grande número de adeptos no Brasil.

Diante dessa realidade, e em consequência dela, há, no português brasileiro, um multilinguismo que deu origem a diversas variantes dialetais com estruturas profundamente modificadas. Essas modificações vão desde o campo semântico até o fonético e morfossintático.

² De acordo com Bagno (2007), quando dizemos que no Brasil se fala português, usamos esse nome simplesmente por comodidade e por uma razão histórica, justamente a de termos sido colônia de Portugal. Do ponto de vista linguístico, porém, a língua falada no Brasil já tem uma gramática. Por isso, os linguistas preferem usar o termo português brasileiro para ser mais claro e marcar bem essa diferença.

Ao reconhecer que a língua portuguesa brasileira apresenta muita diversidade e variabilidade, ressalta-se o pressuposto de que ela não apresenta uma única face. Ela melhor se identifica com um polígono multifacetado composto por uma gama incontável de contribuições linguísticas de povos indígenas, africanos, europeus, asiáticos e diversos imigrantes que em nosso país se instalaram e ainda se instalam, tendo em face o acolhimento de refugiados de diversas partes do mundo.

Além disso, fatores como faixa etária, classe social e localização geográfica ou regional também contribuem para formação de uma língua que apresenta características inerentes e adequadas às necessidades dos usuários. Logo, todas as variedades linguísticas apresentam influências da sociedade na qual elas se constituíram. É nesse sentido que Labov (2008) afirma que o estudo de uma língua é fundamentalmente o estudo da cultura da qual ela é forma e produto. Ou seja, língua e sociedade estão ligadas entre si de um modo inquestionável. Além disso, o autor defende que não há comunidades linguisticamente homogêneas e que a heterogeneidade na formação das línguas é algo natural (LABOV, 2008).

Neste estudo, a natureza variável da língua é um pressuposto fundamental, que o orienta e sustenta, tendo em vista que homogeneidade linguística é, segundo Bagno (2007, p. 17), “um mito que pode ter consequências graves na vida social”. Ademais, é importante lembrar que a língua portuguesa brasileira é marcada por uma heterogeneidade original e, por isso, a não aceitação da diferença é responsável por inúmeros e nefastos preconceitos sociais. Nesse aspecto, o preconceito linguístico tem efeito particularmente negativo.

Dentre as variantes que surgiram no Brasil, ocorreu o aparecimento da LT, na cidade de Bom Despacho, no centro-oeste de Minas Gerais. Na seção seguinte, discorreremos brevemente sobre esta variante linguística.

2.4 O que é a Língua da Tabatinga?

A argila mole, untuosa e branca, foi o que deu origem ao nome do bairro da cidade de Bom Despacho (MG), “Tabatinga”. Contudo, sua cultura foi constituída pelos braços negros e fortes dos antigos escravizados.

Localizado na periferia da cidade de Bom Despacho, Tabatinga é um bairro onde se abrigaram os escravizados libertos, já que a maioria deles veio para Bom Despacho na condição de cativos, juntamente com seus senhores. Assim, como ocorreu em todo o Brasil, após conquistarem a tão esperada liberdade, os ex-escravizados viram-se, apesar de livres, desabrigados.

Dessa forma, muitos indivíduos resolveram fixar residência neste local, que posteriormente viria a ser o bairro Tabatinga. De acordo com relatos concedidos a mim, no ano de 2005, por uma das falantes da língua, descendente de um dos primeiros moradores do bairro, que foi

escravo da lavoura, os escravizados libertos se instalaram no local a fim de trabalhar em um garimpo que havia em frente à sua atual casa. Descobrimo que ali não havia nenhum metal precioso, eles cavaram a terra branca do garimpo (a tabatinga) e barrearam as taperas que ergueram como moradia, passando, em seguida, a tecer cobertas e cortar capim para vender à confecção de colchões.

Hoje ainda residem em Tabatinga muitos descendentes dos negros africanos que ali se instalaram no passado, deixando não somente suas lembranças, mas também sua cultura e sua presença através de diversas manifestações afro-brasileiras que ainda sobrevivem no bairro. Entre as manifestações culturais mais expressivas, podem ser citadas as religiões de matriz africana, o congado, que ainda move muitas pessoas em toda cidade a dançar por Nossa Senhora do Rosário, e, principalmente, a língua, que é formada por um léxico expressivo de uma língua africana e que se constitui como nosso objeto de estudo.

Conforme apontado na seção introdutória, a LT é uma língua afrodescendente e tem provável origem sudanesa (QUEIROZ, 1998). Queiroz (1998, p. 100) afirma que a LT apresenta, em aspecto formal, uma série de características comuns aos *pidgins*, como “vocabulário reduzido, polissemia generalizada, frequente recurso à perífrase, redução de flexões e extrema variação fonética”. Entretanto, do ponto de vista funcional, segundo a autora, ela se afasta dos *pidgins*, pois, de acordo com Castilho ([S.d.], [S.p.]), o “*pidgin* é uma língua de emergência bastante rudimentar, desenvolvida por pessoas interessadas em trocas comerciais”, característica não evidenciada na LT. Assim, Queiroz (1998, p. 100) admite a hipótese de que a LT “venha a ser o resultado da evolução de um *pidgin* ou de um crioulo³ que teria existido efetivamente no período da escravidão”.

Contudo, a autora também observa que a LT muito se aproxima das línguas especiais pela função original que era a de ocultar o que os negros conversavam, impedindo que os senhores de escravos compreendessem a conversa e os castigassem. Queiroz (1998, p. 106) ainda afirma que a LT “partilha com a gíria o fato de funcionar como sinal diacrítico que marca o grupo de negros da Tabatinga”, e, por isso mesmo, segundo a autora, era alvo de preconceito linguístico.

Atualmente, a LT se apresenta como uma língua de tradição e resistência, caracterizando a identidade social de uma comunidade. Ela é apreendida em situações de comunicação informal entre amigos e familiares. Assim, cada vez mais se dissemina entre as práticas de linguagem de toda comunidade bom-despachense. O léxico da LT já é utilizado por logomarcas empresariais de prestígio na cidade e região, além de canções. É possível observar a incorporação de palavras e expressões nas práticas de linguagem de toda a comunidade bom-despachense, deixando de ser segregada ao bairro de origem (QUEIROZ, 1998).

³ Crioulo constitui-se a partir de adaptações de uma língua europeia por falantes de outras línguas, em geral africanas e asiáticas, com as quais os europeus entraram em contato por interesse mercantil (CASTILHO, [S.d.], [S.p.]).

Logo, nota-se que a LT, assim como qualquer outra língua, é um meio de interação, através do qual os indivíduos agem uns sobre os outros e estes com o meio, criando e modificando o contexto e sendo por ele influenciados. Nesse sentido, qualquer tentativa de buscar aprender apenas o invariável, o que muito se observa nos estudos tradicionalistas da língua, significa uma redução na compreensão do fenômeno linguístico de formação de uma língua.

3 Métodos e procedimentos

O *corpus* de estudo totaliza 5.024 *tokens* (itens) e 1.150 *types* (formas), com uma densidade vocabular (*type/token ratio*) de 24%. Ele é oriundo de 10 entrevistas realizadas durante o ano de 2004. As entrevistas estavam gravadas em fitas magnéticas de áudio, recurso disponível na época em que os dados foram coletados, e foram transcritas ortograficamente. Portanto, este *corpus* de estudo constitui-se como um *corpus* oral sincrônico referente à variedade dialetal da LT.

Os entrevistados eram 14 falantes nativos adultos que possuíam, no momento de coleta, entre 18 e 62 anos, sendo 11 deles do sexo masculino e 03 do sexo feminino. Ou seja, foram realizadas 10 entrevistas, nas quais houve 14 participantes, sendo que as entrevistas três, oito e nove tiveram, respectivamente, dois, dois e três participantes simultâneos.

As entrevistas seguiram um roteiro previamente estabelecido, que foi montado a fim de obter o vernáculo dos entrevistados; sendo solicitado que falassem de fatos da infância, situações familiares marcantes, sobre como é a vida na cidade, situações engraçadas ou de risco. Logo, a entrevista sociolinguística direcionada pretendeu envolver afetivamente, durante a entrevista, os participantes, para que eles, ao falarem de suas experiências, esquecessem de monitorar a fala. Houve, portanto, uma redução dos efeitos do paradoxo do observador, já que as perguntas funcionaram como um gatilho e direcionaram a produção de sequências textuais que resultaram em um todo heterogêneo.

A duração das entrevistas é divergente, já que ao ser utilizado o método da entrevista sociolinguística, houve liberdade na expressão dos falantes. Cada entrevista possui uma duração específica e, juntas, totalizam 1h45'57". A seguir, apresentamos a duração de cada uma delas, apresentado-as como T1, T2, T3 e assim por diante.

Quadro 1: Duração das entrevistas sociolinguísticas.

Entrevista	Duração
T1	12'47"
T2	10'32"
T3	11'05"
T4	9'21"
T5	9'49"
T6	8'22"
T7	10'51"
T8	16'56"
T9	8'29"
T10	7'45"
TOTAL	1h 45'57"

Fonte: Elaborado pela autora a partir dos dados da pesquisa.

Conforme mencionado, como houve entrevistas com mais de um participante simultâneo, para a identificação dos turnos de fala, utilizamos a nomenclatura F + nº para garantir o sigilo quanto aos nomes dos participantes da pesquisa. Dessa forma, cada um dos entrevistados foi nomeado, por exemplo, como F1, F2, F3 e assim por diante.

Após transcrição de todas as entrevistas, foi necessário converter cada um dos textos obtidos para o formato .txt, com a codificação ANSI, a fim de preparar o *corpus* para processamento pelo programa *WordSmith Tools 4.0* (SCOTT, 2004). Por meio das ferramentas deste programa, geramos uma lista de palavras e, posteriormente, exploramos as linhas de concordância dos dois itens lexicais mais frequentes no *corpus* de estudo: *cuete* e *ocaia*.

4 Resultados

A lista de palavras foi gerada por meio da ferramenta *WordList*. A partir da análise dessa lista, organizada por ordem decrescente de ocorrência, foi possível identificar as palavras lexicais mais frequentes no *corpus* de estudo. Focalizaremos nossas análises nas duas mais frequentes: *cuete* (117 ocorrências) e *ocaia* (58 ocorrências). Para ilustrar essa primeira aproximação ao *corpus*, apresentamos na Figura 1 um recorte da lista de palavras, em que é possível identificar os itens *cuete* e *ocaia*.

Figura 1: Recorte da lista de palavras.

N	Word	req.	%	exts	%	m
1	E	146	2,91	10	0,00	
2	O	131	2,61	10	0,00	
3	É	127	2,53	10	0,00	
4	CUETE	117	2,33	8	0,00	
5	DE	90	1,79	9	0,00	
6	A	77	1,53	10	0,00	
7	QUE	71	1,41	9	0,00	
8	NÃO	62	1,23	8	0,00	
9	DO	59	1,17	10	0,00	
10	NO	58	1,15	9	0,00	
11	OCAIA	58	1,15	6	0,00	
12	PRA	58	1,15	8	0,00	
13	NUM	54	1,07	8	0,00	
14	AVURA	49	0,98	5	0,00	
15	AÍ	47	0,94	5	0,00	

Fonte: *WordList*.

Após identificar as duas palavras lexicais mais frequentes, utilizamos a ferramenta *Concord* para gerar listas de concordância para esses itens lexicais. Analisando as linhas de concordância de *cuete* e *ocaia*, foi possível notar que uma parcela considerável das palavras lexicais utilizadas durante as entrevistas pertencem à LT, ao passo que as palavras gramaticais (tais como conjunções, artigos, preposições) são pertencentes exclusivamente à variedade da língua portuguesa brasileira falada na cidade de Bom Despacho. À guisa de ilustração, apresentamos a seguir um excerto extraído do *corpus* de estudo. Marcamos em negrito as palavras identificadas como pertencentes à LT.

- **O *cuete* *caxô* uns *cureio* que a *ocaia* dexô no *conjolo*⁴.**

Não poderíamos deixar de mencionar, assim como Queiroz (1998), que a LT possui uma estrutura sintática extraída da língua portuguesa, baseada no padrão SVO (sujeito, verbo, objeto).

- *Aí o cuete caxô um tiquim de ingura na gibera*⁵ (...)

- ✓ *tO cuete*= sujeito
- ✓ *Caxô*= verbo
- ✓ *um tiquin de ingura*= objeto

Realizando uma leitura vertical, permitida pelas duas listas de concordância criadas, observamos como a utilização de substantivos próprios ocorre com baixa frequência na

⁴ O homem comeu a comida que a mulher deixou em casa.

⁵ Aí o homem pegou um pouquinho de dinheiro no bolso.

variedade da LT. Foram identificados apenas dois substantivos próprios, sendo eles “Bené-Pião” e “Garça”. O primeiro deles nomeia um indivíduo e o segundo nomeia uma rua da cidade de Bom Despacho.

Partindo desta constatação, notamos que em nenhum momento durante as entrevistas realizadas o sujeito falante fez uso de nomes próprios para se referir ao outro (3ª pessoa do discurso), mesmo quando esse era personagem de sua narrativa. Em substituição aos nomes próprios, foram utilizados os substantivos simples da variedade da LT, *cuete* e *ocaia*, com comum recorrência à perífrase, que serve, no contexto analisado, para nominalizar e caracterizar o ser a quem o locutor se refere, facilitando a identificação do referente. A seguir, apresentamos mais alguns excertos extraídos das linhas de concordância:

- F1: *ocai/ocaia curimba cum os cuete de covera*⁶.
 - ✓ *ocaia*= 3ª pessoa do discurso, interlocutor, pessoa sobre quem o falante F1 conversa.
 - ✓ *cuete de covera*= 3ª pessoa do discurso, o ser de quem se fala. Perífrase utilizada para se referir a médico.
- F1: *a maioria era cuete cafuvira*⁷.
 - ✓ *cuete cafuvira*= 3ª pessoa do discurso, o ser de quem se fala. Perífrase utilizada para se referir a negro.
- *tipura a ocaia do conjolo ao lado*⁸.
 - ✓ *ocaia do conjolo ao lado* = 3ª pessoa do discurso, o ser de quem se fala. Perífrase utilizada para se referir à vizinha.
- F3: *cuete é práctico, a hora que aperta pro ladu dele ele injira e prontu né?*⁹
 - ✓ *cuete*= 3ª pessoa do discurso, o ser de quem se fala.

Ao longo das análises, percebemos que seria relevante realizar uma limpeza na lista de palavras inicial, gerando, portanto, uma segunda lista que possibilitaria identificar a quantidade exata de palavras da LT presentes no *corpus* de estudo. Esse procedimento nos forneceria um panorama probabilístico do percentual de palavras da LT que os membros da comunidade analisada utilizaram durante as entrevistas. Portanto, a limpeza consistiu na exclusão de todas as palavras pertencentes ao português brasileiro local da lista inicial, resultando em uma lista de palavras formada apenas pelo léxico da LT.

6 Mulher trabalha com os médicos.

7 A maioria era negro.

8 Observa a vizinha.

9 Homem é práctico, na hora que aperta para o lado dele ele sai/corre e pronto, né?

A lista obtida após procedimento mencionado mostrou que 214 palavras utilizadas durante as entrevistas pertencem à variedade da LT. A partir do confronto entre o número de *types* da lista completa (1.150) e o número de *types* dessa segunda lista (214), observamos que o léxico da Tabatinga possui uma ocorrência, dentro da realidade linguística observada, de 18,6% do total de *types*. Na Figura 2, apresentamos um recorte dessa segunda lista de palavras.

Figura 2: Lista de palavras do léxico da LT.

N	Word	Freq.	%	Texts	%	r
1	CUETE	117	2,33	8	0,00	
2	OCAIA	58	1,15	6	0,00	
3	AVURA	49	0,98	5	0,00	
4	INGURA	42	0,84	5	0,00	
5	CAXÁ	28	0,56	6	0,00	
6	CONJOLO	22	0,44	6	0,00	
7	CAXA	19	0,38	6	0,00	
8	MATUABA	18	0,36	7	0,00	
9	TUÉ	16	0,32	5	0,00	
10	CATITA	15	0,30	4	0,00	
11	VIRIANGO	14	0,28	4	0,00	
12	OCORA	13	0,26	4	0,00	
13	TIPURA	11	0,22	5	0,00	
14	CAFUVIRA	10	0,20	5	0,00	
15	CAJUVIRA	10	0,20	4	0,00	

Fonte: *WordList*.

Tal constatação apresenta a relevância das palavras pertencentes à LT no discurso dos indivíduos da comunidade linguística analisada, visto que aproximadamente 1/5 das palavras por eles selecionadas para a realização de suas comunicações orais é decorrente do léxico da Tabatinga. Ademais, esses dados indicam a eventual necessidade de avaliação da relevância da LT para a variedade linguística do português brasileiro falado na cidade de Bom Despacho.

É relevante ainda mencionar que, após uma análise detalhada dessa lista de palavras, observamos que, das 214 palavras, 154 pertencem à classe gramatical dos substantivos. Essa observação nos levou a outra constatação proveniente da observação e análise extensiva dessa lista: todos os substantivos da LT nomeiam entidades concretas do mundo real, ou seja, até o momento não foram identificados substantivos abstratos.

Ao longo da análise de linhas de concordância geradas para outros itens lexicais da LT, outros pontos também nos atraíram a atenção, como a polissemia que acompanha os verbos. Com intuito de exemplificamos tal fenômeno linguístico, escolhemos trabalhar com o verbo *tipurar*

e algumas de suas acepções em seus contextos de ocorrência. Cabe lembrar que, para isso, foi preciso lematizar¹⁰ o verbo *tipurar*, para que nenhuma de suas flexões fosse desconsiderada.

Figura 3: Lematização do verbo *tipurar*.

The screenshot shows a software window titled 'Só palavras da Tabatinga.lst' with a menu bar (File, Edit, View, Compute, Settings, Windows, Help). Below the menu is a table with columns: N, Word, Freq., %, Texts, %, and mas. The table lists 15 words. A 'Lemma forms' window is open over the table, showing a list of verb forms for 'tipurar' with their respective frequencies.

N	Word	Freq.	%	Texts	%	mas
1	CUETE	117	2,33	8	0,00	
2	OCAIA	58	1,15	6	0,00	
3	AVURA	49	0,98	5	0,00	
4	INGURA	42	0,84			
5	CAXÁ	28	0,56			
6	CONJOLO	22	0,44			
7	CAXA	19	0,38			
8	MATUABA	18	0,36			
9	TUÉ	16	0,32			
10	CATITA	15	0,30			
11	VIRIANGO	14	0,28			
12	OCORA	13	0,26			
13	TIPURA	50	0,22			
14	CAFUVIRA	10	0,20			
15	CAJUVIRA	10	0,20	4	0,00	

The 'Lemma forms' window shows the following list:

- TIPURA 11
- TIPURÁ 9
- TIPURÀ 1
- TIPURANO 8
- TIPURAR 2
- TIPURAVA 7
- TIPURO 2
- TIPURÔ 10

Fonte: WordList.

A seguir, apresentamos alguns exemplos extraídos das linhas de concordância obtidas a partir do verbo *tipurar*, bem como seus significados na língua portuguesa brasileira.

- A *ocaia* *falô* que só **tipurava** *cuxipa* na *Marcela*¹¹.

Ao formar um fraseologismo ou agrupamento lexical (verbo + substantivo) com *cuxipa* (órgão sexual masculino e/ou feminino), o verbo *tipurava* assume o valor semântico de "fazer", constituindo, no contexto, a expressão "fazer sexo", "transar".

- O *cuete* **tipurava** *tudo* quanto é *ocaia*¹².

No exemplo acima, o verbo *tipurar* possui o valor semântico de "flertar", "dar em cima", "passar uma cantada".

- O *cuete* **tipurô** o *marcanjo* *avura*¹³.

Já neste exemplo em questão, o verbo *tipurar*, tendo como complemento o substantivo *marcanjo* (cigarro), constrói o sentido de "fumar".

¹⁰ "Um lema ('lemma') é a forma base de uma palavra; é aquela que encabeça um verbete, normalmente a forma morfológica mais simples (singular, infinitivo, etc.). O lema é o conceito organizador do dicionário" (BERBER SARDINHA, 2009, p. 498).

¹¹ A mulher falou que só fazia sexo/transava na Marcela (em referência à dona da casa de prostituição).

¹² O homem flertava tudo quanto é mulher.

¹³ O homem fumou muito.

- *Aí o cuete **tipurô, tipurô, tipurô...***¹⁴

Nesta ocorrência, o verbo *tipurar* aparece em sua acepção mais comumente utilizada na LT, relacionando-se com o ato de “pensar”, “raciocinar”.

- *Os cuetim ia **tipurano** uns pros outro*¹⁵.

Neste caso, o verbo se refere à ação de “contar”, “falar”, “espalhar uma informação”.

5 Considerações finais

A iniciativa apresentada, por meio desta breve análise e descrição da LT, nos concedeu bases para que pudéssemos identificar algumas características inerentes a essa variedade dialetal, evidenciando que uma pesquisa mais abrangente se faz necessária, com vistas ao reconhecimento e valorização linguística e cultural dessa variedade linguística que se instalou em solo nacional em decorrência do encontro entre a língua portuguesa e uma língua de suposta matriz africana.

Além disso, compilar um *corpus* da LT, que se apresenta predominantemente na modalidade oral, colabora para a preservação dessa variedade dialetal e também fornece bases para o estreitamento da relação entre língua, sociedade e cultura.

Referências

BAGNO, M. **Preconceito linguístico: o que é, como se faz**. 49. ed. São Paulo: Loyola, 2007.

BAKHTIN, M. **Marxismo e filosofia da linguagem**. Tradução M. Lahud e Y. F. Vieira. São Paulo: Hucitec, 1999.

BERBER SARDINHA, T. Lexicology and corpus linguistics: an introduction. **D.E.L.T.A.**, São Paulo, v. 25, n. 2, p. 497-510, 2009. DOI: <https://doi.org/10.1590/S0102-44502009000200010>.

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

CASTILHO, A. T. **A hora e a vez do português brasileiro**. São Paulo: Museu da Língua Portuguesa, [S. d.]. Disponível em: <https://www.museudalinguaportuguesa.org.br/wp-content/uploads/2017/09/A-hora-e-a-vez-do-portugues-brasileiro.pdf>. Acesso em: 5 maio 2021.

HALLIDAY, M. A. K. **Language as social semiotic: the social interpretation of language and meaning**. London: Edward Arnold, 1978.

¹⁴ *Aí o homem pensou, pensou, pensou...*

¹⁵ *Os meninos iam falando uns para os outros.*

HALLIDAY, M. A. K. Corpus studies and probabilistic grammar. *In*: AIJMET, K.; ALTENBERG, B. (ed.). **English corpus linguistics**: studies in honour of Jan Svartvik. London: Longman, 1991.

LABOV, W. **Padrões Sociolinguísticos**. Tradução Marcos Bagno; Maria Marta Pereira Scherre e Carolina Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008.

NOVODVORSKI, A. **Notas feitas em aula da disciplina Estudos Descritivos e Linguística de Corpus, do Programa de Pós-graduação em Estudos Linguísticos da Universidade Federal de Uberlândia**. 2020.

QUEIROZ, S. **Pé preto no barro branco**: a língua dos negros da Tabatinga. Belo Horizonte: Ed. UFMG, 1998.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 15 jan. 2022.

TRAVAGLIA, L. C. **Gramática e interação**: uma proposta para o ensino de gramática no 1º e 2º graus. 8. ed. São Paulo: Cortez, 2002.

A identificação de metáforas em *corpus* jornalístico comparável bilíngue de opinião e política

Wagner da Cunha Nunes¹

¹ Mestrando em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/5502387144320996>.
E-mail: wagner.nunes@ufu.br.

1 Introdução

As metáforas estão presentes não só na linguagem cotidiana, mas também nas linguagens científica, filosófica e, principalmente, nos discursos jornalísticos. Estes, em seus relatos, auxiliam a compreensão de mundo a partir de utilizações metafóricas. Deste modo, é por meio das metáforas que nós conceptualizamos o mundo e compreendemos a maioria dos conceitos abstratos.

Na Grécia Antiga, Aristóteles (1973, p. 462) definiu metáfora como “a transferência do nome de uma coisa para outra, ou do gênero para a espécie, ou da espécie para o gênero, ou de uma espécie para outra, ou por analogia”. A definição clássica, com bases aristotélicas, é a de que uma ou mais palavras são usadas fora de seu significado convencional para expressar um outro significado. Outro ponto clássico é que as metáforas se restringem apenas à linguagem poética, não sendo encontradas na linguagem convencional. Lakoff e Johnson (1980), Lakoff (1993) e Sperber e Wilson (2008) vão de encontro às ideias clássicas, afirmando que as metáforas não são somente utilizadas na linguagem poética, mas fazem parte da comunicação cotidiana.

Conforme Berber Sardinha (2009), muita gente acredita que a metáfora é uma figura de linguagem e, como tal, é um acessório linguístico para embelezar a fala e a escrita. Como o exemplo clássico de “Julieta é o sol”, de Shakespeare, em que alguém ser o sol aponta para uma metáfora de vida, luz, jovialidade. Berber Sardinha (2009, p. 39-41) esclarece ainda que:

Atualmente, há outras visões que expandem o escopo de metáfora e a redefinem no conjunto das demais figuras de linguagem. Uma dessas visões é a cognitiva, em que a metáfora aparece como um recurso natural e essencial do ser humano para entender o mundo. Ela não mais significa apenas um recurso linguístico para ornamentar o discurso literário, mas fundamentalmente um tipo de processamento mental que nos permite entender conceitos abstratos, como amor, tempo, vida, entre outros. [...] A metáfora conceptual, conforme dissemos, exprime um conceito novo que surge a partir da aproximação de dois conceitos díspares. A metáfora linguística, por sua vez, são as palavras realmente empregadas no texto e na fala de modo metafórico.

Lakoff e Johnson (2002 [1980], p. 45) sustentam que “a metáfora está infiltrada na vida cotidiana, não somente na linguagem, mas também no pensamento e na ação”. Em outras palavras, mesmo que não tenhamos consciência de seu uso, elas fazem parte do nosso dia a dia e da maneira com que pensamos e falamos sobre o mundo ao nosso redor. No universo do esporte, o futebol, por exemplo, é conceptualizado metaforicamente a partir de conceitos da vida cotidiana, tais como, guerra, festa, negócios, religião etc. Isto posto, este estudo tem como objetivo analisar, com base nos pressupostos metodológicos da Linguística de *Corpus* (doravante LC), duas expressões metafóricas extraídas de nosso *corpus* de estudo.

O *corpus* desta pesquisa é constituído por textos de opinião e política disponibilizados nos *sites* de quatro importantes jornais digitais, dois brasileiros (*Estadão* e *Folha de São Paulo*) e dois argentinos (*Clarín* e *Perfil*). A escolha por jornais desses países decorreu, primeiro, por nossa formação na área de Letras em língua espanhola no Brasil, pela proximidade e relações entre Brasil e Argentina, enquanto países vizinhos com línguas irmãs, pelo interesse em conhecer de que maneira e por meio de quais recursos linguísticos são metaforizados aspectos da política nesses países, para além da representatividade desses países no continente sul-americano. Por outro lado, a motivação pela temática da pesquisa está sustentada no trabalho de Novodvorski e Bevilacqua (2021), em que é abordada a metaforização da política pelo futebol, em *corpus* jornalístico monolíngue de espanhol rio-platense, a partir da seção de *Humor político*.

Este trabalho, em primeiro lugar, obedece ao interesse pela presença de inúmeras representações metafóricas utilizadas na construção do discurso jornalístico de opinião e política, visto que reconhecemos, na leitura desses jornais, em especial os eletrônicos, a recorrência de metáforas na veiculação de informações. Em segundo lugar, este trabalho visa oportunizar, de maneira exploratória, uma aproximação ao universo das metáforas, por meio dos recursos da LC, tendo em conta que a metáfora está presente em todos os momentos de nossas vidas. Como salientam Lakoff e Johnson (2002), definimos a nossa realidade em termos de metáforas e, com isso, passamos a agir baseados nela. Por fim, visamos contribuir para os estudos da metáfora mediados pela LC e para os estudos linguísticos contrastivos entre língua espanhola e portuguesa.

A problematização mais geral que abarca nossa pesquisa de base (mestrado em andamento) e norteia este trabalho é: que evidências de metaforização, no nível textual e cognitivo, podemos identificar num *corpus* jornalístico comparável bilíngue? Que fatos linguísticos são indícios de representações metafóricas no *corpus*? Quais são os mapeamentos implicados entre os domínios fonte e alvo, para a interpretação das metáforas conceptuais, a partir dos dados do *corpus*? Quais são as metáforas linguísticas mais recorrentes e que metáforas conceptuais realizam?

Por meio de uma análise quanti-qualitativa, de cunho interpretativo, procuraremos responder a essas questões, no intuito de verificar o uso e funcionamento das metáforas integradas à vida cotidiana, de acordo com as teorias norteadoras.

2 Fundamentação teórica

A LC se dedica à criação e análise de *corpora* (plural latino de *corpus*), ou seja, conjuntos de textos e transcrições de fala armazenados em arquivos de computador. Segundo Shepherd (2012), ela consiste, simultaneamente, em uma nova metodologia, que se utiliza de textos naturais e das ferramentas da informática, e em uma nova abordagem para descrever a língua.

Ela vem mudando a maneira como se investiga a linguagem, nos seus mais diversos níveis, colocando à disposição do analista quantidades de dados antes inacessíveis. Um dos grandes agentes dessa mudança é a informática; sem ela, a LC contemporânea não poderia existir (BERBER SARDINHA, 2009). Além disso, ela é uma metodologia/abordagem que privilegia o empirismo e a autenticidade dos dados, analisando-os por meio de *softwares* e ferramentas computacionais, como o *WordSmith Tools* (SCOTT, 2012) e o *AntConc* (ANTHONY, 2019), que permitem lidar com técnicas quantitativas e qualitativas de análise.

O presente trabalho também é orientado pela teoria da metáfora conceptual de Lakoff e Johnson (1980). De acordo com alguns estudiosos, como May (*apud* DEIGNAN, 2005), as metáforas são mais frequentes em comunidade e contribuem para um entendimento coletivo de mundo, pois elas ora escondem, ora salientam aspectos da realidade daquele grupo de pessoas. Sendo assim, pode haver evidências de que grupos sociais interpretam acontecimentos de formas distintas de outros.

Kövecses (2005) salienta que o contexto também é fator fundamental a ser levado em consideração no uso das metáforas. Para Silva (2017, p. 217-218):

A história pessoal, profissão e interesses pessoais motivam as escolhas feitas para que determinados domínios-fonte sejam utilizadas para certos domínios-alvo em situações de comunicação. Dessa forma, as metáforas utilizadas na descrição da política poderiam variar não somente no que se refere ao individual, mas também ao cultural. Em um estudo comparativo, a análise das metáforas conceptuais e as respectivas expressões metafóricas por elas licenciadas em textos de língua portuguesa e em língua espanhola seria um indicador da influência cultural na seleção de expressões metafóricas, uma vez que seriam resultado de mapeamentos relevantes para determinadas civilizações e ideologias.

Segundo Deignan (2005), a metáfora, de acordo com a Teoria Conceptual, segue alguns princípios básicos: (1) organiza o pensamento, (2) é central para a linguagem abstrata, (3) é baseada na experiência física e (4) é ideológica. A seguir, apresentamos os procedimentos metodológicos implicados no trabalho.

3 Metodologia

Para Gil (1989), o método científico é um conjunto de procedimentos intelectuais e técnicos utilizados para atingir o conhecimento. Para que seja considerado conhecimento científico, é necessária a identificação dos passos para a sua verificação, ou seja, determinar o método que possibilitou chegar ao conhecimento. Nesse sentido, a seguir, são apresentadas as etapas percorridas relativas ao planejamento e compilação dos *corpora*, bem como à análise, realizada por meio do *Sketch Engine* (KILGARRIFF *et al.*, 2003).

3.1 Planejamento e criação do corpus

O *corpus* de estudo deste trabalho é formado por textos publicados nas seções de opinião e política dos jornais digitais argentinos *Clarín* e *Perfil*, e dos brasileiros *Estadão* e *Folha de São Paulo*. A análise dos dados, realizada por meio da LC, foi baseada na fundamentação teórica existente na área de Metáfora Conceptual.

A fim de alcançar os objetivos pretendidos, inicialmente, entramos em contato com alguns textos isolados, que também estariam no *corpus*, para uma aproximação e análise impressionística (BERBER SARDINHA, 2004), ou seja, para uma primeira impressão acerca dos textos. Posteriormente, adotamos a leitura hipertextual via linhas de concordância, que são fragmentos do conteúdo de vários textos analisados simultaneamente, com o propósito de encontrar indícios de usos metafóricos.

Para a análise dos *corpora*, utilizamos o *Sketch Engine* (KILGARRIFF *et al.*, 2003), doravante SE. Ele é um conjunto de ferramentas *web* que dispõe de diversos *corpora* já anotados e possibilita a criação de novos. O SE existe há mais de 15 anos e é bastante utilizado na Lexicografia, Linguística Computacional, Análise do Discurso e em pesquisas em Tradução e Ensino de Línguas.

Apesar de hoje possuir diversas funções, sua origem está na ferramenta *Word Sketch*, criada em 2002 para auxiliar na elaboração de dicionários. *Word Sketch* é a síntese do comportamento gramatical e das colocações de determinada palavra (KILGARRIFF *et al.*, 2014). Ao selecionar a colocação, também é possível, através da ferramenta *Concordance*, visualizar o contexto em que cada item ocorre. A Figura 1, a seguir, exhibe o painel principal do SE:

Figura 1: Painel do SE.

CORPUS RECIENTES		NUEVO CORPUS
Folha	Portuguese	1.940.029
DGT, Spanish	Spanish	57.311.149
DGT, Portuguese	Portuguese	53.950.705
Portuguese Web 2011 (ptTenTen11)	Portuguese	3.893.392.719
Timestamped JSI web corpus 2014-2016 English	English	18.315.071.361
English Web 2020 (enTenTen20)	English	38.149.437.411
British National Corpus (BNC)	English	95.134.547
Brazilian Portuguese corpus (Corpus Brasileiro)	Portuguese	871.117.178
French Web 2017 (frTenTen17)	French	5.752.261.039

Fonte: www.sketchengine.eu

Para a criação dos *corpora*, utilizamos a função disponibilizada pelo SE de criação de *corpus* a partir dos endereços inseridos, como, por exemplo, <https://www.clarin.com/politica/>, de modo que o SE só compilasse as seções relevantes para este estudo. Esse mesmo processo foi realizado com as seções de opinião e política de cada um dos quatro jornais. A Figura 2 demonstra de forma mais detalhada o ambiente de criação do *corpus*.

Figura 2: Criação do *corpus* a partir da *web*.

The screenshot shows the 'CREAR CORPUS' interface. At the top, there is a search bar with 'Clarin' entered. Below it, the text 'CORPUS: Clarin (Spanish)' is displayed. A progress bar indicates the current step: '1. CREAR CORPUS > 2. AÑADIR TEXTOS > 3. COMPILAR'. The main section is titled 'TEXTOS DE INTERNET' and contains the following configuration options:

- Tipo de entrada:** Three radio buttons are present: 'Búsqueda web' (unselected), 'Las URL' (unselected), and 'Sitio web' (selected).
- URL:** The text 'https://www.clarin.com/politica/' is entered.
- Nombre de carpeta:** The text 'Clarin' is entered.
- Opciones de lista de bloqueo:** A dropdown menu.
- Opciones de lista blanca:** A dropdown menu.
- Restricciones de tamaño:** A dropdown menu.
- Compilar cuando hay terminado:** A checked checkbox.
- Buttons:** 'CANCELAR' and '¡YA!' (highlighted in red).

Fonte: www.sketchengine.eu

3.2 Compilação e armazenamento dos *corpora*

A partir da criação dos *corpora*, iniciamos sua compilação e armazenamento em formato .txt. Para esta pesquisa, foram criados os seguintes *corpora*:

- (i) Clarin.txt (624.761 *tokens* e 533.281 *types*);
- (ii) Estadão.txt (1.433.965 *tokens* e 1.155.405 *types*);
- (iii) Folha.txt (2.255.939 *tokens* e 1.943.776 *types*); e
- (iv) Perfil.txt (1.003.772 *tokens* e 866.819 *types*).

A Figura 3 mostra a compilação do *corpus* pelo SE.

Figura 3: Compilação do *corpus*.



Fonte: www.sketchengine.eu

De acordo com Berber Sardinha (2000), um *corpus* é considerado pequeno se possui menos de 80 mil palavras, médio se possui de 250 a 1 milhão de palavras e grande se possui 10 milhões ou mais palavras. Segundo Sinclair (1991), para ser representativo de uma língua ou variedade linguística, o *corpus* deve ser o mais extenso possível.

Levando-se em conta a data dessas publicações e o crescimento da área da LC nos dias atuais, o *corpus* deste trabalho é classificado como de extensão média, entre 1 e 5 milhões de palavras. Contudo, para análise de metáforas, pela especificidade da análise, trata-se de uma extensão que possibilita inúmeras direções de pesquisa.

Realizados todos os procedimentos de criação, compilação e armazenamento dos *corpora*; o passo seguinte foi analisar os dados utilizando recursos avançados da ferramenta *Concordance* do SE.

3.3 Trabalhando com o *Sketch Engine*

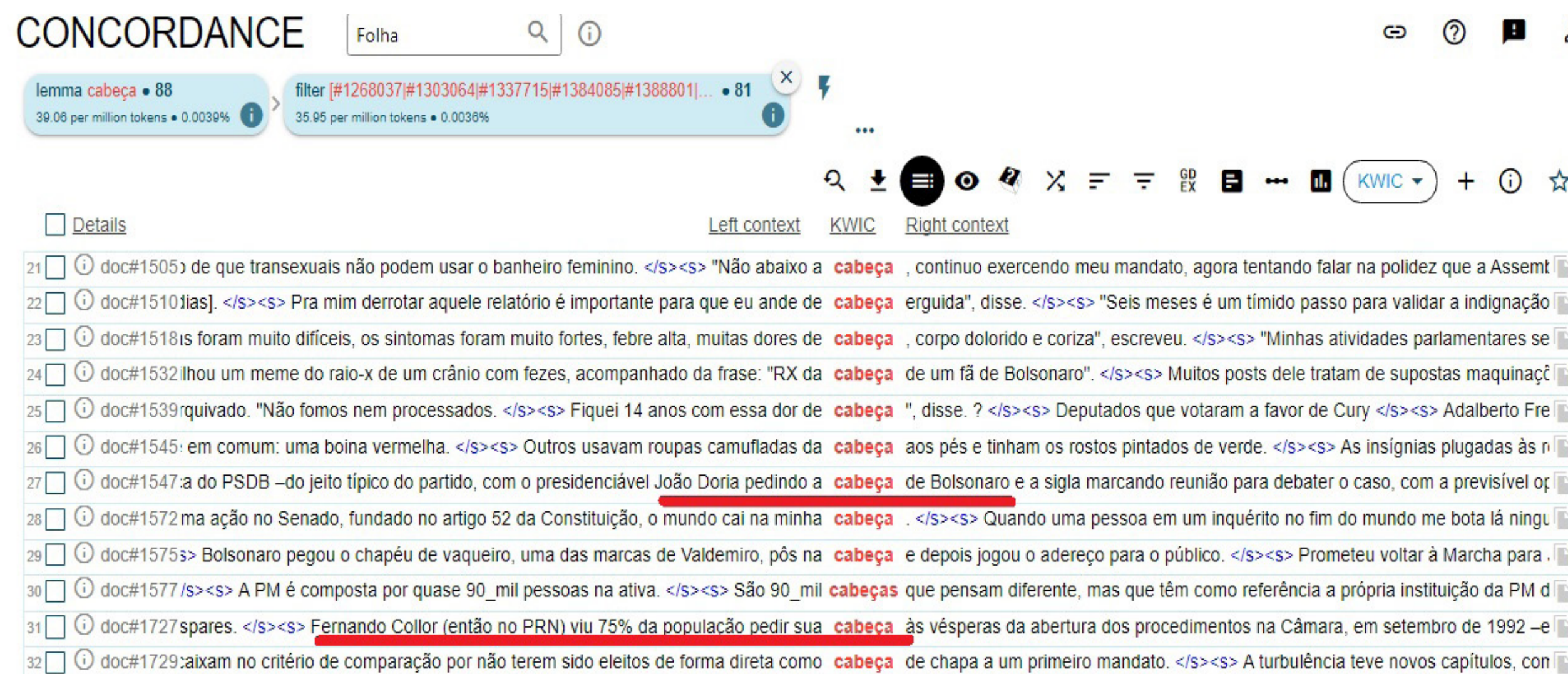
Com os *corpora* devidamente armazenados em arquivos .txt, geramos uma lista de palavras para cada *corpus*, organizadas pelo critério de frequência. As listas foram geradas por meio da ferramenta *Wordlist* do SE.

Em seguida, partindo da observação dos itens lexicais mais frequentes, na ferramenta *Word Sketch* pudemos apreciar candidatos a prováveis usos metafóricos. Com isso, elegemos o item *cabeça* e seu correspondente em espanhol, *cabeza*. A palavra *cabeça* ocorre 71 vezes no Estadão.txt e 86 no Folha.txt. A palavra *cabeza*, por sua vez, ocorre 97 vezes no Clarin.txt e 129 no Perfil.txt.

3.4 Listas de concordância

Com o objetivo de identificar usos metafóricos de uma forma mais detalhada, analisamos as concordâncias das palavras *cabeça* e *cabeza*, por meio da ferramenta *Concordance*. Esse processo analítico foi realizado em todos os *corpora*. A Figura 4 apresenta alguns resultados.

Figura 4: Linhas de concordância de *cabeça* no *corpus* Folha.txt.



Fonte: www.sketchengine.eu

Para o presente estudo exploratório, considerando a necessidade de se fazer um recorte, nos limitamos às análises das unidades fraseológicas² “pedir a cabeça” e “cabeça tem que rodar”, bem como de seus correspondentes em espanhol. Pelo fato de terem ocorrido nos *corpora* de ambas as línguas, essas unidades possibilitam uma perspectiva de análises em contraste, inclusive para conferir se se trataria de uma mesma metáfora conceptual.

4 Resultados e discussão

Após a análise das linhas de concordâncias geradas para todos os *corpora*, apresentamos uma amostra dos trechos que contêm as unidades fraseológicas focalizadas neste estudo:

- Jornal Clarín – “Más allá del llamado de la vicepresidenta, el kirchnerismo duro venía pidiendo la cabeza de Guzmán...” (Para além do chamado da vicepresidenta, o kirchnerismo duro vinha pedindo a cabeça de Guzmán...);
- Jornal Perfil – “Luego que el oficialismo perdiera la elección Primaria, expresó: “Cuando el pueblo pide sangre, alguna cabeza tiene que rodar ante el enojo de la gente” (Logo que o oficialismo perdesse a eleição primaria, expressou: “Quando o povo pede sangue, alguma cabeça tem que rodar diante da indignação das pessoas”);

² “[...] as unidades fraseológicas (UFS) – objeto de estudo da fraseologia – são unidades léxicas formadas por mais de duas palavras gráficas em seu limite inferior, cujo limite superior se situa no nível da oração composta. Estas unidades se caracterizam por sua alta frequência de uso, e de co-ocorrência de seus elementos integrantes; por sua institucionalização, entendida em termos de fixação e especialização semântica; por sua idiomatidade e variação potenciais; assim como pelo grau em que se dão todos estes aspectos nos diferentes tipos” (CORPAS PASTOR, 1996, p. 20).

- c. Estadão – “Internautas compartilharam o vídeo alegando que a PM marchava rumo ao Palácio do Governo pedindo a cabeça do governador baiano, Rui Costa (PT)”;
- d. Folha – “Saindo do campo da esquerda, a quimérica solução agora está na boca do PSDB – do jeito típico do partido, com o presidenciável João Doria pedindo a cabeça de Bolsonaro...”.

Inicialmente, destacamos que “pedir a cabeça” ou “cabeça tem que rodar”, segundo o *Dicionário Brasileiro de Fraseologia* (SILVA, 2013) e o *Dicionário UNESP do Português Contemporâneo* (BORBA, 2005), referem-se ao ato de solicitar/pedir uma punição para alguém. Essas unidades fraseológicas originam-se do direito de executar ladrões e outros criminosos concedido aos nobres pela Coroa francesa no século XVIII.

Todos os exemplos apresentados revelam situações abstratas da política, metaforizadas por meio da imagem mais concreta de uma punição por meio de uma situação extrema que envolve a execução de alguém por decapitação, com uso de guilhotina. O “pedir a cabeça” ou “rolar cabeça”, no âmbito político, significa a destituição do cargo.

Nas referidas expressões metafóricas, os mapeamentos que envolvem o agenciamento do domínio-fonte, a decapitação, para a compreensão do domínio-alvo, demissão do cargo ou função, auxiliam no entendimento da mensagem. A inferência da metáfora conceptual que decorre das expressões focalizadas é: DEMISSÃO DE CARGO OU FUNÇÃO É UMA DECAPITAÇÃO.

Poderíamos discutir um pouco mais sobre as demais unidades fraseológicas encontradas nas listas de concordância das palavras elegidas, mas, devido à limitação de espaço neste capítulo, passemos às considerações finais.

5 Considerações finais

O objetivo desta análise foi explorar *corpora* jornalísticos comparáveis em espanhol e em português, no intuito de verificar evidências de metaforização, no nível textual, a ponto de alcançar a inferência metafórica conceptual no plano cognitivo. Por meio de duas unidades fraseológicas formadas pelo item lexical *cabeça* + verbo, identificados nos *corpora* de estudo, constatamos tanto as metáforas linguísticas quanto as conceptuais.

Aplicando os pressupostos da LC, apoiados na abordagem do estudo da metáfora, desenvolvida inicialmente por George Lakoff e Mark Johnson (1980), e fazendo uso do conjunto de ferramentas do SE, foi possível identificar fatos linguísticos que apresentam indícios metafóricos, mapear os seus domínios e interpretar suas metáforas conceptuais.

| Referências

- ANTHONY, L. **AntConc** (Version 3.5.8) [Computer Software]. Tokyo: Waseda University, 2019. Disponível em: <https://www.laurenceanthony.net/software>. Acesso em: 10 jan. 2022.
- ARISTÓTELES. **Poética**. São Paulo: Abril Cultural, 1973.
- BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.
- BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. **DELTA**, São Paulo, v. 16, n. 2, p. 323-367, 2000. DOI: <https://doi.org/10.1590/S0102-44502000000200005>.
- BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.
- BORBA, F. S. (org.). **Dicionário Unesp do português contemporâneo**. São Paulo: Ed. Unesp, 2005.
- DEIGNAN, A. **Metaphor and Corpus Linguistics**. Amsterdam: John Benjamins, 2005.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 2. ed. São Paulo: Atlas, 1989.
- KILGARRIFF, A. *et al.* **Sketch Engine**. East Sussex: Lexical Computing Limited, 2003. Disponível em: <https://www.sketchengine.eu/>. Acesso em: 10 jan. 2022.
- KILGARRIFF, A. *et al.* The Sketch Engine: ten years on. **Lexicography**, Berlin, v. 1, n. 1, p. 7-36, 2014. DOI: <https://doi.org/10.1007/s40607-014-0009-9>.
- KÖVECSES, Z. **Metaphor in culture: universality and variation**. New York: Cambridge University Press, 2005.
- LAKOFF, G. The contemporary theory of metaphor. *In*: ORTONY, A. (ed.). **Metaphor and thought**. 2. ed. Cambridge: Cambridge University Press, 1993. p. 202-251. DOI: <https://doi.org/10.1017/CBO9781139173865.013>.
- LAKOFF, G.; JOHNSON, M. **Metáforas da vida cotidiana**. Tradução Grupo de Estudos da Indeterminação e da Metáfora (GEIM). São Paulo: Mercado de Letras, 2002.
- LAKOFF, G.; JOHNSON, M. **Metaphors we live by**. Chicago: Chicago University Press, 1980.
- NOVODVORSKI, A.; BEVILACQUA, C. R. De marcar la cancha a una canchereada na metaforização da política pelo futebol: análise de unidades fraseológicas especializadas em *corpus* jornalístico argentino. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 29, n. 2, p. 1191-1228, 2021. DOI: <http://dx.doi.org/10.17851/2237-2083.29.2.1191-1228>.

PASTOR, G. C. **Manual de Fraseologia Espanhola**. Madri: Editorial Gredos, 1996.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012.

SHEPHERD, T. M. G. Panorama da Linguística de Corpus. *In*: SHEPHERD, T. M. G.; BERBER SARDINHA, T.; PINTO, M. V. (org.). **Caminhos da Linguística de Corpus**. Campinas: Mercado de Letras, 2012.

SILVA, C. S. Futebol e Metáfora: um estudo de *corpus* em textos jornalísticos de língua inglesa. *In*: SEMINÁRIO DOS ALUNOS DOS PROGRAMAS DE PÓS-GRADUAÇÃO DO INSTITUTO DE LETRAS DA UFF, 8., 2017, Niterói. **Anais [...]**. Niterói: UFF, 2017. p. 216-230. Disponível em: <http://www.anaisdosappil.uff.br/index.php/VIIISAPPIL-Ling/article/view/671/418>. Acesso em: 10 jan. 2022.

SILVA, J. P. **Dicionário Brasileiro de Fraseologia**: versão preliminar. Rio de Janeiro: [S. n.], 2013.

SINCLAIR, J. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.

SPERBER, D.; WILSON, D. A deflationary account of metaphors. *In*: GIBBS Jr., R. W. (ed.). **The Cambridge Handbook of Metaphor and Thought**. Cambridge: Cambridge University Press, 2008. p. 84-106. DOI: <https://psycnet.apa.org/doi/10.1017/CBO9780511816802.007>.

Elos coesivos na tese e indicação de impessoalidade: estudo exploratório em *corpus* de redações estilo ENEM

Rosena Caixeta Silva Rodrigues de Sousa¹

¹ Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/8029826531413628>. E-mail: rosena@ufu.br.

1 Introdução

Quando tratamos de posicionamento – especificamente no tipo textual dissertativo-argumentativo, ou redação estilo ENEM, como denominamos – há um detalhe importante que precisa ser considerado: a construção da tese no conjunto textual deve ser o mais impessoal possível. Diante dessa questão, é válido analisar como o produtor do texto (chamado de candidato num processo de seleção como o do ENEM) se posiciona de modo impessoal, ou seja, sem o uso da 1ª pessoa do singular ou da 1ª pessoa do plural, como é comum nas demais redações argumentativas opinativas.

Nesse tipo de redação, é possível construir a tese de modo que o autor se posicione por meio de ideias, de fatos e de outras opiniões sem tornar a pessoalidade o ponto primordial de sua construção discursiva e sem deixar sua individualidade, subjetividade e personalidade de lado. Isso, por sua vez, é desafiador para os alunos do Ensino Médio, pois estão acostumados a produzirem seus textos se posicionando de modo mais pessoal, como no caso do artigo de opinião ou do debate escrito, por exemplo.

Nesse contexto, o presente capítulo objetiva investigar, de modo sucinto, a presença de elos coesivos formados por agrupamentos específicos que indicam a impessoalidade dos autores na construção de teses das redações estilo ENEM. O *corpus* explorado nesta análise é composto por sete redações estilo ENEM que alcançaram nota 1000 (nota de referência) neste exame, totalizando 1.978 *tokens* e 373 *types*. As redações foram retiradas de jornais *online* e da *Cartilha do Participante* de cada ano de estudo. Para auxiliar na análise do *corpus*, lançamos mão do *WordSmith Tools 6.0* (SCOTT, 2012), doravante WST.

2 Fundamentação teórica

Nesta seção, são abordados alguns conceitos básicos deste estudo. *A priori*, esclarecemos o que é a coesão, aspecto importante ao se analisar linguisticamente a argumentação. Posteriormente, abordamos a tese na dissertação-argumentativa e sua relação com a argumentação. Por fim, introduzimos a metodologia/abordagem da Linguística de *Corpus*.

2.1 Linguística Textual: coesão

A Linguística Textual é uma subárea da Linguística que estuda a constituição dos textos. Alguns dos elementos textuais abordados dentro da Linguística Textual é a coesão. Tal elemento explora a utilização de elos coesivos na compreensão do texto.

Conforme Koch (1994), a coesão pode ser dividida em: referencial ou remissiva e sequencial ou de sequenciação.

Chamo, pois, de *coesão referencial* aquela em que um componente da superfície textual do texto faz remissão a outro(s) elemento(s) nela presentes ou inferíveis a partir do universo textual. Ao primeiro, denomino *forma referencial* ou *remissiva* e ao segundo, *elemento de referência* ou *referente textual* (KOCH, 2008, p. 31, grifo da autora).

A coesão sequencial diz respeito aos procedimentos linguísticos por meio dos quais se estabelecem, entre segmentos do texto (enunciados, partes de enunciados, parágrafos e sequências textuais), diversos tipos de relações semânticas e/ou pragmáticas, à medida que se faz o texto progredir (KOCH, 1994, p. 53).

A coesão referencial e a sequencial podem ser resumidas como diferentes maneiras de se retomar elementos textuais a fim de evitar repetições e tornar o texto mais fluido, além de propiciar maior entendimento, por parte do leitor, das ideias que ocorrem em sequência, de modo ordenado e bem pensado.

Quando se analisa redações dissertativo-argumentativas, verifica-se que a base para a compreensão textual está na utilização e na diversidade desses elos coesivos, os quais são fundamentais para dar seguimento às partes essenciais do texto, que são: a introdução, o desenvolvimento e a conclusão das ideias, com base em um tema preestabelecido.

Ao se estudar a composição dos argumentos, vê-se o quanto os elementos coesivos auxiliam na organização das informações a serem apresentadas, visto ser essencial mostrar estratégias argumentativas plausíveis que fundamentam o ponto de vista do autor, o qual é candidato à vaga em uma universidade. Essas estratégias são realizadas por meio das escolhas lexicais feitas no decorrer do processo de escrita. À guisa de exemplificação, o uso de “portanto” ou “logo” é feito para indicar a conclusão de uma ideia e ajuda o leitor a saber que se trata de uma finalização do raciocínio, seja na argumentação ou na conclusão do texto. Outro exemplo seria quando há contra-argumentação ou refutação de alguma informação, sendo preciso lançar mão de algum elemento que apresente essa ideia, seja uma locução adverbial, uma conjunção ou um advérbio, como: “Entretanto”, “Todavia” e até mesmo “Mas”, os quais são utilizados para contrariar a ideia do enunciado anterior.

2.2 A construção da tese e da argumentação no texto dissertativo-argumentativo

Do grego *thesis*, a tese significa proposição intelectual. É basicamente a ideia central, a ideia-núcleo, o ponto de vista, a opinião, o posicionamento crítico do autor, portanto, subjetivo, em relação a determinado tema ou assunto proposto. Porém, quando se trata do texto dissertativo-argumentativo, essa proposição deve ser elaborada de modo mais objetivo e impessoal possível,

visto ser uma exigência desse tipo de texto a apresentação das ideias do candidato sem o uso da 1ª pessoa, seja do singular ou do plural. Ou seja, é preciso se posicionar em relação ao que pensa sobre o tema em foco, sem explicitar que é esse o seu posicionamento. Então, como fazê-lo?

Em primeira análise, é preciso entender que, desde o início da humanidade, os indivíduos utilizam-se da linguagem para se comunicar, expressar suas opiniões e modificá-las em determinado momento quando se adquire mais sabedoria. Dessa maneira, por meio da interação no corpo social, se aprende com o outro a trocar opiniões e a argumentar sobre aquilo que se defende. O ser humano precisa dessa interação constante com o outro para criar suas próprias opiniões. Diante disso, é preciso condicionar o ponto de vista ao interlocutor, ou melhor, ao emissor da mensagem que está sendo apresentada, já que ele opina sobre o que pensa, reflete e acredita. Depois, ele formaliza essas informações por meio das palavras, sejam escritas ou faladas. Desse modo, o emissor ou o criador ficam atrelados às ideias que apresentam, por meio das escolhas textuais que expressam sua subjetividade.

No entanto, na construção discursiva há muitas formas de se expressar determinada ideia, e uma delas é expor outros posicionamentos vinculados ao posicionamento inicial do interlocutor, para a relação entre suas palavras e as dele fazer sentido, ornando posicionamentos em uma só frase, em uma ideia completa do que se pensa com base em fatos, opiniões e ideias de outrem. Assim, ao se construir a tese com base em outras ideias já pensadas e fatos analisados, consegue-se produzir um posicionamento sem necessariamente condicionar o uso da 1ª pessoa, visto ser uma ideia consumida por outras ideias e fundamentada em argumentos que serão retomados no desenvolvimento do texto dissertativo, local mais indicado para essa fundamentação.

Quanto ao processo de argumentação, há um problema pontual cunhado por Pécora (1992) que reflete bem essa construção necessária de ideias da produção escrita, a argumentação do dever, como explicita:

A argumentação do dever se esgota nela mesma, isto é, na referência a um padrão "a priori", uma razão oculta e acima do texto, que condena ou aprova esta ou aquela conclusão. Aliás, quando se fala em nome do dever, tudo é conclusão: o texto se fecha em torno de algumas noções veladas que decidem sobre a sorte do tema; ou melhor, retomam uma decisão anterior. *O efeito básico desse tipo de argumentação é, paradoxalmente, a suspensão da criação de argumentos pelo texto.* Em termos mais gerais, poder-se-ia dizer que o seu efeito é o de descaracterizar o ato de linguagem. Ao receber a tarefa de criar os próprios argumentos, de compor um texto capaz de, pela própria especificidade, agir de modo a convencer o seu interlocutor, o aluno tende a interpretá-la de modo a fazer inveja a um escoteiro (PÉCORA, 1992, p. 100, grifo do autor).

Pécora (1992) ainda explica que, basicamente, o que se faz na produção escrita é:

[...] *reduzir o seu texto ao enunciado de uma ordem* – na qual, com certeza, ele não tem lugar como sujeito. Ou seja, essa ordem instaurada pela noção de *dever*, quando não é simplesmente um nome para a ausência de nexos entre as partes do texto, não passa de eco mais ou menos fragmentário de outras ordens, enunciadas por padrões que preexistem a seu usuário e que independem desse texto em particular – portanto, que são incapazes de representar marcas de uma relação interpessoal (PÉCORA, 1992, p. 100, grifo do autor).

Diante disso, percebemos o quanto o processo de escrita é árduo e pressupõe um modo mais aceitável de se construir ideias nesse tipo de texto cobrado em processos seletivos. Assim, chegamos à conclusão de que é possível fundamentar argumentos de modo objetivo, impessoal e, portanto, menos subjetivo e pessoal, visto ser “padronizada” a forma de se construir argumentos na dissertação-argumentativa. E para comprovar essa ideia, é válida a análise de um *corpus* como o deste estudo, pois, assim, é possível compreender de forma mais assertiva a constância dessa moldagem de escrita tão própria desse tipo de texto.

2.3 Linguística de Corpus

A Linguística de *Corpus* (doravante LC) é uma metodologia/abordagem que prevê o trabalho com conjuntos de textos digitais. Ao usarmos tal metodologia/abordagem, contamos com o auxílio de programas e ferramentas computacionais, visto que geralmente lidamos com uma quantidade de textos que dificultaria análises manuais. Esses textos são coletados e organizados de acordo com os critérios de cada pesquisa. É válido ressaltar que, embora a grande quantidade de textos seja algo marcante na LC, ela é uma metodologia/abordagem que não se restringe ao aspecto quantitativo, mas que, em especial, também dá suporte para análises qualitativas, sendo, portanto, usada em pesquisas de natureza quali-quantitativa.

Para o presente trabalho, a LC é fundamental, pois torna mais produtiva a análise textual e a apresentação dos aspectos da construção discursiva de modo mais esclarecedor ao leitor. Nesse sentido, com o intuito de facilitar a análise do *corpus*, utilizamos o WST 6.0 (SCOTT, 2012), definido como um conjunto de ferramentas integradas destinado à análise linguística. Esse *software* permite fazer análises baseadas na frequência e na coocorrência de palavras em *corpora*, conforme explicita Berber Sardinha (2009). Especificamente, utilizamos as ferramentas *WordList* e *Concord*, cujas funções são apresentadas por Berber Sardinha (2009, p. 8):

- *WordList*: produz listas de palavras contendo todas as palavras do arquivo ou arquivos selecionados, elencadas em conjunto com suas frequências absolutas e percentuais. Também compara listas, criando listas de consistência, onde é informado em quantas listas cada palavra aparece.

- *Concord*: realiza concordâncias, ou listagens de uma palavra específica (o 'nóculo', *node word* ou *search word*) juntamente com parte do texto onde ocorreu. Oferece também listas de colocados, isto é, palavras que ocorreram perto do nóculo.

Diante das informações supracitadas, com o auxílio do WST, é possível analisarmos as palavras que são usadas em um *corpus*, a frequência de cada uma delas, o seu posicionamento nos textos do *corpus*, as escolhas lexicais próximas a elas, os agrupamentos, com quais temas ou assuntos associam-se com maior frequência, em quais contextos linguísticos se encontram, dentre outros aspectos. É a partir dessas informações que o pesquisador subsidia suas análises sobre o uso e o comportamento da língua.

3 Metodologia

Para o desenvolvimento desta análise, foi preciso preparar o *corpus* de redações e depois fazer uma breve descrição e análise das redações para a seleção de teses.

A primeira parte do trabalho foi selecionar as redações cujas notas foram máximas no ENEM dos anos de 2011 a 2021. Para isso, os textos foram copiados e transferidos para arquivos .docx, do *Microsoft Word*. Em seguida, salvamos em formato .txt, que é o formato mais produtivo no WST. Depois, revisamos o *corpus* de redações em sua totalidade, por meio de testes.

A segunda parte compreende o processo de descrição e de análise do *corpus* de redações, cujo intuito foi identificar as teses apresentadas pelos autores das redações. Assim, utilizamos o *corpus* para:

- a) Verificar se a tese é apresentada sempre na introdução, local mais adequado para se compor, conforme os critérios de correção desse tipo de redação;
- b) Interpretar e descrever os elos coesivos que indicam a impessoalidade na construção da tese.

4 Análises preliminares

A título de análise, ilustraremos dois exemplos de redações dos sete textos que foram submetidos ao WST. Como mencionado anteriormente, todos os textos em análise são de redações cujas notas foram 1000 no ENEM.

Inicialmente, foi feita a leitura da lista de palavras organizada em ordem de frequência, gerada pela ferramenta *WordList*, em busca de elos coesivos nos sete arquivos de redações nota 1000. Tal busca resultou na identificação de *porém* (conforme consta na Figura 1), conectivo geralmente utilizado para indicar oposição à ideia anterior. Isso nos chamou a atenção, pois

nos mostra que, na elaboração de ideias, o autor/candidato refutou algum aspecto fazendo uso desse elemento coesivo, algo bem comum na construção de teses, proposições, bem como na argumentação desse tipo de texto.

Figura 1: Lista de palavras.

N	Word	Freq.	%	Texts	% Lemmas	Set
123	DESSES	7	0.10	5	25.00	
124	DEVEM	7	0.10	6	30.00	
125	DIANTE	7	0.10	5	25.00	
126	EDUCAÇÃO	7	0.10	6	30.00	
127	ENTÃO	7	0.10	7	35.00	
128	GARANTIR	7	0.10	7	35.00	
129	HÁ	7	0.10	5	25.00	
130	HOJE	7	0.10	6	30.00	
131	INDIVÍDUOS	7	0.10	6	30.00	
132	MUITOS	7	0.10	5	25.00	
133	NECESSÁRIO	7	0.10	6	30.00	
134	PORÉM	7	0.10	6	30.00	
135	SEUS	7	0.10	6	30.00	
136	TAIS	7	0.10	6	30.00	
137	TER	7	0.10	6	30.00	
138	ATENÇÃO	6	0.09	4	20.00	
139	CADA	6	0.09	5	25.00	
140	CIDADÃOS	6	0.09	6	30.00	
141	CONSCIENTIZAÇÃO	6	0.09	4	20.00	
142	DESENVOLVIMENTO	6	0.09	4	20.00	
143	DEVIDO	6	0.09	6	30.00	
144	DISPOSIÇÃO	6	0.09	5	25.00	

Fonte: *WordList* (dados da pesquisa).

Em seguida, geramos linhas de concordância para o conectivo *porém* na ferramenta *Concord* (Figura 2) e observamos que esse elemento coesivo, de fato, seguiu sua função de oposição à ideia anterior apresentada. Contudo, ainda era necessário verificarmos se isso ocorria no início do texto, mais especificamente na introdução, local mais indicado para a apresentação da tese, ou em outras partes do texto, como desenvolvimento ou conclusão. Para fazer essa análise, expandimos as linhas de concordância na própria ferramenta *Concord* (Figura 3).

Figura 2: Linhas de concordância com *porém*.

N	Concordance	Set	Word #	Seri	Seri	Parz	Parz	Hea	Hea	Sec	Sec	File	Dat
1	considerável no número de acessos . Porém, um ponto importante a ser		162	7	13%	0	50%			0	50%	ENEM NOTA M 2021/set/26 01	
2	para exercer determinada atividade . Porém, conduzir um carro é uma		126	7	6%	0	32%			0	32%	ENEM NOTA M 2021/set/26 01	
3	de um ou dois copos de cerveja . Porém, o risco de acidente existe e,		270	9	13%	0	74%			0	74%	ENEM NOTA M 2021/set/26 01	
4	da cultura de diversas civilizações . Porém, com o surgimento do		16	1	11%	0	5%			0	5%	ENEM NOTA M 2021/set/26 01	
5	como instrumento de conscientização , porém sua efetivação encontra		117	4	35%	0	34%			0	34%	ENEM NOTA M 2021/set/26 01	
6	das maravilhas do mundo moderno . Porém é preciso cuidado ao lidar com		37	1	17%	0	11%			0	11%	ENEM NOTA M 2021/set/26 01	
7	do número de acidentes de trânsito . Porém, sua efetividade completa só		242	14	17%	0	74%			0	74%	ENEM NOTA M 2021/set/26 01	

Fonte: *Concord* (dados da pesquisa).

Figura 3: Arquivo de redação – texto 1.



Fonte: Concord (dados de pesquisa).

A partir da análise do *corpus*, foi possível identificar, em dois arquivos, o elemento coesivo *porém* na introdução do texto ao fundamentar a tese, ou seja, essa é uma regularidade interessante para nossa pesquisa, e, para tanto, poderá ser o ponto de partida que nos indicará a necessidade de contraposição para fundamentação de ideias. Todavia, ainda é necessário condicionar o uso desse elemento coesivo à questão da impessoalidade, aspecto fundamental do nosso estudo.

Diante disso, separamos as duas redações que apresentaram essa incidência do *porém* na introdução, na tentativa de analisarmos um pouco mais a fundo as consequências desse uso da construção discursiva e como se estabelece uma relação de impessoalidade nessas produções.

Na Figura 3, vemos o primeiro exemplo de uma redação sobre a Lei Seca – **“Efeitos da implantação da Lei Seca no Brasil”** – tema da aplicação em 2013 do Enem. Nessa redação, a candidata, que obteve nota máxima no exame, fez uso do elemento coesivo *porém* em seu segundo período da introdução.

Para desenvolver o tema em foco, a participante apresentou a ideia de que *“Desde sua invenção, a bebida alcoólica faz parte da cultura de diversas civilizações.”*. E depois contrapôs esse fato elucidando que *“Porém, com o surgimento do automóvel, esse e aquela não podem ser variáveis de uma mesma equação”*. Assim, percebemos que a candidata considera o surgimento do automóvel uma problemática quando se une ao uso da bebida alcoólica. A Lei Seca, sancionada no Brasil no final de 2012, foi criada para diminuir a incidência de pessoas alcoolizadas no trânsito, pois o número de acidentes, dentre eles, muitos considerados graves,

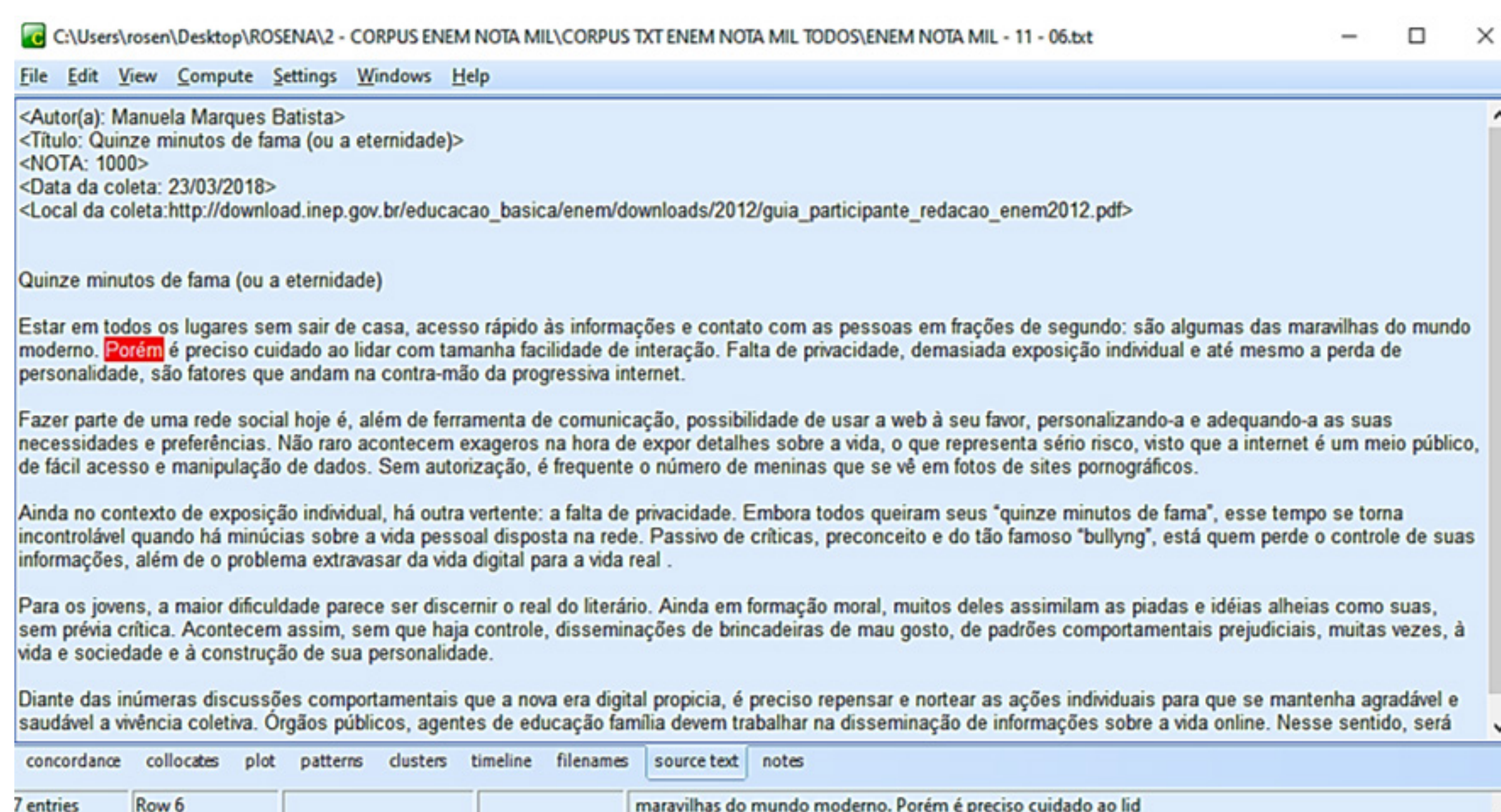
em decorrência dessa combinação, era muito alto. Então, a tese fundamenta-se de modo coerente com o tema, problematizando a questão principal no que tange ao uso do álcool agregado à condução de um automóvel.

Em relação ao aspecto da impessoalidade, observamos que, em nenhum momento da introdução, a autora apresentou elementos que indicassem pessoalidade, como o uso da 1ª pessoa do discurso no singular “Eu” ou da 1ª pessoa do discurso no plural “Nós”, assim como nenhum verbo conjugado que condicionasse essa pessoalidade, como: “concordo”, “penso”, “considero”, “analiso”, dentre outros. Contudo, mesmo sem essa pessoalidade na construção discursiva, foi possível identificar a opinião da autora, baseada em um fato já conhecido e comprovado, por meio de dados explicitados também nos textos motivadores da proposta: a Lei Seca diminuiu consideravelmente o número de mortes decorrentes de acidentes de trânsito.

Ainda na intenção de explorar a hipótese de que as redações com notas favoráveis no exame apresentam com maior frequência a tese de forma impessoal, apresentamos outro exemplo que foi possível observarmos com o auxílio do WST.

Na Figura 4, veiculamos um recorte do segundo e último exemplo de uma redação que faz uso do elemento coesivo *porém* na introdução do texto.

Figura 4: Arquivo de redação – texto 2.



Fonte: *Concord* (dados de pesquisa).

É possível observar que a conjunção adversativa *porém* se localizou no mesmo espaço linguístico-discursivo do outro texto em análise, no início do segundo período da introdução. O tema dessa redação é diferente da anterior, pois aborda a questão da **“Manipulação do comportamento do usuário pelo controle de dados na internet”** e foi cobrado no ENEM de 2018.

De antemão, percebemos que o uso do *porém* condicionou uma oração em contraposição à anterior, em conformidade com o sentido semântico que esse conectivo proporciona. Ao lermos a ideia por completo, compreendemos que o “*acesso rápido às informações*” e o “*contato com as pessoas em frações de segundo*”, para a autora do texto, “*são algumas das maravilhas do mundo moderno*”, já que, atualmente, não é preciso sair de casa para se conectar com indivíduos do outro lado do mundo, bem como para realizar pesquisas de modo rápido e seguro. Contudo, a participante refuta essa ideia ao lançar mão do *porém* para indicar que “**é preciso cuidado ao lidar com tamanha facilidade de interação**”. Ou seja, ela apresentou dois pontos interessantes sobre o uso da *internet* na atualidade (acesso rápido e contato rápido com outras pessoas), mas também reforçou que é necessário cuidado, pois tamanha facilidade também causa outros problemas como “*Falta de privacidade, demasiada exposição individual e até mesmo a perda de personalidade*”. Tudo isso nos mostra que a autora conseguiu relacionar bem a ideia de como os usuários da *internet* se comportam hoje em dia mediante esses fatores explicitados, conectando de forma produtiva a sua tese ao tema explorado.

O uso de elementos que condicionem pessoalidade não foi percebido, assim como no exemplo analisado anteriormente. Além disso, a ideia de apresentar uma contraposição por meio do elemento coesivo *porém* para fundamentação de uma opinião baseada em fatos irrefutáveis, conforme detalhamos, é reforçada.

Diante das análises preliminares compostas, ressaltamos alguns pontos de observação importantes:

- as redações analisadas possuem suas teses no local indicado para elas, que é a introdução do texto;
- sabemos que as teses estão na introdução pelos indicativos linguísticos (nos exemplos explicitados, pelo uso de *porém*, mas poderiam ser verbos conjugados em 3ª pessoa, conjunções que indicam ideias de negação, dentre outros, pois há variação conforme o tema e a personalidade discursiva do autor);
- as teses observadas estabeleceram uma relação de sentido e de coerência com o tema proposto, o que pressupõe planejamento e organização das ideias;
- todas as redações observadas não apresentaram elementos que condicionassem pessoalidade, ou seja, não houve percepção de verbos conjugados em 1ª pessoa, seja do singular ou do plural, e nenhum pronome ou substantivo que demonstrasse esse tipo de pessoalidade.

Por fim, consideramos que a análise desses exemplos foi de grande valia para prosseguirmos com os estudos da pesquisa em andamento, ampliando ainda mais nossa hipótese de que é

possível construir uma técnica eficaz de produção escrita opinativa, mesmo que impessoal, numa redação estilo ENEM.

5 Considerações finais

Diante da proposta de estudar a impessoalidade na construção da tese e da argumentação, com foco em elos coesivos e com base em um *corpus* composto por redações estilo ENEM, foi possível verificar que o uso de determinados elementos coesivos condiciona uma fluidez maior na produção de textos, auxiliando na apresentação de uma escrita opinativa e impessoal.

| Referências

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

KOCH, I. V. **O que é a linguística textual**. In: KOCH, I. V. **A coesão textual**. 7. ed. São Paulo: Contexto, 1994.

PÉCORA, A. **Problemas de redação**. 4. ed. São Paulo: Martins Fontes, 1992.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://lexically.net/wordsmith/version6/>. Acesso em: 14 jan. 2022.

Coletando *corpus* da internet com auxílio de *scripts* de programação

Heitor Carvalho de Almeida Neto¹

¹ Mestrando em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/9429112861487266>. E-mail: heitor@ufu.br.

1 Introdução

A matéria-prima para a Linguística é a língua e para a Linguística de *Corpus* (doravante LC) são os textos, textos estes que tenham sido escritos em linguagem natural e não textos produzidos para atender a fins de pesquisa. Segundo Biderman (2001, p. 79), podemos definir um *corpus* como “uma coletânea de textos em formato eletrônico codificados de modo padronizado e homogêneo. [...] a principal forma de obtenção desses dados da atualidade se dá por intermédio da internet, ou através da digitalização de documentos, tornando-os assim documentos digitais”. Os textos disponibilizados via internet estão comumente em formato de hipertexto² ou em algum outro formato eletrônico.

A captura desses textos pode ser uma tarefa, além de demorada, cansativa. Em se tratando de documentos em formato HTML³, a coleta desse tipo de documento pode ser feita por uma sequência básica de comandos, de seleção⁴, cópia⁵ e colagem⁶. Para a captura de textos em formato PDF, o procedimento acima não se aplica por se tratar de um formato diferente de documento. Para essa tarefa, é preciso que o pesquisador realize de forma manual o *download* do documento, de acordo com as funcionalidades de *download* de arquivos oferecidas pelos navegadores de internet.

Dentre as questões que podem influenciar o tempo e o esforço para a elaboração manual do *corpus*, a intervenção do pesquisador ganha destaque ao estar presente em praticamente todas as fases do projeto (BAKER, 2010). Pensando em tornar essas tarefas menos onerosas, criamos uma ferramenta para fazer esse trabalho de forma mais rápida e prática. Para isso, utilizamos o poder e a flexibilidade que as linguagens de programação nos dão.

Desenvolvemos um *script* que faz o *download* de obras disponibilizadas em *sites*. Em particular, para este trabalho, foram focalizadas as obras do escritor brasileiro Machado de Assis disponíveis em domínio público em formato PDF. O *script* faz o *download* dessas obras e as converte em arquivo de texto simples TXT⁷.

2 Com base em Baker, Hardie e McEnery (2006), Oliveira (2019, p. 45) afirma que “um documento de hipertexto pode conter *links* para outros documentos e formar redes de textos. Os documentos de hipertexto estão presentes na Internet sob o formato HTML, uma derivação do Standard Generalised Markup Language (SGML)”.

3 Linguagem de marcação de hipertexto é um bloco de construção mais básico da internet.

4 Para selecionar texto em navegadores de internet, podemos usar a combinação de teclas CTRL+A.

5 Ao utilizarmos a sequência de teclas CTRL+C, realizamos a cópia de um texto já selecionado.

6 A sequência de teclas CTRL+V é utilizada para realizar o procedimento de colar o texto que esteja na área de transferência do sistema operacional.

7 É um tipo de documento que não preserva a formatação – por exemplo, *links* ou palavras em negrito, itálico, com sublinhado, imagens, cores, diferentes tipos de fonte, tabelas, listas com marcadores/números. Esses arquivos normalmente são salvos como .txt.

2 Fundamentação teórica

O trabalho aqui proposto tem como fundamentação teórica os preceitos da LC que, segundo Berber Sardinha (2004, p. 3), “[...] pode ser definida como uma área que se ocupa da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística”.

Ainda no âmbito da LC, para que um conjunto de dados seja considerado um *corpus*, existem alguns princípios e critérios que devem ser levados em consideração. Segundo Berber Sardinha (2004, p. 18-19), são eles:

- A origem: os dados devem ser autênticos;
- O propósito: o *corpus* deve ter a finalidade de ser objeto de estudo linguístico;
- A composição: o conteúdo do *corpus* deve ser criteriosamente escolhido;
- A formatação: os dados do *corpus* devem ser legíveis por computador;
- A representatividade: o *corpus* deve ser representativo de uma língua ou variedade;
- A extensão: o *corpus* deve ser vasto para ser representativo.

Aluísio e Almeida (2006, p. 157-158, grifos das autoras) discorrem sobre outras quatro características fundamentais para os *corpora*, apontadas por McEnery e Wilson (1996), sendo elas:

a) *amostragem e representatividade (sampling and representativeness)*: um *corpus* deve ter uma amostragem suficiente da língua ou variedade de língua que se quer analisar para obter-se o máximo de representatividade desta mesma língua ou variedade de língua;

b) *tamanho finito (finite size)*: com exceção de *corpus-monitor*⁸, todo *corpus* tem um tamanho finito, por exemplo: 500 mil palavras, 1 milhão de palavras, 10 milhões de palavras, etc.;

c) *formato eletrônico (machine-readable form)*: segundo McEnery e Wilson (1996), atualmente o emprego do termo *corpus* significa admitir necessariamente que os textos estejam no formato eletrônico, diferentemente da idéia que se tinha de *corpus* no passado, a qual se referia somente a textos impressos. Ainda de acordo com McEnery e Wilson (1996), o formato possui vantagens consideráveis: i) os *corpora* podem ser pesquisados e manipulados de forma mais rápida; ii) os *corpora* podem ser mais facilmente enriquecidos com informação extra;

⁸ As autoras definem *corpus-monitor* como “aquele que pode receber novos textos e tornar-se cada vez maior. É um *corpus* útil para Lexicografia, por exemplo, já que é necessário observar palavras novas na língua ou palavras já conhecidas, mas com emprego diferente” (ALUÍSIO; ALMEIDA, 2006, p. 157).

d) *referência padrão (standard reference)*: ainda de acordo com McEnery e Wilson (1996), existe um entendimento tácito de que um *corpus* constitui uma referência padrão para a variedade de língua que ele representa, pressupondo que o *corpus* esteja disponível para outros pesquisadores, em outras palavras, é o que se tem chamado de *reuso do corpus*.

Conforme pontua Dubois *et al.* (1993 *apud* ALUÍSIO; ALMEIDA, 2006, p. 157),

[...] o *corpus* não pode ser considerado como constituindo a língua, mas somente como uma amostra da língua. [...] O *corpus* deve ser representativo, isto é, deve ilustrar toda a gama das características estruturais. Poder-se-ia pensar que as dificuldades serão levantadas se um *corpus* for exaustivo [...]. Na realidade, sendo indefinido o número de enunciados possíveis, não há exaustividade verdadeira e, além disso, grandes quantidades de dados inúteis só podem complicar a pesquisa, tornando-a pesada. O linguista deve, pois, procurar obter um *corpus* realmente significativo. Enfim, o linguista deve desconfiar de tudo o que pode tornar o seu *corpus* não-representativo (método de pesquisa escolhido, anomalia que constitui a intrusão linguística, preconceito sobre a língua).

Ao estudar os conceitos apresentados nesta seção, sobre o que é um *corpus* e algumas das suas principais características, podemos chegar à conclusão que, para a construção de um *corpus*, devemos levar em consideração alguns aspectos importantes: o *corpus* deve ser representativo, os textos devem ser produzidos de forma natural, isto é, não podem ter sido produzidos para a pesquisa, e deve ser possível, a partir do *corpus*, ampliar o conhecimento sobre a língua ou variedade que ele representa.

Além disso, é importante salientar que o computador tem um papel primordial no desenvolvimento da LC. Berber Sardinha (2004, p. xvii) já afirmava que: “o computador pessoal, com memória poderosa e capacidade de armazenamento, começa a desempenhar, nas ciências humanas, o papel transformador que o telescópio teve na física e nas ciências exatas”.

Isso vem demonstrar quão necessária e importante é a utilização do computador para a realização de trabalhos na área da LC. Conforme defendido por Berber Sardinha (2004, p. 18-19), um *corpus* deve ser extenso e representativo. Partindo desse princípio, não conseguimos vislumbrar uma outra forma de analisar uma grande quantidade de textos sem o auxílio de ferramentas computacionais, visto que analisar volumes extensos de textos manualmente torna-se impraticável.

Tendo a sua importância evidenciada em outros trabalhos da área, o computador e as ferramentas advindas dele são de grande relevância para a Linguística. Trataremos neste trabalho de uma ferramenta computacional especificamente voltada à coleta e conversão

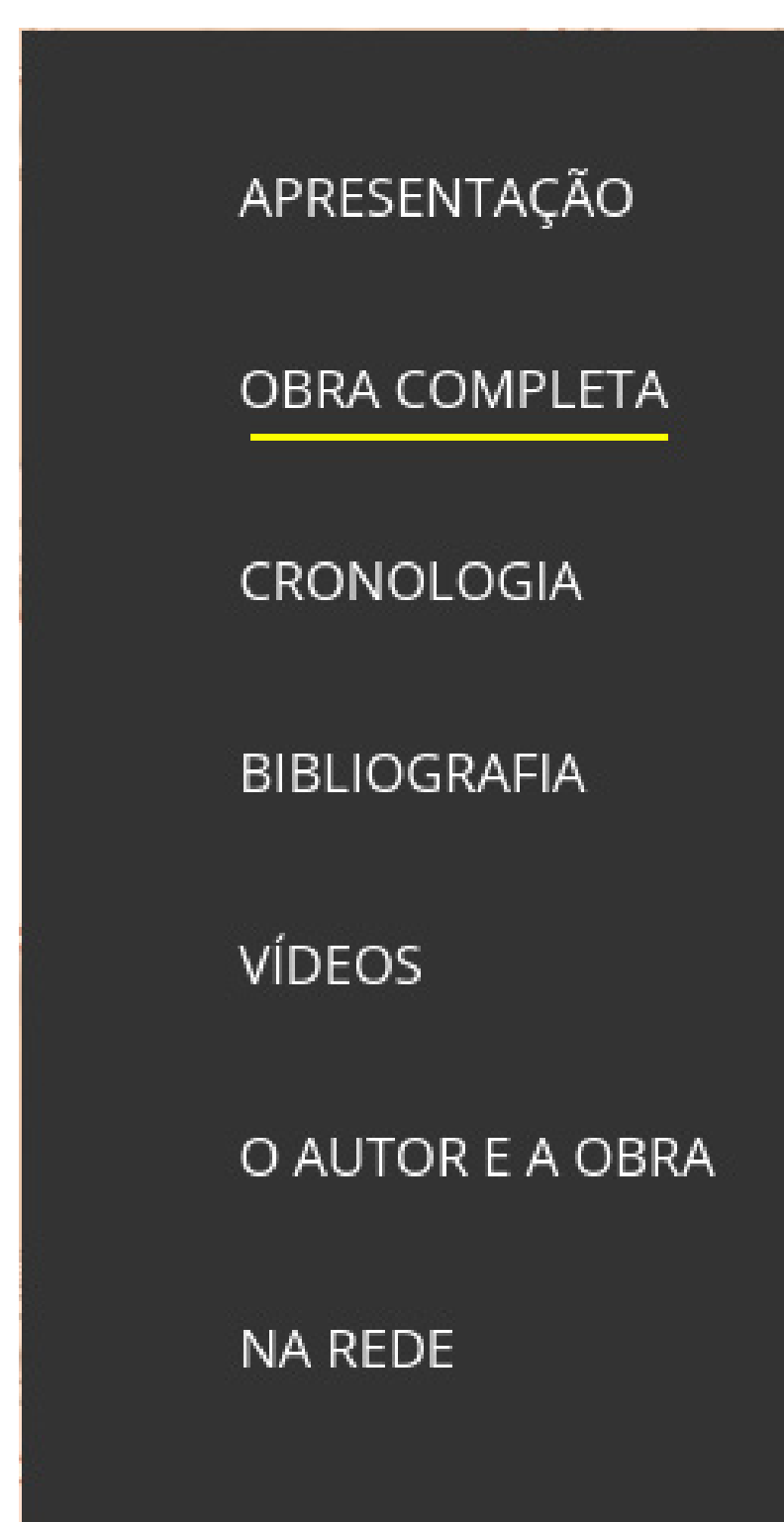
de textos disponíveis na internet. Essa ferramenta foi desenvolvida com o intuito de auxiliar e automatizar a coleta dos textos do escritor brasileiro Machado de Assis. Os textos coletados por essa ferramenta serão utilizados como base para testes da plataforma GEConWeb⁹.

3 Metodologia

Para o desenvolvimento da ferramenta, que vai automatizar a coleta dos textos de Machado de Assis, utilizamos a linguagem de programação *Python*¹⁰. Sua escolha se justifica pela sua simplicidade e grande versatilidade e por ser uma linguagem multiplataforma. Isso quer dizer que os programas desenvolvidos com ela podem ser executados nos principais sistemas operacionais da atualidade (*Linux, Windows e MacOS*). Essa característica da linguagem é importante, pois, uma vez desenvolvido, o programa poderá ser executado em qualquer computador que tenha instalado um dos sistemas operacionais listados acima. *Python* é uma linguagem muito eficiente e versátil. Com ela, é possível desenvolver desde grandes sistemas a ferramentas para automatização de tarefas. Também são criadas ferramentas que envolvem o uso de inteligência artificial.

Para darmos início ao procedimento de extração dos textos, em primeiro lugar, devemos acessar o *site* onde as obras estão hospedadas, <http://machado.mec.gov.br>. Logo na página inicial, temos um item de menu chamado “Obra completa”. Na Figura 1, ilustramos esse menu.

Figura 1: Menu principal do *site*.



Fonte: <http://machado.mec.gov.br>.

⁹ Sistema que será desenvolvido como trabalho final da minha dissertação de mestrado em andamento.

¹⁰ É uma linguagem de programação criada em 1991 por Guido Van Rossumem, com a finalidade de ser uma linguagem simples e de fácil compreensão.

Ao clicarmos nesse item do menu, somos levados para uma outra tela que contém uma lista com as categorias das obras. Cada um desses itens nos leva para uma terceira tela com listagens de todas as obras do autor de acordo com a categoria escolhida. É partir dessa listagem que iniciamos o processo de extração dos textos. A Figura 2 nos dá uma visão da tela de listagem por categorias.

Figura 2: Lista com todas as obras organizadas por categorias.



Fonte: Elaborada pelo autor.

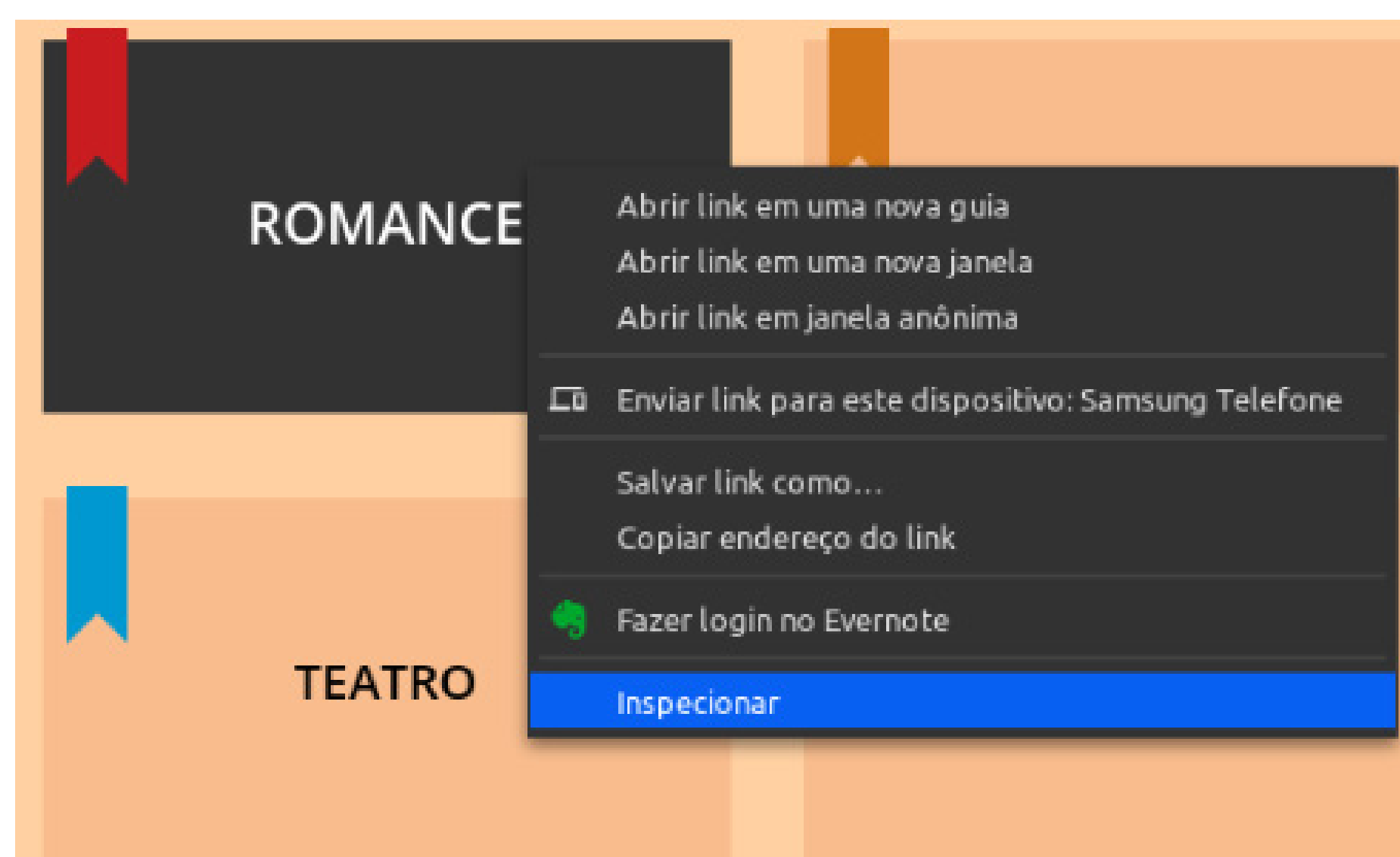
Figura 3: Definindo o endereço do site.

```
1
2 url = "http://machado.mec.gov.br/"
3 response = requests.get(url)
4
5 soup = BeautifulSoup(response.text, 'html.parser')
6
```

Fonte: Elaborada pelo autor.

Conforme pode ser visto na Figura 3, definimos como ponto de partida o endereço principal do site. A partir desse ponto, começamos a procurar os elementos na página que nos levam para uma outra página que contém as obras de acordo com a sua categoria. Precisamos inspecionar um dos elementos da listagem para entendermos como estão organizados. Quando clicamos com o botão direito do mouse e selecionamos a opção de inspecionar, conforme é mostrado na Figura 4, é aberta uma aba no navegador mostrando o código fonte do item que foi selecionado. Na figura 5, ilustramos o código que gera a listagem das categorias.

Figura 4: Inspeccionando o elemento da tela.



Fonte: Elaborado pelo autor.

Figura 5: Código fonte da tela de listagem das obras.

```
▼ <div class="row obras"> == $0
  ::before
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="col-md-3">...</div>
  ▶ <div class="button-ver-mais">...</div>
  ::after
```

Fonte: Elaborada pelo autor.

Conhecendo como foi estruturada a página, podemos começar a ler esses dados. Optamos por utilizar a biblioteca¹¹ de programação chamada *BeautifulSoup*, por ser uma das mais utilizadas no que tange à extração de código HTML. Todos os elementos estão envoltos por um elemento pai. Esse elemento pai tem uma classe de definição de estilo chamada *row obras*. Dentro de cada um desses elementos, temos um *link* que nos leva para a página das obras daquela categoria, conforme ilustra a Figura 6.

Figura 6: Detalhes do elemento HTML que compõe a lista de categorias.

```
▼ <div class="row obras"> == $0
  ::before
  ▼ <div class="col-md-3">
    ▼ <div class="obra romance">
      ▶ <div class="ribbon">...</div>
      ▶ <a href="/obra-completa-lista/itemlist/category/23-romance">...</a>
      <div class="fundo-livro"></div>
    </div>
  </div>
```

Fonte: Elaborada pelo autor.

¹¹ São trechos de código desenvolvidos para realizar uma determinada tarefa em específico.

Demonstramos, na Figura 7, o código que foi desenvolvido para percorrer todos esses elementos e extrair deles os *links* para cada uma das páginas das obras.

Figura 7: Código que percorre todos os elementos da lista de categorias.

```
1 for obras in soup.find_all(class_='row obras'):  
2     pages_of_obras = [a['href']  
3         for a in obras.find_all('a', href=True) if a.text]  
4  
5     for page in pages_of_obras:  
6         # Só interessa para nos os nomes que não terminam com obra-completa-lista  
7         if not page.endswith('/obra-completa-lista'):  
8             url_to_obras.append("http://machado.mec.gov.br" + page)  
9
```

Fonte: Elaborada pelo autor.

Conforme pode ser observado, criamos uma estrutura de repetição que percorre e encontra todos os elementos que estão dentro do elemento que tem a classe “row obras”. Para cada um que for encontrado, buscamos por um atributo chamado “href”. O valor definido nesse atributo é o endereço para a página com a lista das obras por categoria. Exemplo: o componente intitulado “romance” tem um endereço para as obras dessa natureza. Veja um exemplo na figura a seguir.

Figura 8: Lista de obras por categoria.



Fonte: Elaborada pelo autor.

Essa é a página com a listagem das obras por categoria, no caso da Figura 8, das obras presentes na categoria romance. Novamente, é necessário inspecionar um dos elementos para entendermos como a lista com as obras foi estruturada, conforme apresentado na Figura 9.

Figura 9: Lista de obras por categoria.

```
<div class="item"> == $0
  <div class="detalhes">
    <div class="titulo"> Ressurreição </div>
    <div class="detalhe ano"> 1872 </div>
  </div>
  <div class="download">
    <a href="/obra-completa-lista/item/download/20_f90f3d4964f49e34dc473155" title="Download">...</a>
  </div>
</div>
```

Fonte: Elaborada pelo autor.

Conforme se observa na Figura 9, temos dois itens importantes que compõem essa estrutura: um é o componente que está agrupando os detalhes da obra, como título e ano; e o outro é o endereço para *download* do arquivo. Precisamos extrair essas informações e com isso, compor o nome do arquivo e o endereço para *download*. Se observarmos, iremos perceber que o endereço para o arquivo não está completo, está faltando a parte inicial do endereço que nesse caso é o domínio do *site*. Na Figura 10, a seguir, ilustramos o código que foi desenvolvido para realizar essa tarefa, montar o nome e o endereço completo para *download* do arquivo.

Figura 10: Código que percorre todos os elementos da lista de obras.

```
1 for div in soup.find_all(class_='item'):
2     for detalhes in div.find_all(class_='detalhes'):
3         name = detalhes.find_all(class_='titulo')[
4             0].text.strip().replace(" ", "_")
5
6         if name.rfind('/'):
7             name = name.replace("/", "_")
8
9         if name.rfind('-'):
10            name = name.replace("-", "a")
11
12            ano = detalhes.find_all(class_='detalhe ano')[
13                0].text.strip()
14
15            if ano.rfind(' - '):
16                ano = ano.replace(' - ', '_')
17
18            filename = name + "_" + ano + ".pdf"
19
20            get_file("http://machado.mec.gov.br/"+div.find("a",
21                {"title": "Download"}).attrs['href'],
```

Fonte: Elaborada pelo autor.

O código apresentado na Figura 10 percorre a lista de obras e realiza duas tarefas importantes: acessa o elemento com a classe de estilo "detalhes" e, de dentro dela, extrai informações de autoria e ano de publicação. Esse procedimento é necessário para que possamos montar o nome da obra e utilizarmos para isso o padrão: nome e ano de publicação. Com essas informações em mãos, faremos a junção entre o nome da obra e o ano de publicação. Ao compormos o nome, realizamos a substituição dos espaços em branco que separam o nome e o ano por *underline* (por exemplo, Ressurreição_1872.pdf).

Após realizar esse processamento, chamamos a função “bloco de código”, que foi criada especificamente para realizar a tarefa de baixar o arquivo que lhe foi informado.

Figura 11: Função responsável por fazer o *download* do arquivo.



```
1 # Download file from url
2 def get_file(url, filename, category_of_obra):
3
4     with requests.get(url, stream=True) as r:
5         print("Download file: " + filename)
6
7         path_corpus = 'src/corpus/pdf/'+category_of_obra+'/'
8         if not os.path.exists(path_corpus):
9             os.makedirs(path_corpus)
10
11        with open(path_corpus+filename, 'wb', encoding='utf-8') as f:
12            shutil.copyfileobj(r.raw, f)
13            pdf2txt(filename, path_corpus, category_of_obra)
14
```

Fonte: Elaborada pelo autor.

Conforme pode ser observado na Figura 11, a função *get_file*¹² recebe três informações importantes para o seu funcionamento, que são: endereço do arquivo, nome do arquivo e a categoria à qual pertence o arquivo. A função inicia o processo de *download* do arquivo, salvando-o em uma pasta específica.

Ao final do seu processamento, a função responsável por fazer o *download* do arquivo em PDF invoca uma outra função chamada *pdf2txt*, que recebe três informações importantes: nome do arquivo, localização do *corpus* e a categoria a qual pertence o arquivo. A função busca pelo arquivo no *corpus* com base no nome e no caminho do *corpus*, e o converte em arquivo de texto, salvando o novo arquivo na pasta adequada. Na Figura 12, a seguir, temos uma ilustração da função que converte arquivos em PDF para TXT.

¹² Nome dado à função que faz o *download* do arquivo em PDF.

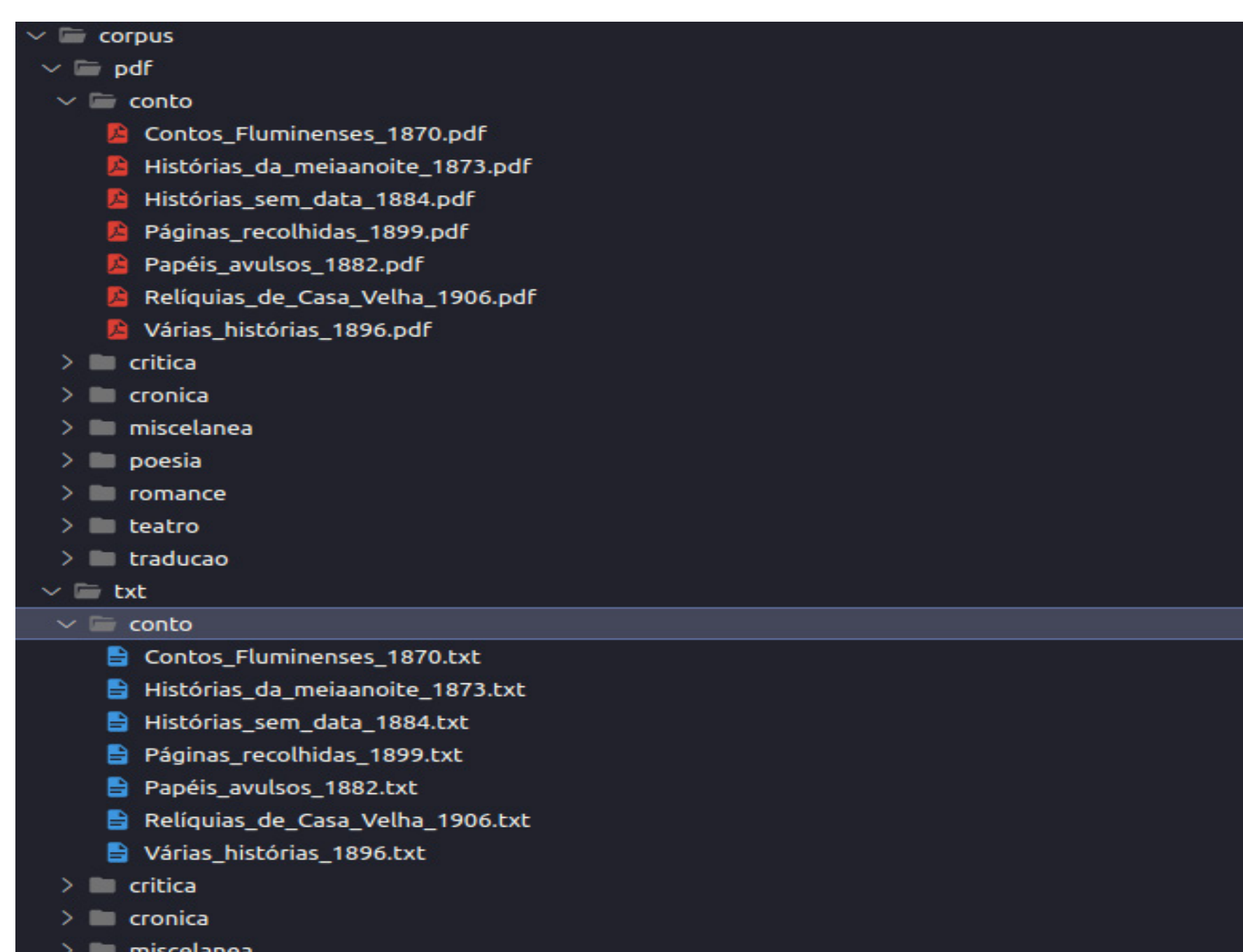
Figura 12: Função responsável por converter PDF em TXT.

```
1 def pdf2txt(filename, path_corpus, category_of_obra):
2     msg = "Converting {file} to txt: ".format(file=filename)
3     print(msg)
4
5     pdf_file_path = path_corpus+filename
6     with open(pdf_file_path, 'rb') as file:
7         pdf = pdftotext.PDF(file)
8
9     if filename.endswith('.pdf'):
10        path_corpus = 'src/corpus/txt/'+category_of_obra+'/'
11
12        if not os.path.exists(path_corpus):
13            os.makedirs(path_corpus)
14
15        new_filename = path_corpus+filename.replace('.pdf', '.txt')
16
17        with open(new_filename, 'w', encoding='utf-8') as output_file:
18            for page in pdf:
19                output_file.write(page)
```

Fonte: Elaborada pelo autor.

Segundo Berber Sardinha (2004, p. 72), “uma vez que os textos tenham sido coletados e limpos, a tarefa seguinte é a organização dos arquivos em uma estrutura coerente”. Tentando seguir essa orientação, dividimos a organização dos arquivos em uma estrutura de pastas, de modo que cada pasta indique quais textos estão armazenados dentro dela. O *corpus* ficou organizado da seguinte forma: uma pasta chamada “corpus”, onde serão armazenados todos os arquivos, e duas outras pastas, uma chamada “pdf” e outra chamada “txt”, ambas para armazenar os arquivos em PDF e TXT, respectivamente. A Figura 13 ilustra como ficou a organização do *corpus* após o *download* e conversão dos arquivos.

Figura 13: Estrutura do *corpus*.



Fonte: Elaborada pelo autor.

Foram apresentados, acima, os procedimentos utilizados para automatizar a tarefa de obtenção e conversão dos dados. A seguir, teceremos algumas considerações acerca dos objetivos traçados para este trabalho.

4 Algumas considerações

Os procedimentos descritos neste estudo demonstram o quanto as ferramentas computacionais podem facilitar o trabalho de pesquisa em LC e poupar os pesquisadores de esforços, por vezes, desnecessários. A utilização de programação para automatizar tarefas repetitivas e laboriosas – como coleta, nomeação, salvamento, conversão, limpeza e normalização de arquivos – viabiliza a economia de tempo e esforço. Fatores como a repetição e o cansaço no desempenho dessas tarefas são indícios para a busca de alternativas para deixá-las, por exemplo, a cargo do computador, cabendo ao pesquisador a tarefa de verificação dos resultados. Por meio dessa automatização, o pesquisador acaba ganhando mais tempo para se dedicar às análises e às discussões dos dados.

| Referências

- ALUÍSIO, S. M.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, São Leopoldo, v. 4, n. 3, p. 156-178, 2006. Disponível em: <http://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>. Acesso em: 28 jan. 2022.
- BAKER, P. Corpus methods in Linguistics. In: LITOSSELITI, L. (ed.). **Research methods in Linguistics**. New York: Continnum International Publishing Group, 2010. p. 93-113.
- BAKER, P.; HARDIE, A.; McENERY, T. **A glossary of Corpus Linguistics**. Edinburgh: Edinburgh University Press, 2006.
- BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.
- BIDERMAN, M. T. C. **Teoria linguística: teoria lexical e linguística computacional**. 2. ed. São Paulo: Martins Fontes, 2001.
- DUBOIS, J. *et al.* **Dicionário de Linguística**. São Paulo: Cultrix, 1993.
- McENERY, T.; WILSON, A. **Corpus Linguistics**. Edinburgh: Edinburgh University Press, 1996.
- OLIVEIRA, F. P. **ToGatherUp**: um protótipo de ferramenta para a construção de *corpora*. 2019. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2019. DOI: <http://dx.doi.org/10.14393/ufu.di.2019.679>.

Exploração do significado do lexema *medo*: uma análise pautada na Análise do Discurso e na Linguística de *Corpus*¹

Thaís dos Santos Souza²

¹ Este trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior – Brasil – CAPES – Código do financiamento 001.

² Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/1690929957314722>. E-mail: thais.souza4@ufu.br.

1 Introdução

Este estudo é um recorte de uma pesquisa de mestrado realizada no programa de pós-graduação em Criminologia da Universidade do Porto, em Portugal (SOUZA, 2018). Nosso objetivo consiste em explorar e compreender os significados atribuídos à palavra *medo* em 21 (vinte e uma) entrevistas semiestruturadas realizadas com brasileiros residentes em quatro cidades do Brasil. Especificamente, pretende-se compreender, por meio da análise do campo semântico desse lexema e de suas derivações no discurso dos entrevistados, de que modo são construídos os significados do sentimento de medo e a sua relação com os contextos físicos e sociais que o faz emergir. Sendo assim, procuramos unir duas áreas de interesse, Linguística Aplicada e Criminologia, para empreender uma abordagem transdisciplinar.

Realizamos uma análise pautada na Análise do Discurso (doravante AD), visto que ela oferece atenção aos processos de produção e interpretação linguística, caracterizadas por tensões sociais e que contemplam a dimensão da mudança social através do discurso. Além disso, utilizamos como abordagem a Linguística de *Corpus* (LC), portanto, partimos de evidências linguísticas empíricas identificadas e analisadas por meio de ferramentas computacionais, em específico a *WordList* e o *Concord* do *WordSmith Tools* 4.0 (SCOTT, 2004), programa utilizado para o processamento do *corpus* e para a análise dos dados.

2 Fundamentação teórica

O arcabouço teórico deste trabalho combina a Criminologia e a Linguística Aplicada. A Criminologia é uma ciência autônoma, empírica e interdisciplinar que estuda o crime, o criminoso, a vítima e o controle social da conduta criminosa (OLIVEIRA, 2018). Já as pesquisas em Linguística Aplicada oferecem um olhar demandado pelo interesse sócio-histórico, envolvendo-se em um processo de renarração ou redescrição da vida social, o que nos direciona à necessidade de compreendê-la.

Estudar a estrutura linguística e a estrutura social, concentrando o olhar nos efeitos de sentidos e na manipulação ideológica nos sistemas de linguagem existentes na sociedade, permite ao pesquisador explorar o que não é visível, mas, de algum modo, perceptível no convívio social, como as manifestações ideológicas que aparelham várias práticas sociais. Nessa perspectiva, o estudo dos efeitos de sentido da língua(gem) não é um estudo voltado a leis e estruturas próprias da língua, mas sim um campo interpretativo que busca o significado. Isso requer olhar para outras ciências. Como afirma Moita Lopes (2006, p. 96), “se quisermos saber sobre a linguagem e vida social nos dias de hoje, é necessário sair do campo da linguagem propriamente dita e ler sociologia, geografia, história, antropologia, psicologia cultural e social, entre outras ciências”.

Diante disso, cruzamos Linguística Aplicada e Criminologia para melhor compreender, através da análise de campo semântico, os significados e os efeitos de sentido que o lexema *medo* evoca nos indivíduos. Mas antes de adentrarmos nesse universo de significados, cabe-nos compreender o que os estudiosos da Criminologia entendem por *medo*.

2.1 Medo na ótica da Criminologia

O dicionário de Francisco Bueno (1898) afirma que *medo* é uma “inquietação que se diferencia da ansiedade, por relacionar-se a objeto determinado”. Por sua vez, o *Dicionário Online de Português* define *medo* como sendo um “estado emocional provocado pela consciência que se tem diante do perigo; aquilo que provoca essa consciência” (DICIO, 2022).

A literatura criminológica nos mostra que a designação de *medo* não é algo simples, pois devemos ampliar o olhar sobre este lexema e compreender o que está em sua volta. Na ótica criminológica, tal fenômeno reflete mudanças de rotina, aquisição de comportamentos de segurança, exclusão de grupos sociais e até mesmo de determinados locais, além do enfraquecimento e desorganização das estruturas sociais (GAROFALO, 1981; HALE, 1996; SKOGAN; MAXFIELD, 1981; FERRARO, 1995).

Estudos já realizados apontam vários fatores que contribuem para a potencialização do medo, e o crime é um fator em destaque na constituição do medo nos sujeitos. O medo, especificamente do crime, tornou-se alvo de pesquisa de grandes estudiosos, como Garofalo (1981), Hale (1996), Skogan e Maxfield (1981), Ferraro (1995), entre outros. Mas cabe ressaltar que o crime, por si só, não agrava o sentimento de medo, pois este sentimento é uma resposta ao significado atribuído a alguns contextos sociais e físicos aos quais ele é relacionado.

Os estudiosos dessa temática apontam que, ao enfrentarem uma ameaça, todos os seres vivos têm uma reação, seja de fugir ou de agredir (BAUMAN, 2008). Não obstante, focalizando somente no ser humano, Bauman (2008, p. 9), pontua que há

[...] uma espécie de medo de “segundo grau”, um medo, por assim dizer, social e culturalmente “reciclado” ou (como o chama Lagrange em seu fundamental estudo do medo) um medo “derivado” que orienta seu comportamento (tendo primeiramente reformado sua percepção do mundo e as expectativas que guiam suas escolhas comportamentais), que haja ou não uma ameaça imediatamente presente. O medo secundário pode ser visto como um rastro de uma experiência passada de enfrentamento da ameaça direta – um resquício que sobrevive ao encontro e se torna um fator importante na modelagem da conduta humana mesmo que não haja mais uma ameaça direta à vida ou à integridade.

Nessa perspectiva, o medo, enquanto componente emocional, trata-se de uma junção de reações emocionais adquiridas em experiências passadas. Assim, o medo do crime, por exemplo, é compreendido por reações emocionais adquiridas em experiências passadas condicionadas pelo crime ou por símbolos associados à criminalidade (FERRARO, 1995; FERRARO; LAGRANGE, 1987).

Por se tratar de um lexema subjetivo, o *medo* é um sentimento difícil de ser mensurado, o que leva o pesquisador a caminhos abstratos da conduta humana, pois tal sentimento é uma resposta ao significado atribuído a alguns contextos sociais e físicos aos quais ele é relacionado, como já demonstrado nos estudos de Brites (2010), Dunstan *et al.* (2005), Guedes (2016), Hunter (1978), Lourenço (2010), Souza (2018), entre outros.

2.2 Medo na perspectiva lexical

Na ótica lexical, Kalverkämper (1983, p. 126 *apud* ZILIO, 2010, p. 131) nos recorda que: “[...] assim como as unidades lexicais estão alojadas em um todo de ordem maior, nominadamente, a oração também, é uma parte completa de um todo de ordem maior e ainda mais complexo, para cuja constituição ela contribui, nominadamente, o texto”.

A partir do trecho veiculado, verifica-se que os lexemas não têm um fim em si, mas estão em orações que, por sua vez, constituem textos que representam um significado muito mais amplo que o do próprio lexema. Em vista disso, compreender, desvelar e descrever os efeitos de sentido e os significados atribuídos ao lexema *medo* em cadeias enunciativas, por meio de métodos empíricos de análise, é um exercício relevante e produtivo.

Perini (2006) ressalta que palavras servem para comunicar significados e que a possibilidade de significados comunicados por uma palavra nunca é infinita. Entretanto, a forma e o contexto no qual a palavra é usada modifica esse sentido. Por exemplo, no presente estudo, o lexema *medo* significa coisas distintas para diferentes indivíduos. Assim, para compreendermos os significados de *medo*, é necessário partirmos do próprio lexema, pois cada lexema possui um conjunto delimitado de significados, que é o que chamamos de área semântica (PERINI, 2006).

Para tanto, lançamos mão da AD, por ela se concentrar no uso amplo e geral da linguagem dentro de e entre grupos específicos de pessoas, considerando as mais diversas formas materiais de significados, além da noção de sujeito, que é aquele ser que fala no mundo (seu contexto social e cultural) e que está interpolado por uma ideologia.

Ademais, realizamos este estudo sob os pressupostos da LC, que, segundo Rajagopalan (2006), é uma vertente muito interessante, pois questiona postulados da Linguística tradicional (teórica), com base em uma perspectiva de ordem eminentemente prática. Assim, tal vertente

é uma esfera de pesquisa científica preocupada com a exploração de *corpora* textuais, ou seja, conjuntos de textos sistematicamente coletados, em formato eletrônico, que servem de objeto de exploração linguística (BERBER SARDINHA, 2004). Esses textos podem ser orais, transcritos ou escritos. O importante é que sejam de linguagem autêntica, que representem a esfera que se pretende estudar e que não tenham sido produzidos apenas com a finalidade de pertencerem a uma pesquisa. Em outras palavras, esses textos precisam refletir utilizações naturais de linguagem. Outro ponto interessante da LC é o uso do computador, que, para ela, é uma ferramenta crucial, já que é necessário ao pesquisador para compilar, armazenar, etiquetar, processar e analisar os *corpora*.

É possível testemunhar a utilidade da LC em diversos campos, como ensino de línguas, política, linguística forense, religião, negócios, entre outros. Ela será utilizada neste estudo por permitir que o pesquisador extraia evidências de uso de língua(gem) natural, e nos auxilia na compreensão de como são dados os efeitos de sentido a tudo que o rodeia, isto é, como os indivíduos usam as palavras.

Ao observarmos o lexema *medo* em utilizações linguísticas autênticas, verificamos que ele pode significar diferentes tipos de medo, por exemplo, “medo do crime”, “medo de estupro”, “medo de lugares escuros”, entre outros. Qualquer outra acepção que a palavra *medo* possa porventura ter é simplesmente tratada ou como extensão ou como desvio daquele significado, permitindo registrar os múltiplos usos dos itens que servem de verbetes.

Posto isto, passemos à apresentação do passo a passo metodológico deste estudo exploratório.

3 Metodologia

Os dados apresentados e analisados neste estudo foram extraídos de um *corpus* de 21 entrevistas semiestruturadas que foram aplicadas em diferentes localidades do Brasil (Castilho-SP; Ouro Preto-MG; Campo Grande-MS e Uberlândia-MG), entre os meses de janeiro e fevereiro de 2018. Como dito anteriormente, esse *corpus*, que contabiliza 53.023 *tokens* (palavras totais), é oriundo de uma pesquisa de mestrado em Criminologia (SOUZA, 2018).

O estudo se debruça no discurso de entrevista semiestruturada, pois, neste tipo de entrevista, os entrevistados se sentem confortáveis ao expressar seus sentimentos, experiências e interpretações em torno do tema discutido. As entrevistas seguiram um guião de perguntas que teve como modelo o adotado no estudo de Guedes (2016). As principais características da entrevista semiestruturada consistem em combinar perguntas abertas e fechadas, podendo o entrevistador improvisar e reconduzir a entrevista aos pontos de interesse. Cabe sublinhar que as questões pré-definidas do guião são seguidas através de uma conversa informal.

Destaca-se ainda que os participantes da entrevista foram selecionados a partir de um critério de heterogeneidade, pois buscou-se considerar a diversidade dos sujeitos para compreender as situações e percepções dos entrevistados acerca do significado associado ao medo, de modo a observar as semelhanças e diferenças de constituir o significado do vocábulo *medo* sob a ótica de quem já foi vítima de algum crime e de quem nunca foi vítima de crimes.

O processo de análise foi subsidiado pela abordagem da LC, utilizando o programa *WordSmith Tools 4.0* (SCOTT, 2004). Como se trata de um recorte, a transcrição das entrevistas já havia sido realizada. Isto posto, a análise ocorreu em duas etapas. A primeira consistiu na identificação e análise das ocorrências do lexema *medo* no *corpus*. Já a segunda etapa consistiu na análise do campo semântico desse lexema e o que ele significa para os entrevistados.

As ferramentas utilizadas para este estudo foram a *WordList* e o *Concord*. No primeiro momento, através da ferramenta *WordList*, foi gerada uma lista de todas as palavras que ocorrem no *corpus*, por ordem de frequência, com suas respectivas ocorrências. A Figura 1 apresenta um recorte dessa lista com o lexema *medo* em destaque.

Figura 1: Lista de palavras com o lexema *medo* destacado.

N	Word	Freq.	%	Texts	%	emmas	Set
23	DO	369	0,68	2	100,00		
24	NO	344	0,63	2	100,00		
25	GENTE	337	0,62	1	50,00		
26	PRA	324	0,60	1	50,00		
27	ACHO	315	0,58	1	50,00		
28	MEDO	313	0,58	1	50,00		
29	AQUI	311	0,57	1	50,00		
30	MUITO	296	0,55	1	50,00		
31	LÁ	288	0,53	1	50,00		
32	PORQUE	278	0,51	1	50,00		
33	ME	276	0,51	1	50,00		
34	AS	263	0,48	2	100,00		
35	CASA	254	0,47	1	50,00		
36	PESSOAS	246	0,45	1	50,00		
37	ENTÃO	240	0,44	1	50,00		
38	ISSO	232	0,43	1	50,00		
39	QUANDO	227	0,42	1	50,00		
40	BEM	219	0,40	1	50,00		
41	JÁ	217	0,40	1	50,00		
42	LUGAR	211	0,39	1	50,00		
43	PARA	210	0,39	1	50,00		
44	ELA	207	0,38	1	50,00		
45	OU	206	0,38	1	50,00		

Fonte: *WordList*.

Em seguida, foi realizada uma busca contextualizada do vocábulo *medo*, utilizando a ferramenta *Concord*, visando incluir as ocorrências de suas derivações e dos contextos no qual o lexema *medo* está inserido. No entanto, sublinhamos que as anotações dos dados foram feitas manualmente. A Figura 2 apresenta um recorte das linhas de concordância do lexema *medo*.

Excerto 1: “Então o que me remete medo é andar sozinho a noite... você está totalmente vulnerável nessas situações”.

Aqui observamos que “andar sozinho” está ligado ao lexema *medo*, portanto, compondo seu campo semântico.

Figura 3: WordList com o processo **SOZINHO** destacado e suas derivações.

N	Word	Freq.	%	Texts	% emmas	Set
3.391	SOMBRA	1		1	100,00	
3.392	SOMENTE	3		1	100,00	
3.393	SOMOS	5		1	100,00	
3.394	SONOLENTO	2		1	100,00	
3.395	SORTE	5		1	100,00	
3.396	SORVETERIA	1		1	100,00	
3.397	SOSSEGADINHO	1		1	100,00	
3.398	SOSSEGADO	1		1	100,00	
3.399	SOU	19	0,04	1	100,00	
3.400	SOZINHA	51	0,10	1	100,00	
3.401	SOZINHAS	2		1	100,00	
3.402	SOZINHO	7	0,01	1	100,00	
3.403	SOZINHOS	1		1	100,00	
3.404	SPRAY	3		1	100,00	
3.405	SUA	192	0,36	1	100,00	
3.406	SUANDO	1		1	100,00	

Fonte: WordList.

Outro aspecto observado no *corpus* que é importante na significação do medo é a familiaridade com os locais. De acordo com os enunciados das entrevistas, conhecer o lugar faz com que o sujeito se sinta mais seguro, o que nos leva a compreender que **lugares desconhecidos** causam medo. Vejamos a seguir dois fragmentos dos discursos de dois entrevistados sobre o assunto:

Figura 4: Linhas de concordância do processo **LOCAL**.

22	se essas pessoas tivessem em outro local também não frequentaria esse	9.088
23	poderia falar um pouco do local dessa sua antiga casa?	6.700
24	acontece aqui. É mais por questão do local. Eu moro aqui, então, as coisas	5.775
25	ENTREVISTADO: Se eu conhecer o local eu não tenho medo.	5.112
26	falei aí é pior. ENTREVISTADOR: E o local é nessas características que falou	5.012
27	primeiro é o local, mas se tiver o local com esses tipo de pessoa que falei	4.998

Fonte: Concord.

Excerto 2: “Se eu conhecer o local eu não tenho medo... eu conhecendo o lugar diminui o medo, porque você conhece as pessoas que normalmente frequentam e as pessoas que frequentam te conhece”.

(Homem, 26 anos, não vítima, Castilho)

Além disso, verificamos um elemento fundamental ligado ao lexema *medo*, que é o **desconhecido**. Uma representação nítida desta compreensão está no excerto a seguir, no

qual *medo* está ligado aos contextos de familiaridade e desconhecimento do local. A partir do campo semântico, os sujeitos demonstram resistência em sair de suas zonas de conforto, tendo em vista que tais zonas trazem sensação de controle e equilíbrio. O desconhecimento, por sua vez, não proporciona essas sensações, pelo contrário, a falta de controle acarreta o aumento do medo nos indivíduos.

Excerto 3: *“Aqui em casa, as vezes passa gente estranha que não é a da vizinhança, mas parece que me sinto mais segura aqui pra cima. Agora umas duas ruas pra baixo já me sinto insegura. Não sei se é porque eu já conheço aqui a rua e ali pra cima e não conheço muito ali pra baixo, não sei... Acho que o fato deu conhecer, de ter familiaridade com o lugar, de conhecer quem vive ali me deixa mais segura”.*

(Mulher, 39 anos, vítima, Uberlândia)

Cabe esclarecer que o desconhecimento não está ligado somente ao local, mas também às pessoas. Além disso, como bem demonstrado no excerto a seguir, o contexto **noite** é uma extensão dada a *medo*, pois o fato de estar de noite ressignifica o sentimento de medo para este entrevistado:

Excerto 4: *“sinto mais medo à noite do que de dia. Mas tem sua ressalva. Dependendo do lugar e tiver umas pessoas estranhas na rua, aí já fico mais atento”.*

Figura 5: WordList com o processo **NOITE** destacado.

N	Word	Freq.	%	Texts	%
2.451	NINGUÉM	42	0,08	1	100,00
2.452	NISSO	3		1	100,00
2.453	NÍVEL	1		1	100,00
2.454	NO	343	0,64	1	100,00
2.455	NOÇÃO	1		1	100,00
2.456	NOIA	3		1	100,00
2.457	NOITE	197	0,37	1	100,00
2.458	NOME	2		1	100,00
2.459	NORA	1		1	100,00
2.460	NORMA	1		1	100,00

Fonte: WordList

Percebe-se, inclusive, que o **estranho**, também aludido como desconhecido, é caracterizado como ameaçador, aumentando o sentimento de medo. Verificamos ainda que a **vestimenta** também está associada à manifestação do medo nos sujeitos. Os discursos de alguns sujeitos reportam a vestimenta para caracterizar que se sentem inseguros. A seguir, apresentamos as linhas de concordância para “boné” (Figura 6) e dois excertos que corroboram o aspecto mencionado.

Figura 6: Linhas de concordância do processo **BONÉ**.

N	Concordance	Set	Tag	Word #	Sent. #	os.	ara. #	os.	#	os.	t. #	os.	File	%
1	generalizando, não é toda pessoa de boné de aba reta, de bermuda, camisa			48.813	3.391	0%	0	0%	0	0%	0	0%	02_merged.txt	92%
2	Geralmente uma pessoa magra, com boné de aba reta, geralmente de tricô,			44.145	3.093	0%	0	0%	0	0%	0	0%	02_merged.txt	84%
3	frio, está calor. Pessoas que usam boné aba reta. Se bem que hoje em dia			41.228	2.904	1%	0	4%	0	4%	0	4%	02_merged.txt	78%
4	com casacos, moletom no calor, de boné ou com capuz, normalmente			35.538	2.488	4%	0	3%	0	3%	0	3%	02_merged.txt	67%
5	adolescentes mais aglomerado, com boné de aba reta, ai fico com um certo			29.784	2.092	9%	0	1%	0	1%	0	1%	02_merged.txt	57%
6	com aquelas roupas largas, com boné aba reta, você fica meio assim,			24.901	1.779	7%	0	1%	0	1%	0	1%	02_merged.txt	47%
7	um rapaz só porque está com um boné de aba reta. Mas sim, para			18.300	1.293	2%	0	7%	0	7%	0	7%	02_merged.txt	35%
8	por exemplo, as pessoas que usam boné aba reta, que utilizam roupas, tipo			16.195	1.188	3%	0	3%	0	3%	0	3%	02_merged.txt	31%
9	fazer tanto mal como uma de brinco e boné pra traz. Pra mim não faz diferença			8.670	688	5%	0	8%	0	8%	0	8%	02_merged.txt	17%
10	suspeito, que usam roupas largas, boné aba reta, chinelo. Evito estar			2.856	220	0%	0	6%	0	6%	0	6%	02_merged.txt	6%
11	uma pessoa que ela anda sem camisa, boné de aba reta, cheia de tatuagem.			2.508	199	1%	0	5%	0	5%	0	5%	02_merged.txt	5%

Fonte: Concord.

Excerto 5: *“dependendo do que está vestindo, tipo, se estiver de dia e a pessoa está de blusa de frio e nem está frio, está calor. Pessoas que usam boné aba reta... é o que me dá mais medo, um certo receio”.*

(Mulher, 26 anos, vítima, Uberlândia)

Excerto 6: *“não é preconceito, mas você vendo uma pessoa que ela anda sem camisa, boné de aba reta, cheia de tatuagem. Infelizmente é um paradigma a ser quebrado? É, mas você vê uma pessoa assim, fica inseguro”.*

(Mulher, 24 anos, não vítima, Castilho)

Reparem que, neste último excerto, a entrevistada aponta um tipo de vestimenta específico e mencionou que seria uma vestimenta que está estereotipada na sociedade, o que indica que há um perfil traçado pela sociedade com base nesses trajes, pois *“todo o bandido tem essas características”*, como explica um dos entrevistados.

Vinculado à vestimenta, outro aspecto conectado ao medo é a **atitude**:

Excerto 7: *“Eu acho que é a atitude da pessoa que faz eu sentir medo. Independente da roupa, do lugar. É a atitude, a forma como ela te aborda, que chega em você, porque as vezes nem sempre a roupa e nem o espaço influencia... então eu acho que é a atitude da pessoa que faz eu me sentir com medo”.*

(Mulher, 50 anos, não vítima, Uberlândia)

Além dos fatores já discutidos, a **toxicodependência** também foi apontada nos discursos como ameaçadora, como explanado na fala de um dos entrevistados: *“Na esquina aqui de casa tem uns maconheiros, a gente fica com medo né”*. Outra figura constante nas falas dos sujeitos é o **motoqueiro**, como demonstrado na figura abaixo:

Figura 7: Lista de palavras com o lexema *motoqueiro* e suas derivações em destaque.

N	Word	Freq.	%	Texts	%
2.374	MOSTROU	2		1	100,00
2.375	MOTIVO	4		1	100,00
2.376	MOTO	14	0,03	1	100,00
2.377	MOTOQUEIRO	8	0,02	1	100,00
2.378	MOTOQUEIROS	5		1	100,00
2.379	MOTOS	1		1	100,00

Fonte: *WordList*.

Figura 8: Linhas de concordância do processo **MOTOQUEIROS**.

N	Concordance	Set	Tag	Word #	Sent. #	os.	. #	a. Pos.	. #	os.	t. #	t. Pos.	File	%
1	tem o terceiro que é a questão dos motoqueiros . Esses sim eu já vejo que			49.120	3.403	0%	0	100%	0	100%	02_merged.txt		93%	
2	da escola com duas amigas e dois motoqueiros param, um ficou na moto e			43.516	3.051	3%	0	89%	0	89%	02_merged.txt		82%	
3	há muitos casos em que ocorre por motoqueiros , fico com medo quando vejo			32.397	2.274	0%	0	66%	0	66%	02_merged.txt		61%	
4	os crimes estão acontecendo muito por motoqueiros . Ai quando passa um			26.982	1.929	0%	0	55%	0	55%	02_merged.txt		51%	
5	quando eu estou na rua e vêem uns motoqueiros doido, muito rápido. A			538	41	2%	0	1%	0	1%	02_merged.txt		1%	

Fonte: *Concord*.

Os entrevistados apontam que o motoqueiro é uma figura marcante para sua sensação de medo:

Excerto 8: “fico com medo quando eu estou na rua e vêem uns motoqueiros doido, muito rápido. A gente fica com medo, temos que ficar esperta quando vemos motoqueiro vindo em nossa direção, porque eles abordam mesmo”.

Tais explanações estão ligadas à percepção de que muitos crimes são cometidos por motoqueiros:

Excerto 9: “Aqui a gente fica mais assim é com motoqueiros, porque os relatos são que os crimes estão acontecendo muito por motoqueiros” / “tem o motoqueiro, que como há muitos casos em que ocorre por motoqueiros, fico com medo quando vejo um”.

Conforme observamos, a construção do significado de medo para os entrevistados envolve alguns contextos sociais e físicos que são apresentados no campo semântico dos discursos, que por vezes estão enraizados na sociedade e no próprio indivíduo. Vimos que algumas palavras e expressões são utilizadas para caracterizar e ressignificar o sentimento de medo e acabam compondo o campo semântico de *medo*.

5 Considerações finais

O presente estudo, situado na interface entre Linguística Aplicada e Criminologia, procurou explorar os significados atribuídos ao lexema *medo*, de modo a compreender o que é sentir medo para os sujeitos cujas entrevistas transcritas compuseram nosso *corpus* de estudo. Foram diversos os lexemas e as expressões utilizados pelos entrevistados para descreverem suas experiências de medo.

Verificamos também que não há uma preocupação entre os entrevistados em diferenciar do medo os estados de alerta e de insegurança, portanto, tanto alerta como insegurança são designados como medo pelos sujeitos. Identificamos que as experiências de medo foram reportadas pelos seguintes lexemas e expressões: 'ficar sozinho', 'noite/escuro', 'familiaridade', 'desconhecimento', 'vestimenta', 'toxicodependência', 'motoqueiro' e 'atitude'. Esse significado dado ao lexema *medo* é fruto da manipulação ideológica nos sistemas de língua(gem).

Referências

BAUMAN, Z. **Medo líquido**. Tradução Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2008.

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

BEVILACQUA, C. R. Por que e para que a Linguística de Corpus na Terminologia. In: TAGNIN, S. E. O.; BEVILACQUA, C. R. (org.). **Corpora na terminologia**. São Paulo: Hub Editorial, 2013. p. 11-27.

BRITES, J. A. Percepção de risco e medo do crime na caracterização do espaço físico e social. **Psychologica**, Coimbra, v. 52, n. 1, p. 315-325, 2010. DOI: https://doi.org/10.14195/1647-8606_52-1_15.

BUENO, F. S. **Dicionário escolar da língua portuguesa**. 11. ed. Rio de Janeiro: FENAME, 1898.

DICIO. Dicionário Online de Português. Disponível em: <https://www.dicio.com.br/medo/>. Acesso em: 5 dez. 2021.

DUNSTAN, F. *et al.* An observation tool to assist with the assessment of urban residential environments. **Journal of Environmental Psychology**, [S. l.], v. 25, n. 3, p. 293-305, 2005. DOI: <https://doi.org/10.1016/j.jenvp.2005.07.004>.

FERRARO, K. **Fear of crime**: interpreting victimization risk. New York: State University of New York Press, 1995.

FERRARO, K.; LAGRANGE, R. The measurement of fear of crime. **Sociological Inquiry**, Syracuse, v. 57, n. 1, p. 70-97, 1987. DOI: <https://doi.org/10.1111/j.1475-682X.1987.tb01181.x>.

GAROFALO, J. The fear of crime: causes and consequences. **Journal of Criminal Law & Criminology**, v. 72, n. 2, p. 839-857, 1981. DOI: <https://doi.org/10.2307/1143018>.

GUEDES, I. M. E. S. **Medo do Crime**: emergência, reações emocionais e discursos. Contributos para a utilização de multi-metodologias. 2016. Tese (Doutoramento em Criminologia) – Faculdade de Direito, Universidade do Porto, Porto, 2016.

HALE, C. Fear of crime: a review of the literature. **International Review of Victimology**, v. 4, n. 2, p. 79-150, 1996. DOI: <https://doi.org/10.1177%2F026975809600400201>.

HUNTER, A. **Symbols of incivility**: social disorder and fear of crime in urban neighborhoods. *In*: ANNUAL MEETING OF THE AMERICAN SOCIETY OF CRIMINOLOGY, 8-12 November, Dallas, 1978.

LOURENÇO, N. Cidades e sentimento de insegurança: violência urbana ou insegurança urbana. *In*: PEREIRA JUNIOR, E. A.; SILVA, J. F.; MARON, J. (org.). **Um toque de qualidade**: eficiência e qualidade na gestão da defesa social. Belo Horizonte: Fundação Guimarães Rosa, 2010. p. 15-39.

MOITA LOPES, L. P. Linguística aplicada e vida contemporânea: problematização dos construtos que têm orientado a pesquisa. *In*: MOITA LOPES, L. P. (org.). **Por uma linguística aplicada indisciplinar**. São Paulo: Parábola Editorial, 2006. p. 85-104.

OLIVEIRA, N. A. **Criminologia**. Salvador: Editoria Juspodivm, 2018.

PERINI, M. **Princípios da linguística descritiva**: introdução ao pensamento gramatical. São Paulo: Parábola Editorial, 2006.

RAJAGOPALAN, K. Repensar o papel da linguística aplicada. *In*: MOITA LOPES, L. P. (org.). **Por uma linguística aplicada indisciplinar**. São Paulo: Parábola Editorial, 2006. p. 149-166.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 31 out. 2021.

SKOGAN, W.; MAXFIELD, M. **Coping with crime**: individual and neighborhood reactions. Beverly Hills: Sage Publications, 1981.

SOUZA, T. S. **Exploração do sentimento de insegurança no Brasil a partir de uma abordagem qualitativa**. 2018. Dissertação (Mestrado em Criminologia) – Faculdade de Direito, Universidade do Porto, Porto, 2018. Disponível em: <https://hdl.handle.net/10216/117298>. Acesso em: 12 jan. 2022.

ZILIO, L. Terminologia Textual e Linguística de Corpus: estudo em parceria. *In*: PERNA, C. L.; DELGADO, H. K.; FINATTO, M. J. (org.). **Linguagens especializadas em corpora**: modos de dizer e interfaces de pesquisa. Porto Alegre: EDIPUCRS, 2010. p. 128-151.

Linguística de *Corpus* e C-ORAL-BRASIL: análise da fala espontânea em Belo Horizonte-MG

Maria de Oliveira Rodrigues¹

¹ Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia. Lattes: <http://lattes.cnpq.br/5030844013340663>. E-mail: maria.rodrigues5@ufu.br.

1 Introdução

Como é sabido, o português é língua oficial no Brasil, em Portugal, em outras nações da África (Angola, Cabo Verde, Guiné-Bissau, Guiné Equatorial, Moçambique e São Tomé e Príncipe), da Oceania (Timor Leste), bem como na Região Administrativa de Macau, na Ásia. Segundo Perini (2016), a língua padrão escrita nessas diferentes localidades é relativamente uniforme, possuindo apenas algumas diferenças de ortografia e estrutura gramatical. Porém, o uso do vocabulário regional pode ocasionar certas dificuldades de compreensão no contato entre pessoas oriundas desses diferentes locais, haja vista que há muitas variações no âmbito da língua falada.

Ainda de acordo com o autor, a fala dos africanos escolarizados aproxima-se do padrão de Portugal. Já no Brasil, onde grande parte da população fala somente o português, a língua vem evoluindo e sofrendo influências de línguas indígenas, africanas e de imigrantes. Fato que essas influências provocam variações na língua, despertando o interesse de pesquisadores, que tomam tais variações como objeto de estudo, como é o caso do estudo apresentado neste capítulo.

Ao adentrarmos nesse campo da linguagem, torna-se relevante abordarmos o termo *norma*. Para tal, apresentamos as proposições de Neves (2017). Para a autora, *norma* é vista como a modalidade linguística “normal”, “comum”, mas também é entendida como uso regrado, como a modalidade “sabida” por alguns falantes e não por outros. Na primeira acepção, seria estabelecida como a frequência de uso, sem que haja valoração, a noção de *norma* se divide por estratos sociais, por períodos e por regiões. Abstrai-se por essa modalidade a frequência e o uso. Já na segunda acepção, *norma* entendida com a inclusão da modalidade “padrão” no uso linguístico, também se divide em diastrática, diacrônica e diatópica², porém com juízo de valor em relação às modalidades, ou seja, umas são mais prestigiadas do que outras.

Nessas duas concepções, o termo *norma* se insere na sociedade. Na primeira acepção, “o que está em questão é o uso, e, então, a relação com a sociedade aponta para a aglutinação social. Na segunda, trata-se do bom uso, e a relação com a sociedade aponta para a discriminação, criando-se, por aí, estigmas e exclusões” (NEVES, 2017, p. 43).

Nessa perspectiva, este estudo trata de uma investigação da variação da concordância verbal, especificamente no recorte diatópico, isto é, de Belo Horizonte, MG. O *corpus* C-ORAL-BRASIL foi a fonte que permitiu a extração do material utilizado em nossas análises. Os postulados da Linguística de *Corpus* (LC) que ancoram este estudo apontam os caminhos trilhados para a

² Vale ressaltar que variação diastrática se refere ao modo de falar de diferentes classes sociais, isto é, por estratos sociais. Por diacrônica entende-se a variação da língua por períodos de tempo; já em se tratando de diatópica refere-se à variação de uso da língua por regiões.

seleção, análise e descrição dos dados, assegurando, dessa forma, a possibilidade de selecionar e analisar dados oriundos de *corpora* orais. Sendo assim, foram selecionados e analisados 9 textos em formato .txt de conversação pública, totalizando 13.493 palavras. As análises foram realizadas com o auxílio do programa *WordSmith Tools* 4.0 (SCOTT, 2004), doravante WST. A escolha por esta versão se justifica pelo fato de esta se encontrar disponibilizada gratuitamente no *site* do programa³.

Com base nas proposições delineadas acima, apresentamos as seguintes questões de pesquisa: (i) que fatos linguísticos relativos à *norma* da diatopia mineira podem ser identificados com o auxílio das ferramentas *WordList* e *Concord*? (ii) que fenômenos são resultantes desses usos? Ao longo deste trabalho, buscaremos responder aos questionamentos levantados.

Este capítulo está organizado da seguinte forma: (i) apresentamos as bases teóricas que ancoram este estudo; (ii) introduzimos o WST, programa utilizado para nossas análises; (iii) apresentamos brevemente o C-ORAL-BRASIL; (iv) descrevemos os procedimentos metodológicos deste estudo; (v) apresentamos nossas análises, discussões e resultados alcançados; e, por fim, (vi) tecemos nossas considerações finais.

2 Bases teóricas

Este trabalho está pautado em autores que tratam da LC, como Berber Sardinha (2004, 2005, 2009), em autores que apresentam o C-ORAL-BRASIL, *corpora* utilizados nesta pesquisa, os quais apontam as características do material utilizado para análise e que será elemento principal deste estudo e, futuramente, irá compor os dados utilizados em nossa tese de doutoramento.

Serão abordados posicionamentos relacionados à concordância verbal na ótica de Cunha e Cintra (2007), por avaliarmos a importância de se estabelecer parâmetros que guiarão a concepção de língua conforme a norma padrão ou também denominada norma culta e com considerações acerca da gramática descritiva na visão de Perini (2016). A escolha por esse autor justifica-se uma vez que serão apresentadas e descritas “porções de linguagem” (BERBER SARDINHA, 2004) capturadas em situações reais de conversação informal e em contextos variados.

Para Cunha e Cintra (2007, p. 510), a concordância verbal pode ser assim definida:

1. A solidariedade entre verbo e o sujeito, que ele faz viver no tempo, exterioriza-se na CONCORDÂNCIA, isto é, na variabilidade do verbo para conformar-se ao número e à pessoa do sujeito.

³ Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 29 jan. 2022.

2. A CONCORDÂNCIA evita a repetição do sujeito, que pode ser indicada pela flexão verbal a ele ajustada:

Eu acabei por adormecer no regaço de minha tia. Quando **acordei**, já era tarde, não **vi** meu pai.
(A. Ribeiro, *CRG*, 257.)

Conforme preceituam os referidos autores, há regras gerais para estabelecimento da concordância verbal, são elas: i) quando ocorrem com um só sujeito ou ii) com mais de um sujeito.

i) Neste caso, o verbo concorda em número e pessoa com seu respectivo sujeito, que pode vir explícito ou de modo subentendido;

ii) de acordo com essa outra regra, o verbo que tem mais de um sujeito (sujeito composto) vai para o plural e dependendo da pessoa, irá: para a 1ª pessoa do plural, se entre os sujeitos figurar um da 1ª pessoa. E, para a 2ª pessoa do plural, se, não existindo sujeito da 1ª pessoa, houver um da 2ª. Assim como exemplificado pelos autores: **Tu ou os teus filhos vereis** a revolução dos espíritos e costumes. (C. Castelo Branco, *J, I*, 21.). E por fim, para a 3ª pessoa do plural, se os sujeitos forem da 3ª pessoa.

Em se tratando de Perini (2016), o autor enfoca a descrição da língua falada padrão, que, segundo ele, é uma variedade altamente uniforme e socialmente aceita em todo o país; ou seja, o autor descreve as construções baseadas na norma urbana e não na norma preconizada pelas gramáticas e ensinada nas escolas. Nesta perspectiva, Perini (2016) postula que “erros de concordância” seriam conforme os seus exemplos que constam na *Gramática Descritiva do Português Brasileiro* que se lê a seguir.

[21] *Eu chegou ontem de Campo Grande.

[22] *A Virgínia fizemos um bolo de chocolate.

Para nós, o que está errado nessas frases é que elas dão indicação contraditória sobre o ocupante de um dos papéis temáticos. No caso de [21], o Tema (quem chega de Campo Grande) é indicado como sendo “eu” pelo sujeito *eu*, mas como sendo uma pessoa diferente de “eu” pelo sufixo; e em [22] o Agente de *fazer* seria “a Virgínia” segundo o sujeito, mas seria “nós” segundo o sufixo. O resultado, nos dois casos, é que a frase é semanticamente malformada, e por isso é inaceitável (PERINI, 2016, p. 100).

É possível perceber, por meio dos exemplos apresentados, a diferença de concepção acerca da concordância verbal pelos autores das gramáticas tradicional e descritiva aqui referidas.

No que diz respeito à LC, pode-se dizer que “é um campo que se dedica à criação e análise de *corpora* (plural latim de *corpus*)” (BERBER SARDINHA, 2009, p. 6), em que os conjuntos de textos e transcrições de fala encontram-se armazenadas em arquivos de computador. Dessa forma, o campo da informática é fator crucial para o estabelecimento da LC contemporânea.

Um dos grandes agentes dessa revolução é a informática; sem ela, a Lingüística de Corpus contemporânea não poderia existir. Assim, o lingüista de corpus depende de programas de computador para lidar com *corpora*. Dentre os vários *softwares* que existem para auxiliar o lingüista de corpus, um deles se destaca: WordSmith Tools (BERBER SARDINHA, 2009, p. 6).

Sendo assim, esta investigação contará com esse conjunto de ferramentas integradas (‘suíte’) destinado à análise linguística. Ainda ancorando em Berber Sardinha (2004, p. 38), esse estudo é de base empírica e se volta à análise de padrões reais de uso em textos naturais e, ainda, é um tipo de pesquisa que se encaixa em outra característica no que se refere ao paradigma informacional baseado em concordâncias.

Na seção seguinte, serão tecidas considerações a respeito deste conjunto de ferramentas integradas utilizado neste estudo.

2.1 WordSmith Tools

Conforme Scott (2015), é possível verificar a ocorrência de uma lista de palavras ou *clusters* de palavras em textos por meio da ferramenta *WordList* e localizar palavras dispostas em ordem alfabética ou por frequência. Ou seja, por meio desse tipo de busca, o usuário consegue saber de quantas palavras são compostos os textos analisados e, ainda, quantas vezes aparece cada palavra que o compõe. Ainda é possível realizar linhas de concordância de palavras utilizando a ferramenta *Concord*.

No que diz respeito ao *Concord*, buscas detalhadas são possíveis de serem realizadas, pois permitem ao usuário ver qualquer palavra ou frase no contexto. E, assim, prosseguir com suas respectivas análises. Outra ferramenta que constitui o programa é a *KeyWords*, que tem a função de encontrar palavras-chave, ou melhor, palavras que são chave em vários textos, segundo Berber Sardinha (2009). Para acionar essa funcionalidade, faz-se necessário ter um conjunto de palavras de referência, pois a ferramenta compara palavras entre textos. Conforme recomenda Berber Sardinha (2005), deve compor-se um *corpus* de referência até 5 vezes maior do que o *corpus* de estudo. Desse modo, para que uma palavra seja considerada chave, ou melhor, ter chavicidade, é preciso que sua frequência relativa seja maior no *corpus* de estudo do que no de referência.

À guisa de conclusão, o WST é um programa que apresenta certa facilidade de compreensão de sua interface e a possibilidade de salvar os resultados em diversos formatos é uma de suas vantagens. Desse modo, pode-se retornar aos arquivos salvos para futuras consultas, o que nem sempre é possível utilizando outros *softwares*.

2.2 C-ORAL-BRASIL: fonte desta pesquisa

No que diz respeito ao C-ORAL-BRASIL, pode-se afirmar que se constitui como um *corpus* de fala do português do Brasil, com ênfase da diatopia mineira, com foco na capital belo-horizontina. Segundo Raso e Mello (2012), os textos que compõem o *corpus* foram obtidos por meio de equipamentos sofisticados com o intuito de garantir a qualidade do material coletado. A coleta ocorreu entre 2006 e 2011, portanto, os autores destacam os anos que compreendem 2008 e 2010 para obtenção dos dados.

Considera-se relevante mensurar que o *corpus* em pauta se encontra alinhado por enunciados; o que, por sua vez, torna-se um aspecto fundamental, conforme Raso (2012, p. 58): “De fato, um *corpus* não alinhado e que assim não permite a correspondência do som com o texto (e eventualmente o espectrograma, como nesse caso, ou a imagem) não pode ser explorado corretamente”. A autora acredita que é imprescindível estudar a fala com consultas constantes às informações acústicas, postulando que, “sem se verificar sobre elas qualquer interpretação e sem categorias linguísticas formadas a partir da observação da modalidade que se quer estudar, e não a partir de outra modalidade” (RASO, 2012, p. 58), tal estudo não é possível.

Diante da abordagem em relação à fala espontânea, convém salientar que essa é planejada enquanto é executada, isto é, uma fala que não realiza um texto planejado anteriormente, conforme ratifica Raso (2012). De acordo com os autores da obra C-ORAL-BRASIL (2012), as situações em que se deu a captura da fala espontânea ocorreram em caráter informal e em contextos diversificados, com o objetivo de aproximar ao máximo da representação de um universo aberto. Afirmam, ainda, que essa seleção foi realizada por não existir um contexto que seja mais típico do que outros, do mesmo modo que acontece quando diz respeito aos contextos formais em que é possível identificar os principais domínios da fala formal.

3 Procedimentos metodológicos

O primeiro procedimento metodológico foi a seleção do *corpus* de estudo, conforme apontado na Introdução. Foram utilizados nove textos em formato txt, provenientes do *corpus* C-ORAL-BRASIL, com o intuito de verificar, neste estudo, como se dá o processo de concordância verbal de falantes da capital mineira. E, ainda, com a finalidade de se realizar algumas descrições

das amostras selecionadas a partir dos seguintes critérios: i) maior incidência de verbos nos textos, a partir de análise impressionística, conforme Berber Sardinha (2009); ii) formas não-padrão e paradigmas reduzidos, por exemplo, e *ficamo* (= ficamos; verbo); iii) Formas dos verbos *estar*, *ir*, por exemplo, *tamo* (= estamos), *tamos* (= estamos), *tão* (= estão), *tar* (= estar), *taria* (= estaria), *tás* (= estás), *tava* (= estava), *tavam* (= estavam), *távamos* (= estávamos), *vamo* (= vamos), *vão* (= vamos).

Após definidos os critérios, partiu-se para o uso do programa *WST* versão 4.0 propriamente dita. A proposta foi a de utilizar as ferramentas *WordList* e *Concord*. Em se tratando da primeira ferramenta, seu emprego destinou-se ao levantamento da quantidade de *tokens* presentes nos textos, isto é, com o intuito de verificar a extensão do *corpus* de estudo. E, ainda, para verificar as palavras que se enquadravam nas formas verbais apontadas anteriormente. Quanto ao *Concord*, essa ferramenta foi empregada com o propósito de gerar linhas de concordância para as palavras. Essa ferramenta pode apontar padrões que poderão ser analisados conforme objetivos do pesquisador. Neste estudo, foram analisados fragmentos, ou melhor, porções de linguagem, como os exemplificados na Figura 1, a seguir:

Figura 1: Recorte das linhas de concordância para *tar*.

125	não só pra mim // [347] cê vai <i>tar</i> fazendo a mimica pra todo	ta*	2.152	0 %	0 %	0 %	bfamcv04.txt
126	assim / é porque aí cê vai <i>tar</i> fazendo não só pra mim //	ta*	2.142	0 %	0 %	0 %	bfamcv04.txt

Fonte: *Concord*.

É possível observar que a forma do verbo *tar* (= estar), apresentada na Figura 1, reproduz o falar característico de falantes da capital mineira, constituintes do banco de dados do C-ORAL-BRASIL. Para se chegar à escolha do *corpus* de estudo, foi realizada uma leitura superficial e geral a fim de encontrar estruturas que caracterizariam ausência de concordância verbal, conforme a Gramática Tradicional (GT) ou que ocorresse alguma variação na estrutura dos segmentos verbais, como já mencionado anteriormente.

Diante dessas considerações, e de posse do material que se enquadrava nesses parâmetros, partiu-se para a inserção dos textos na janela da ferramenta *Concord*. Assim, os procedimentos seguiram-se na respectiva ordem.

Inicialmente foi acessada a janela *Concord*, após sua abertura, utilizou-se a opção *Settings*, em seguida *Choose Texts*, nessa aba a opção “escolher textos” foi acionada. Desse modo, foi possível procurar os arquivos que deveriam ser inseridos para análise. Neste momento, foi feita a seleção e inserção na janela *Files Selected* dos 9 textos em formato txt. Essa ação permitiu que fossem acionados comandos de abertura dos arquivos. Após esse procedimento, foi disponibilizada a tela *Getting Started* em que foi possível inserir arquivo de texto simples composto por lemas, os quais possibilitaram gerar linhas de concordâncias, conforme mostra

a Figura 1 acima. Com essa opção, a busca pelo objeto de estudo foi facilitada e a extração do material para análise foi realizada.

Vale salientar que, ao clicar 2 vezes em cima da linha de concordância selecionada, é possível retornar ao texto para se verificar em qual parte do texto aquele excerto se encontra. E assim, estender as análises e compreendê-las dentro do contexto. Verifica-se, por conseguinte, a confiabilidade dos dados apresentados para a análise e precisão de detalhes.

4 Análise e discussão

Apresentamos os dados coletados após a realização dos procedimentos metodológicos descritos na seção anterior. A Figura 2, a seguir, evidencia o total de palavras abstraídas do *corpus* de estudo.

Figura 2: Estatísticas do *corpus*.

N	Overall	1	2	3	4	5	6	7	8	9
text file	Overall	famcv01.txt	famcv02.txt	bfamcv03.txt	famcv04.txt	bfamcv05.txt	bfamcv06.txt	famcv07.txt	mcv08.txt	mcv09.txt
file size	111.522	11.687	14.368	11.181	15.798	12.597	11.699	11.118	11.058	12.016
tokens (running words) in text	15.150	1.761	2.277	1.930	2.430	1.485	749	1.266	1.509	1.743
tokens used for word list	13.493	1.475	1.845	1.592	1.939	1.475	739	1.226	1.477	1.725
types (distinct words)	1.701	367	339	317	323	315	226	392	341	511
typetoken ratio (TTR)	13	25	18	20	17	21	31	32	23	30
standardised TTR	27,23	30,00	22,00	23,50	21,20	24,80		34,50	27,10	34,70
standardised TTR std.dev.	64,63									
standardised TTR basis	1.000,00	1.000,00	1.000,00	1.000,00	1.000,00	1.000,00	1.000,00	1.000,00	1.000,00	1.000,00
mean word length (in characters)	3	4	3	3	3	3	4	4	4	4
word length std.dev.	1,93	2,06	1,81	1,61	1,64	1,51	1,89	2,29	2,39	2,05
sentences	9,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
mean (in words)	1.499	1.475	1.845	1.592	1.939	1.475	739	1.226	1.477	1.725
std.dev.	358,26									
paragraphs	9,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
mean (in words)	1.499	1.475	1.845	1.592	1.939	1.475	739	1.226	1.477	1.725
std.dev.	358,26									
headings										
mean (in words)	0	0	0	0	0	0	0	0	0	0
std.dev.										
sections	9,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Fonte: *WordList*.

Como se observa, na aba *Estatística* do *corpus*, foram encontrados 13.493 *tokens* e 1.701 *types* nos 9 textos analisados. Por meio dessa opção, é possível verificar ainda a quantidade de palavras existentes em cada texto, o que nos permite afirmar que são textos equilibrados em relação à quantidade de palavras; isto é, verifica-se uma média de 1.600 palavras por texto. Segundo Berber Sardinha (2009), esse equilíbrio é importante. Torna-se ainda relevante ressaltar que, na perspectiva do referido autor, o *corpus* utilizado neste estudo é considerado pequeno. Após essa verificação das 13.493 palavras que constituíam o *corpus* de estudo, foi possível realizar uma leitura atenta na busca por palavras que se enquadravam dentro da classe dos verbos que foram objeto desta pesquisa. Desse modo, despertaram nosso interesse as seguintes palavras: *ficamo, tamo, tão, tar, taria, tás, tava, tavam, távamos, vamo, vão*.

Assim, guiados pelo *corpus*, partimos para a próxima etapa do procedimento metodológico, que foi gerar as linhas de concordância, a fim de se verificar as construções proferidas por falantes investigados, provenientes da capital mineira. Desse modo, após a eliminação de palavras que não se enquadravam em nosso campo de busca, foram encontradas 222 concordâncias para as construções “tar”, “tão”, “tava”, conforme evidencia a Figura 3, a seguir.

Figura 3: Linhas de concordância A.

N	Concordance	et	aq	ord #	#	s	#	s	#	s	File	%
10	que faz faxina lá na casa deles / tava lá / aí / ela virou / e falou	ta*		1.602	0	0	0	0	0	0	%\bfamcv09.txt	67%
11	gay era eu hhh // outro dia / sô // tava + eu cheguei lá de gravata /	ta*		1.576	0	0	0	0	0	0	%\bfamcv09.txt	66%
12	// *GIL: não // a mãe dele tava numa dúvida // porque / eu	ta*		1.512	0	0	0	0	0	0	%\bfamcv09.txt	64%
13	&f [1] infância dos meninos / e tal / aí + e ela falava de mim / e eu	ta*		1.394	0	0	0	0	0	0	%\bfamcv09.txt	60%
14	// cês são casados // *GIL: é // tá certo // *CAM: ô + *GIL: e	tá*		1.323	0	0	0	0	0	0	%\bfamcv09.txt	57%
15	casar comigo // *ADR: mas cês tão casados // cês são íntimos /	o		1.281	0	0	0	0	0	0	%\bfamcv09.txt	55%
16	// *GIL: um é engenharia / o outro tá querendo fazer física // *ADR:	tá*		1.066	0	0	0	0	0	0	%\bfamcv09.txt	46%
17	da &N [2] da Nara // *CAM: também // *ADR: // *GIL: não // ela	ta*		914	0	0	0	0	0	0	%\bfamcv09.txt	39%
18	coisa // *CAM: ô / Fernanda Takai cantando Nara Leão / pra	ta*		816	0	0	0	0	0	0	%\bfamcv09.txt	34%
19	culpa se eu gosto de Fernanda Takai // *ADR: tem // *GIL: gosta	ta*		801	0	0	0	0	0	0	%\bfamcv09.txt	34%
20	*ADR: não / é ele mesmo // ele é tão legal // *CAM: "tenho muitos	o		759	0	0	0	0	0	0	%\bfamcv09.txt	32%
21	*ADR: // da / artes cênicas // que tá formando agora // cê viu	tá*		650	0	0	0	0	0	0	%\bfamcv09.txt	27%
22	/ é namorado da yyy // *GIL: ah / tá / eu sempre esqueço // eu	tá*		628	0	0	0	0	0	0	%\bfamcv09.txt	26%
23	eu // *CAM: // *GIL: esse // *CAM: tanto desse homem // *ADR: // ô /	ta*		611	0	0	0	0	0	0	%\bfamcv09.txt	26%
24	vezes // eu / já tentei dormir no tapete / já tentei dormir no sofá /	ta*		214	0	0	0	0	0	0	%\bfamcv09.txt	8%
25	na nosso sono // a gente tava lá dormindo e de repente	ta*		183	0	0	0	0	0	0	%\bfamcv09.txt	7%
26	/ apagou / Pedro / apagou / e tava / Amanda / Aninha / e eu /	ta*		130	0	0	0	0	0	0	%\bfamcv09.txt	5%
27	/ que ele chama // bom // aí tava nós / tentando dormir / o /	ta*		113	0	0	0	0	0	0	%\bfamcv09.txt	5%
28	última festa / tava tão divertido // tava / Pedro e Amanda dormindo	ta*		13	0	0	0	0	0	0	%\bfamcv09.txt	0%
29	cê nu sabe da última festa / tava tão divertido // tava / Pedro e	o		10	0	0	0	0	0	0	%\bfamcv09.txt	0%

Fonte: Concord.

Como pode se observar, os falantes do C-ORAL-BRASIL, nesta amostra, empregam construções que fogem ao padrão prescritivo da língua. Evidenciamos construções em que a forma verbal “tava” aparece em formas aferéticas (metaplasmos por supressão de fonemas), porém tais construções não trazem prejuízo quanto ao sentido que o emissor quis produzir. E, ainda, se considerarmos a concordância da linha 10, verifica-se que houve concordância com o sujeito da frase. Analisemos a Figura 4: Linhas de concordância B, a seguir, que traz essa porção de linguagem, no contexto.

Figura 4: Linhas de concordância B.

Fonte: Concord

A forma destacada “tava” concorda com o sujeito “menina”. Nessa perspectiva, se considerarmos Perini (2016), podemos concluir que se trata do emprego da norma urbana e não da norma preconizada pelas gramáticas.

Na sequência, apresentamos mais uma linha de concordância, agora em relação aos itens *faz, fazer, fazendo* e os seguintes resultados foram encontrados, como evidencia a Figura 5, a seguir.

Figura 5: Linhas de concordância C.

N	Concordance	Set	Tag	ord #	#	s	#	s	#	Pos	File	%
1	na beira do rio / e tal / os menino fazendo um showzinho lá / um lual	FAZ*		1.944	0	%	0	%	80%	bro\bfamcv09.txt	80%	
2	sei o quê / aí a [1] a menina que faz faxina lá na casa deles / tava	FAZ*		1.595	0	%	0	%	65%	bro\bfamcv09.txt	67%	
3	engenharia / o outro tá querendo fazer física // *ADR: hhh que sexy	FAZ*		1.068	0	%	0	%	44%	bro\bfamcv09.txt	46%	
4	desde seis anos // isso nu fui / fazer há um ano atrás / de /	FAZ*		461	0	%	0	%	19%	bro\bfamcv09.txt	20%	
5	o bastante / pra nu querer fazer mais // *GIL: eu nunca fui	FAZ*		374	0	%	0	%	15%	bro\bfamcv09.txt	16%	
6	Nossa // *FLA: que que a gente faz // eu quero voltar lá // *REN: ô	FAZ*		563	0	%	0	%	26%	bro\bfamcv08.txt	25%	
7	hora assim / pra ver se ea tava fazendo alguma coisa // *REN: / e	FAZ*		492	0	%	0	%	23%	bro\bfamcv08.txt	22%	
8	o mesmo caminho que a gente faz / normalmente / e deu cinco e	FAZ*		182	0	%	0	%	9%	bro\bfamcv08.txt	8%	
9	também / é [1] nu [1] é só ir fazendo as conta aqui // nu	FAZ*		127	0	%	0	%	6%	bro\bfamcv08.txt	6%	
10	/ *FLA: // *CEL: / nós temo que fazer // *SEL: // // *CEL: eu ten	FAZ*		564	0	%	0	%	41%	bro\bfamcv06.txt	35%	
11	// *SEL: / uns tempo / ea nu faz barulho não / sabe // *JOL:	FAZ*		300	0	%	0	%	22%	bro\bfamcv06.txt	20%	
12	agora eu tenho que + *MAR: nu faz vergonha futura esposa tá aí //	FAZ*		1.445	0	%	0	%	60%	bro\bfamcv05.txt	57%	
13	// *JOS: po' apertar e' // *CEL: faz de novo // *JOS: uai / varão //	FAZ*		871	0	%	0	%	36%	bro\bfamcv05.txt	35%	
14	pedir // cê quer que eu / &ba [3] fazer de cabeça / ou quer que eu	FAZ*		571	0	%	0	%	24%	bro\bfamcv05.txt	24%	
15	ir // *JOS: // *CEL: // *JOS: bora fazer o cinco / velho // *MAR: nu	FAZ*		513	0	%	0	%	21%	bro\bfamcv05.txt	22%	
16	filho // *CEL: // *JOS: não // tanto faz // *CAR: uai / mas aí / varão //	FAZ*		453	0	%	0	%	19%	bro\bfamcv05.txt	19%	
17	vacilei / varão // *MAR: / só fazer aí // *CAR: // *JOS: três a	FAZ*		175	0	%	0	%	7%	bro\bfamcv05.txt	7%	
18	// *BRU: [453] agora é ocê que faz // *LUC: [454] pera aí // [455]	FAZ*		2.730	0	%	0	%	96%	bro\bfamcv04.txt	96%	
19	não // *LUC: [424] é ocê que vai fazer // *BRU: [425] // [426] vai //	FAZ*		2.599	0	%	0	%	91%	bro\bfamcv04.txt	91%	
20	/ quero ver como é que cê faz // *BRU: [412] tá // [413] então	FAZ*		2.524	0	%	0	%	89%	bro\bfamcv04.txt	89%	
21	quer começar adivinhando / ou fazendo a mimica // *CEL: [393]	FAZ*		2.404	0	%	0	%	84%	bro\bfamcv04.txt	85%	
22	não só pra mim // [347] cê vai tar fazendo a mimica pra todo mundo	FAZ*		2.153	0	%	0	%	76%	bro\bfamcv04.txt	75%	

Fonte: Concord.

Na Figura 5, observam-se 63 linhas de concordância para essas realizações. As formas aferéticas também ocorrem como se observa nas linhas 2, 15 ao se verificar as formas (*tava* = estava), (*bora* = embora). Analisando as linhas 7, 11, 18, 19, 20 e 22, vê-se ainda a cliticização (transformação de sintagma nominal em pronome átono ou clítico) do pronome sujeito (*ea* = ela), (*ocê* = vocês), (*cê* = vocês). Nas linhas 6 e 8, destacamos as palavras à esquerda de "faz" e notamos o uso do termo *a gente*. Vale ressaltar que há uma tendência a substituir o pronome *nós* por *a gente*, de modo que formas como: *fazemos* 1ª pessoa do plural do modo indicativo, como também, *faremos* 1ª pessoa do plural do futuro do subjuntivo, conforme afirma Perini (2016), têm ocorrências cada vez mais raras. Verifica-se, ainda, que, há na linha 10, na perspectiva de Perini (2016), a simplificação da concordância verbal em (*temo* = temos).

E, por fim, para nutrir as análises e descrições, segue a Figura 6.

Figura 6: Linhas de concordância D.

N	Concordance	Set	Tag	ord #	#	s	#	s	#	Pos	File	%
1	na beira do rio / e tal / os menino fazendo um showzinho lá / um lual	FAZ*		1.944	0	%	0	%	80%	bro\bfamcv09.txt	80%	
2	sei o quê / aí a [1] a menina que faz faxina lá na casa deles / tava	FAZ*		1.595	0	%	0	%	65%	bro\bfamcv09.txt	67%	
3	engenharia / o outro tá querendo fazer física // *ADR: hhh que sexy	FAZ*		1.068	0	%	0	%	44%	bro\bfamcv09.txt	46%	
4	desde seis anos // isso nu fui / fazer há um ano atrás / de /	FAZ*		461	0	%	0	%	19%	bro\bfamcv09.txt	20%	

Fonte: Concord.

Considere, a fim de descrição, a porção de linguagem destacada na linha 1 da Figura 6. Por via de regra, emprega-se o verbo auxiliar “estar” para a concordância do gerúndio, ou com o infinitivo do verbo principal antecedido da preposição *a*, para indicar uma ação durativa, continuada. No caso, como prescreve a Gramática Tradicional, aqui na perspectiva de Cunha e Cintra (2007), desse modo a construção seria “os meninos estavam fazendo um showzinho”.

Nesse caso, houve a omissão do verbo auxiliar, porém ao se recuperar o contexto é possível entender que se trata de “os meninos estão fazendo um showzinho”. Se analisarmos tal porção de linguagem na perspectiva de Perini (2007, p. 237), pode-se inferir que o verbo *fazer* apareceu na construção transitiva, o que fez dele transitivo-ergativo, pois ocorre tanto no coloquial quanto no padrão em frases ergativas, como explicitada anteriormente em “os meninos estão fazendo um showzinho”.

5 Considerações finais

Neste estudo, apresentamos as finalidades das ferramentas *WordList* e *Concord* do WST, as quais demonstraram grande funcionalidade e simplicidade para buscar dados e comprovaram serem eficazes para facilitar o trabalho de investigação. Ao detalhar o procedimento para se gerar uma linha de concordância, cumprimos um elemento que foi proposto inicialmente neste estudo. Ao apresentarmos as análises oriundas da explicitação dos resultados obtidos, apresentamos algumas possibilidades de análises que emergiram das observações acerca da concordância verbal e pautadas nos pressupostos teóricos, como por exemplo, a presença de porções de linguagem representada pela Figura 5, linha 10, que evidencia a simplificação da concordância verbal, como se vê em (temo = temos). Ainda, ao analisarmos outra porção de linguagem apresentada na Figura 6 que evidencia a presença do verbo fazer que apareceu na construção transitiva, o que fez dele um verbo transitivo-ergativo.

As ferramentas *WordList* e *Concord* mostraram-se significativas ao auxiliarem na resolução dos questionamentos desse estudo. Em relação ao questionamento i) permitiram a extração dos fatos linguísticos relativos à norma da diatopia mineira, como se pode ver nos exemplos das Figuras 1, 3, 4, 5 e 6 em que foram geradas linhas de concordância para as realizações de formas não-padrão e paradigmas reduzidos, por exemplo, ficamo (= ficamos; verbo); formas dos verbos estar, ir, por exemplo, tamo (= estamos), tamos (= estamos), tão (= estão), tar (= estar), taria (= estaria), tás (= estás), tava (= estava), tavam (= estavam), távamos (= estávamos), vamo (= vamos), vão (= vamos). Formas essas que representam a realização do falar característico de falantes da capital mineira. Foi possível, ainda, responder ao questionamento ii) do respectivo estudo ao verificarmos fenômenos resultantes desses usos, conforme evidencia a Figura 5. A referida Figura aponta 63 linhas de concordância para construções aferéticas, cliticização e a

simplificação da concordância verbal. Como se pode ver, não nos detivemos apenas em analisar e descrever fatos exclusivamente relacionados à concordância verbal. As possibilidades de análises proporcionadas pelas ferramentas despertaram a curiosidade e intuição do pesquisador que procurou abordar outros elementos que saltaram aos olhos e que aqui não puderam passar despercebidos. Ratificamos a importante contribuição da LC com seu aparato teórico-metodológico, capaz de traçar direcionamentos para estudos acerca da língua, linguagem. E, “como área disciplinar que explora *corpora* computadorizados, embora ainda tímida, tem gradativamente crescido no Brasil nas duas últimas décadas” (MELLO, 2012, p. 31).

Conclui-se que tanto a LC quanto o uso do programa de computador *WST* versão 4.0., configuraram-se como importantes âncoras para a extração e análises dos dados sobre a concordância verbal, numa perspectiva sociolinguística.

Referências

BERBER SARDINHA, T. A influência do tamanho do *corpus* de referência na obtenção de palavras-chave usando o programa computacional WordSmith Tools. **The ESpecialist**, São Paulo, v. 26, n. 2, p. 183-204, 2005.

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

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

CUNHA, C.; CINTRA, L. F. L. **Nova gramática do português contemporâneo**. 4. ed. Rio de Janeiro: Lexikon, 2007.

MELLO, H. Os *corpora* orais e o C-ORAL-BRASIL. In: RASO, T.; MELLO, H. **C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal**. Belo Horizonte: Editora UFMG, 2012.

NEVES, M. H. M. **Que gramática estudar na escola?** Norma e uso na língua portuguesa. São Paulo: Contexto, 2017.

PERINI, M. A. **Estudos de gramática descritiva: as valências verbais**. São Paulo: Parábola Editorial, 2007.

PERINI, M. A. **Gramática descritiva do português brasileiro**. Petrópolis: Vozes, 2016.

RASO, T. O *corpus* C-ORAL-BRASIL. In: RASO, T.; MELLO, H. **C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal**. Belo Horizonte: Editora UFMG, 2012.

RASO, T.; MELLO, H. **C-ORAL-BRASIL I**: *corpus* de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012.

SCOTT, M. **WordSmith Tools Manual**. Stroud: Lexical Analysis Software Ltd., 2015. Disponível em: <https://lexically.net/downloads/version6/wordsmith6.pdf>. Acesso em: 29 jan. 2022.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://www.lexically.net/wordsmith/version4/>. Acesso em: 29 jan. 2022.

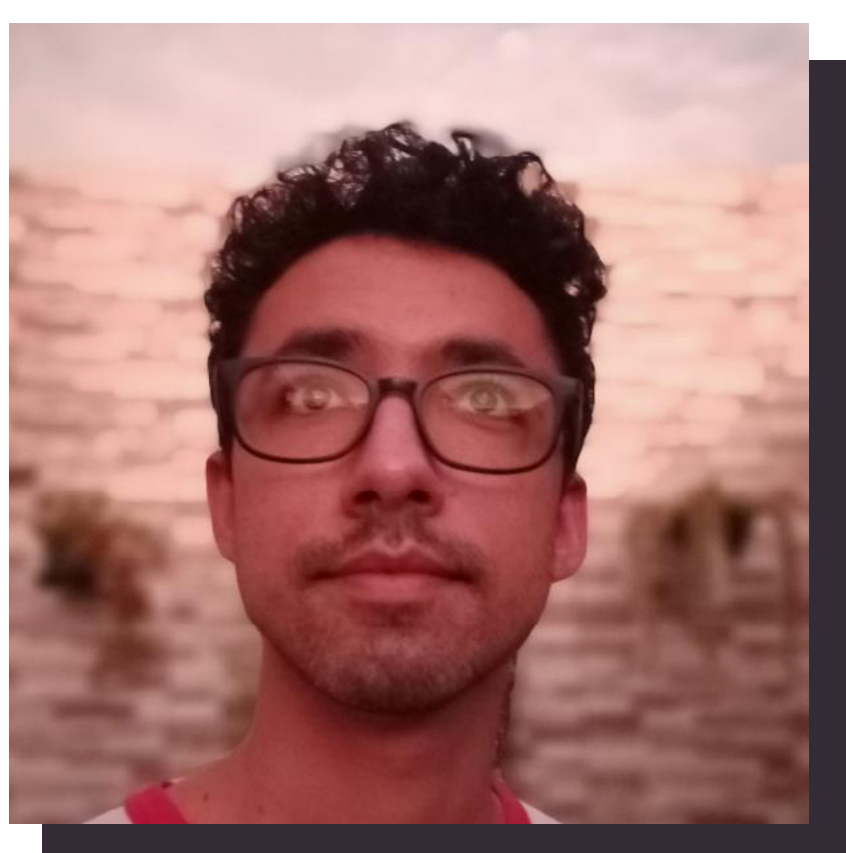
Sobre os autores e organizadores



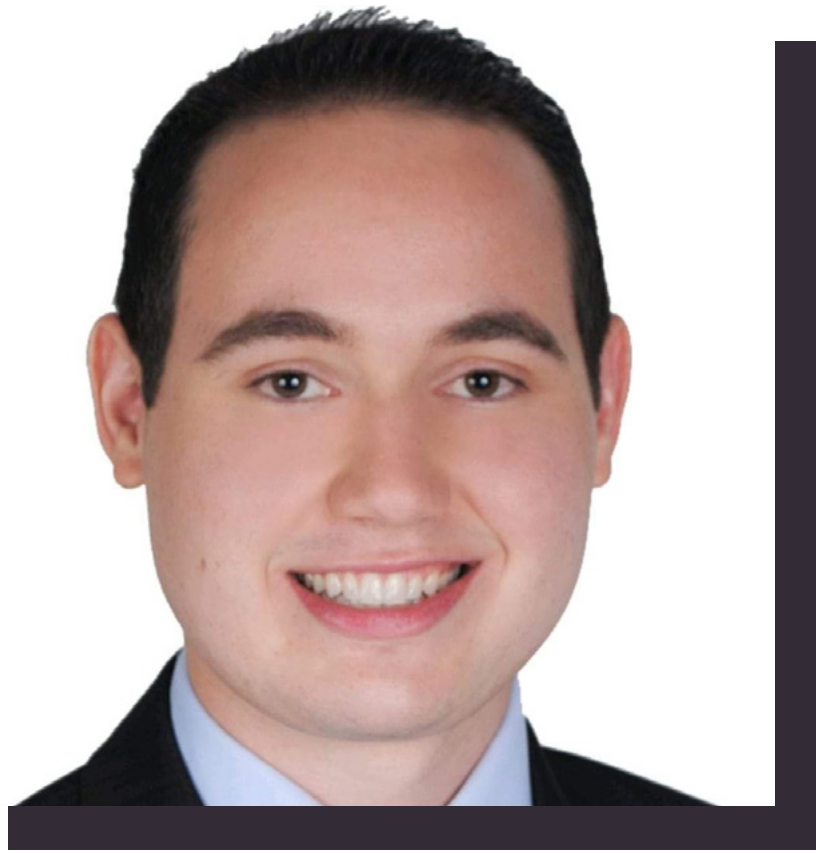
Ariel Novodvorski é Professor Associado do Instituto de Letras e Linguística da Universidade Federal de Uberlândia (ILEEL/UFU). Doutor em Estudos Linguísticos pela Universidade Federal de Minas Gerais (UFMG), com Pós-doutorado pela Universidade Federal do Rio Grande do Sul (UFRGS), sob supervisão da Profa. Dra. Cleci R. Bevilacqua. Como docente, atua no curso de Graduação em Letras Espanhol e no Programa de Pós-Graduação em Estudos Linguísticos (PPGEL). Seus interesses de pesquisa incluem: Estudos Descritivos, Linguística de *Corpus*, Estudos da Tradução, Língua Espanhola, Fraseologia, Terminologia e Linguística Sistêmico-Funcional. Tem experiência de mais de vinte anos na docência, pesquisa e tradução. Conta com publicações em diversos periódicos indexados e em livros. Diretor do Instituto de Letras e Linguística da UFU (2017-2025). Lattes: <http://lattes.cnpq.br/2882362453894798>. E-mail: arivorski@ufu.br.



Joel Victor Reis Lisboa é Doutorando e Mestre em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU) e Graduado em Letras Inglês pela Universidade Federal de Goiás (UFG) – Regional Jataí. É pesquisador nas áreas de Terminologia, Português como Língua Não Materna e Fraseologia, em interface com a Linguística de *Corpus*. É membro do Grupo de Pesquisa e Estudos em Linguística de *Corpus* – GPELC (UFU/CNPq) e do grupo Pesquisas em Léxico – PLEX (UFU/CNPq). Lattes: <http://lattes.cnpq.br/7657573383244824>. E-mail: joelvictorlisboa@gmail.com.



Raphael Marco Oliveira Carneiro é Doutorando e Mestre em Linguística e Linguística Aplicada pela Universidade Federal de Uberlândia (UFU) e Graduado do Curso de Licenciatura Plena em Letras: Habilitação em Inglês e Literaturas de Língua Inglesa pela mesma universidade. Tem experiência no ensino de língua inglesa e seus interesses de pesquisa incluem Estilística, Fraseologia, Linguística de *Corpus* e Tradução. Membro do Grupo de Pesquisa e Estudos em Linguística de *Corpus* – GPELC (UFU/CNPq) e do Grupo em Estudos Contrastivos – GECon (UFU/CNPq). Lattes: <http://lattes.cnpq.br/2014869652199855>. E-mail: raphael.olic@gmail.com.



Lucas Amâncio Mateus é Doutorando em Estudos Linguísticos e Graduado em Administração pela Universidade Federal de Uberlândia (UFU) e Licenciado em Letras Português/Inglês pela Universidade de Franca (UNIFRAN). É pesquisador nas áreas de Terminologia e Linguística de *Corpus* e integrante do Grupo de Pesquisa e Estudos em Linguística de Corpus – GPELC (UFU/CNPq) e do grupo Pesquisas em Léxico – PLex (UFU/CNPq). Lattes: <http://lattes.cnpq.br/0101102792314230>. E-mail: lucas.mateus@ufu.br.



Lidiane Carlos Ramos é Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), Graduada em Letras Português/Francês (UFU) e em Português/Espanhol pela Universidade de Uberaba (Uniube). Especialista em Língua e Literaturas Espanhola e Hispano-americana e em Português Língua Estrangeira e Cultura Brasileira pela UFU. Atuou como Leitora da CAPES na Facultad de Filosofía y Letras da Universidad Nacional de Cuyo em Mendoza, Argentina. É membro do grupo de pesquisas GECon – Grupo de Estudos Contrastivos (UFU/CNPq). Lattes: <http://lattes.cnpq.br/3825836697652282>. E-mail: lidiane.ramos@ufu.br.

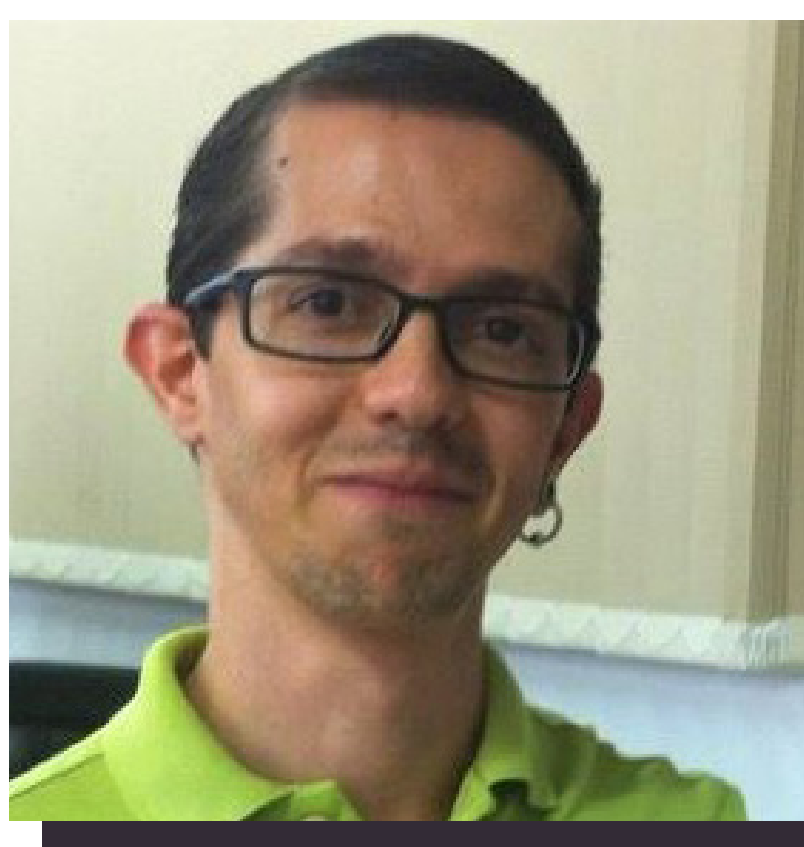


Terezinha de Assis Oliveira é Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), Mestra em Estudos da Linguagem pela Universidade Federal de Goiás (UFG) e Graduada em Letras pela UFU. Atualmente, é professora do Instituto de Estudos da Linguagem da Universidade Federal de Catalão (UFCat). Suas áreas de atuação compreendem Ensino de Línguas Estrangeiras, Língua Inglesa, Culturas de Língua Inglesa e Linguística Aplicada. É membro do Grupo em Estudos Contrastivos – GECon (UFU/CNPq). Lattes: <http://lattes.cnpq.br/3140346526525800>. E-mail: terezinha63@ufcat.edu.br.



Sarah Cristina de Oliveira Sebba é Doutoranda em Estudos da Linguagem, Mestra em Letras e Linguística, Especialista em Letras – Leitura e Ensino e Graduada em Letras Português pela Universidade Federal de Catalão (UFCat). Também é Graduada em Letras Português/Inglês pela Universidade Estadual de Goiás (UEG) e em Administração de Empresas pelo Centro de Ensino Superior de Catalão (CESUC). Tem experiência na área de Letras, com ênfase em Língua Inglesa. Lattes: <http://lattes.cnpq.br/5523573670912856>.

E-mail: sarah_sebba@hotmail.com.



Victor Mariotto Palma é Mestrando em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU). Possui pós-graduação *lato sensu* em Estudos Linguísticos e Literários pela Faculdade Única de Ipatinga, Bacharelado em Tradução e Licenciatura Plena em Letras Inglês/Português, ambos pela UFU. Tem experiência na área de Letras e Tradução. Foi membro e diretor administrativo-financeiro da empresa júnior Babel Traduções do curso de Tradução da UFU. Lattes: <http://lattes.cnpq.br/6432124077851888>.

E-mail: victorm@ufu.br.



Mayra Natanne Alves Marra é Doutoranda e Mestra em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU) e Graduada em Letras Português/Espanhol pela Universidade Federal do Triângulo Mineiro (UFTM). Tem experiência no ensino de línguas, com ênfase na Língua Espanhola. Interessa-se por estudos baseados em *corpus*, trabalhando, principalmente, nos temas: Estudos Descritivos, Linguística de *Corpus* e Linguística Sistêmico-Funcional. Atualmente, é professora efetiva do ensino básico, técnico

e tecnológico do Instituto Federal do Triângulo Mineiro (IFTM) – *Campus Ituiutaba*. Lattes: <http://lattes.cnpq.br/4955128268478795>. E-mail: mayra.marra@ufu.br.



Mariana Souza Santos é Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU) e Graduada em Letras Inglês pela Universidade Estadual de Montes Claros (UNIMONTES). É pesquisadora nas áreas de fonética e fonologia do Inglês. É membro do Grupo de Pesquisa e Estudos em Fonética e Fonologia – GEFONO (UFU/CNPq). Lattes: <http://lattes.cnpq.br/6102463770404857>. E-mail: marianasouza2021@yahoo.com.



Cássia Beatriz de Morais Silva é Mestranda em Estudos Linguísticos pela Universidade Estadual de Goiás (UEG – Campus Cora Coralina), Especialista em Estudos Linguísticos pela mesma universidade, no *campus* de Morrinhos, e Graduada em Letras Inglês/Português pela Instituto Luterano de Ensino Superior (Iles Ulbra). É professora de Língua Portuguesa na Rede Pública do Estado de Goiás e pesquisadora nas áreas de Lexicologia em Língua Portuguesa, em interface com a Linguística de *Corpus*. Lattes: <http://lattes.cnpq.br/3556885127366452>.

E-mail: cassia.de@ufu.br.



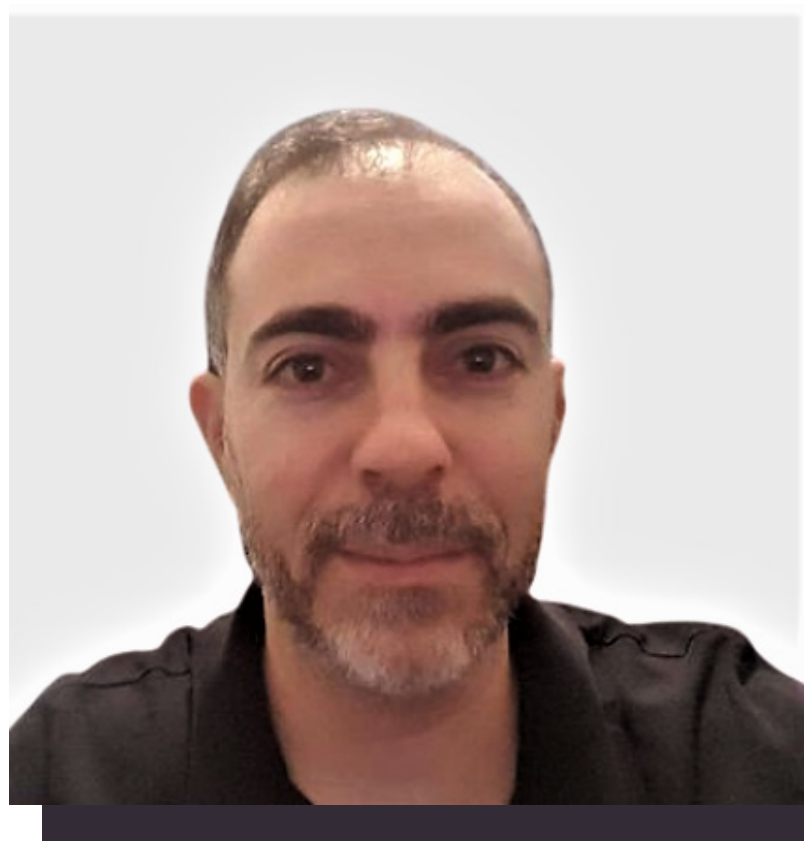
Roberta Gê-Acaiaba é Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), Mestre em Letras pela Universidade Federal do Triângulo Mineiro (UFTM), Especialista em Práticas Pedagógicas pela Universidade Federal de Ouro Preto (UFOP) e Especialista em Educação Inclusiva pela Universidade Castelo Branco (UCB). Atualmente, é professora da rede pública de ensino do estado de Minas Gerais. Lattes: <http://lattes.cnpq.br/0937562879066273>. E-mail: robertage@ufu.br.



Wagner da Cunha Nunes é Mestrando em Estudos Linguísticos, Especialista em Administração de Empresas e Graduado em Letras Espanhol pela Universidade Federal de Uberlândia (UFU). É pesquisador nas áreas de Fraseologia e Metáforas Conceptuais. É membro do Grupo de Pesquisa em Estudos Contrastivos – GECon (UFU/CNPq). Lattes: <http://lattes.cnpq.br/5502387144320996>. E-mail: wagner.nunes@ufu.br.



Rosena Caixeta Silva Rodrigues de Sousa é Mestranda em Estudos Linguísticos e Graduada em Letras Português pela Universidade Federal de Uberlândia (UFU). Atua como professora de Redação na rede privada de ensino, corretora textual e avaliadora em processos seletivos de Redação. É pesquisadora com foco na Linguística de *Corpus*. Faz parte do Grupo de Pesquisa em Estudos Contrastivos – GECon (UFU/CNPq). Lattes: <http://lattes.cnpq.br/8029826531413628>. E-mail: rosena@ufu.br.



Heitor Carvalho de Almeida Neto é Mestrando em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), Especialista em Análise e Desenvolvimento de Sistemas pelo Instituto Federal do Triângulo Mineiro (IFTM) e Graduado em Sistema de Informação pelo Centro Universitário do Triângulo Mineiro (UNITRI). É membro do Grupo de Pesquisa em Estudos Contrastivos – GECon (UFU/CNPq). Lattes: <http://lattes.cnpq.br/9429112861487266>. E-mail: heitor@ufu.br.



Thaís dos Santos Souza é Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU) e Mestre em Criminologia pela Universidade do Porto (U.Porto). É Graduada em Letras Português pela UFU, em Direito pela Universidade de Uberaba (Uniube) e em Pedagogia pela Faculdade Intervale. Possui experiência na área de Letras, com ênfase em Linguística, atuando principalmente nas seguintes temáticas: Linguística Forense, Linguística Aplicada, Análise do Discurso, Teoria Sociológica, Pesquisa Social Qualitativa, Ciências Sociais Aplicada, Sentimento de Insegurança, Medo do Crime e Discurso Midiático. Lattes: <http://lattes.cnpq.br/1690929957314722>. E-mail: thais.souza4@ufu.br.



Maria de Oliveira Rodrigues é Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), Mestre em Letras pelo Programa de Mestrado Profissional em Letras da Universidade Estadual de Montes Claros (UNIMONTES), Especialista em Língua Portuguesa pela Faculdade do Noroeste de Minas (FINOM) e Graduada em Letras Inglês pelo Centro Universitário de Sete Lagoas (UNIFEMM). É integrante do Núcleo de Estudos da Norma Linguística – NormaLi (UFU/CNPq). Lattes: <http://lattes.cnpq.br/5030844013340663>. E-mail: maria.rodrigues5@ufu.br.

Publique com a gente e
compartilhe o conhecimento



www.lettraria.net



 Letraria®