

Estudos exploratórios

em Linguística de Corpus

Ariel Novodvorski
Joel Victor Reis Lisboa
(organizadores)

**Estudos exploratórios
em Linguística de *Corpus***

Ariel Novodvorski
Joel Victor Reis Lisboa
(organizadores)

Estudos exploratórios em Linguística de *Corpus*

Araraquara
Letraria
2021

Estudos exploratórios em Linguística de Corpus

PROJETO EDITORIAL

Letraria

PROJETO GRÁFICO E DIAGRAMAÇÃO

Letraria

CAPA

Letraria

REVISÃO

Letraria

NOVODVORSKI, Ariel; LISBOA, Joel Victor Reis. (org.). **Estudos exploratórios em Linguística de Corpus**. Araraquara: Letraria, 2021.

ISBN: 978-65-86562-57-6

1. Linguística de Corpus. 2. Estudos Exploratórios. 3. Linguística.

CDD: 410 – Linguística

Os textos aqui publicados são de inteira responsabilidade de seus autores e organizadores.

Esta obra ou parte dela não pode ser reproduzida por qualquer meio, sem autorização escrita dos autores e organizadores.

Conselho editorial

Aderlande Pereira Ferraz (UFMG)

Sivelena Cosmo Dias (UFMS/CPTL)

Neubiana S. V. Beilke (UFU)

| Sumário

| | |
|---|-----|
| Estudos descritivos e Linguística de <i>Corpus</i> : da disciplina à pesquisa exploratória <i>Ariel Novodvorski e Joel Victor Reis Lisboa</i> | 7 |
| “Tic” y “discapacidad” en un corpus periodístico: aproximaciones básicas a la luz de la Lingüística de <i>Corpus</i> <i>Yohanna Tamal Hernández Consoró</i> | 16 |
| Representação discursiva do aborto na <i>Revista Azmina</i> : uma análise pautada na Linguística de <i>Corpus</i> e na Linguística Sistêmico-Funcional <i>Bianca Mara Guedes de Souza</i> | 26 |
| Recursos linguístico-discursivos materializados em uma reportagem <i>on-line</i> acerca de estudantes cotistas: uma análise pautada no sistema da avaliatividade e na Linguística de <i>Corpus</i> <i>Conceição Maria Alves de Araújo Guisardi</i> | 37 |
| La exclusión de actores sociales en reportajes periodísticos: un análisis del asesinato de la líder social colombiana María del Pilar Hurtado <i>Laura Alejandra Guerrero Calderón</i> | 51 |
| O AFETO em depoimentos sobre adoção: uma análise da avaliatividade com subsídios da Linguística de <i>Corpus</i> <i>Layane Campos Soares</i> | 61 |
| <i>Saúde de corpo e alma</i> : análise em <i>corpus</i> de programa radial <i>Luana Duarte Silva</i> | 71 |
| A fraseologia em Mario Vargas Llosa: um estudo contrastivo em <i>corpus</i> paralelo bilíngue espanhol/português do jornal <i>El país</i> <i>María del Rosario Mestanza</i> | 81 |
| Libras e linguística de <i>corpus</i> : análise de sinais-termos da indústria automobilística <i>Kássia Mariano de Souza</i> | 92 |
| Português para Falantes de Outras Línguas: uma proposta inicial de árvore de domínio à luz da Linguística de <i>Corpus</i> <i>Joel Victor Reis Lisboa</i> | 102 |
| O léxico <i>kimbundu</i> no português oralizado em Angola: uma análise da frequência em <i>Os contos de Ukamba Kimba</i> <i>Ivone da Silva Santos</i> | 112 |
| Sobre os autores e organizadores | 122 |

Estudos descritivos e Linguística de *Corpus*: da disciplina à pesquisa exploratória

Ariel Novodvorski¹

Joel Victor Reis Lisboa²

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 em Estudos Linguísticos pela Universidade Federal de Uberlândia.

Lattes: <http://lattes.cnpq.br/7657573383244824>. E-mail: joelvictorlisboa@gmail.com.

Esta publicação nasce de nossas experiências acadêmicas em torno da disciplina *Estudos Descritivos e Linguística de Corpus*, ofertada no segundo semestre de 2019, no Programa de Pós-Graduação em Estudos Linguísticos (PPGEL), do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU). Desde 2015, temos ministrado essa disciplina no PPGEL, alternando com outras duas, *Linguística de Corpus aplicada a pesquisas de base empírica* e *Teoria da Avaliatividade e Linguística de Corpus*, em que confluem, também, a descrição linguística e a exploração de *corpora*, por meio de princípios, ferramentas e de diferentes recursos, a partir de determinados recortes e tendências teóricas. Diversas foram as publicações, na forma de artigos ou capítulos de livros, oriundas dos resultados e análises em trabalhos finais de estudantes nessas disciplinas. Inspirados na experiência positiva de um trabalho anterior, que também desenvolvemos no âmbito desse Programa junto ao corpo discente e que derivou na publicação do livro *Ensaio em teorias linguísticas* (NOVODVORSKI; ROSA; CHAGAS, 2016), resolvemos organizar esta publicação, tomando por base os trabalhos finais de nossa disciplina, com o objetivo de dar a conhecer alguns meios e percursos para a pesquisa empírico-descritiva e exploração de diferentes *corpora*.

A idealização da disciplina foi motivada pelo desejo de propiciar aos pós-graduandos conhecimentos gerais e específicos acerca da abordagem, dos procedimentos metodológicos, do potencial de teorização e/ou de constatação de teorias, para além da aplicabilidade da Linguística de *Corpus* (LC) em pesquisas de base empírica. A partir da sensibilização para a observação e percepção de fatos linguísticos, formulação de hipóteses, compilação de *corpora*, levantamento e sistematização de dados, procuramos constituir um caminho para a pesquisa exploratória, guiada por ou baseada em dados. Compreendemos que a disciplina traz importantes contribuições para a formação dos estudantes, tanto no âmbito da pesquisa, do fazer metodológico, como das inúmeras possibilidades de aplicação, em torno dos estudos de descrição linguística. Assim, desde a constituição de uma linguística descritiva até a aplicabilidade em pesquisas exploratórias de base empírica, perpassamos processos voltados para a sensibilização na observação e percepção de fatos linguísticos, hipotetização a partir do levantamento e sistematização de dados, levando em consideração conceitos, princípios básicos e a tecnologização característica de se fazer pesquisa com a LC.

O programa da disciplina tem seguido a estrutura de três eixos norteadores, a saber: (i) Linguística Descritiva e visão geral da LC; (ii) Compilação, procedimentos metodológicos e etiquetagem; e (iii) Aplicabilidade da LC em pesquisas de base empírica. As diferentes temáticas e áreas implicadas nos respectivos segmentos foram, de modo geral: Percepção e observação de fatos linguísticos; Formulação de hipóteses; Introspecção, testagem e análise de *corpus*; Antecedentes da LC; (Re)surgimento, definições, conceitos básicos, planejamento, arquitetura,

compilação e preparação dos *corpora*; Etiquetagem; *Corpora* paralelos; *Corpora* especializados; Tradução; Ensino de língua estrangeira; Metáfora; Papéis temáticos; Fraseologia; *Corpus* e Literatura; e Terminologia. A Metodologia contou com aulas expositivas, discussão das leituras, exposições dialogadas e todas as aulas foram desenvolvidas no laboratório, para utilização de programas e recursos. Dentre os programas computacionais utilizados, seja para análises lexicais, etiquetagem e alinhamento paralelo de *corpus*, destacamos o *WordSmith Tools* 4.0 (SCOTT, 2007), por se tratar da versão completa com registro livre, *AntConc* 3.5.8 (ANTHONY, 2019), *UAM CorpusTool* 3.3 (O'DONNEL, 2019), *ParaConc* 1.0 (BARLOW, 2004) e os recursos presentes, fundamentalmente, nas plataformas *Sketch Engine* (KILGARRIFF, 2019) e *Corpus do Português* (DAVIES, 2019).

Com relação aos aspectos teórico-metodológicos imbricados na abordagem da disciplina e, especificamente, no que tange à constituição da perspectiva descritivista adotada (PERINI, 2006, 2008), voltamos nossa atenção para a percepção, identificação e descrição dos fatos da língua, despreocupados quanto à elaboração ou validação de quaisquer teorias específicas da linguagem, mas com a motivação por promover a reflexão quanto a questões fundamentais de análise, com base nos dados coletados, sistematizados e analisados. Devido à falta de paradigmas que norteiem a interpretação e observação dos fatos, a ênfase recai e potencializa o viés metodológico. Desse modo, questões como “o que deve estar contemplado nas análises?” conduzem à necessidade de recortes, à tomada de decisões e de riscos, à consideração do que seria ou não relevante.

Nessa vertente, Perini (2008) propõe determinados princípios para uma metodologia da descrição gramatical, em que se destaque a exigência de exaustividade (na medida do possível), para evitar a seleção de evidência, voluntária ou involuntária. Conforme o autor, “O pesquisador, na preocupação de encontrar evidência que corrobore sua teoria, seleciona dados favoráveis com muito mais energia do que a que utiliza na procura de dados desfavoráveis, apresentando dessa maneira uma imagem deformada da realidade linguística em estudo” (PERINI, 2008, p. 32-33). Com isso, a pesquisa linguística precisaria estar muito mais baseada em dados e, nesse ponto, o uso de *corpus* é, sem dúvida, um avanço metodológico essencial.

De acordo com Parodi (2010), um *corpus* é uma coleção ou conjunto de textos, que compartilham determinados traços definitórios, compilados no intuito de oferecer dados acerca da língua, mas numa projeção maior, se comparado a uma unidade textual como instância única de fala. Com relação a esse último ponto, um *corpus* não deve ser entendido como uma instância comunicativa em si mesma. Resumidamente, pode-se afirmar que o objetivo da LC reside na análise e descrição da língua em uso, a partir de conjuntos de unidades textuais ou *corpora*. Como também observa Berber Sardinha (2004, 2009), a LC se ocupa da coleta e da

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. Portanto, a exploração da linguagem por meio de evidências empíricas, identificadas com suporte de ferramentas e recursos computacionais, estão no centro da LC.

Dentre os inúmeros benefícios provenientes da utilização de métodos da LC (PARODI, 2010), destacamos, a saber: a disponibilidade de grandes amostras de textos autênticos para indagação empírica de línguas naturais, com acesso a variedades e modos linguísticos heterogêneos; a acessibilidade a análises sistemáticas de grandes quantidades de textos, de modo rápido e com alta confiabilidade; a tecnologização das investigações; a contrastação de hipóteses, por meio de evidência empírica em larga escala; a validação de achados preliminares; a aplicação de tecnologia computacional em todos os níveis da língua; as infinitas possibilidades de exploração em *corpora* etiquetados; e a indagação de variabilidade linguística desde múltiplos registros. Em suma, percebemos que a LC oferece oportunidades revolucionárias para a descrição, análise e ensino de fenômenos linguísticos de todo tipo.

Para além da participação nas discussões durante as aulas, acerca das leituras e da aprendizagem no uso das ferramentas e recursos, as atividades contaram com a elaboração coletiva de um Vocabulário de termos e conceitos, a partir da terminologia e dos conceitos trabalhados nas leituras teóricas. Para a realização dessa tarefa, contamos com os recursos próprios do Moodle institucional da UFU. Entre as premissas para o trabalho em grupo, adotamos alguns critérios como a inserção de um mínimo de vinte termos com os respectivos conceitos por estudante, além de definições de formatação e prazos. A ideia prevaiente foi a de conseguir compilar, até o término do semestre letivo, um *corpus* de conceitos da disciplina *Estudos Descritivos e Linguística de Corpus*, tomado textualmente das referências, que poderia ser utilizado para buscas específicas, posteriormente, pelos próprios estudantes. Para tanto, foi essencial a indicação de autoria e referências de data de publicação e paginação, para viabilizar o uso do Vocabulário como fonte de consulta para citações. Ao término da disciplina, contabilizamos 236 entradas, abarcando os mais diversos aspectos e áreas estudadas. Por exemplo, para os termos *corpus* e seu plural *corpora* foram registradas 26 definições, com alcance tanto às variadas acepções como às nomenclaturas especificativas (*corpora* paralelos, bidirecionais, de aprendiz, de consulta, de referência etc.).

Outra atividade prevista na disciplina foi a Apresentação oral de um projeto de Artigo ou capítulo de livro, que era o trabalho final e que originou a presente publicação. Assim, quando já nos encontrávamos, aproximadamente, na metade do segundo eixo norteador do programa, já de posse de um cabedal de princípios, conceitos e técnicas quanto ao uso de determinadas ferramentas e recursos, e após termos analisado e estudado diferentes possibilidades

metodológicas da LC, realizamos um Seminário. Na ocasião, foram apresentados os objetos de estudo, os *corpora* e as metodologias que iriam sustentar as análises nos respectivos trabalhos, que derivaram nos trabalhos finais. Os critérios adotados para a avaliação tanto das apresentações orais quanto dos trabalhos escritos foram a descrição dos objetos de estudo, a descrição dos procedimentos implicados na compilação e preparação dos *corpora* de estudo, a efetiva utilização das ferramentas e recursos pertinentes, a exposição dos resultados, além da organização e capacidade de síntese. No que tange especificamente aos trabalhos escritos, foram apreciados o domínio do embasamento teórico, o diálogo entre as referências trabalhadas, a profundidade de reflexão crítica, a descrição na apresentação do recorte do objeto de estudo e dos procedimentos metodológicos e resultados alcançados e, também, a qualidade redacional.

Os comentários e sugestões, proferidos durante a apresentação oral dos trabalhos e a partir da primeira versão dos escritos, não só contribuíram para a definição mais pontual de cada projeto como também propiciaram o entendimento de que deveríamos estabelecer critérios para a padronização quanto ao formato dos textos, já pensando na possibilidade de publicação dos resultados em conjunto, como um *e-book*. Nesse sentido, adotamos algumas normas e fixamos a extensão, a estrutura interna dos trabalhos finais e propusemos prazos por meio de um cronograma. Após a entrega dos textos, avaliação, conclusão da disciplina, sugestões, indicação de revisões e correções necessárias, este livro começou a tomar forma. Ainda que já em plena pandemia pela COVID-19 e lidando com inúmeras dificuldades, estamos conseguindo realizar nosso desejo de reunir os trabalhos resultantes da disciplina e materializá-los nesta publicação.

Os estudos exploratórios apresentados neste *e-book* iniciam-se com o capítulo intitulado “‘TIC’ y ‘discapacidad’ en un *corpus* periodístico: aproximaciones básicas a la luz de la Lingüística de *Corpus*”, de autoria de Yohanna Tamal Hernández Consoró. Neste capítulo, a autora parte de um *corpus* de estudo composto por 82 notícias *on-line* publicadas entre 2008 e 2019, compiladas a partir da base de dados do jornal dominicano *Listín Diario*. A análise exploratória empreendida é realizada com base na frequência de ocorrência, nos contextos linguísticos e nas tendências de coocorrência de duas palavras-chave, “TIC” e “discapacidad”, com o objetivo de analisar como elas são veiculadas nas notícias que compõem o *corpus* de estudo. As análises são realizadas por meio das ferramentas *WordList* e *Concord* do programa de análise lexical *WordSmith Tools* (doravante, WST).

Na sequência, Bianca Mara Guedes de Souza, no capítulo intitulado “Representação discursiva do aborto na *Revista Azmina*: uma análise pautada na Lingüística de *Corpus* e na Lingüística Sistêmico-Funcional”, analisa a reportagem intitulada “Como é feito um aborto

seguro?”, assinada por Helena Bertho e publicada na *Revista AzMina* em 2019, com o objetivo de explorar, em termos linguísticos, como o aborto é representado nesta reportagem. Neste estudo exploratório, a análise é embasada na Linguística Sistêmico-Funcional e empreendida por meio das ferramentas *WordList*, *KeyWords* e *Concord* do WST.

No capítulo intitulado “Recursos linguístico-discursivos materializados em uma reportagem *on-line* acerca de estudantes cotistas: uma análise pautada no Sistema da Avaliatividade e na Linguística de *Corpus*”, Conceição Maria Alves de Araújo Guisardi apresenta uma análise exploratória das marcas de avaliações referentes a estudantes negros e estudantes oriundos de periferia na reportagem intitulada “As pessoas não acham que alguém como eu possa ser inteligente’: a vida dos alunos da periferia na USP”, assinada por Letícia Mori e publicada no *BBC News Brasil*. Esta análise, pautada no Sistema da Avaliatividade, foi realizada por meio das ferramentas *WordList* e *Concord* do WST.

Laura Alejandra Guerrero Calderón, no capítulo intitulado “La exclusión de actores sociales en reportajes periodísticos: un análisis del asesinato de la líder social colombiana María del Pilar Hurtado”, com base em um *corpus* de reportagens, apresenta uma análise exploratória das formas de representação/exclusão de atores sociais que poderiam estar envolvidos no assassinato da líder social colombiana María del Pilar Hurtado. O *corpus* de estudo é composto por seis reportagens *on-line*, publicadas nos jornais *El Tiempo*, *El Espectador* e *El Heraldo*. A análise apresentada pela autora é embasada pela Teoria de Representação de Atores Sociais e pela Análise Crítica do Discurso, e foi realizada por meio das ferramentas *WordList* e *Concord* do WST.

No capítulo seguinte, intitulado “O AFETO em depoimentos sobre adoção: uma análise da Avaliatividade com subsídios da Linguística de *Corpus*”, Layane Campos Soares apresenta uma análise de um *corpus* composto por dezessete depoimentos sobre as experiências de adoção, compilados do site *Quero uma família* – projeto desenvolvido pelo Ministério Público do Rio de Janeiro, com o intuito de verificar como crianças e adolescentes adotados são avaliados nos depoimentos dos adotantes. A análise exploratória é pautada no Sistema da Avaliatividade e foi realizada por meio do WST, em específico, por meio das ferramentas *WordList* e *Concord*.

Na sequência, Luana Duarte Silva, no capítulo intitulado “*Saúde de Corpo e Alma*: análise em *corpus* de programa radial”, apresenta a análise exploratória de um *corpus* oral (transcrito) composto pelas duas primeiras edições do programa da *Rádio Globo* intitulado *Saúde de Corpo e Alma*. Neste capítulo, a autora verifica, por meio das ferramentas *WordList* e *Concord* do WST, se a missão do programa sob análise é refletida pelas unidades lexicais mais frequentes no *corpus* de estudo.

Em “A fraseologia em Mario Vargas Llosa: um estudo contrastivo em *corpus* paralelo bilíngue espanhol/português do jornal *El País*”, María del Rosario Mestanza analisa, com base no arcabouço teórico da Fraseologia e da Tradução, um *corpus* paralelo bilíngue (espanhol/português) composto por cinco ensaios publicados por Mario Vargas Llosa em 2015 na coluna de opinião “Pedra de Toque” do jornal *El País* e por suas respectivas traduções para o português brasileiro. Neste estudo exploratório guiado por *corpus*, a autora identifica e analisa, por meio das ferramentas *WordList* e *Concord* do WST, unidades fraseológicas presentes nos ensaios originais em espanhol e explora como essas unidades são traduzidas.

Kássia Mariano de Souza, no capítulo intitulado “Libras e Linguística de *Corpus*: análise de sinais-terminos da indústria automobilística”, identifica e analisa sinais-terminos presentes em um *corpus* constituído por entrevistas (gravadas em áudio e vídeo, e posteriormente transcritas), realizadas com funcionários surdos de uma empresa do setor automobilístico localizada em Catalão-Goiás. Neste estudo exploratório, situado na intersecção entre Terminologia, Libras e LC, a autora investiga o modo como os sinais-terminos são constituídos e utilizados pelos funcionários entrevistados, focalizando na análise da motivação e não-motivação na criação destes sinais-terminos.

No capítulo seguinte, intitulado “Português para Falantes de Outras Línguas: uma proposta inicial de árvore de domínio à luz da Linguística de *Corpus*”, Joel Victor Reis Lisboa apresenta uma proposta inicial da árvore de domínio da área de Português para Falantes de Outras Línguas (PFOL). Esta proposta baseia-se na análise quali-quantitativa de candidatos a unidades fraseológicas especializadas designativas de potenciais subáreas do PFOL. A análise exploratória apresentada neste capítulo, realizada por meio da ferramenta *Concord* do WST, é feita com base em um *corpus* especializado composto por 66 artigos publicados na *Revista SIPLE*, periódico brasileiro especializado na área.

Por fim, Ivonete da Silva Santos, no capítulo intitulado “O léxico *kimbundu* no português oralizado em Angola: uma análise da frequência em *Os contos de Ukamba Kimba*”, parte de um *corpus* composto pelos 24 contos da obra angolana *Os contos de Ukamba Kimba* para explorar, em termos quali-quantitativos, a presença de unidades do léxico *kimbundu* na obra em questão, de modo a atestar a existência de um léxico misto na variedade angolana do português, representada pela referida obra. As análises exploratórias foram realizadas por meio das ferramentas *WordList* e *Concord* do WST.

A coleção de estudos exploratórios que compõe este livro demonstra as possibilidades e proficuidade da utilização da LC como metodologia e abordagem em diferentes pesquisas no escopo de áreas, teorias e abordagens linguísticas diversas, como Análise Crítica do Discurso, Linguística Sistêmico-Funcional, Teoria da Avaliatividade, Terminologia, Tradução, Fraseologia

e Lexicologia. Ademais, os estudos aqui apresentados evidenciam a relevância do treinamento de pesquisadores para a realização de pesquisas baseadas e/ou guiadas por *corpora*.

Cabe registrar que, na autoria desta publicação, convergem outras quatro nacionalidades, para além da brasileira: Ariel Novodvorski (Argentina), Laura Alejandra Guerrero Calderón (Colômbia), María del Rosario Mestanza (Peru) e Yohanna Tamal Hernández Consoró (República Dominicana). Por essa razão, o livro apresenta dois de seus capítulos e três dos *corpora* de estudo em língua espanhola. Ainda é oportuno mencionar que alguns dos autores tiveram seu primeiro contato com a LC na disciplina *Estudos Descritivos e Linguística de Corpus* e que isso foi marcante a tal ponto que a LC passou a fazer parte, de algum modo, de pesquisas de mestrado e doutorado em que não estava prevista anteriormente. Alguns dos estudos exploratórios originados nessa disciplina acabaram firmando as bases para pesquisas de mestrado.

Desejamos uma ótima e enriquecedora leitura!

| Referências

ANTHONY, L. **AntConc** (3.5.8) [Programa computacional]. Tokyo, Japan: Waseda University, 2019. Disponível em: <http://www.antlab.sci.waseda.ac.jp/>. Acesso em: 31 maio 2021.

BARLOW, M. **ParaConc**, 1.0 (Build 269). Parallel Concordance Software. Houston, USA: Programming, ELF, Ltd., 2004.

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

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

DAVIES, M. **Corpus do Português**, 2019. Disponível em: <https://www.corpusdoportugues.org/>. Acesso em: 15 out. 2019.

KILGARRIFF, A. **Sketch Engine**, 2019 Disponível em: <http://sketchengine.co.uk/>. Acesso em: 03 nov. 2019.

NOVODVORSKI, A.; ROSA, G.; CHAGAS, L. (org.). **Ensaio em teorias linguísticas**. Uberlândia: EDUFU, 2016.

O'DONNELL, M. **UAM Corpus Tool**, versão 3.3v2. [Programa computacional]. 2019. Disponível em: <http://www.corpustool.com/index.html>. Acesso em: 31 maio 2021.

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

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

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

SCOTT, M. **WordSmith Tools** (4.0) [Programa computacional]. Liverpool: Lexical Analysis Software, 2007. Disponível em: <https://lexically.net/wordsmith/version4/index.htm>. Acesso em: 31 maio 2021.

“Tic” y “discapacidad” en un *corpus* periodístico: aproximaciones básicas a la luz de la Lingüística de *Corpus*

Yohanna Tamal Hernández Consoró¹

¹ Doctoranda en Educación por la Universidad Federal de Uberlândia, bajo la dirección de la Profa. Dra. Lázara Cristina da Silva. Bolsista por la Organización de Estados Americanos (OEA) y el Grupo Coimbra de Universidades Brasileñas (GCUB). Lattes: <http://lattes.cnpq.br/0868459113121304>. E-mail: tamara11.yh@gmail.com.

1 Introducción

Vivimos en una sociedad donde constantes mudanzas tecnológicas están aconteciendo con tan solo un abrir y cerrar de ojos, lo que es muy evidente para todos nosotros de una u otra manera. De igual forma, las computadoras también incorporan determinadas características que “acomodan” la vida de sus usuarios, o por lo menos, es lo esperado. Pero no es la computadora en sí que se está evolucionando; en realidad es el hombre quien está desarrollando otras formas de responder a determinadas situaciones que procuran mejorar la calidad de vida de los seres humanos y cuyas marcas son evidentes en todo lo que su mano toca, como es el caso de la computadora.

En el área de la Lingüística también se han desarrollado importantes y notables avances de la mano con la tecnología. Manejar grandes cantidades de textos (*corpus*) para pensar sobre cómo se está utilizando la lengua en un determinado contexto, cómo se comporta, atendiendo a determinados criterios, preguntas de investigación, etc. (BERBER SARDINHA, 2004), es uno de esos aportes y que precisamente me propongo enfocar en este trabajo.

Se trata específicamente de la Lingüística de *Corpus* (en adelante LC), la cual es definida por Berber Sardinha (2004), referencia en Brasil al introducir los estudios lingüísticos basados en *corpus*, de la siguiente manera:

La LC se ocupa de la recolección e indagación de corpus, este entendido como un conjunto de datos lingüísticos textuales recogidos criteriosamente, bajo el propósito de servir para la investigación de una lengua o variedad lingüística. En ese sentido, se dedica a la exploración del lenguaje por medio de evidencias empíricas, extraídas por computadora. (BERBER SARDINHA, 2004, p. 3)².

Una de las grandes ventajas que pude observar con relación a la implementada por la LC, con sus criterios y su forma en cómo aborda la lengua, basándose en *corpus* desde una perspectiva funcional, es precisamente el “dejar que el *corpus* hable”³. Ese es el ejercicio que me propongo realizar en este trabajo. En ese sentido, el objetivo general es: observar lo que el *corpus* de estudio nos “dice” de manera específica con relación a las frecuencias y concordancias de las

2 Traducido de “A LC se ocupa da coleta e da 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. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador”.

3 “Dejar el *corpus* hablar”, (i.e., observar lo que un *corpus* nos puede mostrar) fue una de las frases que más me impactó y que pude rescatar de las clases en la disciplina “Estudios Descriptivos y Lingüística de *Corpus* (2.º semestre, 2019) con el Prof. Dr. Ariel Novodvorski en la Universidad Federal de Uberlândia, MG, Brasil.

palabras-clave⁴ "TIC⁵" y "discapacidad", mediante el uso de herramientas de *WordSmith Tools* (en adelante WST).

Para alcanzar este objetivo general han sido trazados los siguientes objetivos específicos: (i) Planificar la recolecta del *corpus* de estudios a partir de los criterios de la LC; (ii) Recolectar y organizar el *corpus* de estudio propiamente dicho y (iii) Utilizar la suite de WST para observar las frecuencias y concordancias de las palabras-clave mediante la herramienta del software:

- a) WordList:** Lista de palabras frecuentes en el *corpus* de estudio. Serán lematizadas⁶ algunas de las palabras más repetidas, conforme la herramienta, específicamente las relacionadas con "TIC" y "discapacidad" para posteriormente observarlas;
- b) Concord:** Patrones de la lengua en uso con relación a las palabras-clave, teniendo en consideración el contexto en el cual dichos patrones están inseridos, forman parte, (i. e., su hábitat), y *clusters*.

El *corpus* de este artículo está compuesto por 82 textos, noticias *online*, en un período desde el año 2008 al 2019 y que están alojados en la base de datos del periódico *Listín Diario* en la República Dominicana, recolectados mediante búsqueda simple con las palabras "TIC" y "discapacidad". La razón por la cual fue escogido ese *corpus* es porque forma parte de una pesquisa de doctorado más amplia, para identificar cómo los medios de comunicación de masa construyen sus noticias, es decir, retratan la "realidad" sobre la relación entre las TIC y las personas con discapacidad. En este caso fue elegido el *Listín Diario*, uno de los periódicos más leídos⁷ en su versión impresa en la República Dominicana..

Para el desarrollo del análisis se tomaron como subsidios los proporcionados por el referencial teórico que nos ofrece la LC (BERBER SARDINHA, 2004), a través del tratamiento de textos usando la suite de WST, por tratarse de un primer acercamiento a la LC a partir de los aprendizajes adquiridos en el transcurso de la disciplina "Estudios Descriptivos y Lingüística de *Corpus*" en la Universidad Federal de Uberlândia (UFU).

Los resultados obtenidos nos apuntan que hay un discurso sobre el uso y acceso de/a las TIC por parte de personas con discapacidad visual, sin embargo, hay otras inquietudes suscitadas al manipular el *corpus* de estudio como: ¿De cuáles tecnologías digitales se está hablando? ¿En cuántos textos hay frecuencias de estas cuestiones? ¿Son de carácter nacional o internacional a pesar de haber sido publicadas en el periódico nacional?

4 Es importante resaltar que las palabras-clave "TIC" y "discapacidad", fueron escogidas intencionalmente por la autora debido a cuestiones relacionadas con un discurso muy proliferado en su país, República Dominicana, sobre la inclusión de las tecnologías en los procesos de enseñanza-aprendizaje. En ese sentido, la intencionalidad de la autora estuvo ligada a procurar identificar en el pequeño *corpus* de estudio, qué se dice al respecto de dichas palabras.

5 Tecnologías de la Información y Comunicación.

6 *Lematizar* es un proceso lingüístico que, dada una palabra flexionada (ej., comiendo), encuentra su lema (ej., comer).

7 Según el sitio web *Prensa Escrita* de la Rep. Dom. Acceso en: <https://www.prensaescrita.com/america/dominicana.php>. En fecha: 15 nov. 2019.

2 Fundamentación teórica

Con el objetivo de observar cómo el *corpus* de estudio se comporta y qué nos puede mostrar con relación a las palabras “TIC” y “discapacidad”, este artículo se fundamenta en los principios y criterios de la LC a partir de Berber Sardinha (2004), asumiendo una postura de ver la LC como “una perspectiva, una manera de llegar al lenguaje” (BERBER SARDINHA, 2004, p. 357).

Para llegar al lenguaje desde una perspectiva funcional es importante decidir, planificar y trabajar con un *corpus* de estudio. En ese sentido, hay criterios que precisamos considerar antes de iniciar un trabajo con el *corpus* propiamente dicho.

El autor nos presenta algunos prerequisites acerca de la selección del *corpus* de estudio, que están resumidos a continuación: i) El *corpus* debe ser compuesto por textos auténticos, en lenguaje natural; ii) Autenticidad de los textos escritos por hablantes nativos; iii) El *corpus* de estudio debe ser escogido criteriosamente y iv) Representativo: cabe preguntar: ¿representativo de qué y para quién? (BERBER SARDINHA, 2004, p. 19-20).

Es muy importante tener en mente el objetivo de la investigación que se pretende realizar. Solo así se podrán escoger textos de manera criteriosa, o sea, de acuerdo con las pretensiones de la pesquisa. Otros criterios como tipología, extensión, especificidades, adecuación, etc., del *corpus* también son necesarios considerar.

Una vez planificado cómo el *corpus* será tratado, se procede a su organización y posterior almacenamiento, siendo recomendable almacenarlo en espacios diferentes al escritorio de la computadora, como área de trabajo, muchas veces susceptible a ciertos desórdenes como consecuencia del trabajo que ahí se realiza muchas veces.

Las tecnologías digitales han causado una revolución en los diferentes estratos de la sociedad. Con el surgimiento de los medios de comunicación digitales de masa, surgieron nuevas formas de presentar la información y distribuirla. Tal es el caso de las noticias en su versión en línea. Los nuevos soportes digitales han llevado a las editoriales a re(plantear) la forma y el fondo (contenido) de lo que se distribuye con tan solo un clic a través de la gran *World Wide Web*. En ese sentido, manipular textos a través de la computadora también es efectivamente un considerable avance. Para el tratamiento y análisis de *corpus* disponemos, entre otros programas, de la suite WST 7.0 (SCOTT, 2016), “[...] escrito por Mike Scott y publicado por la Oxford University Press [...], el programa coloca a disposición del analista una serie de recursos que, si bien usados, son extremadamente útiles y poderosos en el análisis de varios aspectos del lenguaje [...]” (BERBER SARDINHA, 2004, p. 86)⁸.

⁸ Traducido de: “[...] escrito por Mike Scott e publicado pela Oxford University Press [...] o programa coloca à disposição do analista uma série de recursos que, bem usados, são extremamente úteis e poderosos na análise de vários aspectos da linguagem [...]”.

El *software* consta de varias herramientas, siendo las herramientas *WordList* y *Concord* las que se utilizaron en este trabajo y con las cuales se pueden realizar análisis observando, entre otros aspectos, la lista de palabras individuales que la herramienta ofrece (*WordList*) y líneas de concordancia (*Concord*) respectivamente.

En la próxima sección se muestran resumidamente los pasos dados para alcanzar el objetivo general.

3 Metodología

La metodología de investigación, para la producción del conocimiento, no se da por acaso ni procura “encajar” los objetivos simplemente en un patrón a ser seguido, sino que nos ayuda a organizar el conocimiento que vamos adquiriendo. En ese sentido, se presentan los pasos dados con relación a la planificación del *corpus* de estudio, su recolección y la utilización de las herramientas de WST, respectivamente.

3.1 Planificación del *corpus*

Antes de seleccionar el *corpus* de estudio fue necesario responder a preguntas como las siguientes: ¿Qué pretendo hacer con el *corpus*? ¿Qué quiero probar con él? ¿En qué formato están los textos que quiero trabajar? ¿Son legibles por el computador? etc. Esas preguntas tienen relación directa con una de las preguntas de investigación de la pesquisa en curso del doctorado que estoy realizando, de manera que esas preguntas ya fueron respondidas. El Cuadro 1 resume en parte las respuestas a los cuestionamientos arriba mencionados:

Cuadro 1: Tipología del *corpus* de estudio.

| Criterios | Características |
|--|---|
| Lengua | Español |
| Modo | Escrito |
| Tiempo | Sincrónico (Noticias del 2008-2019) |
| Selección | Muestreo (muestra de noticias en el periódico) Estático (selección no renovable) |
| Contenido | Especializado (noticias en el periódico <i>Listín Diario</i> , Rep. Dom.) |
| Autoría | Varios autores (lengua nativa) |
| Finalidad | De estudio |
| Fecha recolección <i>corpus</i> | Septiembre, 2019. |

Fuente: Elaborado por la autora.

3.2 Compilación, organización y almacenamiento del corpus

El siguiente paso fue la compilación del *corpus*. Compilar un *corpus* no es simplemente ir a una base de datos y proceder a *copiar y pegar* en un procesador de textos. En realidad, es un proceso que lleva en consideración las preguntas de investigación, además: ¿De dónde y cómo compilar? ¿Cómo voy a organizar e identificar (nominalizar) los archivos de mi *corpus* de estudio? ¿Cuál es el tamaño ideal del *corpus* considerando el/los análisis que se pretende(n) realizar? ¿Cuál será la fuente de recolección del *corpus* y si será necesaria autorización?

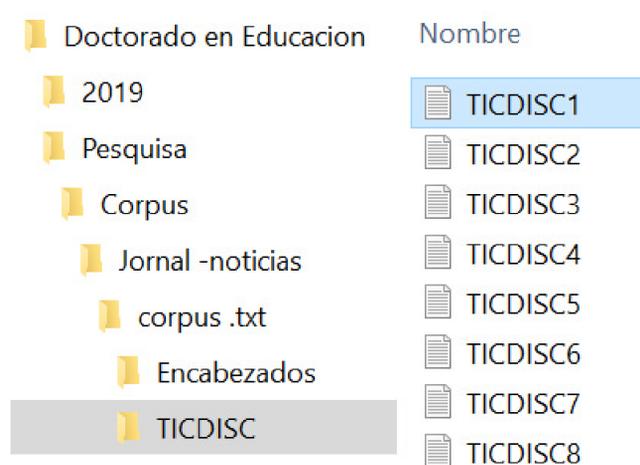
Los textos que componen el *corpus* de esta investigación son noticias extraídas de la base de datos del periódico *Listín Diario* de la República Dominicana, comprendiendo el período del 2008 al 2019 y fueron extraídos mediante “búsqueda simple” por medio de las palabras-clave “TIC” y “discapacidad”. Es, por consiguiente, representativo, lo necesariamente extenso y adecuado para las finalidades de esta investigación.

El proceso de compilación incluyó el copiar, en primer lugar, toda la información referente a los encabezados en una tabla en el programa *Word* de Microsoft: número de la noticia (esto sirvió para después nominalizar cada texto) fecha, sección, autor/a, título, línea fina, URL, fecha de consulta, y página en la cual la noticia apareció al momento de realizar la búsqueda.

El texto de cada noticia fue copiado manualmente en un programa de extensión *.txt*; en este caso fue utilizado “Bloc de notas”, programa al cual WST puede hacer una lectura y posteriormente nominalizado con la unión de las palabras-clave y el número de la noticia a seguir, por ejemplo: “TICDISC1”.

Para mantener el orden y preservación de los datos organizados con rigor y cautela, fue creada una carpeta en el “disco C” del computador con las subcarpetas donde fue alojado el *corpus* de estudio tal y como se muestra en la Figura 1:

Figura 1: Arquitectura y almacenamiento del *corpus*.



Fuente: Elaborado por la autora.

Realizados todos los procedimientos relacionados con la planificación, compilación, organización y almacenamiento del *corpus*, el siguiente paso fue el trabajo con WST, que será explicado en la siguiente sección.

3.3 Trabajando con *WordSmith Tools*

3.3.1 Lista de palabras

El primer acercamiento con WST fue generar, mediante la herramienta *WordList*, una lista de palabras con el objetivo de observar la frecuencia de palabras que más nos interesaban, o sea, "TIC" y "discapacidad". Las primeras 20 palabras son mostradas en la Figura 2 con sus respectivas frecuencias, porcentaje de ocurrencias con relación a la totalidad del *corpus* y el número de textos en los cuales ocurren, mostrando un total de 82 textos.

Figura 2: Lista de las primeras 35 palabras del *corpus* de estudio lematizadas.

| N | Word | Freq. | % | Texts | % | Dispersion | Lemmas | Set |
|----|--------------|-------|-------|-------|---------|------------|----------------|-----|
| 1 | DE | 2.858 | 7.19% | 82 | 100,00% | 0,97 | | |
| 2 | LA | 1.672 | 4,21% | 82 | 100,00% | 0,96 | | |
| 3 | QUE | 1.352 | 3,40% | 81 | 98,78% | 0,95 | | |
| 4 | Y | 1.216 | 3,06% | 82 | 100,00% | 0,97 | | |
| 5 | EL | 1.189 | 2,99% | 82 | 100,00% | 0,95 | | |
| 6 | EN | 1.075 | 2,70% | 82 | 100,00% | 0,95 | | |
| 7 | A | 859 | 2,16% | 82 | 100,00% | 0,99 | | |
| 8 | LOS | 696 | 1,75% | 78 | 95,12% | 0,95 | | |
| 9 | CON | 646 | 1,63% | 82 | 100,00% | 0,93 | | |
| 10 | LAS | 586 | 1,47% | 78 | 95,12% | 0,92 | | |
| 11 | PARA | 533 | 1,34% | 80 | 97,56% | 0,95 | | |
| 12 | SE | 425 | 1,07% | 76 | 92,68% | 0,93 | | |
| 13 | UN | 420 | 1,06% | 76 | 92,68% | 0,93 | | |
| 14 | DEL | 377 | 0,95% | 79 | 96,34% | 0,95 | | |
| 15 | UNA | 350 | 0,88% | 72 | 87,80% | 0,91 | | |
| 16 | POR | 323 | 0,81% | 70 | 85,37% | 0,96 | | |
| 17 | DISCAPACIDAD | 298 | 0,75% | n/a | n/a | n/a | disca...d[255] | |
| 18 | ES | 242 | 0,61% | 67 | 81,71% | 0,94 | | |
| 19 | PERSONAS | 288 | 0,72% | n/a | n/a | n/a | perso...[239] | |
| 20 | NO | 236 | 0,59% | 58 | 70,73% | 0,88 | | |
| 21 | SU | 208 | 0,52% | 64 | 78,05% | 0,88 | | |
| 22 | COMO | 188 | 0,47% | 70 | 85,37% | 0,97 | | |
| 23 | AL | 181 | 0,46% | 62 | 75,61% | 0,92 | | |
| 24 | O | 149 | 0,37% | 52 | 63,41% | 0,88 | | |
| 25 | MÁS | 146 | 0,37% | 47 | 57,32% | 0,87 | | |
| 26 | TECNOLOGÍA | 235 | 0,59% | n/a | n/a | n/a | tecn...[125] | |
| 27 | LO | 116 | 0,29% | 44 | 53,66% | 0,85 | | |
| 28 | EDUCACIÓN | 158 | 0,40% | n/a | n/a | n/a | educ...113 | |
| 29 | HA | 107 | 0,27% | 40 | 48,78% | 0,77 | | |
| 30 | ESTE | 106 | 0,27% | 43 | 52,44% | 0,88 | | |
| 31 | PAÍS | 104 | 0,26% | 46 | 56,10% | 0,92 | | |
| 32 | SUS | 100 | 0,25% | 51 | 62,20% | 0,88 | | |
| 33 | SON | 95 | 0,24% | 45 | 54,88% | 0,88 | | |
| 34 | NIÑOS | 113 | 0,28% | n/a | n/a | n/a | niños[91] | |
| 35 | TAMBIÉN | 80 | 0,20% | 49 | 59,76% | 0,92 | | |

Fuente: Elaborada por la autora por medio de la herramienta *WordList* de WST 7.0.

Las palabras que aparecen en la Figura 2 son las primeras que el programa muestra cuando la ventana se expande 100%, ocupando toda la pantalla. Fueron lematizadas las palabras "discapacidad", "personas", "tecnología", "educación" y "niños" por presentar otras variantes, como por ejemplo en el caso de "discapacidad": "discapacidades, discapacitante, discapacitado(s)...". Otras palabras frecuentes que suscitan curiosidad son: educación, uso y acceso, digital, personas, pacientes, trabajo, Ministerio... (que aparecen más adelante en la lista de palabras).

En la parte inferior de la ventana de *WordList* hay cinco pestañas, y una de ellas, *Statistics*, ofrece información acerca del tamaño del *corpus* en su totalidad, de manera que nos encontramos con los siguientes datos: número total de palabras en el *corpus* (*tokens*: 39,749) y palabras distintas (*types*: 6,811). El texto TICDISC19 es el menos denso con un tamaño de 882 y el más denso es el texto TICDISC20 con una densidad total de 8,488.

3.3.2 Lista de concordancias

Con el interés de identificar, aunque no de manera profunda y detallada, cómo las palabras “TIC” y “discapacidad” son utilizadas en el *corpus* y sus respectivos contextos de uso, fue realizada una búsqueda mediante la herramienta *Concord*, generada a partir de la lista de palabras (Shift+Ctrl+C) como se muestra en la Figura 3:

Figura 3: Lista de líneas de concordancias de la palabra *Tecnología*, las primeras 35 palabras del *corpus* de estudio.

| N | Concordance |
|----|---|
| 1 | La revolución científico tecnológica ha generado grandes |
| 2 | La Oficina Presidencial de Tecnologías de la Información y |
| 3 | Una novedosa tecnología basada en ondas acústicas |
| 4 | Lo último en tecnología antifalsificación son las |
| 5 | La tecnología puede cambiar el mundo y |
| 6 | La transformación tecnológica y generacional trae una |
| 7 | proyectos utilizando instrumentos tecnológicos . El aula cuenta con diversas |
| 8 | que el acceso a servicios y utilización de tecnología de apoyo pueden hacer que |
| 9 | . Y es que a través de los aparatos tecnológicos modernos las personas no |
| 10 | por estar a la vanguardia con las tecnologías de la información y |
| 11 | ley destinada a facilitarles el acceso a tecnologías que se han convertido en |
| 12 | están accesando al universo de las tecnologías y la internet. Una biblioteca |
| 13 | no tenía acceso a ningún tipo de tecnología de la información y la |
| 14 | que dificulta el acceso y uso de la tecnología es la edad, según el estudio |
| 15 | 20 años el acercamiento y manejo de la tecnología va disminuyendo. De 20 a 24 |
| 16 | . De acuerdo con el informe, las tecnologías adaptadas son aliadas de los |
| 17 | en las Américas (POETA) de los Centros Tecnológicos Comunitarios (CTC) de la |
| 18 | . También se analizó la incidencia de la tecnología en la calidad de la educación |
| 19 | , el aspecto del acceso a estas tecnologías es primordial cuando se |
| 20 | audifono , puedo escuchar”. “las nuevas tecnologías me permiten comunicarme |
| 21 | con los últimos adelantos tecnológicos . No solo se trata de buscar |
| 22 | de los grandes avances que ha tenido la tecnología . La reunión de consulta es la |
| 23 | de Belgrado , que han desarrollado la tecnología Teslagram, capaz de |
| 24 | ser que los beneficios de la revolución científico-tecnológica lleguen en |
| 25 | se espera cambios en los segmentos de tecnología de la información, condiciones |
| 26 | obras. El campo de aplicación de esa tecnología es amplio, y en un futuro |
| 27 | . En el caso de Jiménez, las herramientas tecnológicas les permiten realizar |
| 28 | sano o estar más cerca de sus hijos...la tecnología puede ver de los lados”. |
| 29 | empresa china , apuestan por “utilizar la tecnología para construir puentes entre |
| 30 | Claro . Elizondo explicó que la tecnología vino a revolucionar el mercado |
| 31 | al cliente vía telefónica de Coraasan. Tecnología La sala cuenta con siete |
| 32 | Computador , impartido por el Instituto Tecnológico Las Americas (ITLA). El |
| 33 | , tabletas y otras herramientas tecnológicas , decenas de profesores que |
| 34 | Comunitarios (CTC) para llevar la tecnología a puntos lejanos del país y |
| 35 | v servicios”. concluyó . EL USO DE LA TECNOLOGÍA Para iPhone: Jorae |

concordance collocates plot patterns clusters timeline filenames source text

212 entries T S Help

Fuente: Elaborada por la autora por medio de la herramienta *Concord* de WST 7.0.

Para finalidades del presente artículo, solo me limitaré a esas breves informaciones metodológicas y en la siguiente sección discutiré algunos de los hallazgos encontrados.

4 Resultados y discusión

Aplicando los criterios y principios de la LC en la selección del *corpus* de estudio y el uso de las herramientas *WordList* y *Concord*, fue posible observar que las palabras “TIC” y “discapacidad” se repiten más frecuentemente tal y como se obtuvo el *corpus* en la base de datos del periódico *Listín Diario*.

Según los patrones observados, cuando se habla de “tecnología”, palabra, con una frecuencia de (235) incluyendo las palabras lematizadas, se está refiriendo a las TIC efectivamente. Eso nos lleva a pensar en ¿qué pasaría si se recolecta otro *corpus* en el mismo periódico con las palabras “tecnología” y “discapacidad” en lugar de utilizar “TIC” y “discapacidad”? Se trata de nuevas tecnologías, tecnologías de la información.

Con relación a la lista de concordancias de la palabra “discapacidad”, se observa que a mano izquierda de esa palabra, con cierta frecuencia concurren las palabras “personas con”, y es algo que pude observar también en la lista de *clusters*. Aunque el enfoque en la pesquisa de doctorado, por el momento, no es identificar cómo las “personas con discapacidad” son representadas, podemos ver que es muy recurrente su uso. Tal vez eso tenga cierta incidencia con relación al uso y acceso (palabras también frecuentes y que llamaron mi atención) de las TIC por parte de las personas con discapacidad, siendo que, es lo que defiende. La discapacidad es una condición de vida de una persona y no la reducción de su identidad. Hay también una frecuencia considerable, en algunos textos, de mencionar “los discapacitados” y “personas discapacitadas”.

Se observa también que algunos textos no nos permiten identificar de cuál discapacidad se está hablando, ya que “personas con discapacidad” no nos ofrece esa información de manera puntual y específica, lo cual es una limitación, en cierto sentido, de la herramienta. En ese sentido, a partir de la observación del *corpus*, son recurrentes las siguientes discapacidades: “visual, física, auditiva”, y efectivamente la lista de colocaciones nos muestra una frecuencia de (35) para las palabras “con discapacidad visual”.

Podría(mos) decir/discutir mucho más con relación al observado en el *corpus* de estudio criteriosamente seleccionado y manipulado mediante herramientas tecnológicas como las ofrecidas por WST, sin embargo, en este artículo la limitación es hasta aquí, pasando en la próxima sección a las consideraciones finales.

5 Consideraciones finales

Aplicando los criterios y principios de la LC, y haciendo uso de las herramientas *WordList* y *Concord*, fue posible observar que las palabras-clave “TIC” y “discapacidad” nos proveen interesante información sobre lo que de ellas se dice, hace, opina, etc., discursivamente en la República Dominicana o fuera de ella, divulgado a través de noticias en línea en el periódico *Listín Diario*. Este trabajo suscitó más inquietudes que respuestas, lo que es un hallazgo que ciertamente nos permitió “dejar el *corpus* hablar” y responder a nuestro objetivo general.

Es muy satisfactorio observar cómo se está utilizando la lengua en determinados contextos, cómo se construyen significados y, de manera específica, cómo el discurso del actual gobierno de la República Dominicana, a través de su interés por crear una “República Digital”, se está construyendo discursivamente a través de los periódicos. Además podemos ver también, lo que es de dominio público y del aquí y el ahora, una vez que los periódicos nos “ofrecen luz” sobre lo que está aconteciendo. El deseo y esfuerzo por crear una escuela y sociedad inclusivas en la Rep. Dom., se ve en la promulgación de la Política Educativa de Educación Especial en la perspectiva inclusiva, de manera que es importante recordar que eso incluye las tecnologías digitales para TODOS. Pero ¿será que es así?

| Referencias

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

SCOTT, M. **WordSmith Tools version 7**. Stroud: Lexical Analysis Software, 2016. Disponible em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 20 out. 2019.

Representação discursiva do aborto na *Revista Azmina*: uma análise pautada na Linguística de Corpus e na Linguística Sistêmico-Funcional¹

Bianca Mara Guedes de 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 Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia, sob orientação da Profa. Dra. Maria Aparecida Resende Ottoni. Lattes: <http://lattes.cnpq.br/8803872713238286>. E-mail: biancamgsouza@gmail.com.

1 Introdução

Este capítulo é o recorte de uma pesquisa de mestrado em andamento no Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia. O objetivo da análise aqui apresentada é entender como o aborto é representado na reportagem “Como é feito um aborto seguro?”³, publicada em 18 de setembro de 2019 na *Revista AzMina*⁴, e assinada pela jornalista Helena Bertho. Sendo uma temática controversa, é normal que as diferentes representações da prática social do aborto sejam debatidas socialmente. A grande reportagem⁵ em questão é parte da seção de reportagens especiais no site de jornalismo independente *Revista AzMina*. Em 23 de setembro de 2019, essa reportagem foi denunciada ao Ministério Público pela Ministra da Mulher, da Família e dos Direitos Humanos, Damare Alves, por apologia ao crime.

Escolhemos realizar uma análise pautada na Linguística Sistêmico-Funcional (LSF) e na abordagem conhecida como Linguística de *Corpus* (LC). A LC “ocupa-se da coleta e exploração de *corpora*, ou conjuntos de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística” (BERBER SARDINHA, 2004, p. 3). Especialmente ligada à revolução tecnológica, a LC parte de evidências linguísticas empíricas identificadas e analisadas por meio do uso de programas computacionais. Utilizamos o programa *WordSmith Tools 7.0* (SCOTT, 2016) para o processamento do *corpus*⁶ e para a análise dos dados.

Pensando na questão da representação, fizemos um recorte e trabalharemos com a metafunção ideacional, mais especificamente com o sistema da TRANSITIVIDADE, proposto na LSF. De acordo com Fuzer e Cabral (2014, p. 33), “A metafunção ideacional é realizada por duas funções distintas: a experiencial e a lógica”. A função experiencial está relacionada às nossas representações do mundo; a função lógica refere-se aos grupos lexicais e oracionais (FUZER; CABRAL, 2014). A partir do que o *corpus* nos guiou, o foco da análise será na função experiencial do sistema da TRANSITIVIDADE.

3 Disponível em: <https://azmina.com.br/especiais/como-e-feito-um-aborto-seguro/>. Acesso em: 19 maio 2020.

4 A *Revista AzMina* é uma revista eletrônica independente. Ela pratica um jornalismo feminista especializado em mulheres e compromete-se com o ativismo de causas sociais. O acesso ao site é gratuito e ele funciona por meio de parcerias com instituições financiadoras e doações de leitoras/es.

5 Segundo Criado (2006, p. 17), “A grande-reportagem é o espaço privilegiado para a incorporação dos diversos modos de falar. Se a grande reportagem tem a ambição de aprofundar um tema, lançar uma luz sobre um fenômeno, desvendar uma realidade, ela é o gênero jornalístico por excelência para que afluam as maneiras de falar de setores excluídos econômica e culturalmente”.

6 Embora nos refiramos no decorrer do texto a *corpus*, a nomeação mais adequada seria *minicorpus* (GOULD, 2009), pois é um único texto que foi recortado dos *corpora* de nossa pesquisa de mestrado.

2 Fundamentação teórica

A LSF está preocupada em compreender a natureza e as funções da linguagem. Segundo Fuzer e Cabral (2014, p. 19), “Ela é sistêmica porque vê a língua como redes de sistemas linguísticos interligados, das quais nos servimos para construir significados”. E ela é funcional “porque explica as estruturas gramaticais em relação ao significado, às funções que a linguagem desempenha em textos” (FUZER; CABRAL, 2014, p. 19). Não podemos esquecer que todo uso que fazemos da linguagem está relacionado à nossa convivência em sociedade (HALLIDAY; MATTHIESSEN, 2014).

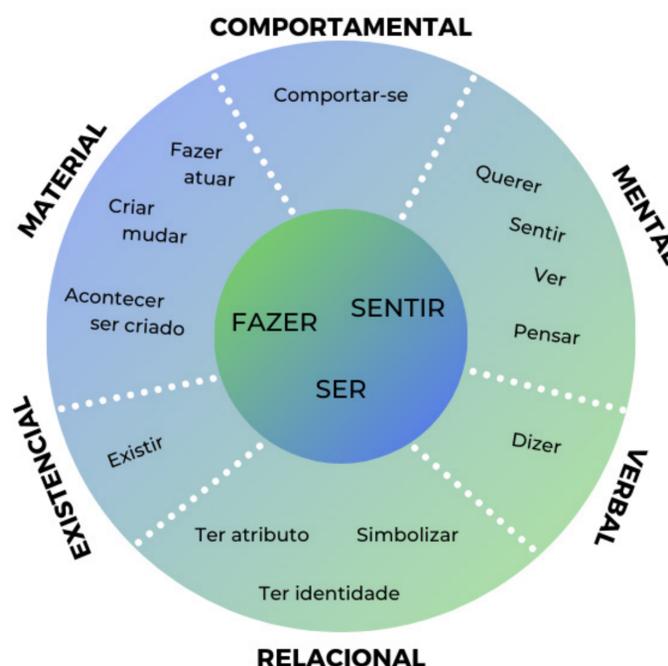
Nessa questão, ressaltamos a relação existente entre texto e contexto; a linguagem apresenta uma conexão com as variáveis contextuais. Para Halliday (1982, p. 47), “toda língua funciona em contextos de situação e pode vincular-se a esses contextos”⁷. Dessa forma, podemos afirmar que o texto sempre estará inserido em dois contextos: o de situação e o de cultura. De acordo com Fuzer e Cabral (2014, p. 27), “O contexto de situação é o ambiente imediato, no qual o texto está de fato funcionando”. Em relação ao contexto de cultura, ele representa “não só as práticas mais amplas associadas a diferentes países e grupos étnicos, mas também a práticas institucionalizadas em grupos sociais, tais como, a família, a igreja, a justiça, etc.” (FUZER; CABRAL, 2014, p. 27). Ainda em relação ao contexto de situação, Fuzer e Cabral (2014, p. 29), com base em Halliday e Hasan (1989), o descrevem “como um modelo conceitual formado por três variáveis: campo, relações e modo”. Quanto ao campo, ele equivale à atividade que está sendo realizada, ou seja, o que está acontecendo. As relações envolvem os participantes e a natureza dos papéis que eles desempenham. E, por fim, o modo refere-se à função que a linguagem desempenha e ao veículo utilizado naquele contexto situacional (HALLIDAY; MATTHIESSEN, 2014).

Após a organização do nosso *corpus*, percebemos a frequência de orações como representação, ligada à metafunção ideacional, ou seja, a experiencial. Essa metafunção se relaciona ao sistema da TRANSITIVIDADE. Segundo Fuzer e Cabral (2014, p. 40), “A transitividade é um sistema de descrição de toda a oração”, diferenciando dos estudos da gramática tradicional que estuda a relação entre os verbos e seus complementos. A oração é composta por processo, participantes e circunstâncias. O processo indica a experiência que se desenrola no decorrer do tempo. Os participantes são pessoas, coisas, seres animados ou inanimados que fazem parte do processo ou são afetados por ele. Os participantes recebem classificações específicas para cada tipo de processo (FUZER; CABRAL, 2014). E, por último, as circunstâncias que indicam o modo, o tempo, o lugar, a causa etc. (HALLIDAY; MATTHIESSEN, 2014).

⁷ No original: “Todo lenguaje funciona en contextos de situación y puede vincularse a esos contextos” (tradução nossa).

A partir dos pressupostos apresentados, decidimos analisar os processos a serem identificados no nosso *corpus*, já que esses “representam eventos que constituem experiências, atividades humanas realizadas no mundo; representam aspectos do mundo físico mental e social” (FUZER; CABRAL, 2014, p. 41). Sobre os tipos de processos, veja a Figura 1, a seguir:

Figura 1: Tipos de processos.



Fonte: Adaptado de Fuzer e Cabral (2014).

Sobre a LC, podemos colocar que essa entende a linguagem como um sistema probabilístico e está intimamente ligada “à exploração da linguagem por meio de evidências empíricas, extraídas por computador” (BERBER SARDINHA, 2004, p. 3). A concepção probabilística da linguagem é parte da espinha dorsal da LC, e parte do entendimento que, embora as possibilidades da língua sejam infinitas, o que realmente tem relevância para o estudo é a realidade linguística, ou seja, como realmente essa língua é usada por seus falantes. A compilação de um *corpus* eletrônico é basilar na pesquisa em LC, sendo *corpus* entendido como “uma parte da biblioteca eletrônica, construído a partir de um desenho explícito, com objetivos específicos” (BERBER SARDINHA, 2004, p. 16). Para realizar nossa análise, fizemos uma articulação entre a LSF e a LC.

3 Procedimentos

A escolha do texto a ser analisado foi intencional. O procedimento de coleta foi buscá-lo no *site* de origem e copiá-lo para arquivo *.doc*. Em seguida, convertemos o texto escolhido para arquivo *.txt*, e o salvamos em codificação Unicode, para que fosse lido pelo programa *WordSmith Tools 7.0* (SCOTT, 2016). O *corpus* de estudo é a grande reportagem “Como é feito um aborto seguro?”. O mesmo procedimento foi feito para o *corpus* de referência, compilado a partir das outras oito grandes reportagens da *Revista AzMina* de 2019 publicadas até o mês

de setembro. O *corpus* de estudo possui em sua totalidade 2.296 *tokens* e 719 *types*. O *corpus* de referência possui 19.522 *tokens* e 4.207 *types*.

Em seguida, o programa *WordSmith Tools 7.0* foi configurado para língua portuguesa e geramos a primeira lista de palavras, utilizando a ferramenta *WordList*. A *WordList* é uma ferramenta pré-definida “para produzir, a cada vez, duas listas de palavras, uma ordenada alfabeticamente [...] e outra classificada por ordem de frequência” (BERBER SARDINHA, 2004, p. 91). A primeira lista gerada tinha apenas a função de nos mostrar quais seriam os *lexemas*⁸ mais frequentes para a criação de uma lista de lemas, em *.txt*, para a automatização da lematização. Para a lista de lemas, foram incluídas todas as formas dos processos **ser, ter, poder, fazer e usar**, além da lematização das palavras derivadas de aborto (abortamento, abortos, abortar, aborta), mulher (mulheres), anticoncepcional (anticoncepcionais), clandestina (clandestina, clandestinas, clandestino, clandestinas), contracepção (contraceptivo, contraceptivos), Brasil (brasileiro, brasileira, brasileiros, brasileiras) e algumas palavras gramaticais.

A lista de lemas foi incorporada ao programa e, então, foram geradas as listas de palavras tanto do *corpus* de estudo quanto do *corpus* de referência. Em seguida, a ferramenta *KeyWords*, que “permite a seleção de itens de uma lista de palavras (ou mais) por meio da comparação de suas frequências com uma lista de referências” (BERBER SARDINHA, 2004, p. 96), foi utilizada para cruzar dados dos *corpora* mencionados para a apresentação de uma lista de palavras-chave. Ao escolher um *corpus* de referência de diferentes temas, e, no entanto, pertencente à mesma revista e ao mesmo gênero do discurso (grande reportagem), conseguimos um olhar ainda mais apurado sobre o que é realmente chave no texto em questão. A ferramenta *KeyWords* nos mostra que “as palavras cujas frequências no *corpus* de estudo forem significativamente maiores, segundo o resultado da prova estatística, são consideradas chave” (BERBER SARDINHA, 2004, p. 97), assim cumprindo a função de ser “representativo de certa variedade” (BERBER SARDINHA, 2004, p. 25) que, nesse caso, são as reportagens da revista digital escolhida para análise. A seguir, apresentamos um recorte da lista de palavras-chave.

⁸ Segundo Perini (2008, p. 30-32), “É o resultado de uma opção teórica – ou seja, um construto hipotético [...] diferentes formas (“palavras”) [podem] ser reunidas em uma só entidade linguística de determinado nível, que se caracteriza pelas propriedades comuns a todas elas. [...] É necessário distinguir palavras e *lexemas*, porque, em certo nível, não é possível expressar o comportamento gramatical de um *lexema*”.

Figura 2: Lista de palavras-chave.

| N | Key word | freq. in CORPUS DE | % | Texts | freq. in CORPUS DE | Rc. % | BIC | Log_L | Log_R | P |
|----|--------------|-----------------------|-------|-------|-----------------------|-------|--------|--------|--------|--------------|
| 1 | ABORTO | 55 | 2,40% | 1 | 0 | | 237,68 | 247,67 | 143,62 | 0,0000000000 |
| 2 | FAZER | 32 | 1,39% | 1 | 68 | 0,35% | 23,86 | 33,85 | 2,00 | 0,0000000030 |
| 3 | MISOPROSTOL | 20 | 0,87% | 1 | 0 | | 80,07 | 90,06 | 142,16 | 0,0000000000 |
| 4 | USAR | 15 | 0,65% | 1 | 10 | 0,05% | 26,13 | 36,12 | 3,67 | 0,0000000001 |
| 5 | PROCEDIMENTO | 14 | 0,61% | 1 | 4 | 0,02% | 34,87 | 44,86 | 4,90 | 0,0000000000 |
| 6 | OMS | 12 | 0,52% | 1 | 2 | 0,01% | 33,01 | 43,00 | 5,67 | 0,0000000000 |
| 7 | LEGAL | 10 | 0,44% | 1 | 0 | | 35,04 | 45,03 | 141,16 | 0,0000000000 |
| 8 | DREZETT | 8 | 0,35% | 1 | 0 | | 26,03 | 36,03 | 140,84 | 0,0000000001 |
| 9 | ASPIRAÇÃO | 8 | 0,35% | 1 | 0 | | 26,03 | 36,03 | 140,84 | 0,0000000001 |
| 10 | GRAVIDEZ | 8 | 0,35% | 1 | 0 | | 26,03 | 36,03 | 140,84 | 0,0000000001 |
| 11 | ÚTERO | 8 | 0,35% | 1 | 0 | | 26,03 | 36,03 | 140,84 | 0,0000000001 |
| 12 | GESTAÇÃO | 7 | 0,30% | 1 | 0 | | 21,53 | 31,52 | 140,64 | 0,000000168 |

Fonte: KeyWords.

Considerando a lista de palavras-chave, decidimos usar a lista de concordância gerada a partir do lema “aborto”, em decorrência de sua frequência e chavicidade⁹ atestadas na Figura 2. Portanto, geramos, em seguida, a lista de concordância do lema “aborto”, com a ferramenta *Concord*.

4 Análise

Buscando observar como a reportagem escolhida representa o aborto, analisamos as linhas de concordância do lema “aborto”. Embora o *corpus* de estudo tenha uma pequena dimensão, percebemos algumas estruturas que têm maior recorrência, como: **aborto + Atributo** (16 – dezesseis ocorrências), sendo **aborto legal**, 9 (nove) delas; **aborto é** (4 – quatro ocorrências); **aborto no/em + circunstância de localização** (7 – sete ocorrências). Por outro lado, algumas estruturas pouco usadas chamam a atenção, como **processo + aborto** (2 ocorrências) e **o aborto + indicativo de posse** (1 ocorrência), conforme a figura a seguir:

Figura 3: Parte da lista de concordância do lema aborto.

| | |
|----|--|
| 20 | não podem fazer denúncia de mulheres que fizeram aborto ilegal – no entanto, existem casos de mulheres |
| 21 | ou não). Nela é explicado como proceder caso haja um aborto incompleto , infecção e hemorragia, sem julgamento ou |
| 22 | uma dramática (a palavra é essa) redução de complicações por aborto inseguro ”, diz o especialista. Segundo ele, utilizar um |
| 23 | bem diferente no Brasil. Para ajudar na redução dos danos do aborto inseguro , reunimos aqui informações da Organização |
| 24 | que o de uma menstruação. A Fundação Orientame, que realiza abortos legais na Colômbia, explica que é comum ter enjoo, |
| 25 | com o uso somente do Misoprostol. No Brasil, nem para o aborto legal a Mifepristone está disponível. Também |
| 26 | em realização de aborto legal no país. A mulher que busca um aborto legal aqui no Brasil tem a possibilidade de fazer uma |
| 27 | é feita no Brasil e como o aborto é crime, fora dos casos de aborto legal é difícil ter certeza de qual método uma clínica |
| 28 | procurar um dos hospitais da rede de referência que realiza o aborto legal . Há 176 hospitais cadastrados no Ministério da |
| 29 | legal para vítimas de estupro no Brasil, segundo o Mapa do Aborto Legal . Mas ainda assim existem dificuldades para |
| 30 | Pérola Byington, em São Paulo, referência em realização de aborto legal no país. A mulher que busca um aborto legal aqui |
| 31 | de Saúde Pública da USP, Jefferson Drezett, explica que o aborto legal no Brasil é feito de acordo com essas orientações. |
| 32 | no Ministério da Saúde como provedores de serviço de aborto legal para vítimas de estupro no Brasil, segundo o |

Fonte: *Concord*.

⁹ De acordo com Gonçalves (2008, p. 390), a chavicidade “Dá a medida em que uma palavra é original e característica do *Corpus* de Estudo, em relação ao *Corpus* de Referência”.

Seguindo as orientações da LSF e com o que o nosso *corpus* nos guiou, escolhemos algumas orações para analisá-las. Levando em consideração que o sistema da TRANSITIVIDADE está relacionado à variável de registro campo, importa explicarmos o que está acontecendo no momento de interação entre os participantes:

Excerto 1: Essa é a história de como **aconteceu o aborto** da estudante de direito Rebeca Mendes, 32 anos, **em uma clínica na Colômbia**, onde o **aborto é legalizado**.

O uso da linguagem permite a produção de atribuições valorativas. Referindo-se ao *corpus* de estudo, temos o Atributo **legal**, caracterizando o aborto. Isso nos remete para uma reflexão acerca da legalidade do ato de abortar. Em específico, em se tratando de uma reportagem denunciada por apologia ao crime, podemos observar que, ao falar sobre o aborto, a jornalista faz questão de pontuar de qual tipo ela fala usando a junção **aborto + Atributo** (legal, ilegal, seguro, inseguro, clandestino, cirúrgico, incompleto), ou ainda pela utilização da estrutura **Processo Relacional + Atributo**.

No Excerto 1, a jornalista Helena Bertho retoma a narrativa realizada nos parágrafos anteriores para localizá-la no espaço, tempo e condição, por meio de uma circunstância de lugar (em uma clínica na Colômbia) e por meio da estrutura Processo Relacional (é) + Atributo (legalizado), deixando claro que trata sobre uma experiência particular da estudante Rebeca. Ao ressaltar a legalidade do que foi descrito, a jornalista já faz uma ressalva, ainda que implícita, nesse momento do texto, de que o país no qual ela escreve não compartilha da mesma situação. Além disso, o uso de um Processo Existencial (aconteceu) + Existente (o aborto da estudante de direito Rebeca Mendes, 32 anos) representa algo que existe ou acontece (HALLIDAY; MATTHIESSEN, 2014) e, em se tratando da grande reportagem em questão, é essencial esta demarcação logo no início do texto: que o aborto é um fato realizado, não somente uma ideia abstrata, e que ele faz parte da vivência da mulher. É comum que orações existenciais venham acompanhadas de circunstâncias de localização (em uma clínica na Colômbia).

Excerto 2: Nesses casos, a mulher pode procurar um dos **hospitais da rede de referência que realiza o aborto legal**.

Na continuação da reportagem, o Excerto 2 retoma a exposição feita em outros parágrafos sobre os casos permitidos por lei. O Art. 128 do Código Penal (BRASIL, 1940) indica que “não se pune o aborto praticado por médico: I – se não há outro meio de salvar a vida da gestante; II – se a gravidez resulta de estupro e o aborto é precedido de consentimento da gestante ou, quando incapaz, de seu representante legal” e a Arguição de Descumprimento de Preceito Fundamental (ADPF) nº 54, aprovada em 2012, que permitiu o aborto em caso de feto sem desenvolvimento encefálico, ou seja, anencéfalo (ARTIGO 19, 2018). Os **hospitais da rede de referência** citados

são o ator da oração; esses 76 hospitais da rede de saúde pública brasileira realizam (Processo Material Criativo) o procedimento de **aborto legal** (Meta). O aborto é representado, nesse excerto, como algo que é um direito, mas também de difícil acesso.

É importante pontuar que, só em 2017, mais de 30 mulheres viajaram de seus estados até a cidade de São Paulo (PESSOA, 2019), em busca de atendimento no Hospital Pérola Byington para casos de aborto permitidos por lei e realizados pelo Sistema Único de Saúde (SUS). Já que, muito embora, a lei postule que qualquer hospital com setor de ginecologia e obstetrícia é capaz de realizar o procedimento do aborto, muitos se recusam a fazê-lo, e outros sequer têm ciência do que aponta a legislação (ARTIGO 19, 2018).

Excerto 3: Em todas as demais situações, fazer um aborto ou ajudar em um **aborto é crime** previsto no Código Penal brasileiro e pode ser punido com prisão de um a três anos para a mulher que aborta e de até 10 anos para quem realiza o procedimento.

No parágrafo seguinte, do qual foi retirado o Excerto 3, Bertho precisa **aborto é crime** (Portador + Processo Relacional Atributivo Intensivo + Atributo), mais uma vez deixando claro sobre qual aborto ela fala. O uso do Processo Relacional Atributivo marca uma situação que não é reversível (FUZER; CABRAL, 2014). Quando pensamos na estabilidade da lei, que legisla sobre o assunto desde 1940, e as burocracias envolvidas em quaisquer mudanças legislativas, entendemos o uso dessa estrutura pela jornalista.

Excerto 4: “Independentemente do tipo de aborto, elas são maltratadas, passavam por dificuldade no atendimento. Meu estudo fez uma observação entre mulheres negras e brancas e mesmo **sendo o aborto um estigma** que **atinge todas as mulheres**, atinge de forma mais potente as mulheres negras, sobretudo as pretas”, conta.

O Excerto 4 foi escolhido para a análise, pois apresenta uma citação direta da fala de Emanuelle Goes, enfermeira e epidemiologista, que pesquisou o atendimento dado às mulheres pós-aborto. Esse tipo de fonte é classificado como fonte especialista (LAGE, 2003), contribuindo com a interpretação de contexto sobre o evento do qual se fala. Lembramos que o uso da linguagem permite a produção de atribuições valorativas e, aqui, mais uma vez temos um Processo Relacional Atributivo (sendo) + Portador (o aborto) + Atributo (um estigma). O Atributo **estigma**, utilizado por Goes para caracterizar o aborto, é de extrema relevância, por ser usado para se referir a “um atributo profundamente depreciativo” (GOFFMAN, 1988, p. 6). Além disso, na continuidade de sua fala, Goes nos diz que esse estigma **atinge todas as mulheres**. O uso do Processo Material Transformativo **atingir** é assim classificado, pois realiza uma mudança no mundo físico, reforçando a ideia de mudança no tratamento social dedicado às mulheres que passaram por um aborto, seja ele legal ou ilegal.

5 Considerações finais

O objetivo da análise apresentada foi entender como a grande reportagem “Como é feito um aborto seguro?”, do jornal digital *Revista AzMina*, representa o aborto, especialmente considerando a denúncia feita ao MP, que sugere que o texto em questão faz uma apologia ao crime. O Art. 287 do Código Penal (BRASIL, 1940), que legisla sobre isso, diz que é crime “fazer, publicamente, apologia de fato criminoso ou de autor de crime: Pena – detenção, de três a seis meses, ou multa”. O entendimento jurídico de apologia ao crime costumeiramente diz respeito a quem elogia, exalta ou enaltece crimes que já aconteceram, e não dos que podem vir a ocorrer (VIANNA, 2012). Isso posto, voltamos à análise desenvolvida na seção anterior.

A partir da observação da lista de palavras-chave, gerada pela ferramenta *KeyWords*, entendemos o que salta como diferencial no texto selecionado e o que ele tem de relevante. Embora a análise apresentada seja derivada de uma exploração inicial, com um *corpus* de pequena extensão, podemos observar a aplicabilidade da abordagem empírica na realização de análises consistentes e relevantes. A leitura das linhas de concordância permitiu a observação de ocorrências do lema **aborto** e como essas impactam a narrativa criada pela jornalista, que tem o cuidado de situar o leitor a todo momento, já que aborda um tema controverso, que esbarra em questões legais.

Os excertos escolhidos para a análise resumem as construções de representação do aborto que aparecem no decorrer do texto de Helena Bertho. Ao longo da análise, vimos que a autora representa o aborto, primeiramente, como algo legal e parte dos direitos das mulheres na Colômbia. Em seguida, Bertho começa a falar sobre a realidade brasileira, atribuindo ao aborto a qualidade legal quando relacionado aos casos permitidos por lei dentro dos hospitais da rede de referência realizado por médicos. No terceiro momento, a jornalista usa o Atributo crime, representando todos os casos de aborto fora da lei. Por fim, o Excerto 4 representa o aborto muito fortemente ligado à realidade das mulheres e como algo intrinsecamente ruim.

Ao mesmo tempo, a grande reportagem como um todo apresenta a faceta legal do aborto no Brasil, como algo de difícil acesso e, de certa forma, critica o estigma e incita a libertação das mulheres, quando pensamos em como a primeira parte (Excerto 1) se encaixa no corpo textual. A emancipação das mulheres e o interesse pelos direitos das mulheres também fazem parte da linha editorial do *site* de jornalismo independente *Revista AzMina*. Em uma sociedade na qual a condenação por apologia ao crime tem sido usada “rotineiramente como fundamento para calar a boca de quem defende mudanças nas leis” (VIANNA, 2012, p. 2), não é surpresa a denúncia da grande reportagem de Helena Bertho.

| Referências

ARTIGO 19. **Breve panorama sobre aborto legal e transparência no Brasil.** São Paulo:

ARTIGO 19, 2018. Disponível em: https://artigo19.org/wp-content/blogs.dir/24/files/2018/12/AbortoLegalTransparência_Reduzido.pdf. Acesso em: 7 maio 2020.

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

BERTHO, H. Como é feito um aborto seguro? **Revista AzMina**, [s.l.], 18 set. 2019. Especiais. Disponível em: <https://azmina.com.br/especiais/como-e-feito-um-aborto-seguro/>. Acesso em: 19 maio 2020.

BRASIL. Código Penal Brasileiro. **Decreto-lei nº 2.848**, de 7 de dezembro de 1940. Código Penal: Diário Oficial da União, 1940. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em: 10 nov. 2019.

CRIADO, A. **Falares:** a oralidade como elemento construtor da grande-reportagem. 2006. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2006. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27142/tde-04082009-212731/publico/484148.pdf>. Acesso em: 15 nov. 2019.

FUZER, C.; CABRAL, S. R. S. **Introdução à gramática sistêmico-funcional em língua portuguesa.** Campinas: Mercado de Letras, 2014.

GOFFMAN, E. **Estigma:** notas sobre a manipulação da identidade deteriorada. 4. ed. São Paulo: LTC, 1988.

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.

GOULD, T. Assessing lexical production in NNS-NNS casual conversations: a mini-corpus approach. **Sophia Junior College Faculty Journal**, Hadano, v. 29, n. 1, p. 25-45, 2009. Disponível em: <https://www.jrc.sophia.ac.jp/pdf/research/bulletin/ver2902.pdf>. Acesso em: 1 jun. 2021.

HALLIDAY, M. A. K.; HASAN, R. **Language, context, and text:** aspects of language in a social-semiotic perspective. Oxford: Oxford University Press, 1989.

HALLIDAY, M. A. K. **El lenguaje como semiótica social**: la interpretación social del lenguaje y del significado. Traducción de Jorge Ferrero Santana. México: Fondo de Cultura Económica, 1982.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **An introduction to Functional Grammar**. 4th. ed. London: Routledge, 2014.

LAGE, N. **A reportagem**: teoria e técnica de entrevista e pesquisa jornalística. 3. ed. Rio de Janeiro: Record, 2003.

PERINI, M. A. **Estudos de Gramática Descritiva**: as valências verbais. São Paulo: Parábola, 2008.

PESSOA, G. S. Mulheres têm que viajar a São Paulo por aborto legal. **Folha de São Paulo**, São Paulo, 23 fev. 2019. Cotidiano. Disponível em: <https://www1.folha.uol.com.br/cotidiano/2019/02/mulheres-tem-que-viajar-a-sao-paulo-por-aborto-legal.shtml?loggedpaywall>. Acesso em: 13 nov. 2019.

SCOTT, M. **WordSmith Tools version 7**. Stroud: Lexical Analysis Software, 2016. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 10 set. 2019.

VIANNA, T. L. O crime de apologia como instrumento de censura. **Jus.com.br**, Teresina, 25 set. 2012. Artigos. Disponível em: <https://jus.com.br/artigos/22684/o-crime-de-apologia-como-instrumento-de-censura>. Acesso em: 17 nov. 2019.

Recursos linguístico-discursivos materializados em uma reportagem *on-line* acerca de estudantes cotistas: uma análise pautada no sistema da avaliatividade e na Linguística de *Corpus*¹

Conceição Maria Alves de Araújo Guisardi²

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, sob orientação da Profa. Dra. Maria Aparecida Resende Ottoni. Professora de Língua Portuguesa e de Língua Inglesa da Secretaria de Educação do Distrito Federal. Lattes: <http://lattes.cnpq.br/0826370701965497> E-mail: cvguisardi@gmail.com

1 Introdução

O presente capítulo tem por objetivo geral apresentar uma análise, pautada nos constructos teórico-metodológicos do Sistema da Avaliatividade e da Linguística de *Corpus*, de uma reportagem *on-line*, que versa sobre estudantes oriundos de periferia e que ingressaram na universidade por meio de cotas sociais e raciais. E tem como objetivos específicos: a) identificar marcas de avaliações e suas recorrências no *corpus* analítico e b) empreender uma análise acerca das escolhas lexicogramaticais realizadas pelo produtor da reportagem, com o intuito de jogar luz nas avaliações de Atitude e seus efeitos sociais (considerando a categoria mais frequente).

Sabemos que o produtor do texto tem ao seu dispor uma variedade de recursos linguísticos que servem para avaliar coisas, eventos, comportamentos, pessoas etc. Partindo disso, empreendemos uma análise de uma reportagem que apresenta a seguinte manchete: “As pessoas não acham que alguém como eu possa ser inteligente’: a vida dos alunos da periferia na USP”³. O tema discutido na reportagem é o mesmo da minha tese de doutorado, que versa sobre a prática social que envolve o sistema de cotas. As políticas educacionais brasileiras que visam a garantir o acesso ao ensino superior precisam avançar muito, principalmente, no que tange à oferta de vagas para as minorias⁴, tais como, pobres, pretos e indígenas (PPI). No entanto, acreditamos que apenas a garantia do acesso ao ensino superior não é suficiente, é preciso garantir também a permanência desses alunos, pois existe um grande abismo social entre os alunos oriundos de escolas públicas e aqueles oriundos de escolas particulares.

A análise da reportagem está pautada no Sistema da Avaliatividade (SA) (MARTIN; WHITE, 2005) e na Linguística de *Corpus* (LC) (BERBER SARDINHA, 2004), pois acreditamos que explorar o SA em conjunto com a LC é produtivo para análises como a que será apresentada neste capítulo. Isto, porque a LC compreende a linguagem como um sistema probabilístico, parte de uma configuração sistêmico-funcional do significado nos contextos situacionais e culturais e permite um tratamento estatístico e qualitativo dos dados.

Para o processamento do *corpus*, utilizamos o *WordSmith Tools*, versão 7.0 (SCOTT, 2016), e utilizamos ferramentas desse programa para criação de lista de palavras, lista de concordância, assim como para a realização da lematização e para a inserção de etiquetas que facilitaram a

3 Disponível em: <https://www.bbc.com/portuguese/salasocial-48060977>. Acesso em: 19 maio 2020.

4 Entendemos como minorias tudo aquilo que se refere a grupos marginalizados dentro de uma sociedade e que possui ligação com aspectos econômicos, sociais, culturais, físicos ou religiosos, de gênero ou orientação sexual. Concordamos com Freitas (2007, p. 194), quando ele defende que “as minorias são, quase sempre, numericamente maiores que as majorias [...]. Por isso, o termo deve ser entendido em seu sentido sócio-antropológico, que diz ser o grupo que se encontra excluído das bases hegemônicas para limites identificatórios. Tais limites, no mais das vezes, são identificados pelos grupos dominantes e detentores de poder, privilégio e prestígio. As minorias estariam, assim, em desigualdade de direitos e oportunidades em relação aos grupos majoritários, sendo, freqüentemente, alvos de discriminação, preconceito, exclusão ou invisibilidade. Vale ressaltar que tais ações (discriminatórias) não são causa da categorização, mas consequência”.

análise dos dados. Como o *corpus* nos guiou para uma quantidade significativa do processo **ser**, seguido de atributos e circunstâncias, denotando Julgamento, é nessa categoria que centramos a nossa análise.

2 O Sistema da Avaliatividade (SA) e a categoria julgamento

O SA configura-se como uma abordagem que explora, descreve e explica a maneira pela qual a língua é utilizada para avaliar, adotar uma postura, construir *personas* textuais e lidar com posicionamentos interpessoais (WHITE, 2004). O SA foi inspirado na Linguística Sistêmico-Funcional (LSF) de Halliday (1985), que considera a linguagem como um sistema semiótico que se organiza por meio de estratos: fonologia/grafologia, lexicogramática e semântica. Segundo Vian Junior (2010, p. 21), o SA “localiza-se no estrato da semântica do discurso, e é realizado, em termos léxicos e gramaticais, no estrato da lexicogramática, oralmente ou escrito, de acordo com a interação que se desenvolve, pelo estrato grafonológico”.

Esse sistema é um dos principais recursos semântico-discursivos que está centrado na metafunção interpessoal⁵. Essa metafunção é a responsável pelas relações dos atores sociais uns com os outros. Nos diferentes contextos de interação, nós desempenhamos funções sociais, estabelecendo relações interpessoais que são responsáveis por constituir nossa identidade e nossa identificação com o outro. Ainda sobre o SA, ele é constituído por três subsistemas: ATITUDE, GRADAÇÃO e ENGAJAMENTO⁶. Para esse trabalho, o foco recaiu no subsistema ATITUDE, mais especificamente na categoria Julgamento.

Para Martin e White (2005, p. 1), o SA está “preocupado com a forma como os escritores/falantes interpretam para si identidades ou personalidades particulares, com a forma como se alinham ou se desalinham como inquiridos ou potenciais⁷”. Além disso, esses pesquisadores ressaltam que o SA procura evidenciar “como eles constroem para seus textos uma intenção de interlocutor ideal⁸” (MARTIN; WHITE, 2005, p. 1), posicionando-se de acordo com suas crenças, ideologias e valores. Em outras palavras, o escritor/falante se posiciona: (i) em relação à própria atitude, quando ele deixa marcada a sua opinião, seja negativa ou positiva, acerca de pessoas, lugares, coisas etc.; (ii) em relação ao outro, quando o produtor do texto tem a possibilidade de inserir recursos linguísticos que representam uma atitude responsiva e, por fim, (iii) em

5 A metafunção ideacional serve para representar nossas experiências, o modo como compreendemos o mundo; a metafunção interpessoal diz respeito às relações que mantemos uns com os outros, o modo como agimos em termos interacionais. E, por último, a metafunção textual está a serviço da organização dos textos, de modo a torná-los coerentes e coesos (HALLIDAY; MATTHIESSEN, 2014).

6 Para saber mais sobre o SA e sobre os subsistemas ATITUDE, GRADAÇÃO e ENGAJAMENTO, ler Martin e White (2005).

7 No original: “It is concerned with how writers/speakers construe for themselves particular authorial identities or personae, with how they align or disalign themselves with actual or potential respondents” (todas as traduções são de autoria própria).

8 No original: “[...] how they construct for their texts an intended or ideal audience”.

relação ao texto, quando é trazido do discurso do outro para o momento da interação (MARTIN; WHITE, 2005).

Em relação ao Julgamento, categoria semântica explorada nesse estudo, ele emprega recursos com o objetivo de julgar, seja de maneira positiva ou negativa, os comportamentos dos diferentes atores sociais. O Julgamento pode ser de estima social (que não possui implicações legais) ou de sanção social (que pode ter punições legais ou religiosas). No quadro a seguir, temos a explicação dos tipos de Julgamento:

Quadro 1: Categoria Julgamento.

| JULGAMENTO | |
|---|--|
| <p>ESTIMA SOCIAL “A estima social tende a ser controlada na cultura oral, por meio de bate-papo, fofocas, piadas e histórias de vários tipos – com humor frequentemente tendo como papel desempenhar uma crítica (Eggins & Slade 1997). Compartilhar valores nessa área é fundamental à formação de esferas sociais (família, amigos, colegas, etc.)⁹” (MARTIN; WHITE, 2005, p. 52).</p> | <p>1. Normalidade (o quão comum, o quão normal, alguém é). 2. Capacidade (o quão capaz alguém é). 3. Tenacidade (o quão determinado, o quão disposto alguém é).</p> |
| <p>SANÇÃO SOCIAL “A sanção social, por outro lado, é mais frequentemente codificada por escrito, como decretos, regras, regulamentos e leis sobre como se comportar conforme determinado pela igreja e pelo estado – com multas e punições contra aqueles que não cumprem o código. Compartilhando valores nesta área, é possível sustentar o dever cívico e as observâncias religiosas¹⁰” (MARTIN; WHITE, 2005, p. 52).</p> | <p>1. Veracidade (o quão verdadeiro, o quão honesto alguém é). 2. Propriedade (o quão ético, o qual moral alguém é).</p> |

Fonte: Inspirado em Martin e White (2005).

Com o intuito de mapear os recursos linguísticos utilizados para avaliar/julgar na reportagem que se constitui como nosso *corpus* analítico, recorreremos à LC. Sobre essa abordagem, discorreremos a seguir.

9 No original: “Social esteem tends to be policed in the oral culture, through chat, gossip, jokes and stories of various kinds – with humour often having a critical role to play (Eggins & Slade 1997). Sharing values in this area is critical to the formation of social networks (family, friends, colleagues, etc.)”.

10 No original: “Social sanction on the other hand is more often codified in writing, as edicts, decrees, rules, regulations and laws about how to behave as surveilled by church and state – with penalties and punishments as levers against those not complying with the code. Sharing values in this area underpins civic duty and religious observances”.

3 A Linguística de Corpus e o programa *WordSmith Tools*

A LC não é um novo tipo de Linguística, mas é tanto uma metodologia quanto uma abordagem teórica diferenciada (NOVODVORSKI¹¹, 2019). Quando examinamos a linguagem de forma extensiva, temos mais possibilidades de garantir uma análise mais consistente, mais produtiva, o que nos leva a afirmar que a LC é muito mais do que uma ferramenta, como muitos a definem. A LC “vem revolucionando a maneira como se investiga a linguagem, nos seus mais diversos níveis, colocando à disposição do analista quantidades de dados antes inacessíveis” (BERBER SARDINHA, 2009, p. 8). Nesse sentido, a LC “dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador” (BERBER SARDINHA, 2004, p. 3). E é seu caráter probabilístico, “o conhecimento da probabilidade de ocorrência de traços lexicais, estruturais, pragmáticos e discursivos [que] está no cerne da Linguística de Corpus” (BERBER SARDINHA, 2004, p. 24).

Para empreender uma análise, a LC tem a seu serviço vários programas, dentre eles, o *WordSmith Tools* (SCOTT, 2016). Segundo Berber Sardinha (2009, p. 8), “o programa *WordSmith Tools* é um conjunto de programas integrados (‘suíte’) destinado à análise linguística. Mais especificamente, esse *software* permite fazer análises baseadas na frequência e na co-ocorrência de palavras em *corpora*”. Para além disso, esse programa “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” (BERBER SARDINHA, 2009, p. 8). Para apresentarmos as funcionalidades desse programa, elaboramos um quadro, contemplando apenas as principais ferramentas utilizadas.

Quadro 2: Ferramentas *WordList*, *Concord* e *KeyWords* e suas funcionalidades.

| | |
|-------------------|--|
| <i>WordList</i> : | “Essa ferramenta “produz listas de palavra 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” (BERBER SARDINHA, 2009, p. 8) |
| <i>Concord</i> : | Por meio dessa ferramenta, podemos realizar “concordâncias, ou listagens de uma palavra específica (o ‘nódulo’, <i>node word</i> ou <i>search word</i>) juntamente com parte do texto onde ocorreu”. Além disso, ela oferece “listas de colocados, isto é, palavras que ocorreram perto do nódulo” (BERBER SARDINHA, 2009, p. 8). |
| <i>KeyWords</i> : | Por meio dessa ferramenta, é possível extrair “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” (BERBER SARDINHA, 2009, p. 8) |

Fonte: Elaborado a partir de Berber Sardinha (2009, p. 8).

¹¹ Enunciado do professor Dr. Ariel Novodvorski, em momentos de interação, na aula da disciplina intitulada Estudos Descritivos e Linguística de Corpus, realizada no âmbito do PPGEL-UFU (2019).

Tendo em vista as potencialidades do WST e nossos interesses de investigação, apresentamos os procedimentos metodológicos utilizados.

4 Procedimentos metodológicos: da apresentação à análise do corpus

Para empreendermos a nossa análise, seguimos os seguintes passos:

Passo 1: Seleção da reportagem *on-line*;

Passo 2: Limpeza e conversão do texto jornalístico para a extensão .txt;

Passo 3: Criação de uma lista de processos e atributos no bloco de notas, para realização da lematização automática;

Passo 4: Inserção dos lemas na ferramenta *WordList* (*settings – Stop-Lemma-Matchlists*);

Passo 5: Inserção do texto no programa *WordSmith Tools 7.0*;

Passo 6: Geração da lista de palavras lematizada, por meio da ferramenta *WordList*;

Passo 7: Mensuração do *corpus* por meio do recurso *Statistics* da ferramenta *WordList*;

Passo 8: Geração de linhas de concordância por meio da ferramenta *Concord*;

Passo 9: Inserção de etiquetas nas linhas de concordância – Separação das ocorrências em categorias do subsistema da Atitude: Apreciação (A) e Julgamento (J), a fim de verificarmos a frequência e escolhermos qual categoria iríamos empreender a análise;

Passo 10: Análise da categoria com maior ocorrência: Julgamento.

O *corpus* analítico possui 3.265 *tokens* e 1.105 *types* (extensão analisada por meio do recurso *Statistics*). Quando criamos a lista de palavras, percebemos uma grande frequência do processo **ser**, no entanto, apenas a análise da lista de palavras não revelou muito a respeito das avaliações da reportagem sobre o jovem de periferia que ingressa no ensino superior público por meio de cotas. Para analisarmos essas avaliações mais a fundo, geramos as linhas de concordância do processo **ser** (lematizado), por meio da ferramenta *Concord*. A lista de concordância nos forneceu todos os contextos linguísticos que antecederam ou precediam o processo **ser**, fator que contribuiu para explicitar marcas do contexto no qual esse processo ocorreu. Foi, então, que percebemos uma quantidade significativa de atributos e algumas circunstâncias¹². Assim, com o intuito de destacá-los, atribuímos cores diferentes para as duas palavras que ocorreram logo após todas as coocorrências do processo-alvo (procedimento realizado por meio do recurso *Concord Sort*). Dessa forma, conseguimos deixar em evidência os

12 As orações relacionais permitem representar seres no mundo em termos de características e identidades. No Processo Relacional, tem-se o participante que em um dado momento configura-se como um atributo, e, em outro, como um identificativo. As orações relacionais atributivas são aquelas que colaboram para construir relações abstratas de membros de uma classe. E as orações relacionais identificativas são responsáveis por estabelecerem a identidade de um participante; um dos participantes tem uma identidade determinada. As circunstanciais são aquelas que estabelecem relação de tempo, lugar, modo, causa, acompanhamento, dentre outras (HALLIDAY; MATTHIESSEN, 2014).

processos e possíveis atributos e/ou circunstâncias que antecederiam ou precediam o processo **ser**, de modo a auxiliar a análise.

Além disso, como apresentado anteriormente, inserimos etiquetas em cada linha de concordância, com o intuito de facilitar a análise dos dados. Fizemos isso porque, em alguns momentos, os atributos eram indexadores do processo **ser**, mas nem sempre se referiam à avaliação de Julgamento, mas sim de apreciação. Todas essas decisões nos guiaram com mais certeza sobre qual categoria iríamos centrar a análise. Por fim, a confirmação de que o nosso *corpus*, de fato, estava guiando para a categoria Julgamento do SA, deu-se a partir do acesso ao texto completo, que acessamos por meio de cliques duplos nas linhas de concordância do processo lematizado, verificando os recursos léxicos-gramaticais em destaque. Para direcionar nossa análise, inserimos as etiquetas “J”, para Julgamento, e “A”, para Apreciação¹³. Com isso, foi possível construir sentidos acerca de como os alunos de periferia são julgados na reportagem-alvo.

A seguir, na Figura 1 apresentamos um recorte da *WordList*, com o processo **ser** (lematizado) destacado, e, na Figura 2, apresentamos um recorte das linhas de concordância do mesmo processo **ser**, lematizado, com as etiquetas inseridas.

Figura 1: *WordList* com o processo **ser** (lematizado) destacado.

| N | Word | Freq | % | Texts | % Dispersion | Le...s Set |
|----|------|------|-------|-------|--------------|------------|
| 1 | DE | 169 | 5,18% | 1 | 1... | 0,92 |
| 2 | QUE | 116 | 3,55% | 1 | 1... | 0,88 |
| 3 | E | 90 | 2,76% | 1 | 1... | 0,94 |
| 4 | SER | 88 | 2,70% | 1 | 1... | n/aser[9] |
| 5 | A | 84 | 2,57% | 1 | 1... | 0,87 |
| 6 | TEM | 56 | 1,72% | 1 | 1... | n/ate...3 |
| 7 | O | 52 | 1,59% | 1 | 1... | 0,85 |
| 8 | NÃO | 48 | 1,47% | 1 | 1... | 0,89 |
| 9 | NA | 48 | 1,47% | 1 | 1... | 0,85 |
| 10 | É | 45 | 1,38% | 1 | 1... | 0,84 |
| 11 | EM | 43 | 1,32% | 1 | 1... | 0,86 |
| 12 | PARA | 42 | 1,29% | 1 | 1... | 0,84 |
| 13 | DA | 38 | 1,16% | 1 | 1... | 0,82 |
| 14 | EU | 31 | 0,95% | 1 | 1... | 0,78 |
| 15 | COMO | 31 | 0,95% | 1 | 1... | 0,90 |
| 16 | UMA | 30 | 0,92% | 1 | 1... | 0,67 |
| 17 | SE | 30 | 0,92% | 1 | 1... | 0,73 |
| 18 | DIZ | 30 | 0,92% | 1 | 1... | 0,83 |
| 19 | IR | 28 | 0,86% | 1 | 1... | n/a ir[3] |
| 20 | DO | 28 | 0,86% | 1 | 1... | 0,88 |
| 21 | UM | 26 | 0,80% | 1 | 1... | 0,71 |
| 22 | NO | 26 | 0,80% | 1 | 1... | 0,82 |
| 23 | AS | 26 | 0,80% | 1 | 1... | 0,84 |

Fonte: *WordList*.

¹³ Desde um trabalho inicial, exposto durante a aula sobre Linguística de *Corpus*, e da qual esse estudo é parte, identificamos que a categoria AFETO não teve uma frequência significativa no nosso *corpus*, por isso, optamos em etiquetar somente a Apreciação e o Julgamento.

Figura 2: Linhas de concordância do processo **ser** lematizado.

| N | Concordance | Set | Tag | Word # | S... | S... | P... | P... | H... | H... | S... |
|----|--|-----|-----|--------|------|------|------|------|------|------|------|
| 1 | fala, se alguém me incomoda eu sou direta ", diz ela. "Aqui não, | J | | 2... | 695 | 14 | 0..4 | | | | 0 |
| 2 | . "Sim, eu estou aqui, mas eu sou a exceção . Você não pode | J | | 2...7 | ... | 7 | 0..5 | | | | 0 |
| 3 | . "Com meus amigos eu sou muito objetiva , falo o que | J | | 2...5 | 94 | 5 | 0..3 | | | | 0 |
| 4 | "No meu caso é bem nítido (que sou da periferia) pelo meu | J | | 667 | 24 | 8 | 0... | | | | 0 |
| 5 | não são necessárias e de eles seriam o exemplo de que | J | | 2...3 | ... | 38 | 0..1 | | | | 0 |
| 6 | forma - então decidi que seria engenheira civil . Quando | J | | 90 | 126 | 088 | | | | | 0 |
| 7 | , de R\$ 400 por mês, não seria suficiente para ele parar | A | | 1...3 | 81 | 12 | 0..1 | | | | 0 |
| 8 | Direito - também estão longe de serem ideais: em algumas, | A | | 1...7 | 84 | 17 | 0..5 | | | | 0 |
| 9 | o lado positivo das pessoas serem muito unidas . "A gente | J | | 2...6 | ... | 18 | 0..4 | | | | 0 |
| 10 | que alguém como eu possa ser inteligente : a vida dos | J | | 38 | 036 | 036 | | | | | 0 |
| 11 | que alguém como eu possa ser inteligente : a vida dos | J | | 10 | 0 | 8 | 0 | 8 | | | 0 |
| 12 | ela em um texto sobre como é ser filha de uma diarista na USP, | J | | 1...2 | 55 | 16 | 0..0 | | | | 0 |
| 13 | you entrar na universidade e ser um dos únicos negros. A | J | | 2...2 | ... | 9 | 0..0 | | | | 0 |
| 14 | perder, um Bolsa Família pode ser cortado, mas a educação | A | | 1...0 | 69 | 11 | 0..8 | | | | 0 |
| 15 | como eu, com meu estilo, pode ser inteligente, pode estar | J | | 715 | 26 | 16 | 0... | | | | 0 |
| 16 | baixa renda], porque costumam ser as primeiras na família a ter | J | | 2...8 | ... | 18 | 0..6 | | | | 0 |
| 17 | no vestibular deste ano. Como é 'ser da quebrada' e estudar na | J | | 378 | 12 | 3 | 0... | | | | 0 |
| 18 | bastante simbólico de como é "ser da quebrada" e estudar na | J | | 422 | 1247 | 0... | | | | | 0 |
| 19 | de Sorocaba, no interior de São Paulo, a discriminação | A | | 2...7 | ... | 20 | 0..5 | | | | 0 |
| 20 | nas redes sociais. Em São Paulo, além de estudar, | A | | 1...4 | 76 | 2 | 0..2 | | | | 0 |
| 21 | sua cidade, Jacareí (SP), para São Paulo. Aluno de escola | ... | | 1...1 | 73 | 21 | 0..9 | | | | 0 |
| 22 | bônus de programas de inclusão são iguais ao dos outros alunos, | A | | 2...1 | ... | 22 | 0..9 | | | | 0 |

Fonte: Concord.

A Figura 1 apresenta o processo **ser** lematizado, bem como sua frequência (88 ocorrências) e a sua porcentagem em relação ao total de *tokens* do *corpus* (2,70%). Na Figura 2, o processo está marcado em azul, e alguns atributos e circunstâncias estão marcados em roxo e verde. Nem tudo em roxo e verde são atributos ou circunstâncias, mas procuramos nos guiar pela palavra de busca e pelas palavras que ocorrem duas casas à direita e uma casa à direita, respectivamente, porque mostraram mais a seguinte ordem: processo **ser** + Atributo e/ou circunstâncias (ex: "ser inteligente", "seriam o exemplo", "sou da periferia"). Partimos, então, para a nossa análise.

5 Análise do corpus

O *corpus* lematizado apresenta a frequência de 88 (oitenta e oito) vezes do processo **ser**. Desses 88 processos, 66 corroboram avaliações, estando acompanhados de atributos ou de atributos + circunstâncias, configurando em orações relacionais atributivas, relacionais identificativas e relacionais circunstanciais. Das 66 orações que apresentam marcas avaliativas, 22 são de Apreciação e 44 são de julgamento. Vejamos alguns exemplos:

Quadro 3: Exemplos de ocorrências.

| CATEGORIA | EXEMPLOS DE OCORRÊNCIAS |
|---|---|
| APRECIÇÃO (Avaliação de pessoas, coisas, lugares, eventos) | 1. A USP é muito mais do que eu esperava. 2. O ‘capital social’ é uma das grandes barreiras enfrentadas por esses alunos [de baixa renda], 3. A USP é como se fosse outro país. |
| JULGAMENTO (Avaliação de pessoas e comportamentos) | 1. Se alguém me incomoda, sou direta . 2. No meu caso, é bem nítido, que sou de periferia . 3. Os negros ainda são poucos na universidade . |

Fonte: Elaborado pela autora; excertos extraídos do *corpus* analítico.

Essas avaliações de Atitude, realizadas por escolhas lexicogramaticais, contribuem para a criação e troca de significados e para evidenciar os efeitos sociais dos recursos linguísticos que utilizamos nos textos que produzimos. Por uma questão de dimensão espacial, apresentamos apenas a análise de quatro excertos, evidenciados a seguir, mas que são suficientes para demonstrar como os alunos cotistas são julgados na reportagem-alvo. Afirmamos isso, porque, no *corpus* analisado, há uma presença expressiva de Julgamentos negativos (de estima social) em termos de capacidade (**quão capaz** os alunos cotistas são) e de tenacidade (**quão dispostos** eles estão em permanecer na universidade, diante de tantos desafios), no que se refere aos alunos de periferia que ingressaram na USP. Vejamos como os alunos cotistas são avaliados no *corpus* da reportagem.

Excerto 1: Thais afirma que, como os negros ainda **são poucos na universidade**, existe o lado positivo das pessoas **serem** muito unidas. “A gente se ajuda muito.”

A partir do Excerto 1, verificamos que há uma pouca representatividade do negro no ensino superior, marcada pelas escolhas lexicogramaticais: “são poucos na universidade”. Devemos ressaltar que somente em 2018 é que começa a ter uma mudança da representatividade do negro na universidade. Pela primeira vez, segundo pesquisa do Instituto Brasileiro de Geografia e Estatística (IBGE), o índice de alunos pardos e negros matriculados em universidades públicas alcançou o índice de 50,3%, superando, então, – embora ainda não seja o ideal – a taxa de alunos brancos (BRASIL, 2019).

A oração analisada apresenta um Processo Relacional Circunstancial/de Extensão e de Localização (são poucos na universidade), revelando uma avaliação de Julgamento de capacidade. Esse ator social é avaliado como se questionássemos: quão capaz o negro é de conseguir ingressar no ensino superior e vencer o cenário de desigualdade social? Trata-se de um Julgamento negativo, embora, tenhamos, logo após, o enunciado da participante Thaís, um contraponto em relação a essa pouca representatividade: “existe o lado positivo das pessoas serem muito unidas”. Isso é revelado por meio de um Julgamento de tenacidade.

Pelo que podemos analisar, essa estima social, esse Julgamento de tenacidade, marca quão dispostas as pessoas estão em ajudar umas às outras; assim, essas pessoas são avaliadas como solidárias, caracterizando um Julgamento positivo. Essa avaliação positiva, a certeza de que poderão ter apoio, é uma das garantias de permanência dos negros na universidade, pois essa é majoritariamente branca, como podemos perceber no discurso de vários estudantes, e que foi evidenciado no nosso *corpus* analítico, conforme apontamos na figura a seguir:

Figura 3: Recorte das linhas de concordância do processo **ser**.

| N | Concordance | Set | Tag | Word # | S | S | P | P | H |
|----|--|-----|-----|--------|-----|----|------|---|---|
| 45 | de uma universidade que ainda é majoritariamente branca. A | A | | 324 | 922 | 0 | ... | | |
| 46 | importante, diz o analista, é o capital cultural: conhecer o | A | | 2...2 | ... | 7 | 0..0 | | |
| 47 | um pouco deslocado", conta. É bem chocante você entrar na | ... | | 2...4 | ... | 1 | 0..2 | | |
| 44 | em uma fábrica na Vila Formosa. Era bem puxado, no primeiro | É | | 1...1 | 77 | 1 | 0..9 | | |
| 45 | de uma universidade que ainda é majoritariamente branca. A | A | | 324 | 922 | 0 | ... | | |
| 46 | importante, diz o analista, é o capital cultural: conhecer o | A | | 2...2 | ... | 7 | 0..0 | | |
| 47 | um pouco deslocado", conta. É bem chocante você entrar na | A | | 2...4 | ... | 1 | 0..2 | | |
| 48 | e vê essas duas realidades, é um choque tão grande. A USP | A | | 2...4 | ... | 14 | 0..2 | | |

Fonte: Concord.

São avaliações de Apreciação, que, embora essa categoria não tenha sido escolhida para ser discutida nesse estudo, contribuem para percebermos como a universidade pode ser representada, pois mostram o choque que os alunos têm ao começarem a fazer parte desse espaço, e, a própria avaliação da universidade, em termos de Apreciação, considerada majoritariamente branca, influencia no Julgamento dos cotistas negros. Passamos, então, para a análise do segundo excerto.

Excerto 2: "É por isso que todos os estudantes entrevistados pela BBC dizem ficar muito incomodados quando suas histórias são usadas por quem defende a ideia de 'meritocracia', de que políticas de inclusão não são necessárias e de que eles **seriam o exemplo** de que "quem quer consegue".

No Excerto 2, notamos que, apesar de os estudantes serem julgados como capazes, já que servirão de **exemplo** para outros graduandos, isso não é uma situação confortável para esses cotistas. Eles se sentem incomodados, pois muitas pessoas utilizam as suas histórias, as suas identidades, as suas narrativas para defenderem a meritocracia. Temos um Julgamento de capacidade (estima social), marcado pelo Processo Relacional Identificativo (**seriam o exemplo**). Esses estudantes, por terem sido capazes de conquistar uma vaga na universidade, são identificados e representados como "exemplos" para aqueles que estão tentando esse ingresso.

Excerto 3: “No meu caso é bem nítido (que **sou da periferia**) pelo meu modo de vestir. Mas faço questão de me vestir do modo da quebrada mesmo, nesse estilo chavoso (boné de aba larga, correntes, estilo típico de funkeiros)”, diz ele. ‘As pessoas de classe média não acham que alguém como eu, com meu estilo, pode **ser inteligente**, pode estar nesse espaço”.

No Excerto 3, percebemos, de forma explícita, um Processo Relacional Identificativo “que **sou da periferia**”. O uso desse processo potencializa o Julgamento de capacidade, já que o fato de ele se identificar com a periferia, de se considerar um estudante “da quebrada”, faz com que as pessoas não acreditem na sua inteligência, no seu direito de conquistar uma vaga na universidade: “As pessoas de classe média não acham que alguém como eu, com meu estilo, pode **ser inteligente**, pode estar nesse espaço”. O “peso” que esses estudantes carregam por serem oriundos de escolas de periferia, ou por residirem em periferia, é bem evidenciado em todo o *corpus* analisado. Notamos, também, um Julgamento de tenacidade, quando o participante afirma que ele está **disposto** a continuar se vestindo no que ele chama de “estilo chavoso”: “Mas faço questão de me vestir do modo da quebrada mesmo”. O estudante enfatiza o desejo de que as pessoas respeitem a sua identidade e com o que ele se identifica.

Por fim, tomamos o Excerto 4 como mais um exemplo de Julgamento negativo desses estudantes, que, por carregarem as marcas do contexto que vivem ou que estudaram, acabam sendo expurgados pela elite brasileira.

Excerto 4: “Para mim foi bem simbólico das barreiras que quem **é pobre, da periferia**, enfrenta. E se eu não **fosse aluno**, não poderia entrar? A universidade não **é pública?**”, diz ele à BBC News Brasil”.

Sabemos que esse contexto de exclusão social é marcado pela globalização. De acordo com Camino (2011, p. 7), “para entender as formas dos processos de exclusão social, devemos, portanto, analisar o contexto contemporâneo onde se desenvolvem”. Esse pesquisador defende ainda que esse “contexto é dominado pela globalização, que deve ser entendida como um conjunto de processos que se estruturam em níveis diferentes (cultural, econômico e social)” (CAMINO, 2011, p. 7). No Excerto 4, notamos uma comprovação do poder simbólico que existe entre os alunos que são oriundos de escolas de periferia e entre os alunos oriundos de escolas particulares. Esse espaço que, por muito tempo, foi “ocupado” pela elite brasileira, abre suas portas, por meio da ação afirmativa que envolve o sistema de cotas, para que atores sociais pobres, de periferia, tal qual como se identificam e são identificados, possam ingressar na universidade pública, como é questionado pelo participante do excerto analisado: “A universidade não é pública?”. Enfim, é como se dissesse: Esse espaço é nosso, alunos de escolas públicas, negros, pobres.

Enfim, esse contexto de exclusão, de luta por justiça social, não é exclusividade da USP, representada na reportagem analisada. Em várias instituições públicas de ensino superior do país, temos visto diversos depoimentos, materializados em textos que circulam socialmente, de jovens que se sentem inferiores em relação a outros estudantes e que se sentem julgados o tempo todo. Ressaltamos que os textos contêm traços e pistas de questões sociais e são permeados de relações de poder (VAN DIJK, 2018), de avaliações (MARTIN; WHITE, 2005), e podemos perceber tudo isso por meio da linguagem, que é uma forma de prática social, que evidencia uma relação bidirecional entre textos e sociedade (FAIRCLOUGH, 1992).

6 Considerações finais

A partir da análise apresentada ao longo deste capítulo, percebemos avaliações negativas acerca de estudantes oriundos de periferia e de negros na universidade. Sabemos que existem várias pesquisas que dão destaque para essa questão do racismo e preconceito no ensino superior, no entanto, seria inocente da nossa parte se desconsiderássemos o papel da linguagem nisso tudo. Segundo van Dijk (2015, p. 34), “A maior parte das pesquisas sobre racismo focaliza formas de discriminação e exclusão, ou preconceitos e ideologias, mas ignora o papel fundamental da linguagem, do discurso e da comunicação na reprodução do racismo nas sociedades modernas”. Por meio da reportagem analisada, assim como em tantos outros textos que circulam na mídia, é possível perceber imagens preconceituosas, estereotipadas, sexistas ou racistas (VAN DIJK, 2018).

A geração de uma lista de palavras, a lematização e a inserção de etiquetas nas linhas de concordância nortearam e foram profícuas para a percepção da escolha e da frequência de recursos linguístico-discursivos na reportagem analisada, auxiliando diretamente na compreensão e análise dos enunciados. Portanto, acreditamos que os estudos pautados na LC e na LSF, que têm caráter probabilístico, nos auxiliaram sobremaneira, afinal, a quantidade de atributos negativos na reportagem analisada evidenciou o quanto os cotistas são julgados, em um espaço que democraticamente seria deles. Se fôssemos realizar o tratamento do *corpus*, apenas com o olhar humano, sem a ajuda de uma ferramenta, encontraríamos mais dificuldades e seria exigido mais tempo para se deleitar sobre o que estava sendo investigado.

Apesar de este capítulo apresentar somente uma análise exploratória, acreditamos que o constructo teórico e os procedimentos metodológicos adotados podem contribuir para outras pesquisas, motivando-nos a investir em novas empreitadas que corroborem diferentes temáticas sociais sob o viés da LSF, do SA e da LC, possibilitando análises quali-quantitativas dos dados.

| Referências

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

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

BRASIL. Instituto Brasileiro de Geografia e Estatística. **Desigualdades sociais por cor ou raça no Brasil**. Rio de Janeiro: IBGE, 2019. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101681_informativo.pdf. Acesso em: 25 maio 2020.

CAMINO, L. Prefácio. In: TECHIO, E. M.; LIMA, M. E. O. (org.). **Cultura e produção das diferenças: estereótipos e preconceito no Brasil, Espanha e Portugal**. Brasília: Technopolitik, 2011. p. 7-12.

FAIRCLOUGH, N. **Discourse and social change**. Cambridge: Polity Press, 1992.

FREITAS, R. O. A periferia da periferia: mídias alternativas e cultura de minorias em ambientes não-metropolitanos. **Especiaria**, Ilhéus, v. 10, n. 17, p. 191-212, 2007. Disponível em: <http://periodicos.uesc.br/index.php/especiaria/article/view/866>. Acesso em: 1 mar. 2021.

HALLIDAY, M. A. K. **An introduction to Functional Grammar**. London: Arnold, 1985.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **An introduction to Functional Grammar**. 4th. ed. London: Routledge, 2014.

MARTIN, J. R.; WHITE, P. R. R. **The language of evaluation: appraisal in English**. London: Palgrave Macmillan, 2005.

MORI, L. 'As pessoas não acham que alguém como eu possa ser inteligente': a vida dos alunos da periferia na USP. **BBC News Brasil**, São Paulo, 29 abr. 2019. Sala social. Disponível em: <https://www.bbc.com/portuguese/salasocial-48060977>. Acesso em: 19 maio 2020.

SCOTT, M. **WordSmith Tools version 7**. Stroud: Lexical Analysis Software, 2016. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 16 maio 2020.

VAN DIJK, T. A. Discurso das elites e racismo institucional. In: LARA, G. P.; LIMBERTI, R. P. (org.). **Discursos e (des)igualdade social**. São Paulo: Contexto, 2015. p. 31-48.

VAN DIJK, T. A. **Discurso e poder**. Tradução Judith Hoffnagel e Karina Falcone. São Paulo: Contexto, 2018.

VIAN JUNIOR, O. O Sistema da Avaliatividade e a linguagem da avaliação. *In*: VIAN JUNIOR, O.; SOUZA, A. A.; ALMEIDA, F. S. D. P. (org.). **A linguagem da avaliação em língua portuguesa: estudos sistêmico-funcionais com base no sistema de avaliatividade**. São Carlos: Pedro & João, 2010. p. 19-30.

WHITE, P. R. R. Valoração: a linguagem da avaliação e da perspectiva. **Linguagem em (Dis)curso**, Tubarão, v. 4, n. esp., p. 177-205. 2004. Disponível em: http://www.portaldeperiodicos.unisul.br/index.php/Linguagem_Discurso/article/view/295/314. Acesso em: 17 maio 2020.

La exclusión de actores sociales en reportajes periodísticos: un análisis del asesinato de la líder social colombiana María del Pilar Hurtado¹

Laura Alejandra Guerrero Calderón²

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

² Alumna de maestría en Estudios Lingüísticos por la Universidade Federal de Uberlândia, bajo la dirección de la Profa. Dra. Maria Aparecida Resende Ottoni. Lattes: <http://lattes.cnpq.br/5977709827888478>. E-mail: laaguerrero@gmail.com.

1 Introducción

El presente trabajo es un análisis de las formas de representación/exclusión de actores sociales identificadas en un corpus de seis reportajes periodísticos, tomados de tres periódicos digitales colombianos, que presentan uno de los ya más de 700 casos de homicidio de líderes sociales ocurridos en el país en los últimos años. Las medidas que se han tomado por parte del actual gobierno han sido muy cuestionadas, puesto que las cifras continúan en aumento y los resultados de las investigaciones han sido nulos, la mayoría de los casos permanecen impunes.

El objetivo de este trabajo es analizar las formas de representación, **Supresión** y/o **Encubrimiento** en reportajes periodísticos colombianos de actores sociales que podrían estar involucrados en el caso de asesinato de María del Pilar Hurtado, líder social de Tierralta, Córdoba, asesinada el 21 de julio de 2019. Dentro de los posibles implicados señalados en los textos, se identificaron dos tipos: entidades ilegales (narcotráfico, grupos armados ilegales, disidencias de las FARC – Fuerzas Armadas Revolucionarias de Colombia) e instituciones del estado (el ejército, la policía, la alcaldía municipal), esos últimos fueron investigados por posible relación con el crimen y rápidamente absueltos de forma poco justificada.

De este modo, se pretende analizar una muestra de cada tipo para identificar cómo los tres periódicos nacionales: *El Tiempo*³, *El Espectador*⁴ y *El Heraldo*⁵ representan a dichos actores. Para alcanzar ese objetivo se utilizará la categoría *exclusión* y sus dos variantes, **Supresión** y **Encubrimiento** de la Teoría de Representación de Actores Sociales (RAS) de Van Leeuwen (2008). Así como los presupuestos teórico-metodológicos de la Lingüística de Corpus (LC), según Berber Sardinha (2009) y Shepherd, Berber Sardinha y Pinto (2012); y el Análisis Crítico del Discurso (ACD), planteado por Fairclough (2003; 2006) y Chouliaraki y Fairclough (1999).

2 Presupuestos teóricos

Las tres bases teórico-metodológicas de este análisis tienen afinidades entre sí, siendo la principal de ellas su carácter crítico en el campo de la investigación. Además, todas consideran el discurso como su objeto de estudio, ya que, como apuntan Chouliaraki y Fairclough (1999), forma parte de las prácticas sociales y, por lo tanto, es una forma de aproximarse a la investigación en el ámbito social. Esos autores entienden el discurso como uno de los momentos de las prácticas sociales, donde interactúa de forma dialéctica con otros elementos como son, las creencias, conocimientos, experiencias de los participantes, entre otros (CHOULIARAKI; FAIRCLOUGH,

3 Disponible en: <https://www.eltiempo.com/>. En fecha: 10 jul. 2019.

4 Disponible en: <https://www.elespectador.com/noticias/>. En fecha: 10 jul. 2019.

5 Disponible en: <https://www.elheraldo.co/>. En fecha: 10 jul. 2019.

1999). El ACD y la RAS también establecen un diálogo en el sentido que ambas reconocen un gran valor a la Lingüística Sistémico-Funcional (LSF) de Halliday (1985), de la cual se sirven en sus análisis de la RAS. Fairclough (2003) define a los actores sociales como aquellos que participan en el texto de forma activa, es decir, un objeto directo no podría considerarse actor social en una oración.

Por otro lado, van Leeuwen (2008) define las representaciones como formas de incluir o excluir actores sociales, entendidos por su importancia socio semántica, y así, su distribución social. Ambos reconocen que las formas de representación responden a intereses relacionados con la información que se quiere o no que reciba el lector. Van Leeuwen (2008) hace una clasificación mucho más exhaustiva que denomina categorías socio semánticas. Su inventario se compone de 23 categorías, divididas en dos sistemas, *INCLUSIÓN/EXCLUSIÓN*, que conducen a varios sub-sistemas. Para los propósitos de este análisis se adoptará únicamente la categoría de *Exclusión* y sus sub-sistemas **Supresión** y **Encubrimiento**. Considerando esos sub-sistemas, las categorías socio semánticas dispuestas para este análisis, se tomará como guía la adaptación al español y portugués hecha por Novodvorski (2013).

En cuanto a la LC, se asume como el método de investigación base de este análisis. Por medio de la lectura de líneas de concordancia, se espera identificar y analizar los actores que presenten mayor frecuencia e intentar demarcar patrones de su representación en el corpus escogido. Según Shepherd (2012) ese proceso empírico que recurre al corpus en busca de respuestas es la práctica que dio origen a la LC. La misma autora señala que la LC tiene preferencia por el análisis de corpus extensos, ya que ofrecen resultados más confiables a nivel cuantitativo. Sin embargo, también puede utilizarse con corpus pequeños, siempre y cuando estos sean una muestra significativa del fenómeno que se estudia, como es el caso de este análisis. La LC también ha guardado una estrecha relación con el ACD y le ha brindado un gran aporte, la rigurosidad analítica. Shepherd (2012) expone que los estudios basados en el ACD son comúnmente cualitativos, sin embargo, los usos instrumentales de la LC permiten identificar la frecuencia y el tipo de léxico utilizado en determinados textos y grupos sociales, lo cual puede apuntar a trazos de naturaleza ideológica.

3 Metodología

Siguiendo los presupuestos teóricos ya presentados, se diseñó una propuesta que incluye: el componente crítico y las categorías de análisis del ACD y la teoría de RAS; apoyadas en el *software* de lectura de corpus, *WordSmith Tools* (WST en adelante), en su versión 7 (SCOTT, 2016), y sus herramientas para listado de palabras (*WordList*) y listado de líneas de concordancias (*Concord*). El proceso fue el siguiente: Primero, se hizo recopilación del corpus. Se escogieron

seis reportajes de noticias de tres periódicos nacionales: *El Tiempo*, *El Espectador* y *El Herald*. Dichos textos se escogieron según los siguientes criterios: su fecha de publicación (la semana después del evento); popularidad de las fuentes; y su naturaleza de reportaje, todos los textos presentan un análisis de los hechos, no apenas la noticia del ocurrido. Para lograr un equilibrio en el corpus, se escogieron dos textos de cada periódico, los cuales fueron todos encontrados en las páginas *web* oficiales de las fuentes. Se escogieron diferentes periódicos para facilitar la identificación de elementos excluidos de los textos, ya que, como menciona Van Leeuwen (2008), en el caso de la supresión total del actor social, no se encuentran trazos de él en un solo texto y para su identificación debe recurrirse a la comparación del mismo evento en diferentes textos.

Una vez seleccionados los textos, es necesario hacer la limpieza del *corpus*, donde se convierte cada texto a formato *.txt*, retirando todas las imágenes e *hiperlinks*, se rotula con su título y fuente para facilitar su comparación. Luego se suben los textos al programa WST, donde se crean la lista de palabras y, a partir de esta, las listas con líneas de concordancia de cada uno de los actores sociales identificados, según la delimitación de categorización ya planteada. Antes de la creación de esas últimas listas, se realizó un proceso de lematización en la lista de palabras, donde se crean conjuntos de lexemas mayores como son: (narcotráfico, narcotraficantes, narcos), también se eliminaron palabras gramaticales (artículos, preposiciones y conjunciones) con altos números de ocurrencia que se consideraron irrelevantes para el análisis. El tamaño del corpus es de 3.835 ítems (*tokens*) y 1.105 vocablos (*types*). Si bien se identificaron alrededor de 12 actores sociales involucrados en el corpus, se ha decidido delimitar este análisis, por cuestiones de extensión, a solo dos listas de concordancia (“orden público” y “grupos ilegales”). A continuación, se presenta el análisis de esas listas.

4 Análisis

Este análisis se apoyará en las categorías socio-semánticas de *Supresión* y *Encubrimiento* planteadas por Van Leeuwen (2008) y adaptadas al español y al portugués por Novodvorski (2013). El siguiente Cuadro expone la composición de dichas categorías:

Cuadro 1: Recorte de la teoría RAS adaptada al español

| | Categorías Socio-semánticas | Definición | Cómo se realiza |
|--|------------------------------------|--|---|
| E X C L U S I Ó N | Supresión | Exclusión total del actor social | -Eliminación del agente de la pasiva. -oraciones infinitivas que funcionan como un participante gramatical |
| | Encubrimiento | Exclusión parcial del actor social (existen trazos del actor en alguna otra parte del texto) | -Eliminación de los beneficiarios -nominación o nombres procesuales -adjetivos -voz media -elipsis (aplicada únicamente para encubrimiento) |

Fuente: Van Leeuwen (2008) e Novodvorski (2008)

Como apunta Novodvorski (2013), las representaciones lingüísticas de **Supresión** y **Encubrimiento** que han demostrado ser aplicables al español y el portugués son: 1) la eliminación del agente de la pasiva; 2) las oraciones infinitivas; 3) la eliminación de los beneficiarios; 4) la nominación o nombres procesuales; 5) los adjetivos; y 6) la perspectiva de ergatividad, la voz media. Esos serán, entonces, los trazos lingüísticos que se esperan identificar en este análisis.

El primer grupo analizado es el relativo a los actores vinculados a la legalidad, para su análisis se hizo una lista utilizando la herramienta *Concord* con el término "orden público", que presentó siete líneas de concordancia.

Figura 1: Lista de concordancia: orden público.

| | |
|--|-----------------------|
| detalles de la situación de orden público en la localidad, donde la | el tiempo, críticas a |
| que se derivan del problema de orden público en el sur de Córdoba. "Las | el tiempo, críticas a |
| administración en el manejo del orden público habrían derivado en el | el espectador, |
| irregularidades en el control del orden público originadas por la invasión de | el espectador, |
| irregularidades en el control del orden público originadas por la invasión de | El tiempo, investigan |
| hubo mal manejo en el control del orden público por invasión de lotes. El alcalde | El tiempo, investigan |
| negligencia en el manejo del orden público que derivó en la muerte de | el tiempo, críticas a |

Fuente: Programa WST.

Llamó la atención ese término por la relación directa que establece con el homicidio y porque es en sí misma una nominación de encubrimiento de otros actores relacionados al crimen, ¿quién alteró el orden público? El término también representó una alta frecuencia (siete ocurrencias) en comparación con otros actores; se tomaron cuatro fragmentos, de forma aleatoria, para hacer un análisis más detallado.

Cuadro 2: Ejemplos extraídos de lista de concordancia “orden público”

| | |
|---|---------------|
| La Procuraduría indaga si hubo mal manejo en el control del orden público por invasión de lotes. | El Tiempo |
| ...alcalde de Tierralta Fabio Otero, a quien la Procuraduría abrió investigación disciplinaria por supuesta negligencia en el manejo del orden público que derivó en la muerte de María del Pilar Hurtado. | El Tiempo |
| El Ministerio Público busca determinar si fallas en la administración en el manejo del orden público habrían derivado en el homicidio de la lideresa social | El Espectador |
| La Procuraduría abrió investigación en contra del alcalde de Tierralta, Córdoba, Fabio Leonardo Otero Avilés, por presuntas irregularidades en el control del orden público originadas por la invasión de unos lotes en el municipio, para analizar si estas habrían derivado en el asesinato de la mujer. | El Espectador |

Fuente: corpus de investigación.

Se encontraron varias formas de exclusión, los cuales se listan a continuación:

- **Nombres procesuales:** el término “orden público” es un **Nombre Procesual** que en términos generales se entiende como el normal funcionamiento de una sociedad, garantizado por los límites que imponen entidades del estado, como son la policía o el ejército. Sin embargo, se observa una supresión parcial del agente (Encubrimiento), ya que esas entidades se mencionan en otras partes de los textos, pero no en las oraciones relacionadas con las acusaciones. Además, el término se encuentra acompañado de las frases “negligencia en el manejo de...”, “mal manejo en el control de...”, “fallas en el manejo de...”, “presuntas irregularidades en el control de...”, lo cual se identifica como otra estrategia de encubrimiento representada con **nombres procesuales**, que indican una actividad, igualmente, sin agente. En tres ocurrencias, “orden público” está sucedida de los verbos “derivó en/que habría derivado en (el asesinato/homicidio/muerte de la líder social)”. Es decir, que el actor social que se investiga como culpable del homicidio es representado como tan solo una actividad hecha de forma irregular, sin ningún agente específico.
- **Adjetivos:** Se identificaron los adjetivos “supuesta” y “presuntas”, utilizados para calificar la operatividad irregular que la alcaldía le dio a los hechos. El autor que atribuye esos calificativos no es claramente representado, ¿es el periódico?, ¿las unidades de investigación? En ese sentido, es otra manifestación de exclusión.

Además, es importante mencionar que solo dos de los tres periódicos hicieron referencia al “orden público”. *El Herald*o, por su parte, se refirió a la situación de la siguiente manera

“... responsabilidades disciplinarias frente **al manejo que**, desde la Alcaldía, presidida por Fabio Otero Avilez, **se le haya dado al tema** de entrega de lotes a familias invasoras.”

Allí, se omite la prosodia semántica negativa de las palabras “fallas”, “irregularidades” o “problemas”, como sugirieron los otros medios, optando por el término “manejo (dado al tema)”. Además, acusa de responsable a “la alcaldía” (referencia a una entidad colectiva), “presidida por Fabio Avilez” (referencia específica del actor social), pero ese actor es alejado del proceso con el uso del morfema **SE de pasiva refleja** en la frase, “al manejo que... se le haya dado al tema”. Es decir, alguien es el responsable por el “manejo... dado al tema”. Aunque el actor social implicado sí se menciona, es representado de forma encubierta, donde el actor se presenta junto a lo sucedido, pero es ambiguo si él ejecutó la acción o no, el “se” lo convierte en un agente indeterminado.

Seguido de eso, la frase preposicional que se presenta inmediatamente después “...de entrega de lotes a familias invasoras” es también ambigua, ya que utiliza un Nombre Procesual “entrega de lotes” donde se excluye también el agente, mientras el beneficiario “familias invasoras” se representa con un adjetivo negativo. Así, el texto sugiere que la acción realizada (por una entidad no mencionada, podría inferirse que es la alcaldía) es un acto positivo (entrega de lotes) y los beneficiarios (familias desplazadas) acarrean un atributo negativo (“invasores”).

Se considera importante resaltar que, según fuentes externas como el noticiero del CANAL 1, *Noticias 1*, las irregularidades en el orden público mencionadas por esos reportajes se refieren principalmente a dos aspectos: 1) la población desplazada (llamados “invasores”) que ocupaba un terrero, propiedad del padre del alcalde, fue atacada por el Escuadrón Móvil Antidisturbios (ESMAD), que actuó bajo la orden del alcalde e incumplió protocolos de seguridad. Esas acciones fueron motivo de investigación, considerando que haber utilizado ese escuadrón para desalojar una población vulnerable que incluía mujeres y niños y que además estaban reclamando su derecho legítimo de reubicación de vivienda, fue una acción desproporcionada; y 2) Hubo inoperancia de la alcaldía, ya que la defensoría del pueblo ya le había informado sobre la existencia de panfletos que amenazaban a esa población desplazada, sin embargo, no se tomó ninguna acción preventiva ni de protección.⁶

El segundo grupo para analizar es el relacionado con las acusaciones hechas a actores sociales vinculados a la ilegalidad, para eso se realizó una lista de concordancia lematizada semánticamente con los términos “ilegal(es)/ilícito(s)” de la cual se obtuvieron ocho ocurrencias, como se observan en la siguiente imagen.

⁶ Véase en: <https://www.youtube.com/watch?v=w-DfxGjxZpA>. En fecha: 10 jul. 2019.

Figura 2: Lista de concordancia lematizada semánticamente: “ilegal(es)/ilícito(s)”

| N | Concordance | File |
|---|---|-----------------------|
| 1 | 'Clan del Golfo' para mover rentas ilícitas y es por eso que en el último mes | el heraldo, |
| 2 | para poder mantener los negocios ilegales". "El compromiso del | el heraldo, comisión |
| 3 | también repudió el accionar de grupos ilegales relacionados con el narcotráfico | el tiempo, críticas a |
| 4 | y la vinculación de todos esos grupos ilegales, empezando por el 'Clan del | el heraldo, comisión |
| 5 | exponencial de los cultivos ilícitos que han pasado de 400 | el heraldo, comisión |
| 6 | la presencia de grupos armados ilegales en esa zona, fue | el espectador, se |
| 7 | en la capital no hay grupos armados ilegales, sino bandas de delincuencia | el espectador, se |
| 8 | Colombia (AGC), la estructura armada ilegal que hace presencia en esta región | el espectador, |

Fuente: Programa WST.

La primera observación que salta a la vista es que, contrario al caso anterior, *El Heraldo* es la fuente que más utiliza ese término, siendo autor de cuatro de las ocho ocurrencias, *El Espectador* menciona tres veces ese término y *El Tiempo* tan solo una vez. Al observar detenidamente el contenido del texto, se encontró que eso se debe a la prioridad que *El Heraldo* da a las citas textuales de la senadora Paloma Valencia, quien pertenece al partido político afín al del alcalde del municipio, el mismo del presidente Iván Duque. Los siguientes ejemplos muestran cómo su discurso está cargado de referencias acusativas a los grupos ilegales y de defensa a las autoridades, en las cuales se incluye a ella misma, creando una clara distinción entre **ellos** (todos los vinculados a la ilegalidad) y **nosotros** (las autoridades, la legalidad). Por el contrario, el alcalde, quien se supone hace parte del grupo de la legalidad, pero tiene una investigación en su contra por su posible vinculación al crimen, es completamente excluido en el discurso de la senadora.

- "nos preocupa mucho el clamor de la ciudadanía por seguridad y la vinculación de todos esos **grupos ilegales**, empezando por el '**Clan del Golfo**', **las disidencias de las Farc** y **cantidades de bandas** que están en el territorio".

- generan que todas esas **estructuras criminales** ocupen el territorio, asesinan los líderes sociales para tomar control de las comunidades e impiden la entrada del Estado para poder mantener los **negocios ilegales**".

El Tiempo también se refiere al comentario de la senadora sobre los grupos ilegales, sin embargo, se presenta en discurso indirecto utilizando los verbos (ella) "repudió" y "culpó" (el accionar de esos grupos). En cuanto a *El Espectador*, sus referencias, "estructura armada ilegal" y "grupos armados ilegales", no establecen ningún vínculo directo con el homicidio, sino que se mencionan en una sección anexa al reportaje que intenta hacer un análisis más extenso de la propagación que han tenido esos grupos en otras zonas del país.

5 Consideraciones finales

Después de hacer un análisis que buscaba contrastar las representaciones de actores sociales de dos posiciones opuestas en el ámbito político, el “orden público” (representación de la legalidad) y “grupo(s) ilícito(s)/ilegal(es)”, se puede concluir que, a pesar de tratarse de un análisis aún en estadio inicial, permite observar una clara diferenciación entre las formas de representación de ambos grupos. Por un lado, las categorías de exclusión, *Encubrimiento* y *Supresión* fueron muy recurrentes en la lista de concordancia de orden público, donde *los nombres procesuales* y *adjetivos* fueron las formas más utilizadas para encubrir actores como son el alcalde del municipio, las fuerzas militares y demás autoridades investigadas por el homicidio de la líder social. Por otro lado, la representación de actores vinculados a lo ilícito y lo ilegal se manifiestan de forma reiterativa, lo cual se entiende como otra forma de encubrimiento que busca distraer la atención del lector y apartarlo de las investigaciones que vinculan a las autoridades ya mencionadas.

Además, es importante resaltar la gran contribución de las herramientas del programa WST y los principios metodológicos de la LC. Esos permitieron una identificación rápida y efectiva de los actores representados y, a su vez, facilitaron la delimitación de este análisis, siguiendo marcas cuantitativas, como fue la frecuencia de los términos reportada en los textos, lo cual da mayor fiabilidad a los resultados. Se espera continuar ahondando en este análisis y seguir explorando las demás categorías de van Leeuwen (2008), bajo las pautas del empirismo de la LC.

| Referencias

BERBER SARDINHA, T. **Pesquisa em Lingüística de Corpus com WordSmith Tools**.

Campinas: Mercado de Letras, 2009.

CHOULIARAKI, L.; FAIRCLOUGH, N. **Discourse in late modernity**: rethinking critical discourse analysis. Edinburgh: Edinburgh University Press, 1999.

FAIRCLOUGH, N. **Analysing discourse**: textual analysis for social research. London/New York: Routledge, 2003.

FAIRCLOUGH, N. **Language and globalization**. London/New York: Routledge, 2006.

HALLIDAY, M. A. K. **An introduction to Functional Grammar**. London: Arnold, 1985.

NOVODVORSKI, A. **A representação de atores sociais nos discursos sobre o ensino de espanhol no Brasil em corpus jornalístico**. 2008. Dissertação (Mestrado em Estudos Linguísticos) - Faculdade de Letras. Universidade Federal de Minas Gerais, Belo Horizonte, 2008.

NOVODVORSKI, A. Representação de atores sociais. *In*: MAGALHÃES, C. M. (org.). **Representação social em corpus de tradução e mídia**. Belo Horizonte: Editora UFMG, 2013. p. 13-48.

SCOTT, M. **WordSmith Tools version 7**. Stroud: Lexical Analysis Software, 2016. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 17 maio 2019.

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 das Letras, 2012. p. 15-30.

SHEPHERD, T. M. G.; BERBER SARDINHA, T.; PINTO, M. V. (org.). **Caminhos da Linguística de Corpus**. Campinas: Mercado das Letras, 2012

VAN LEEUWEN, T. **Discourse and practice: new tools for critical discourse analysis**. New York: Oxford University Press, 2008.

O AFETO em depoimentos sobre adoção: uma análise da avaliatividade com subsídios da Linguística de *Corpus*¹

Layane Campos Soares²

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

2 Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia, sob orientação da Profa. Dra. Maria Aparecida Resende Ottoni. Lattes: <http://lattes.cnpq.br/2461261839263924>. E-mail: layanecsoares@gmail.com.

1 Introdução

A adoção³ é uma problemática bastante desafiadora no Brasil, conforme comprovam os dados disponibilizados no Cadastro Nacional de Adoção⁴ (2019). Em virtude dessa complexidade, é muito comum nos depararmos com práticas de incentivo e apoio à adoção tardia, por exemplo. Uma dessas práticas se inscreve no contexto do projeto *Quero uma família*, desenvolvido pelo Ministério Público do Rio de Janeiro, que contou com a colaboração de diferentes atores sociais que enviaram depoimentos⁵ acerca do processo de adoção, especialmente, aqueles que adotaram crianças e adolescentes que não se enquadravam no perfil usualmente desejado pela maior parte dos adotantes ou que utilizaram a busca ativa enquanto recurso para adoção.

Desse modo, o intuito principal da publicação desses depoimentos no *site* do projeto foi o de incentivar outras pessoas a adotarem crianças e adolescentes cujo perfil não se enquadra dentro do que é geralmente desejado, dando visibilidade, de certa forma, a essa prática. Por essa razão, ficamos instigados a compreender: i) Como essas crianças e adolescentes são avaliados nesses depoimentos pelos adotantes? ii) Essas avaliações são marcadas por AFETOS ou JULGAMENTOS? iii) Como a Linguística de *Corpus* pode contribuir para a identificação dessas marcas?

Diante dessas indagações, propomos empreender uma análise acerca dos depoimentos disponibilizados no *site Quero uma família*, com a finalidade de compreender como as crianças e os adolescentes adotados são avaliados pelos adotantes nos depoimentos publicados no *site*. Para tanto, ancoramos no Sistema da Avaliatividade de Martin e White (2005), tendo em vista que esse sistema apresenta técnicas sistemáticas para a análise de textos, fato que pode contribuir para um melhor entendimento das funções sociais, sentimentos e posições assumidas pelos atores em determinados contextos interacionais. Além disso, procuramos identificar qual categoria do Subsistema da ATITUDE aparece com mais frequência no *corpus* de análise, a partir do uso do programa *WordSmith Tools 7.0* (SCOTT, 2016). Para realizarmos essa identificação, recorreremos à Linguística de *Corpus*, mais especificamente, aos escritos de Berber Sardinha (2004, 2009). Assim, somente a partir da coleta e exploração de nosso *corpus* é que conseguimos verificar a recorrência de determinados processos e atributos, o que nos levou a perceber que os depoimentos analisados estavam carregados de AFETO expressos pelos pais em relação aos seus filhos adotados.

3 A temática abordada neste capítulo é um recorte de uma pesquisa de doutorado em desenvolvimento.

4 Dados disponíveis para consulta em: <https://www.cnj.jus.br/cnanovo/pages/publico/index.jsf>. Acesso em: 20 maio 2020.

5 Depoimentos disponíveis em: <http://queroumafamilia.mprj.mp.br/depoimentos>. Acesso em: 20 maio 2020.

2 Sistema da Avaliatividade

O Sistema da Avaliatividade é uma teoria que tem o intuito de compreender os domínios semânticos que operam no discurso, adotando categorias que englobam diversas estruturas gramaticais dentro de um grupo semântico discursivo (WHITE, 2004). Segundo White (2004), esse sistema procura entender as posições de valor adotadas pelos indivíduos socialmente, de modo a perceber suas filiações ou distanciamentos em determinados contextos comunicacionais.

Para a realização da categorização dessas avaliações, Martin e seus colaboradores propuseram o Sistema da Avaliatividade (*Appraisal Framework*), que nos permite perceber, por meio da perspectiva sistêmico-funcional, os vários recursos linguísticos que podem ser empregados e os possíveis caminhos para se analisar o modo como esses recursos ocorrem na língua. Ressaltamos que o Sistema da Avaliatividade é um dos componentes da metafunção interpessoal da LSF de Halliday (1985) que, junto ao Envolvimento e à Negociação, compõem a Semântica do discurso. Martin e White (2005) tomaram como base essa metafunção e criaram o Sistema da Avaliatividade, a fim de oferecer subsídios para explorar o discurso no contexto das avaliações, revelando os tipos de atitudes que são negociados em um texto, bem como a força dos sentimentos envolvidos (MARTIN; ROSE, 2003). Em razão do nosso objeto de estudo, focaremos no Subsistema da ATITUDE, que está relacionado aos sentimentos e emoções que expressamos, aos julgamentos de caráter e às avaliações de valor, podendo ser categorizadas em três campos semânticos diferentes: AFETO (emoção), JULGAMENTO (ética) e APRECIÇÃO (estética). Neste capítulo, daremos destaque ao AFETO, pois foi a categoria mais recorrente no *corpus* analisado.

Nesse sentido, podemos apontar que o AFETO é uma categoria de análise que denota valores que são expressos na forma de qualidades (atributos), de processos e de circunstâncias. O AFETO também pode ser realizado por meio de entidades virtuais, ou seja, por nominalizações (WHITE, 2004). Enquanto categoria, o AFETO envolve a presença de um participante consciente, que, de alguma forma, é afetado emocionalmente por um fenômeno ou entidade, sendo responsável por desencadear a emoção (MARTIN; WHITE, 2005). Desse modo, o AFETO é visto como um recurso semântico empregado para a realização das emoções linguisticamente no discurso. Para Martin e White (2005), o AFETO pode ser identificado sob três aspectos: como qualidade, como processo e como comentários.

De acordo com White (2004, p. 186, grifos nossos), “as emoções podem ser reunidas em três grandes grupos ligados à **in/felicidade, in/segurança e in/satisfação**”. Assim, a variável in/felicidade está relacionada às emoções ligadas aos “assuntos do coração” – felicidade, amor, tristeza e ódio, por exemplo. A variável da in/segurança, por sua vez, está relacionada

às emoções ligadas ao bem-estar ecossocial – medo, ansiedade e confiança, dentre outros. Por último, temos a variável in/satisfação, que está relacionada às emoções ligadas ao *telos*, isto é, a busca de objetivos – curiosidade, respeito, tédio, desprezo, dentre outros sentimentos (WHITE, 2004).

3 Linguística de Corpus

A Linguística de *Corpus* é uma abordagem que está inscrita em um quadro conceitual formado por uma abordagem empirista, assumindo uma visão da linguagem como sistema probabilístico (BERBER SARDINHA, 2004). Segundo Berber Sardinha (2004), no contexto dos estudos linguísticos, o empírico significa que os dados são decorrentes da observação da linguagem, sendo geralmente reunidos sob a forma de *corpus*.

Ao adotar uma visão de linguagem enquanto um sistema probabilístico, a Linguística de *Corpus* se filia à linguística hallidayana, tendo em vista que ela se pauta na descrição dos sistemas linguísticos, em função dos contextos em que os falantes os empregam (BERBER SARDINHA, 2009). De acordo com Berber Sardinha (2004, p. 30-31), essa visão de linguagem “pressupõe que, embora muitos traços lingüísticos sejam possíveis teoricamente, não ocorrem com a mesma frequência”. Isso implica no fato de que a diferença de frequência entre os traços não é algo aleatório (BERBER SARDINHA, 2004). Em virtude de a diferença de frequência entre os traços ser significativa, é feito um mapeamento regular entre a frequência maior ou menor de um traço, conjuntamente com o seu contexto de ocorrência, conforme aponta Berber Sardinha (2004).

Para a realização desse mapeamento, o linguista de *corpus* explora os *corpora* por meio de programas e ferramentas computacionais. Existem vários *softwares* que podem auxiliar o linguista de *corpus*, dentre eles, o *WordSmith Tools*. Segundo Berber Sardinha (2009, p. 8), “o programa *WordSmith Tools* é um conjunto de programas integrados (“suíte”) destinado à análise linguística”. O uso desse programa permite ao linguista de *corpus* fazer análises quali-quantitativas com base na frequência e na coocorrência de palavras em *corpora*. Esse programa dispõe de algumas ferramentas, a saber: *WordList* (gerador de lista de palavras), *Concord* (gerador de linhas de concordância) e *KeyWords* (gerador de palavras-chave). Essas ferramentas são fundamentais para se realizar o mapeamento da frequência e da ocorrência de algumas palavras no *corpus* de análise, tornando-se, assim, um importante mecanismo de exploração e descrição linguísticas. Dentre as ferramentas disponíveis pelo programa, desenvolvemos nossa análise a partir do uso da *WordList* e do *Concord*, conforme veremos na próxima seção.

4 Recursos

O *corpus* de análise é composto por dezessete (17) depoimentos que foram produzidos por diferentes atores sociais com a finalidade de narrar as suas experiências sobre a adoção. Esses depoimentos, como indicado na seção introdutória, foram retirados do site *Quero uma família*.

Para a realização do procedimento de análise, foram necessárias algumas etapas:

- ⇒ Etapa 1: Seleção dos textos que constituiriam o *corpus* de análise.
- ⇒ Etapa 2: Limpeza e conversão dos textos para o formato .txt, utilizando o bloco de notas nessa etapa.
- ⇒ Etapa 3: Utilização do programa *WordSmith Tools 7.0*.
- ⇒ Etapa 4: Leitura do *corpus* por meio das ferramentas e funcionalidades do *WordSmith Tools* para a extração de dados e análise de resultados.

Após a inserção dos textos no programa, foi possível verificar a extensão do *corpus*, que é de 10.214 palavras totais e de 2.468 palavras diferentes. Para delimitarmos o enfoque da nossa análise, realizamos um exame preliminar dos dados, por meio da geração de uma lista de palavras, com a finalidade de percebermos quais eram as palavras mais recorrentes no *corpus*, bem como as suas frequências de ocorrência. Em seguida, lematizamos o *corpus* semiautomaticamente, a partir da elaboração de uma lista de lemas em formato .txt, com base em nomes (substantivos), processos (verbos) e atributos (adjetivos) mais recorrentes no *corpus*, sendo agrupadas em conformidade com suas flexões e variações.

A lematização foi elementar para visualizarmos, de fato, quais eram os processos, atributos e nomes mais frequentes. Posteriormente a essa etapa, utilizamos a ferramenta *Concord* para visualizarmos melhor o modo como essas palavras foram empregadas no *corpus*, juntamente com a parte do texto onde elas ocorreram.

5 Análise

Com base no nosso objetivo de análise, a partir do exame preliminar do *corpus* de estudo, decidimos focalizar em um único processo. Dessa forma, ao gerarmos a lista de palavras lematizada, percebemos que o processo “ser” é o mais recorrente no nosso *corpus*, tendo uma frequência de 180 ocorrências. Apesar de termos delimitado nossa análise ao processo mais recorrente, acreditamos ser relevante descrevermos o *corpus* de forma geral, apontando outras palavras que também tiveram uma frequência significativa. Assim, o segundo processo mais recorrente é “ter”, com 145 ocorrências, seguido dos processos “ir”, com 114 ocorrências,

e “adotar”, com 112 ocorrências. Além disso, temos na sequência a circunstância de modo “não”, que ocorreu 93 vezes no *corpus*. Sob o nosso ponto de vista, a presença constante de alguns processos pode estar relacionada com a própria construção de narrativas acerca do processo de adoção, além de contribuir para a construção de identificações e representações sobre o(a) filho(a).

Sobre a presença do processo “ser” no nosso *corpus*, notamos que ele pode indicar tanto avaliações de AFETO quanto de JULGAMENTO, pois os Processos Relacionais Atributivos, a depender do uso, podem materializar esses dois tipos de avaliações. Assim, quando esse processo está relacionado ao campo das emoções e sentimentos, ele denota AFETO como qualidade; por outro lado, quando ele está baseado em JULGAMENTOS de caráter, ele se materializa como JULGAMENTO.

De modo a identificarmos essas avaliações, focamos somente nas ocorrências que estavam relacionadas ao nosso objetivo de análise, excluindo as demais. Como resultado, das 180 ocorrências do processo “ser”, somente 34 se referiam às crianças e aos adolescentes adotados. A partir disso, fizemos uma análise dessas ocorrências com a finalidade de percebermos se as avaliações mais recorrentes eram de AFETO ou JULGAMENTO. Posteriormente à análise, passamos para o procedimento de inserção de etiquetas na coluna “Set” da lista de concordâncias do processo “ser” gerada pela ferramenta *Concord*, na qual a letra A se refere à categoria AFETO e a letra J ao JULGAMENTO, conforme podemos notar na Figura 1.

Figura 1: Lista de concordâncias do processo “ser” com as etiquetas inseridas

| N | Concordance | Set | Word3... | ... | ... |
|----|--|-----|--------------|-----|-----|
| 1 | mais em nós e, principalmente, percebia o quanto era amada. Nossa filha, agora Leticia Sofia, nossa Lelê, | A | 3662413 | 0 | |
| 2 | Após certa hesitação, visto que o menino - Alyson - não era da idade que tínhamos imaginado, aceitamos e | J | 5652612 | 0 | |
| 3 | em 2009 pela 4 VIJ Capital quando tinha a época 5 anos. Era uma criança que já tinha sido devolvida três vezes e | J | 29 2 1 02 | | |
| 4 | então sentir o mesmo que eu já tinha sentido, de que era sim, a nossa Isabela, nossa filha. Passamos o final | A | 210 637 | 0 | |
| 5 | seis anos de idade, pesando apenas 15 Kg, meu menino era tão frágil que eu tinha medo de machucá-lo, se | J | 125 513 | 0 | |
| 6 | e carinhosa. Naquele momento já sabíamos que aquela era a nossa filha, independente das questões médicas, | A | 28510 7 | 0 | |
| 7 | e na mesma hora que a peguei em meu colo, senti que era a minha filha. Apesar do meu marido se sentir com | A | 170 519 | 0 | |
| 8 | talvez pelo fato de serem negros, e a minha filha Sabrina ser deficiente física (não possui o antebraço e escoliose | J | 86 348 08 | | |
| 9 | e com baixa autoestima, vivia assombrado pelo medo de ser abandonado. Eu o coloquei em terapia pouco depois | A | 199 712 | 0 | |
| 10 | . Por fim, novamente perguntei se ainda tinha medo de ser devolvida ao abrigo e ela deu de ombros e disse que | A | 6644010 | 0 | |
| 11 | sido chamados como por conta própria. Apesar de ser criativo, vivaz e líder, Alyson também é curioso, tem | J | 3...6 ... 3 | 0 | |
| 12 | não havia pretendentes para eles, talvez pelo fato de serem negros, e a minha filha Sabrina ser deficiente | J | 79 341 07 | | |
| 13 | sonhos, eu e Ernani de sermos pais e Léo e Sabrina de serem filhos! Nossa família extensa aceitou super bem | A | 176 615 | 0 | |
| 14 | com seus 3, 4 anos que estão destituídas, prontas para serem adotadas. Isto nos fez buscar informações, | J | 147 365 | 0 | |
| 15 | deu ao ler essas palavras e falei pro meu marido "olha são nossas filhas!". Embora a postagem fosse do mês | A | 318 865 | 0 | |
| 16 | mais cor e mais fé, pq ela é a prova viva de que o amor é o mais forte dos laços de uma família. Il | A | 3241319 | 0 | |
| 17 | crescimento pessoal tem sido intenso. Sua docilidade é um encanto. Aos domingos, vamos a evangelização que | A | 50832 3 | 0 | |
| 18 | vi ela receber nosso sobrenome, vi ela me ensinar o que é amor incondicional, me ensinar a ser mais tolerante e | A | 143 331 | 0 | |
| 19 | filha chegou... Hoje tudo tem mais cor e mais fé, pq ela é a prova viva de que o amor é o mais forte dos laços de | A | 3161311 | 0 | |
| 20 | . Apesar de ser criativo, vivaz e líder, Alyson também é curioso, tem dificuldade em se concentrar e "se | J | 3...3 ... 10 | 0 | |
| 21 | se lembrar. Então afirmamos sempre: Você é nossa filha e a queremos muito! As vezes até | A | 57626 5 | 0 | |

Fonte: WordSmith Tools 7.0.

Após os procedimentos de análise, inserção de etiquetas e classificação dos resultados a partir de categorias, foi possível respondermos as nossas duas inquietações apresentadas na seção introdutória deste capítulo. De acordo com a análise, encontramos no nosso *corpus* 34 ocorrências do processo “ser” que estavam relacionados a avaliações dos adotantes acerca de seus filhos adotados. Dessas 34 ocorrências, 25 dizem respeito a avaliações de AFETO e 9 a

avaliações de JULGAMENTO. Dessa forma, podemos afirmar que essas crianças e adolescentes são avaliados, predominantemente, a partir de emoções ligadas aos assuntos do coração, com 12 ocorrências, seguida de emoções ligadas ao *telos*, com 9 ocorrências, e, também, de emoções ligadas ao bem-estar ecossocial, com apenas 4 ocorrências.

Todas as avaliações de AFETO ligadas aos assuntos do coração foram de felicidade; as avaliações relacionadas ao *telos*, por sua vez, foram todas de satisfação; enquanto as avaliações voltadas ao bem-estar ecossocial foram de segurança e de insegurança. Apesar de a ocorrência de AFETOS materializados pela segurança e insegurança ter uma menor frequência, consideramos que eles são bastante representativos em nosso *corpus*. Isso se dá em virtude de duas razões: a) a necessidade de se sentir seguro é fundamental no contexto de adoções bem-sucedidas; b) a insegurança pode estar relacionada, na maior parte dos casos, aos traumas vivenciados por muitas crianças e adolescentes que estiveram em situação de adoção. Acreditamos que esses traumas têm uma relação direta com a violação de direitos dos menores por parte de seu tutor, fazendo com que eles fiquem sob a responsabilidade do Estado, passando a morar em instituições de acolhimento. Esses traumas marcam substancialmente a vida das crianças e adolescentes, culminando em sentimentos, por exemplo, insegurança e medo do abandono. Para exemplificar essa situação, vejamos a Oração 1, a seguir: “[...] vivia assombrado pelo **medo** de ser **abandonado**” (DEPOIMENTO 1).

Na Oração 1, podemos perceber que o AFETO de insegurança é uma avaliação negativa, que se realiza como qualidade atribuída ao participante por meio de um Processo Relacional Atributivo, em que se tem o processo “ser” e o atributo “abandonado”. Além disso, nessa mesma oração há a escolha linguística do nome “medo”, sendo visto como algo que caracteriza o Processo Relacional Atributivo “ser abandonado”, descrevendo, de certa forma, o participante. O uso desse nome confere uma situação que aterroriza a vida dessa criança e/ou adolescente que foi adotado, fazendo com que ele se sinta inseguro. Essa insegurança é decorrente, sob o nosso ponto de vista, dos traumas vivenciados na própria infância, ocasionando emoções que abarcam sentimentos de ansiedade e incerteza não somente em relação ao ambiente familiar, mas também acerca das pessoas que o rodeiam.

Conforme mencionamos anteriormente, a maior parte das avaliações é de emoções ligadas aos assuntos do coração, denotando sentimentos de felicidade. Acreditamos que as crianças e os adolescentes adotados foram avaliados a partir de sentimentos que caracterizam o amor, carinho, alegria, dentre outros, com a finalidade de mostrar o quão realizados estão os pais com o processo de adoção e quão amáveis são os filhos “do coração”. Nesse contexto, podemos afirmar que essas avaliações mantêm uma relação intrínseca com o próprio ato de adotar, tendo em vista que ele implica uma reconstrução de laços familiares para crianças e adolescentes

que foram privados de permanecerem com as suas famílias biológicas, o que proporciona o estabelecimento de relações paterno-filiais. O estabelecimento dessas relações no âmbito da adoção se dá por meio de uma família substituta/adotiva que, segundo Lisboa (2013, p. 310), é aquela que passará a “constituir laços de afinidade e de afeto com o menor”.

Com base nesse preceito, notamos que o vínculo afetivo é fundamental nas relações familiares, podendo ser visto como um importante pilar no processo de adoção, uma vez que é por meio do afeto que se constrói laços efetivos que, por sua vez, contribuem para a solidificação de novos territórios familiares. Para exemplificar essa situação, vejamos as Orações 2 e 3, a seguir: Oração 2 - “[...] é uma filha **amorosa** [...]” (DEPOIMENTO 14); Oração 3 - “Sua **docilidade é um encanto**” (DEPOIMENTO 3).

Na Oração 2, podemos perceber que se trata de um Processo Relacional Identificativo, no qual a filha é identificada pelo uso do atributo “amorosa”, o que caracteriza uma avaliação de AFETO de felicidade, pois esse atributo faz parte da região de sentimentos ligados ao “coração”. Ao avaliar afetivamente a filha adotiva como uma pessoa amorosa, os pais externalizam simbolicamente o sentimento que construíram em relação à filha, de forma a mostrar que filhos adotivos são amáveis, dando visibilidade ao processo de adoção.

Em relação à Oração 3, temos uma avaliação de AFETO que é materializada por meio de um Processo Relacional Atributivo, no qual qualifica-se a filha adotada como uma pessoa que encantou os pais pela sua doçura. Assim, notamos que os pais adotivos construíram uma avaliação positiva em relação à filha, mostrando-nos a consolidação de um vínculo afetivo que, sob o nosso ponto de vista, é algo fundamental no processo de recolocação do menor em uma família substituta.

No contexto da adoção, podemos apontar que o estabelecimento e a manutenção de vínculos afetivos são elementares para a consolidação de uma adoção bem-sucedida, tendo em vista que, quando há um rompimento desse vínculo, os filhos adotivos são encaminhados novamente para instituições de acolhimento. Nesse processo de construção de vínculos, é importante que os pais saibam lidar com a história e com os traumas do filho adotado, de modo a compreender sua trajetória de vida ao invés de negá-la.

6 Considerações finais

A realização da análise descrita na seção anterior nos possibilitou refletir acerca da nossa terceira indagação, que diz respeito às contribuições da Linguística de *Corpus* para a identificação das marcas de AFETO no *corpus* analisado. Diante do percurso trilhado, notamos que a Linguística de *Corpus* foi elementar para a constituição e descrição do *corpus* analisado, indicando-nos possíveis caminhos para realizarmos a análise. Por meio do programa utilizado, o

WordSmith Tools 7.0, conseguimos delimitar a categoria a ser empregada na análise do *corpus*. É importante salientarmos que o *WordSmith Tools* é um programa que nos auxilia a visualizar não somente as ocorrências que aparecem com maior frequência, mas também aponta para o quão significativas podem ser as ocorrências pouco frequentes. Assim, o uso das ferramentas *WordList* e *Concord* foram fundamentais no procedimento de identificação da categoria mais produtiva para a análise, segundo o objetivo principal da nossa investigação.

Empreendemos uma análise exploratória sobre o modo como as crianças e adolescentes adotados são avaliados nos depoimentos retirados do site *Quero uma família*. Para tanto, nos baseamos no processo que ocorreu com maior frequência no *corpus*, o que nos levou a perceber a predominância de avaliações marcadas pelo AFETO. Com base nessas questões, construímos nossa análise de modo a elaborarmos uma reflexão acerca da avaliação marcada pelo AFETO no contexto da adoção, baseando-nos na expressividade das ocorrências diante da problemática analisada.

A partir do estudo realizado, acreditamos que a articulação entre a Linguística de *Corpus* (BERBER SARDINHA, 2004, 2009) e o Sistema da Avaliatividade (MARTIN; WHITE, 2005) foi muito profícua, pois nos possibilitou construir importantes considerações sobre a prática de adoção no Brasil.

| 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.

BRASIL. **Conselho Nacional de Justiça** (CNA). Cadastro Nacional de Adoção. Disponível em: <https://www.cnj.jus.br/cnanovo/pages/publico/index.jsf>. Acesso em: 20 maio 2020.

HALLIDAY, M. A. K. **An introduction to Functional Grammar**. London: Arnold, 1985.

LISBOA, R. S. **Manual de Direito Civil: direito de família e sucessões**. 8. ed. São Paulo: Saraiva, 2013.

MARTIN, J. R.; ROSE, D. **Working with discourse: meaning beyond the clause**. London/New York: Continuum, 2003.

MARTIN, J. R.; WHITE, P. R. R. **The language of evaluation: appraisal in English**. London: Palgrave Macmillan, 2005.

MINISTÉRIO PÚBLICO DO RIO DE JANEIRO. **Quero uma família**. Disponível em: <http://queroumafamilia.mprj.mp.br/apresentacao>. Acesso em: 20 maio 2020.

SCOTT, M. **WordSmith Tools version 7**. Stroud: Lexical Analysis Software, 2016. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 17 maio 2019.

WHITE, P. R. R. Valoração: a linguagem da avaliação e da perspectiva. **Linguagem em (Dis)curso**, Tubarão, v. 4, n. esp., p. 177-205. 2004. Disponível em: http://www.portaldeperiodicos.unisul.br/index.php/Linguagem_Discurso/article/view/295/314. Acesso em: 17 maio 2020.

Saúde de corpo e alma: análise em corpus de programa radial

Luana Duarte Silva¹

¹ Doutoranda em Estudos da Linguagem pela Universidade Federal de Catalão, sob a orientação da Profa. Dra. Vanessa Regina Duarte Xavier.
Lattes: <http://lattes.cnpq.br/8436774877328143>.
E-mails: luana_duarte@ufg.br; duarte.luana01@gmail.com.

1 Introdução

Neste capítulo, trabalhamos com a aplicação do programa *WordSmith Tools* versão 4.0 (doravante WST) (SCOTT, 2004), no estudo exploratório de um *corpus* oral (transcrito), que aborda a temática da saúde física, mental e espiritual, compilado do programa da Rádio Globo intitulado *Saúde de Corpo e Alma*². Esta análise é oriunda das indagações suscitadas pela minha pesquisa de doutoramento “*Saúde de Corpo e Alma: análise de campos lexicais*”, que está em andamento.

Recorremos à Linguística de *Corpus* para verificar a ocorrência das unidades lexicais presentes nos dois primeiros episódios do programa radial em análise, com o propósito de verificar a maior frequência das palavras, em especial, substantivos, adjetivos e verbos, que podem carregar sentidos referentes à temática tratada no programa de rádio em questão.

A apresentadora descreve a missão do programa, em sua primeira edição, da seguinte forma: “A gente vai falar aqui de saúde física, vai, vai sim, vai falar de saúde do corpo. E vai perceber como essa saúde tá muito mais ligada do que a gente pensa com aquilo que a gente sente, com aquilo que tá na nossa cabeça e com aquilo que tá no nosso coração, na nossa alma” (RÁDIO GLOBO, 2017). Frente a esta proposição, surgem algumas questões: será que a saúde física tem relação com a mente e com o espiritual? É preciso equilíbrio para viver com qualidade de vida? Os sentimentos podem interferir na saúde física e mental? As palavras podem evidenciar essa relação da saúde física, mental e espiritual?

Dessa forma, derivada dessas questões, formulamos a seguinte hipótese: a missão do programa (saúde do corpo, da mente e do espírito) é evidenciada nas ocorrências mais frequentes das unidades lexicais. A frequência das unidades lexicais foi observada a partir da ferramenta *WordList* do WST, que contribuiu para o processamento, leitura e análise dos dados coletados. Acreditamos também que poderemos compreender, por meio de tais ocorrências, se realmente é preciso ter um equilíbrio de mente, corpo e espírito, para viver com qualidade.

2 Fundamentação teórica

A Linguística de *Corpus* é uma área que tem crescido nos últimos anos, exercendo “grande influência na pesquisa lingüística” (BERBER SARDINHA, 2004, p. 5). No Brasil, “a pesquisa em *corpus* se dá em centros mais voltados ao Processamento de Linguagem Natural, à Lexicografia e à Lingüística Computacional” (BERBER SARDINHA, 2004, p. 6). Podemos observar atualmente que a Linguística de *Corpus* está presente em outros estudos, como em Linguística Forense, área pouco conhecida no Brasil.

² Programa disponível em: <https://radioglobo.globo.com/programas/programa/216/saude-de-corpo-e-alma.htm>. Acesso em: 26 maio 2020.

A Linguística de *Corpus*, na perspectiva de Berber Sardinha (2004, p. 3),

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

Tão logo, é uma área que permite estudos analíticos fonéticos, discursivos, variacionistas, lexicográficos, lexicológicos, entre outros, a depender do que objetiva o pesquisador. Um mesmo *corpus*, por exemplo, pode servir para análises diversificadas. Programas computacionais, como o WST, auxiliam no desenvolvimento dessas investigações, de modo que ocorram com mais rapidez, eficácia, e possibilitando que o investigador identifique fatos que, sem esse auxílio computacional, se tornariam de difícil verificação.

Cabe ressaltar que essa perspectiva também está alinhada à linguística descritiva, que busca apresentar, de forma sistemática, “os fatos da língua – não a elaboração ou validação de alguma teoria específica da linguagem” (PERINI, 2008, p. 13). Nesse tipo de trabalho, o pesquisador precisa ser treinado para uma percepção dos fatos, de forma que a análise seja realizada a partir do *corpus*, e, com base nos dados obtidos, analisá-los segundo alguma teoria. Então, nós precisamos respeitar tais fatos lingüísticos, já que os dados que determinam o caminho a ser seguido pela pesquisa, considerando os aspectos formais e os fatos semânticos, de modo que o contexto de enunciação seja central, já que “é decisivo para que o receptor possa decodificar corretamente a frase” (PERINI, 2008, p. 25) e “reflete a realidade sem preconceitos teóricos” (PERINI, 2006, p. 38).

Nesse sentido, é a partir dos fatos da língua, observados no *corpus*, que realizamos a inventariação e análise das unidades lingüísticas das duas primeiras edições do programa radial sob análise, que foram transcritas para a realização deste estudo. O estudo foi também norteado por meio de hipóteses formuladas a partir dessas observações.

Neste caso, observamos as unidades lexicais mais frequentes. Optamos por verificar apenas os substantivos, os adjetivos e os verbos. Como se trata de um programa radial que se propõe a falar da saúde do ser humano como um todo (física, mental e espiritual), nos atentamos também para este fator, como forma de conferir se as ocorrências reproduzem tal proposta e se elas evidenciam de fato o que se deve fazer para ter qualidade de vida.

Para Perini (2006, p. 28), “é um fato que algumas palavras designam coisas e outras atribuem qualidades, mas a classificação construída pelo lingüista com base nesse fato é uma hipótese”. Por isso que propor observar apenas os substantivos, adjetivos e verbos já é uma hipótese nossa

de que a organização de dados, dessa maneira, nos permitiria identificar o que é apresentado pelo programa radial, nas duas edições em análise.

Ainda seguindo a perspectiva de Perini (2006, p. 31), “os fatos são diretamente observáveis através do uso que os falantes fazem da língua; as hipóteses são explicitadas pelos lingüistas, e pretendem representar o conhecimento que os falantes têm (sem saber), e que controla o seu uso da língua”. As escolhas lexicais feitas pelos falantes evidenciam seu conhecimento, seu contexto social, histórico, político. E as hipóteses do pesquisador guiam a observação de tais fatos empíricos dos enunciadores. Compreende-se a relevância de o pesquisador determinar hipóteses, a partir dos fatos linguísticos, que são também sociais, históricos, científicos etc.

Segundo Berber Sardinha (2004, p. 23),

[...] a linguagem é um sistema probabilístico, no qual certos traços são mais freqüentes que outros. No caso do léxico, pode-se diferenciar as palavras entre aquelas de maior freqüência e as de menor freqüência, sendo que a diferença entre elas é relativa [...]. Portanto, quanto maior a quantidade de palavras, maior a probabilidade de aparecerem palavras de baixa freqüência.

Quando o falante escolhe ao enunciar certas palavras em detrimento de outras, ele está, portanto, dando preferência a algumas unidades lexicais, deixando em evidência a freqüência de uso do que se está dizendo. Por exemplo, no caso do programa radial em análise, que se tem como temática a saúde humana, espera-se que a freqüência maior seja de palavras que dizem da saúde física, mental ou espiritual de uma pessoa. Além disso, como afirma Berber Sardinha (2004), quanto mais extenso o *corpus* é, maior a probabilidade de observação de palavras diferentes com graus de ocorrência menores. O que faz as palavras serem mais ou menos frequentes é o contexto de fala, o enunciador. Se o programa radial fosse sobre educação pedagógica, por exemplo, provavelmente, a freqüência maior não seria de palavras da área da saúde e sim da educação, mais especificamente da pedagógica.

Portanto, inferimos que “é pelo conhecimento da freqüência atestada que se pode estimar a probabilidade teórica” (BERBER SARDINHA, 2004, p. 32) de ocorrência de traços linguísticos. Quando identificamos pelos fatos linguísticos que a freqüência maior é de um determinado assunto ou que apresenta a maior ocorrência linguística de determinado evento de fala, temos maior fiabilidade para fazer afirmações acerca de determinado *corpus* de estudo. Por fim, para criar uma base de dados sistematizada e eficiente para análises linguísticas, é preciso uma metodologia que dê base para a coleta e organização dos dados, que, neste estudo, foi a Linguística de *Corpus*.

3 Metodologia

Inicialmente, como o *corpus* de análise é originalmente oral, foi necessário realizar sua transcrição, de modo que fosse possível o processamento do *corpus* pelo WST. Para esta etapa, elaboramos métodos de transcrição, construídos a partir das diretrizes de Du Bois (1991). A transcrição realizada objetiva apresentar, em formato escrito, o texto oral da forma como a entrevistadora e os participantes falam. Desta maneira, determinamos como critérios:

- Transcrever a forma falada sem intervenção, por exemplo: “mas, então, **vamo** lá, vamos supor que tem gente que *tá* ouvindo a gente” (excerto retirado do *corpus*);
- As marcas de pontuação são realizadas com base na entonação dos falantes;
- Risos são marcados entre colchetes: [risos];
- Pausa na fala é marcada por reticência;
- O intervalo dos blocos é marcado entre colchetes: [intervalo];
- Quando há sobreposição de falas, identificamos com a fala entre colchetes, por exemplo: “AVEZUM: Então, veja, um indivíduo que tem o traço de raiva já está dizendo ‘eu não estou grato a nada, eu não tolero, [FERRÃO: Humrum.] né? Eu não perdoo’. [BABA: Exato.] Esses indivíduos no prazo médio, cinco a dez anos, e alguns estudos até dezoito anos [FERRÃO: Hummm.] eles têm mais propensão para desenvolver diabetes e insuficiência cardíaca. Então, isso chama a atenção, porque...” (excerto retirado do *corpus*);
- Quando há a citação a algum referente, como nome de livros, o colocamos em itálico;
- Em caso de dúvida na escrita de alguma palavra, se faz necessário a consulta a dicionário da língua (português, inglês, sânscrito etc.), quando for o caso.

Mediante tais critérios, realizamos as transcrições, utilizando os áudios que estão disponíveis na página da *Rádio Globo*³, não sendo possível baixá-los. De posse do material textual, geramos os arquivos em formato .txt, para que conseguíssemos trabalhar com os textos no programa WST.

Segundo Berber Sardinha (2009, p. 8), o WST

[...] é um conjunto de programas integrados (suíte) destinado à análise linguística. 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 de *corpus* (retirar partes indesejadas de cada texto, organizar o conjunto de arquivos, inserir e remover etiquetas etc.), antes da análise propriamente dita.

³ Áudios disponíveis em: <https://radioglobo.globo.com/programas/programa/216/saude-de-corpo-e-alma.htm>. Acesso em: 21 maio 2020.

É o que foi feito neste estudo. Após o processamento dos arquivos pelo programa, recorreremos à ferramenta *WordList* para gerar a lista de unidades lexicais. Analisamos apenas os substantivos, adjetivos e verbos, haja vista que o objetivo foi observar as palavras que poderiam carregar sentidos tangentes à saúde física, mental e/ou espiritual.

Primeiramente, geramos a lista de palavras por meio da ferramenta *WordList*, obtendo 2.612 palavras distintas, conforme apresentado na Figura 1, a seguir:

Figura 1: Dados estatísticos do *corpus* em análise.

| | N | Overall | 1 | 2 |
|--------------------------------|---|---------|---------------|--------------|
| text file | | Overall | 04-09-17 .txt | 09-09-17.txt |
| file size | | 100.751 | 45.650 | 55.101 |
| tokens (running words) in text | | 17.483 | 7.897 | 9.586 |
| tokens used for word list | | 17.463 | 7.888 | 9.575 |
| types (distinct words) | | 2.612 | 1.538 | 1.760 |
| type/token ratio (TTR) | | 15 | 19 | 18 |

Fonte: Ferramenta *WordList*.

Em sequência, realizamos a limpeza da lista de palavras, mantendo apenas os substantivos, os verbos e os adjetivos. Nesta etapa, foi necessário recorrermos ao *Concord*, para fazer a seleção adequada dos itens lexicais, tendo em vista que esta ferramenta nos permite ver a palavra em todos os seus contextos linguísticos no *corpus* de estudo. Logo, o contexto de enunciação ajuda a identificar a classe de palavras de uma unidade lexical. Assim, obtivemos 2.228 substantivos, verbos e adjetivos. A Figura 2, a seguir, apresenta um recorte da lista de palavras após o procedimento de limpeza.

Figura 2: Lista de palavras após a limpeza.

| N | Word | Freq. | % | Texts | % | Lemmas | Set |
|---|-----------|-------|------|-------|--------|--------|-----|
| 1 | É | 391 | 2,24 | 2 | 100,00 | | |
| 2 | GENTE | 169 | 0,97 | 2 | 100,00 | | |
| 3 | TEM | 113 | 0,65 | 2 | 100,00 | | |
| 4 | COMPaixÃO | 107 | 0,61 | 2 | 100,00 | | |
| 5 | TÁ | 105 | 0,60 | 2 | 100,00 | | |
| 6 | GRATIDÃO | 86 | 0,49 | 1 | 50,00 | | |
| 7 | VAI | 77 | 0,44 | 2 | 100,00 | | |
| 8 | PODE | 64 | 0,37 | 2 | 100,00 | | |

Fonte: Ferramenta *WordList*.

Realizamos, em sequência, a junção dos lemas que apresentavam palavras no singular e no plural, como forma de melhor organizar os dados, obtendo o seguinte resultado:

Figura 3: Lista de palavras lematizada.

| N | Word | Freq. | % | Texts | % | Lemmas | Se |
|----|-----------|-------|------|-------|--------|--------------|----|
| 1 | É | 391 | 2,24 | 2 | 100,00 | | |
| 2 | GENTE | 169 | 0,97 | 2 | 100,00 | | |
| 3 | TEM | 113 | 0,65 | 2 | 100,00 | | |
| 4 | COMPAIXÃO | 107 | 0,61 | 2 | 100,00 | | |
| 5 | TÁ | 105 | 0,60 | 2 | 100,00 | | |
| 6 | GRATIDÃO | 86 | 0,49 | 1 | 50,00 | | |
| 7 | PESSOA | 86 | 0,19 | 2 | 100,00 | ssoas[52] | |
| 8 | VAI | 77 | 0,44 | 2 | 100,00 | | |
| 9 | PODE | 64 | 0,37 | 2 | 100,00 | | |
| 10 | SER | 62 | 0,35 | 2 | 100,00 | | |
| 11 | BEM | 53 | 0,30 | 2 | 100,00 | | |
| 12 | VIDA | 52 | 0,30 | 2 | 100,00 | | |
| 13 | SAÚDE | 51 | 0,29 | 2 | 100,00 | | |
| 14 | DIA | 48 | 0,23 | 2 | 100,00 | 411 dias[71] | |

Fonte: Ferramenta *WordList*.

Verificamos, com esta etapa, mudanças na ordem da lista (organizada em ordem decrescente de frequência). Por exemplo, a palavra “pessoa” passa para a sétima posição, o que evidencia a facilitação da visualização, o que é um benefício da limpeza dos resultados, de modo que os dados fiquem mais bem organizados para análise, facilitando o trabalho do pesquisador.

4 Resultados e discussão

Os princípios da Linguística de *Corpus* e o programa WST nos ajudaram a identificar as unidades lexicais mais ocorrentes no *corpus* em análise. A Tabela 1, a seguir, apresenta algumas das palavras mais frequentes em nosso *corpus* de estudo.

Tabela 1: Dados extraídos do WST.

| Ord. | Substantivos/Adjetivos | Quantidade | Ord. | Verbos | Quantidade |
|------|------------------------|------------|------|-------------------|------------|
| 1 | Gente | 169 | 1 | É | 391 |
| 2 | Compaixão | 107 | 2 | Tem | 113 |
| 3 | Gratidão | 86 | 3 | Tá | 105 |
| 4 | Pessoa(s) | 86 | 4 | Vai | 77 |
| | Ser* | 62 | | Ser* ⁴ | 62 |

⁴ A palavra “ser”, marcada com o asterisco, está apresentada nas duas colunas, porque ora aparece como substantivo ora como verbo. Para quantificar as ocorrências de “ser” como verbo e como substantivo, seria necessário utilizar a ferramenta *Concord* para analisar todas as ocorrências. Contudo, optamos por não realizar este procedimento neste momento, haja vista que, pela extensão do *corpus*, demandaria um tempo maior, que não tivemos neste momento.

| | | | | | |
|----|-------------|----|----|-------|----|
| 6 | Bem | 53 | 6 | Pode | 64 |
| 7 | Vida | 52 | 7 | Tenho | 41 |
| 8 | Saúde | 51 | 8 | Foi | 39 |
| 9 | Dia/Janaína | 48 | 9 | Ter | 34 |
| 10 | Coisa | 43 | 10 | Fazer | 32 |

Fonte: Elaborada pela autora.

Observamos que, na coluna de verbos da Tabela 1, há a ocorrência de um mesmo verbo em registros separados, como é o caso dos verbos “ter” (ter, tenho, tem), “ser” (é, ser) e “ir” (vai, foi). Se executássemos a junção dos verbos conjugados em um verbo só, processo que realizamos com o singular e plural, com certeza notaríamos mudanças nessa ordem apresentada.

Os verbos apresentados aqui não carregam, por si só, sentidos referentes à temática do programa radial, portanto, compreendemos que, neste caso, os verbos visualizados não apresentam relevância para a hipótese deste estudo, porque dependem de complementos para sua compreensão. Na coluna de substantivos/adjetivos da Tabela 1, observamos que as unidades lexicais se referem ao ser humano (gente, ser (ora como verbo, ora como substantivo) e pessoa) e a “instrumentos”, que no programa de rádio são tratados como forma de promover qualidade de vida, que é gratidão e compaixão, temas dos dois primeiros programas, respectivamente.

Quando fazemos a exclusão dos verbos e observamos apenas os substantivos e adjetivos, as palavras mais frequentes são:

Tabela 2: Dados extraídos do WST.

| | | | | | |
|----|-----------|-----|----|------------|----|
| 1 | Gente | 169 | 11 | História | 31 |
| 2 | Compaixão | 107 | 12 | Alma | 29 |
| 3 | Gratidão | 86 | 13 | Bom | 29 |
| 4 | Pessoa | 86 | 14 | Mundo | 26 |
| 5 | Bem | 53 | 15 | Sufrimento | 23 |
| 6 | Vida | 52 | 16 | Humano | 21 |
| 7 | Saúde | 51 | 17 | Meditação | 21 |
| 8 | Dia | 48 | 18 | Momento | 21 |
| 9 | Coisa | 43 | 19 | Palavra | 21 |
| 10 | Corpo | 41 | | | |

Fonte: Elaborada pela autora.

Conforme demonstrado na Tabela 2, os substantivos e adjetivos mais ocorrentes se referem à saúde do ser humano e às práticas que promovem qualidade de vida. Portanto, se referem àquilo que sentimos, ao que está na nossa cabeça, no nosso coração e na nossa alma. Podemos identificar ainda palavras que, na perspectiva dos participantes do programa radial, se referem a práticas que podem melhorar a qualidade de vida: gratidão, compaixão, meditação.

5 Considerações finais

Neste estudo, a Linguística de *Corpus* e o programa WST nos permitiram identificar, por meio da análise dos dados obtidos pela ferramenta *WordList*, que os substantivos e adjetivos mais frequentes no *corpus* em análise representam o que é mais abordado nas duas primeiras edições do programa *Saúde de Corpo e Alma*, gratidão e compaixão, além de terem auxiliado na identificação das unidades lexicais que remetem para a missão do programa.

As escolhas lexicais dos falantes no programa radial evidenciam que a qualidade de vida pode ser vivenciada com gratidão e compaixão, duas ferramentas sugeridas pelos participantes, que podem ser utilizadas para vivermos com mais qualidade. Muitos cientistas reforçam essa tese. A perspectiva de Freitas, Silveira e Pieta (2009, p. 244) nos leva à compreensão, a partir do relato de testes científicos, de como a gratidão ocorre em crianças de cinco a doze anos, qual o sentimento que elas desenvolvem e qual o efeito de sua prática: “vários autores têm apontado o fato de que a gratidão pode contribuir para promover o bem-estar e melhorar a qualidade de vida daqueles que a experienciam”.

| 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.

DU BOIS, J. W. Transcription design principles for spoken discourse research. **Pragmatics**, Antwerp, v. 1, n. 1, p. 71-106. 1991. Disponível em: <https://journals.linguisticsociety.org/elanguage/pragmatics/article/view/464/396.html>. Acesso em: 17 maio 2020.

FREITAS, L. B. L.; SILVEIRA, P. G.; PIETA, M. A. M. Sentimento de gratidão em crianças de 5 a 12 anos. **Psicologia em Estudo**, Maringá, v. 14, n. 2, p. 243-250, 2009. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/80047/000723182.pdf?sequence=1>. Acesso em: 17 maio 2020.

PERINI, M. A. **Estudos de Gramática Descritiva**: as valências verbais. São Paulo: Parábola, 2008.

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

RÁDIO GLOBO. **Saúde de Corpo e Alma**. Rio de Janeiro: Rádio Globo, 4 de setembro de 2017. Programa de Rádio. Disponível em: <https://radioglobo.globo.com/programas/programa/216/saude-de-corpo-e-alma.htm>. Acesso em: 26 maio 2020.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://lexically.net/wordsmith/downloads/>. Acesso em: 17 maio 2020.

A fraseologia em Mario Vargas Llosa: um estudo contrastivo em corpus paralelo bilíngue espanhol/português do jornal *El país*

María del Rosario Mestanza¹

¹ Mestranda em Estudos Linguísticos pela Universidade Federal de Uberlândia, sob orientação do Prof. Dr. Ariel Novodvorski. Lattes: <http://lattes.cnpq.br/3183051238718994>. E-mail: mrmestanza@yahoo.com.br.

1 Introdução

Neste capítulo, estuda-se as unidades fraseológicas (UFs) em cinco ensaios publicados em 2015 na coluna de opinião *Pedra de Toque* do jornal *El País*, escritos pelo prêmio Nobel Mario Vargas Llosa em sua versão original em espanhol, bem como analisa-se suas correspondentes traduções para o português brasileiro², com o intuito de compreender e explicar como essas UFs foram traduzidas.

Ao analisar a importância dos fraseologismos, verificou-se a complexidade de sua identificação e tradução, devido ao fato de estes estarem ligados a dizeres e costumes próprios e inseridos em contextos culturais de um povo, em sua história, costumes, crenças e estados afetivos. É claro que, enquanto falantes de uma língua, utilizamos naturalmente inúmeros fraseologismos nas interações diárias, contudo, não é uma utilização consciente. Devido a isso, considera-se que os aspectos culturais presentes nos fraseologismos remetem a um processo complexo, em que estes são construídos e naturalmente utilizados. Em decorrência disso, há dificuldades em sua percepção e identificação.

Com essa inquietação, e a partir de um estudo contrastivo bilíngue, cujo foco consiste no estudo de aspectos lexicogramaticais por meio de uma abordagem descritiva, desenvolve-se este estudo, com o objetivo de identificar e analisar aspectos das UFs originalmente escritas em espanhol e suas correspondentes traduções, em *corpus* paralelo e à luz da Linguística de *Corpus* (LC). Apresenta-se os resultados da análise contrastiva de cinco textos em relação tradutória, que formam parte de um *corpus* bilíngue espanhol/português que vem sendo desenvolvido no âmbito de uma pesquisa de mestrado em andamento, que envolve um número maior de ensaios.

Considera-se importante o desenvolvimento de estudos como o que se propõe neste capítulo, especificamente a partir da língua espanhola, uma vez que “a tradução de fraseologismos ou colocações não forma parte das explicações encontradas nos dicionários mais utilizados na língua espanhola, fato que dificulta a compreensão tanto de estudantes como de tradutores” (NOVODVORSKI; ALVES, 2014, p. 203). Essa dificuldade com a tradução de fraseologismos também foi contemplada por Timofeeva (2008). Segundo esta autora, pode-se falar de intraduzibilidade de alguns deles, devido às referências históricas, socioculturais e idiossincráticas presentes neles. Assim, por exemplo, algumas das dificuldades do tradutor surgem em relação ao processo de tradução dos heterossemânticos (*falsos amigos*) presentes no âmbito do fraseologismo na língua espanhola. Também em relação à tradução dos fraseologismos, Oliveira (2009) manifesta que, por mais difícil que seja traduzir os fraseologismos, os tradutores precisam buscar perífrases

² Edição espanhola disponível em: <https://elpais.com/autor/mario-vargas-llosa/>. Edição brasileira disponível em: <https://brasil.elpais.com/autor/mario-vargas-llosa/>. Acesso em: 15 mar. 2019.

ou equivalentes para tais expressões, e não se contentar com meras explicações. Evidentemente, os aspectos culturais que estão por detrás de cada fraseologismo irão dificultar o processo, mas essa consciência associada a um bom trabalho poderá garantir o sucesso da tradução.

Ainda em relação à tradução, Newmark (2006) enfatiza que o objetivo de toda tradução é produzir no leitor o mesmo efeito que produz o texto original. Por sua vez, Hurtado Albir (2008, p. 28), ao destacar algumas reflexões sobre a tradução, afirma: “se traduz porque as línguas e culturas são diferentes, para comunicar e superar as barreiras das diferenças linguísticas e culturais; se traduz para alguém que não conhece a língua e a cultura em que está formulado um texto seja escrito, oral ou audiovisual”³. Isto é, destaca-se a finalidade comunicativa da tradução.

Para uma abordagem específica sobre a Fraseologia, referencia-se os estudos de Corpas Pastor (1996), ao apontar que, independentemente da língua, a Fraseologia está sendo reconhecida como uma disciplina ou subdisciplina da Lexicologia, que se ocupa das combinações de, pelo menos, duas palavras, cujo limite é a oração, com estabilidade comprovada pela frequência de uso e de caráter idiomático.

Neste estudo, utilizamos os princípios básicos da LC, apresentados em Berber Sardinha (2004), para a compilação e tratamento do *corpus*, bem como para a análise quali-quantitativa dos dados. O *corpus* foi preparado e sistematizado para emprego das ferramentas presentes no programa *WordSmith Tools – WST*, em sua versão 6.0 (SCOTT, 2012).

2 Procedimentos metodológicos

Nesta seção, descreve-se os procedimentos metodológicos adotados para a realização deste estudo, desde a compilação até a análise, passando pela conversão do *corpus*, elaboração dos cabeçalhos e metadados, uso das ferramentas do programa WST, para finalizar com a análise do *corpus* e interpretação das traduções. Na análise do *corpus*, utilizaremos a fundamentação teórica presente na área da Fraseologia e da Tradução.

Utilizou-se um *corpus* paralelo, composto por pares de textos em suas versões originais (espanhol) e suas correspondentes traduções (português brasileiro). Os ensaios foram retirados do jornal espanhol *El País* da coluna de opinião *Pedra de Toque* em sua versão livre *on-line* em espanhol e português no ano de 2015. As Figuras 1 e 2 ilustram as páginas do jornal *El País* obtidas em sua versão *on-line*.

³ No original: “Se traduce porque las lenguas y las culturas son diferentes, para comunicar, para traspasar la barrera de incomunicación debida a esa diferencia lingüística y cultural. Se traduce para alguien que no conoce la lengua y generalmente la cultura, en que está formulado un texto escrito, oral o audiovisual” (tradução nossa).

Figura 1: Texto extraído do jornal *El País*.

PIEDRA DEL TOQUE



La felicidad, ja, ja

MARIO VARGAS LLOSA | 22-02-2015 - 18:01

Fonte: https://elpais.com/elpais/2015/02/19/opinion/1424359367_178330.html.

Figura 2: Texto extraído do jornal *El País*.

PEDRA DE TOQUE



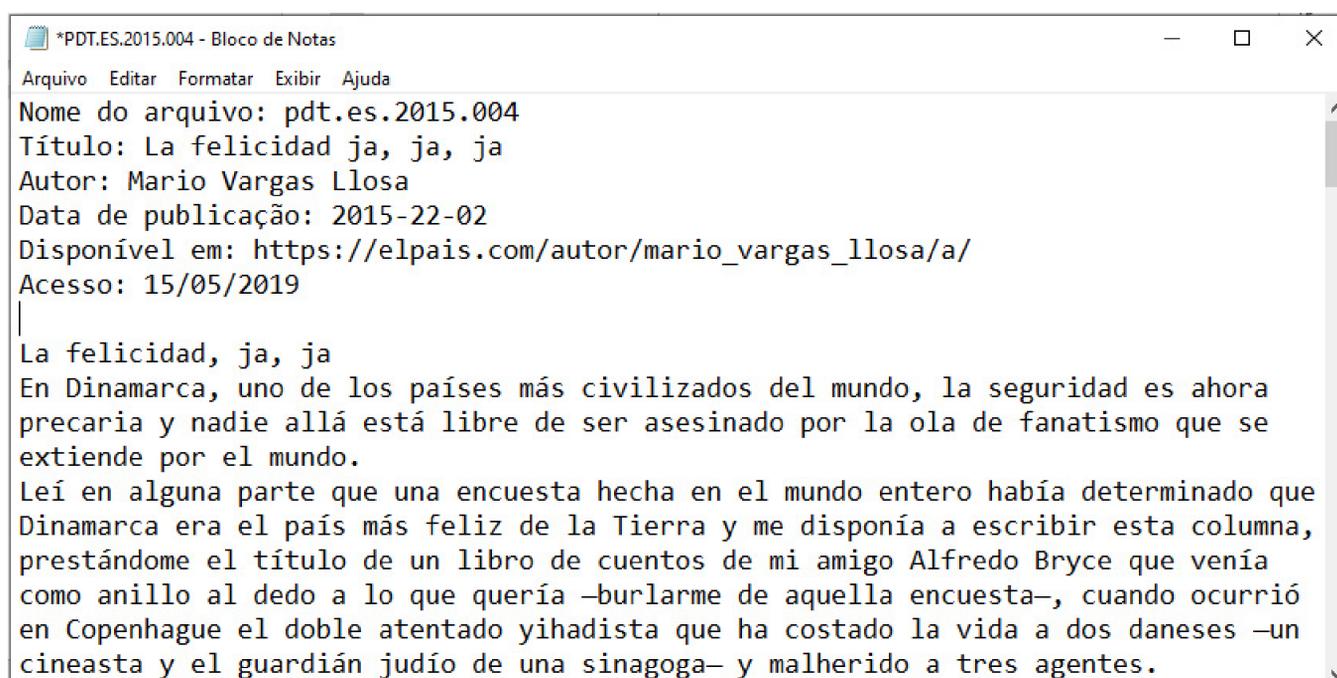
A felicidade, rá, rá, rá

MARIO VARGAS LLOSA | 22-02-2015 - 18:01

Fonte: https://brasil.elpais.com/brasil/2015/02/19/opinion/1424359367_178330.html

O passo seguinte consistiu na conversão do *corpus* para o formato .txt (Bloco de notas). Na sequência, foi inserido um cabeçalho, para uma melhor identificação de cada texto, possibilitando a obtenção de informações no momento do uso das ferramentas do WST, como mostrado na Figura 3.

Figura 3: Texto em formato .txt com cabeçalho para identificação do *corpus* em espanhol.



Fonte: Bloco de Notas.

Para além do cabeçalho, foi criado um arquivo em .doc para registro dos metadados em paralelo de cada bitexto, possibilitando o planejamento de compilação e balanceamento do *corpus*, ou seja, percebendo a quantidade de textos escritos por ano, sua distribuição e saber

quais deles foram traduzidos. Essas informações presentes nos metadados de cada texto são: (i) Nome do arquivo: usamos uma nomenclatura própria para identificar os textos; em espanhol optamos por nomeá-los como “pdt.es” (*pedra de toque* em espanhol) e nomeamos os textos em português como “pdt.pt” (*pedra de toque* em português), seguido do ano de publicação e uma numeração sequencial; (ii) Título do texto conforme aparece no jornal; (iii) Nome do autor; (iv) Data de publicação; (v) Endereço da página eletrônica de publicação e (vi) Data da coleta. Nos metadados do *corpus* traduzido, também, pode-se registrar o nome do tradutor que no caso deste *corpus* não há. No Quadro 1, apresenta-se os metadados dos *corpora* original e traduzido.

Quadro 1: Metadados do *corpus* paralelo.

| Corpus Original | Corpus Traduzido |
|--|---|
| Nome do arquivo: pdt.es.2015.001 Título: Cusco en el tiempo Autor: Maria Vargas Llosa Data de publicação: 2015-10-01 Disponível em: https://elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 | Nome do arquivo: pdt.pt.2015.001 Título: Cusco no tempo Tradutor: não indicado no jornal Data de publicação: 2015-11-01 Disponível em: https://brasil.elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 |
| Nome do arquivo: pdt.es.2015.002 Título: El regreso de las ideas Autor: Maria Vargas Llosa Data de publicação: 2015-24-01 Disponível em: https://elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 | Nome do arquivo: pdt.pt.2015.002 Título: A volta das ideias Tradutor: não indicado no jornal Data de publicação: 2015-25-01 Disponível em: https://brasil.elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 |
| Nome do arquivo: pdt.es.2015.003 Título: El Harakiri Autor: Mario Vargas Llosa Data de publicação: 2015-07-02 Disponível em: https://elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 | Nome do arquivo: pdt.pt.2015.003 Título: O Haraquiri Tradutor: não indicado no jornal Data de publicação: 2015-07-02 Disponível em: https://brasil.elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 |
| Nome do arquivo: pdt.es.2015.004 Título: La felicidad ja, ja, ja Autor: Mario Vargas Llosa Data de publicação: 2015-22-02 Disponível em: https://elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 | Nome do arquivo: pdt.pt.2015.004 Título: A felicidade ra,ra,ra Tradutor: não indicado no jornal Data de publicação: 2015-22-02 Disponível em: https://brasil.elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 |
| Nome do arquivo: pdt.es.2015.005 Título: Al borde del abismo Autor: Mario Vargas Llosa Data de publicação: 2015-03-07 Disponível em: https://elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 | Nome do arquivo: pdt.es.2015.005 Título: A beira do abismo Tradutor: não indicado no jornal Data de publicação: 2015-03-07 Disponível em: https://brasil.elpais.com/autor/mario_vargas_llosa/a/ Acesso: 15/05/2019 |

Fonte: Elaboração da autora.

A seguir, com o *corpus* de estudo convertido em .txt e configurado para codificação Unicode, se procedeu à análise com o auxílio das ferramentas *WordList* (listador de palavras) e *Concord* (gerador de linhas de concordância) do programa WST. Com essas ferramentas, levantamos os dados gerais do *corpus* e procedemos a sua leitura, descrição e análise.

3 Análise dos dados

Na Tabela 1 e na Figura 4, apresentam-se os dados quantitativos do *corpus*, fornecidos pela ferramenta *WordList*, aplicando a função de dados estatísticos dos respectivos *corpora*. As colunas Total de palavras (*tokens*) e Palavras diferentes (*types*) correspondem aos dados estatísticos do *corpus*, conforme o total de palavras e a quantidade de palavras diferentes presentes nos textos, respectivamente.

Tabela 1: Textos originais e textos traduzidos.

| Jornal <i>El País</i> | Quantidade de textos | Total de palavras (<i>tokens</i>) | Palavras diferentes (<i>types</i>) |
|-------------------------------|----------------------|-------------------------------------|--------------------------------------|
| Textos originais (espanhol) | 5 | 6.760 | 2.385 |
| Textos traduzidos (português) | 5 | 6.371 | 2.413 |

Fonte: Elaboração da autora.

Figura 4: Ferramenta *WordList* – dados estatísticos relativos ao *corpus* em espanhol e português.

CORPUS-ESP-MVLLOSAL_files.lst

| N | text file | file size | tokens (running words) in | tokens used for word list | sum of entries | types (distinct words) | type/token ratio | standard TTR |
|---|---------------------|-----------|---------------------------|---------------------------|----------------|------------------------|------------------|--------------|
| 1 | Overall | 83.764 | 6.768 | 6.757 | | 2.383 | 35,27 | 48,56 |
| 2 | PDT.ES.2015.001.txt | 18.262 | 1.486 | 1.485 | | 699 | 47,07 | 50,60 |
| 3 | PDT.ES.2015.002.txt | 16.282 | 1.288 | 1.288 | | 610 | 47,36 | 49,70 |
| 4 | PDT.ES.2015.003.txt | 16.446 | 1.347 | 1.341 | | 610 | 45,49 | 48,50 |
| 5 | PDT.ES.2015.004.txt | 15.270 | 1.198 | 1.194 | | 541 | 45,31 | 46,80 |
| 6 | PDT.ES.2015.005.txt | 17.504 | 1.449 | 1.449 | | 627 | 43,27 | 47,20 |

CORPUS-PORT-MVLLOSA_files.lst

| N | text file | file size | tokens (running words) in | tokens used for word list | sum of entries | types (distinct words) | type/token ratio | standard TTR |
|---|---------------------|-----------|---------------------------|---------------------------|----------------|------------------------|------------------|--------------|
| 1 | Overall | 80.858 | 6.383 | 6.371 | | 2.413 | 37,87 | 51,80 |
| 2 | PDT.PT.2015.001.txt | 17.620 | 1.395 | 1.394 | | 718 | 51,51 | 54,60 |
| 3 | PDT.PT.2015.002.txt | 15.722 | 1.211 | 1.211 | | 621 | 51,28 | 52,80 |
| 4 | PDT.PT.2015.003.txt | 15.728 | 1.260 | 1.254 | | 613 | 48,88 | 51,30 |
| 5 | PDT.PT.2015.004.txt | 14.816 | 1.147 | 1.143 | | 560 | 48,99 | 50,30 |
| 6 | PDT.PT.2015.005.txt | 16.972 | 1.370 | 1.369 | | 637 | 46,53 | 50,00 |

Fonte: Programa WST 6.0.

Utilizando a ferramenta *WordList*, aplicou-se a função ordem de frequência no *corpus* em espanhol e em português, para obter os dados referentes à quantidade de palavras diferentes. Esse número obedece à variedade lexical e gramatical presente em cada língua, isso significa que o processamento em cada língua é diferente, no uso de preposições, junções de preposições e artigos. Na Figura 5, mostra-se o detalhe da função ordem de frequência.

Figura 5: Lista de palavras do *corpus* em espanhol (à esquerda) e em português (à direita).

| N | Word | Freq. | % | Texts | % Dispersion | Le...s Set |
|----|------|-------|-------|-------|--------------|------------|
| 2 | Y | 272 | 4,02% | 5 | 1... | n/a |
| 3 | QUE | 260 | 3,84% | 5 | 1... | n/a |
| 4 | LA | 253 | 3,73% | 5 | 1... | n/a |
| 5 | LOS | 180 | 2,66% | 5 | 1... | n/a |
| 6 | EN | 179 | 2,64% | 5 | 1... | n/a |
| 7 | EL | 178 | 2,63% | 5 | 1... | n/a |
| 8 | A | 158 | 2,33% | 5 | 1... | n/a |
| 9 | LAS | 100 | 1,48% | 5 | 1... | n/a |
| 10 | POR | 79 | 1,17% | 5 | 1... | n/a |
| 11 | DEL | 68 | 1,00% | 5 | 1... | n/a |
| 12 | NO | 62 | 0,92% | 5 | 1... | n/a |
| 13 | UN | 60 | 0,89% | 5 | 1... | n/a |
| 14 | UNA | 59 | 0,87% | 5 | 1... | n/a |
| 15 | ES | 57 | 0,84% | 5 | 1... | n/a |
| 16 | SE | 56 | 0,83% | 5 | 1... | n/a |
| 17 | SU | 53 | 0,78% | 5 | 1... | n/a |
| 18 | LO | 45 | 0,66% | 5 | 1... | n/a |
| 19 | CON | 43 | 0,63% | 5 | 1... | n/a |
| 20 | MÁS | 39 | 0,58% | 5 | 1... | n/a |
| 21 | AL | 36 | 0,53% | 5 | 1... | n/a |
| 22 | COMO | 36 | 0,53% | 5 | 1... | n/a |
| 23 | O | 36 | 0,53% | 4 | 80...% | n/a |
| 24 | HA | 34 | 0,50% | 5 | 1... | n/a |
| 25 | PARA | 25 | 0,37% | 5 | 1... | n/a |

| N | Word | Freq. | % | Texts | % Dispersion | Le...s Set |
|----|------|-------|-------|-------|--------------|------------|
| 1 | E | 290 | 4,54% | 5 | 1... | n/a |
| 2 | DE | 279 | 4,37% | 5 | 1... | n/a |
| 3 | QUE | 235 | 3,68% | 5 | 1... | n/a |
| 4 | A | 205 | 3,21% | 5 | 1... | n/a |
| 5 | O | 145 | 2,27% | 5 | 1... | n/a |
| 6 | OS | 91 | 1,42% | 5 | 1... | n/a |
| 7 | SE | 82 | 1,28% | 5 | 1... | n/a |
| 8 | DA | 78 | 1,22% | 5 | 1... | n/a |
| 9 | DO | 78 | 1,22% | 5 | 1... | n/a |
| 10 | EM | 68 | 1,06% | 5 | 1... | n/a |
| 11 | NÃO | 63 | 0,99% | 5 | 1... | n/a |
| 12 | UM | 63 | 0,99% | 5 | 1... | n/a |
| 13 | UMA | 61 | 0,96% | 5 | 1... | n/a |
| 14 | É | 55 | 0,86% | 5 | 1... | n/a |
| 15 | DOS | 52 | 0,81% | 5 | 1... | n/a |
| 16 | PARA | 50 | 0,78% | 5 | 1... | n/a |
| 17 | AS | 49 | 0,77% | 5 | 1... | n/a |
| 18 | COM | 49 | 0,77% | 5 | 1... | n/a |
| 19 | POR | 43 | 0,67% | 5 | 1... | n/a |
| 20 | COMO | 40 | 0,63% | 5 | 1... | n/a |
| 21 | NA | 40 | 0,63% | 5 | 1... | n/a |
| 22 | NO | 39 | 0,61% | 5 | 1... | n/a |
| 23 | OU | 38 | 0,60% | 5 | 1... | n/a |
| 24 | MAIS | 36 | 0,56% | 5 | 1... | n/a |

Fonte: Programa WST 6.0.

Em seguida, com base na lista de palavras organizadas em ordem decrescente de frequência, analisamos a preposição “com” no *corpus* em espanhol e em português, usando a ferramenta *Concord*, no intuito de explorar os contextos linguísticos e identificar expressões lexicais. Na Figura 6, mostra-se o resultado desta aplicação.

Figura 6: Linhas de concordância de “con/com” no corpus em espanhol (à esquerda) e em português (à direita).

| N | Concordance | Set |
|----|--|-----|
| 4 | trabajos en los que Mahoma aparecía con el cuerpo de un perro. Hombre | |
| 5 | Vilks ha explicado que no hizo aquello con el ánimo de ofender las creencias | |
| 6 | o delirantes sofismas que, ataviados con una pretenciosa retórica | |
| 7 | arbitraria y falaz. Por otra parte, basta con pasar revista a las manifestaciones | |
| 8 | ya muy pocas y hay que ir a buscarlas —con lupa— en los más alejados | |
| 9 | y periódicos, con la celeridad con que esos mismos Gobiernos | |
| 10 | británico llegó alguna vez a codearse con los tres ilustres exiliados que | |
| 11 | tres ilustres exiliados que coincidieron con él en aquella ciudad: Joyce, Lenin | |
| 12 | relaciones diplomáticas y comerciales con ellos y denunciarlos en el plano | |
| 13 | deja de serlo y llega a confundirse con los Estados totalitarios y | |
| 14 | cívica y moral. Sanguinetti contrasta, con mucha razón, la actitud de esos | |
| 15 | de la libertad de expresión y de crítica, con el argumento multiculturalista de | |
| 16 | el dibujante Lars Vilks se enfrentan con coraje al desafío del terror, luchan | |
| 17 | que las financian, están equipadas con armas modernísimas y tienen | |
| 18 | a cazar en la noche regresan a la gruta con la boca llena de un sangriento | |
| 19 | y filmaba semejante hazaña con lujo perverso de detalles— y | |
| 20 | el revoltoso Michel Houellebecq con su última novela, Sumisión, y por | |
| 21 | es decir, dejar de identificarse con el Estado. Todo esto parece | |
| 22 | y por las modas, y la insolencia con que ha escrito siempre las | |
| 23 | iglesias, palacios, están levantados con las piedras monumentales, | |
| 24 | del imperio. Hoy ocurre lo mismo, con la diferencia de que las lenguas | |
| 25 | ayudan, se hallan vivos y operantes, con escuelas, talleres, bibliotecas, | |
| 26 | y se emborrachaban los parroquianos con la brava chicha de maíz | |
| 27 | canales, radioemisoras y periódicos, con la celeridad con que esos mismos | |
| 28 | los fanáticos islamistas persiguen con saña desde que, en el año 2007, | |

| N | Concordance | Set |
|----|--|-----|
| 11 | relações diplomáticas e comerciais com eles e denunciá-los no plano | |
| 12 | obra, Tom Stoppard nos confronta com um tremendo dilema de | |
| 13 | de sê-lo, chegando a se confundir com os Estados totalitários e | |
| 14 | civil e moral. Sanguinetti contrapõe, com muita razão, a atitude desses | |
| 15 | de expressão e de crítica, com o argumento multiculturalista | |
| 16 | sofismas que, disfarçados com uma pretenciosa retórica | |
| 17 | — inteligente, estudioso, cordial e "com vontade de ajudar os demais", | |
| 18 | os paroquianos se embebedavam com a forte chicha de milho | |
| 19 | cartunista Lars Vilks, que enfrentam com coragem o desafio do terror, | |
| 20 | cada vez mais enfurecida com os atropelos desenfreados da | |
| 21 | que as financiam, estão equipadas com armas muito modernas e têm | |
| 22 | e filmava semelhante façanha com perversa riqueza de detalhes | |
| 23 | de rádio e jornais são fechados com a celeridade com que esses | |
| 24 | para caçar à noite voltam à gruta com a boca cheia de um sangrento | |
| 25 | de levar ao palco uma história com uma temática que combina a | |
| 26 | ou seja, deixar de se identificar com o Estado. Tudo isso parece | |
| 27 | é, lamentavelmente, inevitável, com a condição de que tais | |
| 28 | e pelas modas e a insolência com que sempre escreveu as | |
| 29 | igrejas e palácios estão levantados com as pedras monumentais, | |
| 30 | pela forma enérgica como lidaram com a crise provocada pelo desafio | |
| 31 | Vilks explicou que fez aquilo não com a intenção de ofender a | |
| 32 | uma ternura compassiva para com todas as extravagâncias e | |
| 33 | os fanáticos islâmicos perseguem com sanha desde que, em 2007, | |
| 34 | poucas e é preciso procurá-las — com lupa — nos subúrbios mais | |
| 35 | Está cuidadosamente pronunciado, com erres e jotas vibrantes, e é | |

Fonte: Programa WST 6.0.

Segundo Corpas Pastor (1996), optou-se pelo critério gramatical para a identificação das colocações e enunciados fraseológicos, bem como de sua relação sintática e semântica. Assim, identificou-se a colocação com presença do adjetivo intensificador: “*hazaña con lujo perverso de detalles*”, na tradução, “*façanha com perversa riqueza de detalhes*”. O uso do adjetivo “perverso”, como colocado na locução “*con lujo de detalles*”, ilustra a caracterização do indivíduo pela escolha lexical na UF. Nas outras expressões lexicais (“*ya muy pocas y hay que ir a buscarlas con lupa en los más alejados*” e “*y se emborrachaban los parroquianos con la brava chicha de maíz*”), observa-se a presença de locuções verbais de cunho sociocultural.

Na Figura 7, mostra-se o resultado da procura por concordâncias de expressões lexicais analisando o verbo “estar”, pelo uso da ferramenta *Concord*. Identifica-se um enunciado fraseológico de cunho cultural, com significado denotativo idiomático (*la lengua de los incas está viva y coleando*). É idiomático porque obedece a uma convencionalidade e seu uso é frequente, tornando-se muitas vezes padrões de comunicação em países de língua espanhola.

Figura 7: Linhas de concordância de “está” no *corpus* em espanhol (à esquerda) e em português (à direita).

| N | Concordance |
|---|--|
| 1 | la lengua de los incas está viva y coleando. |
| 2 | Huaro y Andahuaylillas está terminada y es |
| 3 | de Quispicanchi. Lo indio está tan presente que a |
| 4 | en el Instituto Krohl. Está rodeada de |
| 5 | de que el enemigo está retrocediendo , que |

| N | Concordance |
|---|--|
| 1 | é precária, e ninguém está a salvo de ser |
| 2 | é precária, e ninguém lá está a salvo de ser |
| 3 | saem da cinemateca onde está acontecendo um |
| 4 | totalmente o que está acontecendo , |
| 5 | onde a língua dos incas está bem viva. Como não |
| 6 | que sua popularidade está caindo e que, ao |

Fonte: Programa WST 6.0.

No caso destacado na Figura 7, trata-se de uma idiomaticidade parcial, pois um dos elementos apresenta significado idiomático, formado pelo binômio “viva y coleando”, que significa “cheia de vida e muito ativa”. Na tradução “a língua dos incas está bem viva”, o tradutor adequou uma tradução para manter os referenciais idiomáticos do texto original; entretanto, perde-se a possibilidade de oferecer ao leitor a mesma experiência de quem lê na língua original, como afirma Newmark (2006).

Figura 8: Linhas de concordância de “viva y coleando”.

| Corpus del Español: Web/Dialectos | | | | | | | | | | | | | | | | | | | | |
|--------------------------------------|---|-----------|-----------|---------|-------|----------|------------|------|---------|---------|--------|-----------|----------|--------|-----------|--------|------|-------------|----------|-------------|
| BUSCAR | | FREQUENCY | | | | CONTEXT | | | | HISTORY | | | | | | | | | | |
| SEE CONTEXT: CLICK ON WORD [HELP...] | | | | | | | | | | | | | | | | | | | | |
| | | ALL | Argentina | Bolivia | Chile | Colombia | Costa Rica | Cuba | Rep Dom | Ecuador | España | Guatemala | Honduras | México | Nicaragua | Panamá | Perú | Puerto Rico | Paraguay | El Salvador |
| 1 | VIVA Y COLEANDO | 24 | 3 | 1 | 1 | | | | | 4 | 3 | 1 | 2 | | | 2 | | | | |
| (SHUFFLE) | | | | | | | | | | | | | | | | | | | | |
| CLICK FOR MORE CONTEXT | | | | | | | | | | | | | | | | | | | | |
| 1 | Es cierto que la vieja religión está aún viva y coleando en algunos sitios, pero el Dios de los siglos pasados está condenado a | | | | | | | | | | | | | | | | | | | |
| 2 | Shahid! Parece la rivalidad de Fashion sigue viva y coleando . Después de que Priyanka Chopra luciera un look peli pelirrojo en Love Story | | | | | | | | | | | | | | | | | | | |
| 3 | Pero eureka hoy por hoy la oligarquía esta viva y coleando y encarnada por los hijos de los propios revolucionarios. En tonces que paso | | | | | | | | | | | | | | | | | | | |
| 4 | al contrario y la Feria de Julio sigue viva y coleando . ETA tira a dar con toda la mala hostia, lo prepara todo | | | | | | | | | | | | | | | | | | | |
| 5 | más lejos de la realidad. Increíblemente sigue viva y coleando . Porque increíble resulta que un fármaco indicado en la DM2 - cuya principal | | | | | | | | | | | | | | | | | | | |
| 6 | , pero la teoría de cuerdas sigue tan viva y coleando como siempre, pues se ha aceptado que no se entendía bien el modelo | | | | | | | | | | | | | | | | | | | |
| 8 | seguridad. La privacidad no solamente está viva y coleando , sino que también es un tema en el que pensamos todo el día | | | | | | | | | | | | | | | | | | | |
| 9 | y en el futuro. La privacidad está viva y coleando | | | | | | | | | | | | | | | | | | | |
| 10 | incluidas como femeninas. Esta cultura que sigue viva y coleando es la civilización de los hombres, creada por ellos y para ellos. | | | | | | | | | | | | | | | | | | | |

Fonte: *Corpus do Espanhol: Web/Dialetos*.

De acordo com o *Corpus do Espanhol: Web/Dialetos* de Mark Davies⁴ (Figura 8), os usos dessa UF são comuns em países, tais como, Argentina, Espanha, Guatemala, México e Peru. Ainda que se observe baixa frequência, deve-se levar em consideração que todo *corpus* é um recorte da realidade, portanto, uma mostra parcial dos usos linguísticos, e que a busca foi feita por uma forma específica, sem contemplar as flexões possíveis de gênero e número do adjetivo “viva”.

⁴ Disponível para buscas em: <https://www.corpusdelespanol.org/web-dial/>. Acesso em: 22 maio 2020.

4 Considerações finais

Neste capítulo, apresentou-se identificação de fraseologismos e uma análise contrastiva num *corpus* paralelo bilíngue (espanhol/português) jornalístico, formado por cinco ensaios publicados em 2015 na coluna de opinião *Pedra de Toque* do jornal *El País*, originalmente escritos por Mario Vargas Llosa. A análise foi fundamentada pelos princípios da LC e realizada por meio do programa WST.

A identificação de algumas UFs, colocações e locuções, encontradas no *corpus* bilíngue estudado, corroboraram a presença de aspectos socioculturais dentro de uma abordagem descritiva, que possibilitou a compreensão das combinações linguísticas e palavras peculiares de cunho sociocultural. Este estudo faz parte de uma pesquisa de mestrado em andamento no Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia (PPGEL-UFU), que toma como referência um *corpus* paralelo bilíngue espanhol/português de maior extensão, visando identificar as UFs com características sociais, históricas e culturais, próprias da região do autor, e suas respectivas traduções.

| Referências

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

CORPAS PASTOR, G. **Manual de fraseología española**. Madrid: Gredos, 1996.

HURTADO ALBIR, A. **Traducción y Traductología**: introducción a la Traductología. 4. ed. Madrid: Cátedra, 2008.

MARK, D. **Corpus do Espanhol**. Disponível em: <https://www.corpusdelespanol.org/>. Acesso em: 20 dic. 2019.

NEWMARK, P. **Métodos de traducción**. In: NEWMARK, P. **Manual de traducción**. Traducción de Virgilio Moya. 5. ed. Madrid: Cátedra, 2006. p. 69-80.

NOVODVORSKI, A.; ALVES, M. L. A tradução de fraseologismos no jornal *El País*: um estudo contrastivo em espanhol e português. **Domínios de Lingu@gem**, Uberlândia, v. 8, n. 2, p. 202-218. 2014. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/27737/15772>. Acesso em: 12 nov. 2019.

OLIVEIRA, S. T. **Comparação de fraseologismos franceses em dicionários bilíngues brasileiros**. 2009. Dissertação (Mestrado em Linguística Aplicada) – Instituto de Letras, Universidade de Brasília, Brasília, 2009. Disponível em: https://repositorio.unb.br/bitstream/10482/4275/1/2009_SirleneTerezinhaOliveira.pdf. Acesso em: 12 nov. 2019.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 17 maio 2020.

TIMOFEEVA, L. **Acerca de los aspectos traductológicos de la fraseología española**. 2008. Tesis (Doctorado en Filosofía y Letras) – Facultad de Filosofía y Letras, Universidad de Alicante, Alicante, 2008. Disponível em: https://rua.ua.es/dspace/bitstream/10045/7707/1/tesis_doctoral_larissa_timofeeva.pdf. Acesso em: 27 maio 2020.

Libras e Linguística de *Corpus*: análise de sinais-termos da indústria automobilística

Kássia Mariano de Souza¹

¹ Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia, sob orientação do Prof. Dr. Ariel Novodvorski. Lattes: <http://lattes.cnpq.br/7285881366227137>. E-mail: kassia_mariano@ufg.br.

1 Introdução

Este capítulo apresenta um estudo sobre o léxico especializado em Libras (Língua Brasileira de Sinais) de uma empresa do setor automobilístico localizada em Catalão-Goiás, isto é, analisamos termos que são utilizados dentro desse setor, materializados em forma de sinais. Sabemos que cada área do conhecimento possui termos técnico-científicos, e o modo como estes são constituídos e utilizados pelas pessoas surdas é o que buscamos investigar. O *corpus* de estudo utilizado foi coletado para nossa pesquisa de mestrado (SOUZA, 2018), a partir de um conjunto de entrevistas realizadas na empresa supracitada. As entrevistas foram gravadas em áudio e vídeo e transcritas para a Língua Portuguesa, a fim de identificar os sinais-termos presentes nas sinalizações dos entrevistados.

Krieger e Finatto (2004) pontuam que a terminologia é um elemento inerente às chamadas comunicações especializadas, permitindo ao homem denominar objetos, processos e conceitos específicos de um campo tecnológico e/ou científico. Desse modo, buscamos identificar, nas entrevistas dos participantes, os termos técnicos que integram a prática profissional e estabelecer uma análise de motivação e não-motivação para a criação dos sinais-termos em questão.

Apresentamos neste capítulo, portanto, uma pesquisa fruto da interação entre Terminologia, Libras e Linguística de *Corpus* (LC), cujo objetivo principal é identificar os termos presentes nos relatos sinalizados dos participantes surdos e, posteriormente, analisá-los de acordo com a motivação ou não-motivação do signo linguístico terminológico na Libras. No que diz respeito à interação entre Terminologia e LC, Bevilacqua (2013) considera uma parceria altamente produtiva, uma vez que as ferramentas disponíveis pela LC possibilitam a identificação de termos em seus contextos de uso.

2 Fundamentação teórica

Considerando que a língua utilizada pelos surdos é de modalidade visual-espacial e a dos ouvintes oral-auditiva, propusemo-nos a observar a maneira pela qual se efetiva a comunicação mediante o uso de terminologia no ambiente de trabalho da empresa automobilística. Isto é, analisamos a maneira pela qual a terminologia da área ganha vida e sentido em forma de sinais, uma vez que é por meio dela que a comunicação entre surdos e ouvintes flui dentro do ambiente de trabalho. No que tange à comunicação, também buscamos compreender, a partir das entrevistas com os participantes surdos, de que modo é efetivada, uma vez que esse processo envolve duas línguas diferentes, no caso, a Língua Portuguesa e a Libras.

Sabemos que cada área do conhecimento possui termos técnico-científicos, isto é, terminologias, e sendo a Libras um sistema linguístico organizado e utilizado pela comunidade surda brasileira, deve, portanto, ser capaz de, assim como qualquer outra língua, expressar tais terminologias.

A Libras é estruturada linguisticamente de modo a possibilitar a comunicação de pessoas surdas e ouvintes. Sua origem é a Língua de Sinais Francesa (LSF), que foi trazida ao Brasil em 1816 e, a partir de então, sofreu as influências culturais brasileiras até se tornar uma língua de sinais própria do país, que tem o seu *status* linguístico reconhecido graças aos diversos aspectos que a caracterizam como uma língua. A formação dos sinais é um deles, pois é resultado de uma estrutura gramatical que se articula sistematicamente, de modo similar ao que acontece com o léxico nas línguas orais. Na Libras, os sinais existem a partir da combinação dos cinco parâmetros existentes: Configuração de Mãos (CM), Ponto de Articulação (PA), Orientação (O), Movimento (MO) e Expressão Facial e Corporal (EFC) (GESSER, 2009).

Estudos linguísticos são essenciais para qualquer língua e, no que se refere à Libras, consideramos avanços nas pesquisas que buscam elucidar os fenômenos linguísticos de uma língua de modalidade visual-espacial. No tocante à Libras e Terminologia, Faulstich (2014), criadora do sinal-termo no Brasil e, conseqüentemente, precursora dos estudos de Terminologia e Libras no país, define o termo em Libras como sinal-termo, pois se configura como um termo em forma de sinal, criado para denotar conceitos de linguagem especializada em língua de sinais.

Reconhecemos a necessidade de mais abordagens lexicais e terminológicas da Libras e, neste estudo em especial, coadunamos outra importante área, que é a LC. Para Berber Sardinha (2004), a LC trata da coleta e da exploração de *corpora*, com o objetivo de pesquisar uma língua ou variedade linguística. Neste estudo, os recursos advindos da LC foram fundamentais para a identificação dos termos presentes nas entrevistas dos participantes, conforme será demonstrado posteriormente na seção de metodologia. No que diz respeito às análises dos sinais-termos, esclarecemos que estas se deram a partir das noções de iconicidade e arbitrariedade do signo linguístico na Libras, isto é, a motivação ou não motivação no processo de criação dos sinais.

2.1 Iconicidade e arbitrariedade dos sinais terminológicos

O signo linguístico na Libras pode ser concebido por duas vertentes, sendo uma delas embasada em Saussure (2006), que defende a arbitrariedade entre significado e significante, e a outra em Peirce (1995), que considera os elementos icônicos que permeiam o signo linguístico.

A partir dessas discussões, redirecionaremos as teorias aqui apresentadas para o plano da Língua de Sinais, a fim de possibilitar as análises quanto à motivação ou não-motivação do signo linguístico terminológico na Libras.

Saussure (2006), ao teorizar sobre signos linguísticos, utiliza o termo “arbitrário” para definir o laço que une o significado e o significante. Para este teórico, não há relação óbvia entre um e outro. Na Libras, a arbitrariedade é representada pelos signos cuja sinalização não possui alguma relação de semelhança com o seu objeto, isto é, os sinais em si não carregam traços que remetam à palavra. Por outro lado, dentre as abordagens do signo feitas por Peirce (1995), encontramos o ícone, que traz em si as características do significado em seu significante.

Desse modo, a iconicidade na Libras é representada por um signo que mantém semelhança com o seu objeto e, devido a isso, os sinais na Libras que trazem essa representatividade são considerados icônicos. Por sua vez, o arbitrário não traz consigo relação alguma entre o sinal e o seu referente.

3 Metodologia

O projeto de pesquisa, quando da época de realização do mestrado (SOUZA, 2018), foi submetido ao Comitê de Ética em Pesquisa (CEP) da Universidade Federal de Goiás (UFG) e, após a aprovação, iniciamos os procedimentos para a coleta dos dados². Segundo Vergara (2000), a coleta de dados é a forma pela qual se obtêm os dados necessários para responder ao problema. Por isso, para obtermos as informações pertinentes à catalogação e à sistematização do léxico especializado do setor automobilístico, especificamente, como se dá a utilização desse conjunto de termos em Libras, realizamos a coleta dos dados por meio de entrevistas semiestruturadas, gravadas em vídeo devido à modalidade linguística dos participantes ser visual-espacial, nesse caso, a Libras, e não oral-auditiva, como no caso dos usuários de línguas orais.

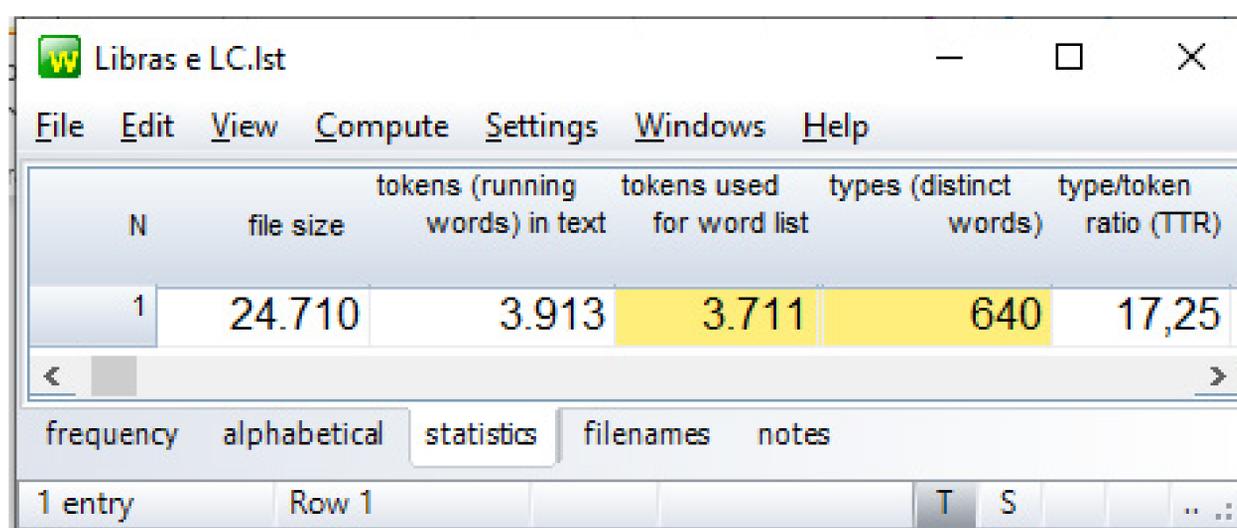
As gravações tiveram entre trinta e cinco (35) e quarenta e cinco (45) minutos de duração, com exceção de uma, que durou cinquenta e seis (56) minutos. Para a realização dessa pesquisa, foram selecionados dez (10) participantes, sendo seis (06) do sexo masculino e quatro (04) do sexo feminino. A idade dos participantes variou entre vinte e seis (26) e quarenta e nove (49) anos, e o tempo de trabalho na área automobilística foi de quatro (04) a quatorze (14) anos, com exceção de uma entrevistada, que havia sido demitida recentemente, tendo trabalhado por nove (09) anos no ramo. Os setores de lotação dos entrevistados foram: pintura, montagem, produção e logística.

² O *corpus* de transcrições das entrevistas do mestrado está sendo utilizado neste estudo a fim de estabelecer relação com a LC, enquanto trabalho da disciplina de Doutorado (PPGEL-UFU).

As entrevistas foram transcritas da Libras para a Língua Portuguesa e optamos por mantê-las na sintaxe da Língua de Sinais, de modo a respeitar a estrutura de sinalização dos participantes. Os textos transcritos totalizaram 3.711 palavras que, após a transcrição, foram salvos em formato .txt, para viabilizar o processamento do *corpus* pelo programa *WordSmith Tools 6.0* (SCOTT, 2012).

Inicialmente, geramos a lista de palavras do *corpus*, por meio da ferramenta *WordList*, com o intuito de identificar os termos presentes nas entrevistas. Berber Sardinha (2004) define a lista de palavras como uma lista que contém todas as palavras do arquivo, apresentadas por frequência ou ordem alfabética. A ferramenta *WordList* possibilitou o acesso aos dados estatísticos do *corpus*, que assim se representa: *tokens* (itens): 3.711 e *types* (formas): 640, como mostra a Figura 1.

Figura 1: Dados estatísticos do *corpus*.



The screenshot shows a window titled "Libras e LC.lst" with a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a table of statistics. The table has columns for N, file size, tokens (running words) in text, tokens used for word list, types (distinct words), and type/token ratio (TTR). The data row shows: N=1, file size=24.710, tokens in text=3.913, tokens used=3.711, types=640, and TTR=17,25. Below the table are tabs for frequency, alphabetical, statistics (selected), filenames, and notes. The status bar at the bottom indicates "1 entry" and "Row 1".

| N | file size | tokens (running words) in text | tokens used for word list | types (distinct words) | type/token ratio (TTR) |
|---|-----------|--------------------------------|---------------------------|------------------------|------------------------|
| 1 | 24.710 | 3.913 | 3.711 | 640 | 17,25 |

Fonte: *WordList – Statistics*.

Após ter gerado a lista de palavras, realizamos o processo de limpeza da lista, deixando apenas os termos.

4 Resultados e discussão

Após a realização do procedimento de limpeza da lista de palavras, identificamos sessenta e seis (66) termos presentes nas entrevistas. A Figura 2 apresenta um recorte dessa lista dos termos.

Figura 2: Lista de termos

| N | Word | Freq. | Texts |
|----|---------------|-------|-------|
| 1 | FERRAMENTAS | 30 | 1 |
| 2 | PINTURA | 27 | 1 |
| 3 | CARRO | 24 | 1 |
| 4 | SOLDA | 17 | 1 |
| 5 | RADIADOR | 16 | 1 |
| 6 | MONTAGEM | 15 | 1 |
| 7 | PORTA | 14 | 1 |
| 8 | ELEVADOR | 13 | 1 |
| 9 | PARAFUSADEIRA | 13 | 1 |
| 10 | MITSUBISHI | 12 | 1 |
| 11 | MOTOR | 11 | 1 |
| 12 | FERRO | 10 | 1 |
| 13 | MÁQUINAS | 10 | 1 |
| 14 | LINHA | 9 | 1 |
| 15 | CARROCERIA | 8 | 1 |
| 16 | FIOS | 8 | 1 |
| 17 | CHAVE | 7 | 1 |
| 18 | EMPILHADEIRA | 7 | 1 |
| 19 | PARAFUSO | 7 | 1 |
| 20 | PEÇAS | 7 | 1 |

Fonte: WordSmith Tools.

Os sinais-termos identificados foram categorizados em seis (06) campos lexicais que designam as etapas no processo de fabricação dos automóveis, os instrumentos utilizados diariamente pelos funcionários da empresa, o maquinário utilizado no setor, os materiais, as peças e os procedimentos comuns à prática profissional do setor automobilístico. No Quadro 1, apresentaremos os termos em suas respectivas categorias.

Quadro 1: Listagem dos sinais-termos coletados distribuídos em categorias.

| CATEGORIA | UNIDADES TERMINOLÓGICAS |
|-----------------------------|---|
| Etapas de fabricação | Abastecimento, liberação, linha de produção, pátio de automóveis, pintura, teste, túnel. |
| Instrumentos | Apertadeira, chave de fenda, ferro de solda, lixadeira elétrica, palete, parafusadeira, parafuso, pinça de solda, poltriz, revólver de pintura, <i>spray</i> de pintura. |
| Máquinas | Elevador de carro, elevador da linha de produção, empilhadeira, laser, rebocador, robô de peças, robô de computador, robô de pintura, transportador. |
| Materiais | Borracha, caixa, ferro, fita adesiva isolante, lixa, máscara, plástico, vidro. |
| Peças de automóveis | Alto falante, ar-condicionado, assoalho, borracha de porta, câmbio, carroceria, chassi, conector da mangueira do radiador, farol, chicote, freio, mangueira inferior do radiador, mola, motor, painel, para-choque, peça, porta de carro, pneu, radiador, tampa do radiador, ventilador do radiador, volante. |
| Procedimentos | Calafetação, inspeção mecânica, lixar, montar, parafusar, pintar, polir, solda. |

Fonte: Elaborado pela autora.

Dentre os sinais coletados, foi possível apreender que, entre um entrevistado e outro, alguns sinais-termos apresentaram mais de um significante para o mesmo referente, como foi o caso de "carroceria", "chassi", "motor", "solda", "palete" e "pátio de automóveis", que apresentaram duas (02) variantes cada um, e "empilhadeira", que apresentou três (03) variantes.

Conforme mostra o Quadro 1, as categorias que mais apresentaram itens terminológicos foram as de instrumentos e peças de automóveis. Nossa hipótese para o número maior de sinais nestes grupos em relação aos outros é que, como foi relatado por alguns participantes, a transitividade deles dentro da empresa é restrita aos seus setores de lotação, o que os impossibilita de conhecer todas as etapas da fabricação dos automóveis, assim como máquinas, materiais e procedimentos comuns a outras áreas. No que diz respeito aos instrumentos e peças de automóveis, acreditamos que o conhecimento por parte dos participantes seja maior, devido a cinco (05) funcionários trabalharem na linha de produção, onde lidam com uma variedade maior de peças e instrumentos, o que lhes possibilitou relatar um número maior de unidades terminológicas.

No que se refere às análises dos sinais-termos, tendo em vista a quantidade de itens coletados, optamos por analisar um sinal que apresente características de iconicidade e um de arbitrariedade, para que as discussões tecidas a esse respeito sejam elucidadas nas análises. Elegemos o procedimento "calafetação" para representar os sinais icônicos e a peça "pneu" para representar os arbitrários.

Os sinais-termos serão apresentados em imagens e, posteriormente, faremos a descrição fonológica do sinal. A partir daí, partiremos para a análise motivacional do signo. As descrições serão embasadas nos cinco parâmetros para a formação dos sinais: Configuração das Mãos (CM), Orientação da Palma (OP), Ponto de Articulação (PA), Movimento (M) e Expressões não Manuais (ENM).

5 Análises



Sinal 1:

CM: 01. 🖐️
OP: para a esquerda.
PA: em espaço neutro frente ao corpo.
M: retilíneo para baixo.
ENM: não há.

Sinal-termo – Calafetação



Sinal 2:

CM: 42. 🖐️

OP: para a esquerda.

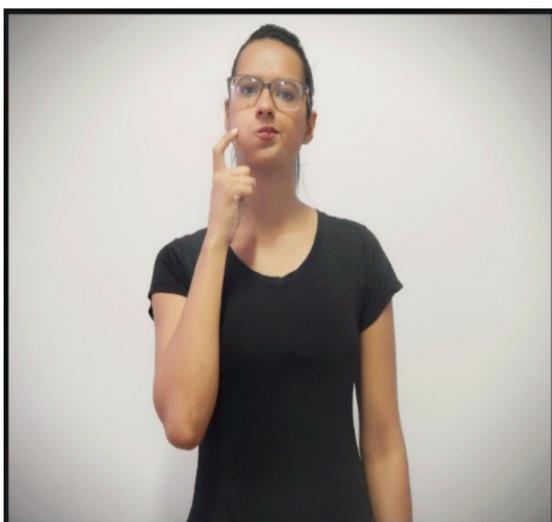
PA: em espaço neutro frente ao corpo.

M: encostar os dedos polegar e médio, duas repetições.

ENM: não há.

O procedimento “calafetação” tem a definição de “Ato ou efeito de calafetar; tapar ou vedar fendas e buracos” (FERREIRA, 2012, p. 388). O sinal-termo criado para este procedimento é icônico, por fazer alusão a sua operacionalização por meio da Configuração de Mão (CM) e Movimento (M) denotados no primeiro sinal, acrescido do sinal referente à cola (segundo sinal), que é o material utilizado no processo de vedação das superfícies dos automóveis. Desse modo, o sinal-termo enquadra-se no rol dos sinais motivados por fatores extralinguísticos.

Sinal-termo – Pneu



Descrição:

CM: 13 🖐️

OP: em direção ao rosto.

PA: dedo indicador fixado na bochecha.

M: não há.

ENM: bochecha inflada.

O sinal-termo que designa a unidade terminológica “pneu”, que se define como “uma estrutura resistente, chamada carcaça, formada por várias telas sobrepostas e cruzadas, que permitem a movimentação do automóvel” (FIAT, 2012, p. 63), foi analisado e considerado arbitrário, por não fazer nenhuma alusão quanto à forma, movimento ou utilidade do objeto. Consideramos existir diversas possibilidades de sinalizar este termo iconicamente, de modo a representar um objeto circular ou movimento rotativo, no entanto, os participantes foram unânimes em sinalizar o seu equivalente em Libras com o dedo indicador direito tocando na bochecha, o que o enquadra na categoria dos sinais-termos imotivados.

Dentre todos os sinais coletados, apenas três foram considerados arbitrários: “ferro”, “plástico” e “pneu”. Os demais foram todos analisados e categorizados como icônicos.

6 Considerações finais

Esta pesquisa buscou realizar um estudo terminológico em Libras com o auxílio da ferramenta *WordList* do programa *WordSmith Tools 6.0*, que possibilitou o tratamento do *corpus* especializado em Libras, permitindo a quantificação dos sinais-termos. A ferramenta *Concord*, que proporciona a visualização das ocorrências em linhas de concordância, ainda que não explorada no presente trabalho, será levada em consideração no desdobramento deste e para próximos trabalhos, fundamentalmente, no que tange à análise da sintaxe de agrupamentos lexicais e como suporte tanto para a descrição fonológica dos sinais quanto para eventuais padrões de sua origem motivacional.

Conforme observado, as análises nos possibilitaram apreender que a maioria dos sinais-termos coletados foram considerados icônicos, sendo sessenta e dois (62) sinais-termos icônicos, ou seja, 94% do total apresentou traços de semelhança entre o objeto e seu referente, enquanto somente três (03) foram considerados arbitrários, correspondendo a 4,5% do total, e apenas um (01) foi considerado relativamente icônico, correspondendo a 1,5% dos termos.

Nossa hipótese para a quantidade de sinais-termos icônicos encontrada é a falta de conhecimento e domínio da Libras dentro da empresa, o que dificulta a comunicação dos funcionários surdos com os seus pares ouvintes, sendo, muitas vezes, mais fácil estabelecer comunicação por meio de desenhos no ar. Desse modo, a falta de conhecimento e domínio da Libras por parte dos funcionários ouvintes do setor - que, para estabelecerem comunicação com os funcionários surdos, veem a necessidade de sinalizar de modo icônico para se fazerem entendidos em situações cotidianas - também contribui para a formação icônica dos sinais-termos.

Concluimos que a LC, inicialmente desenvolvida para estudos linguísticos de línguas orais, pode servir também às línguas de sinais, viabilizando pesquisas lexicais e terminológicas no âmbito da Libras, como é o caso do estudo que pretendemos desenvolver na tese de Doutorado, que diz respeito à compilação de um *corpus* de descrição fonológica dos sinais toponímicos do estado de Goiás, a fim de permitir, por meio da etiquetagem dos Parâmetros, as análises quanto à motivação do signo linguístico toponímico em Libras.

| Referências

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, 2013. p. 11-27.

FAULSTICH, E. Características conceituais que distinguem o que é de para que serve nas definições de terminologias científica e técnica. In: ISQUERDO, A. N.; DAL CORNO, G. O. M. (org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. v. VII. Campo Grande: Editora UFMS, 2014.

FERREIRA, A. B. H. **Dicionário Aurélio básico da língua portuguesa**. 5. ed. Rio de Janeiro: Nova Fronteira, 2012.

FIAT. **Dicionário técnico automotivo**. Impresso nº. 53001362. julho de 2012. Disponível em: www.infotecfiat.com.br/Por/Anexos/Geral/53001362%20-%20Dicionario%20Tecnico%20Automotivo%20montado.pdf. Acesso em: 20 maio 2020.

GESSER, A. **Libras? Que língua é essa?**: crenças e preconceitos em torno da língua de sinais e da realidade surda. São Paulo: Parábola, 2009.

KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à terminologia: teoria e prática**. São Paulo: Contexto, 2004.

PEIRCE, C. S. **Semiótica**. 2. ed. Tradução José Teixeira Coelho Neto. São Paulo: Perspectiva, 1995.

SAUSSURE, F. **Curso de lingüística geral**. 27. ed. Tradução Antônio Chelini, José Paulo Paes e Izidoro Blikstein. São Paulo: Cultrix, 2006.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://lexically.net/wordsmith/downloads/>. Acesso em: 17 maio 2020.

SOUZA, K. M. **Libras e Terminologia: estudo de sinais-termos do setor automobilístico de Catalão-GO**. 2018. Dissertação (Mestrado em Estudos da Linguagem) – Unidade Acadêmica Especial de Letras e Linguística, Universidade Federal de Goiás, Catalão, 2018. Disponível em: https://files.cercomp.ufg.br/weby/up/570/o/Dissertação_Kassia_Mariano.pdf. Acesso em: 20 maio 2020.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. 3. ed. São Paulo: Atlas, 2000.

Português para Falantes de Outras Línguas: uma proposta inicial de árvore de domínio à luz da Linguística de *Corpus*¹

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. Uma versão ampliada, com alterações metodológicas e discussões mais aprofundadas sobre os dados oriundos do *corpus* deste estudo exploratório, foi publicada em Lisboa (2020).

2 Doutorando em Estudos Linguísticos pela Universidade Federal de Uberlândia, sob orientação do Prof. Dr. Guilherme Fromm. Lattes: <http://lattes.cnpq.br/7657573383244824>.
E-mail: joelvictorlisboa@gmail.com.

1 Introdução

Português para Falantes de Outras Línguas (doravante PFOL) é uma área de atuação científico-profissional que se ocupa do ensino e aprendizagem da língua portuguesa em diversos contextos e para diferentes públicos-alvo com necessidades e objetivos linguísticos distintos. Em virtude disso, cada contexto de atuação do PFOL apresenta especificidades que complexificam a efetividade dos mesmos modelos de ensino nos demais contextos. Essa situação prevê diferenciações de materiais e métodos pedagógicos e, por conseguinte, estimula desdobramentos da área de PFOL.

Haja vista a diversidade tangente à área, este capítulo objetiva apresentar um estudo exploratório cuja finalidade foi realizar um mapeamento inicial dos possíveis campos de atuação do PFOL e, a partir disso, realizar uma proposta inicial de árvore de domínio da área. Para isso, partimos da compilação de artigos publicados em um periódico brasileiro especializado em PFOL – a *Revista SIPLE*³ – e, por meio dos princípios metodológicos da Linguística de *Corpus* e da utilização do programa de análise lexical *WordSmith Tools 6.0* (SCOTT, 2012), fizemos uma análise quali-quantitativa do *corpus* de estudo.

As próximas seções estão organizadas do seguinte modo: (i) apresentação da heterogeneidade tangente ao ensino e aprendizagem de PFOL, a justificativa deste estudo exploratório e da escolha de PFOL para designar a área; (ii) introdução da Terminologia, área em que se insere este estudo, e da Linguística de *Corpus*, utilizada como metodologia nesta pesquisa; (iii) descrição dos procedimentos metodológicos para a identificação dos possíveis campos de atuação do PFOL; (iv) discussão dos resultados obtidos e apresentação da proposta de árvore de domínio e (v) considerações finais sobre o estudo realizado.

2 Fundamentação teórico-metodológica

O interesse de falantes de outras línguas pela língua portuguesa vem crescendo exponencialmente no âmbito mundial devido a diversos fatores, como: (i) o estatuto da língua em diversos países e blocos econômicos; (ii) acordos bi e/ou multilaterais entre empresas e instituições de ensino; (iii) turismo; (iv) fluxo migratório que vem crescendo vertiginosamente, dentre outros fatores⁴. Acreditamos que esse cenário complexo e heterogêneo aumenta a demanda de ensino de PFOL e estimula ramificações nesta área, em função dos contextos de ensino, de modo a buscar atender às necessidades e/ou aos propósitos de ensino e aprendizagem da língua da melhor maneira possível, levando em conta as características socio

³ *Revista da Sociedade Internacional de Português Língua Estrangeira*, disponível em: <https://assiple.org/index.php/revista-siple/>. Acesso em: 21 out. 2019.

⁴ Para um aprofundamento sobre essa discussão em um panorama do português no mundo, ver Moita Lopes (2013), dentre outros.

e psicolinguísticas dos públicos-alvo atendidos pela área. Isto posto, é relevante evidenciar e diferenciar as modalidades de atuação no âmbito da área de PFOL, de modo a promover maior conscientização sobre as esferas de atuação da área.

É pertinente justificar a escolha da unidade fraseológica especializada⁵ (doravante UFE) “PFOL” em face à UFE comumente empregada nos estudos da área, “Português (como) Língua Estrangeira” (doravante PLE). Assim como Santos (2018, p. 806), consideramos que “se o aprendizado da língua alvo ocorrer fora da comunidade nativa desse idioma, diz-se que se trata de aprendizagem de língua estrangeira (LE)”. Sendo assim, consideramos PLE como uma das modalidades de ensino e aprendizagem da área sob análise que, via de regra, acontece em contextos não lusófonos onde não há imersão na língua-alvo, dentre outras especificidades. Em vista disso, no escopo da área em questão, acreditamos que PFOL é uma UFE “guarda-chuva” mais apropriada para designá-la, visto que não engendra confusões em relação aos contextos de ensino das subáreas, como acontece com a utilização de PLE. Ademais, segundo as análises apresentadas em Lisboa (2020), dentre as UFEs utilizadas para designar a área como um todo, PFOL é a menos variante em termos formais e conceituais no *corpus* analisado.

Por lidar diretamente com UFEs, este estudo insere-se no âmbito da Terminologia, subárea da Linguística voltada às comunicações especializadas e que possui como objeto de estudo central os termos, abrangendo as UFEs e as definições terminológicas (KRIEGER; FINATTO, 2004). Grosso modo, a Terminologia ocupa-se da identificação, análise, descrição e, por vezes, sistematização de terminologias das áreas técnico-científicas, bem como é voltada às considerações teóricas sobre os processos designativos e sobre a natureza das terminologias no âmbito das áreas de especialidade (CABRÉ, 1999).

Para realizar o mapeamento dos possíveis campos de atuação do PFOL, representados por UFEs, utilizamos a Linguística de *Corpus* (doravante LC), que consiste em uma metodologia/ abordagem quanti-qualitativa, de base empírica e estatística, voltada à análise e descrição linguística por meio da investigação em *corpora* computadorizados. No escopo da LC, *corpora* são “conjuntos de dados lingüísticos textuais [em formato legível por computador] coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística” (BERBER SARDINHA, 2004, p. 3).

Por meio da utilização de ferramentas computacionais, a LC permite a exploração de grandes quantidades de texto com rapidez impossível de ser alcançada manualmente. Ademais, as ferramentas possibilitam a organização dos dados de acordo com os propósitos da pesquisa, facilitando análises quali-quantitativas, bem como geram resultados cuja precisão promove

⁵ Unidades fraseológicas especializadas (UFEs) são, em suma, agrupamentos sintagmáticos de valor terminológico, geralmente marcados pela frequência de (co)ocorrência e pela presença de termos entre seus constituintes, que representam nódulos conceituais das áreas das quais fazem parte e que, por conseguinte, veiculam conhecimento especializado (KRIEGER; FINATTO, 2004).

“maior confiabilidade e certeza nas análises quantitativas dos traços [e padrões] linguísticos em grandes amostras textuais⁶” (PARODI, 2010, p. 22). As características mencionadas nos parágrafos anteriores justificam a utilização da LC como metodologia/abordagem neste estudo, visto que faremos o levantamento de candidatos a UFEs que, possivelmente, designam as subáreas do PFOL por meio da análise da coocorrência de itens lexicais no *corpus* de estudo.

3 Procedimentos metodológicos

O primeiro procedimento metodológico foi a delimitação da fonte de compilação dos arquivos que comporiam o *corpus* de estudo. A *Revista SIPLE* foi selecionada devido aos seguintes fatores: (i) é um periódico inteiramente voltado à área de PFOL; (ii) por ter nove edições e um número considerável de artigos escritos por diversos autores, o procedimento de compilação do *corpus* seria realizado de maneira mais célere e o *corpus* teria tamanho suficiente para a realização da análise exploratória⁷.

Delimitada a fonte de compilação do *corpus*, os próximos procedimentos metodológicos são subdivisíveis em duas grandes etapas, a saber: (i) compilação e preparação do *corpus* para processamento pelo *WordSmith Tools* 6.0 (SCOTT, 2012), doravante WST; (ii) processamento e análise do *corpus* de estudo.

3.1 Compilação e preparação do corpus

A primeira etapa metodológica contou com os seguintes procedimentos:

- (i) Compilação dos artigos publicados nas nove edições da *Revista SIPLE*;
- (ii) Atribuição de códigos aos arquivos para facilitar sua futura identificação;
- (iii) Conversão dos arquivos compilados em .pdf⁸ para .docx⁹, por meio do procedimento de OCR¹⁰ do programa *ABBYY FineReader*¹¹;
- (iv) Revisão e limpeza de cada arquivo, excluindo informações pós-textuais;
- (v) Conversão dos arquivos em .docx para .txt¹².

6 No original: “[...] mayor confiabilidad y certeza en los análisis cuantitativos de los rasgos lingüísticos en grandes muestras de textos” (tradução nossa).

7 Sabemos que um *corpus* composto apenas por artigos oriundos de um só periódico não é representativo da área como um todo, mas como o estudo relatado neste capítulo é apenas exploratório, o *corpus* compilado é suficiente para a realização dessa primeira análise.

8 Sigla de *Portable Document Format*, formato portátil de documento.

9 Extensão de arquivo desenvolvida pela *Microsoft Corporation*, legível pela ferramenta *Microsoft Word* do *Microsoft Office Suite*.

10 Sigla de *Optical Character Recognition*, reconhecimento óptico de caracteres.

11 Disponível em: <https://www.abbyy.com/pt-br/finereader/>. Acesso em: 19 maio 2020.

12 Documento de texto sem formatação. Esse procedimento é necessário, haja vista que, em geral, arquivos em .txt são os mais satisfatoriamente processados pelos programas de análise lexical.

Ao final dos procedimentos descritos, o *corpus* de estudo foi mensurado com 66 arquivos, cada arquivo representando um artigo das nove edições da *Revista SIPLE*. Ademais, o *corpus* possui ao todo 230.081 *tokens* e 18.731 *types*.

3.2 Processamento e análise do corpus

Nessa segunda etapa metodológica, o primeiro procedimento foi a utilização do utilitário *Text Converter* do programa WST para a conversão em massa da codificação dos arquivos em .txt (que via de regra é UTF-8) para a codificação utilizada pelo WST (Unicode). Feito isso, carregou-se o *corpus* no WST e, por meio da ferramenta *Concord*, buscou-se por “português” e “portuguesa”, a fim de encontrarmos candidatos a UFEs possivelmente utilizados para designar campos de atuação do PFOL. Esse procedimento resultou na geração do total de 1.693 linhas de concordância, das quais 477 apresentaram candidatos a UFEs contendo “português” ou “portuguesa” como um dos constituintes.

Os dados foram tabelados conforme o código dos arquivos e o número de vezes que cada UFE ocorreu em cada arquivo. Foi necessário fazer esse levantamento de ocorrências, pois, em virtude da alta variação terminológica, estabelecemos como critério para determinar se um agrupamento lexicogramatical seria considerado como possível UFE designadora de subárea ou não a sua ocorrência em, pelo menos, dois artigos de autoria diferente. Isto porque alguns candidatos a UFEs foram utilizados mais de uma vez, porém em um único artigo do *corpus* ou em dois artigos do mesmo autor, não sendo representativos e, por conseguinte, se constituindo como escolhas lexicais particulares de determinado autor.

Nos casos em que foram encontrados candidatos a UFEs como “português como LE/LA”, os separamos em duas UFEs distintas, “português como LE” e “português como LA”, haja vista que seria relevante quantificar a utilização de cada UFE e de cada forma variante. Ademais, a variação entre letras maiúsculas e minúsculas foi ignorada, por não ser pertinente para este estudo, no momento.

Antes de passarmos à discussão dos resultados, é relevante pontuar que os candidatos a UFEs extraídos do *corpus* não necessariamente foram utilizados pelos autores para denominar os campos de atuação, mas podem ter sido utilizados como citações de nomes de cursos, de livros, de departamentos etc. Entretanto, como possuem potencial valor designativo das diferentes subáreas de atuação do PFOL, foram selecionadas como candidatos a UFEs designativas neste estudo.

4 Resultados e discussão

A partir da análise do *corpus* e aplicação dos critérios estabelecidos (cf. Subseção 3.2), foram identificados cinco candidatos a UFEs designativas de possíveis subáreas de atuação do

PFOL. A Tabela 1 apresenta os candidatos a UFEs identificados, a quantidade de ocorrências de cada candidato e de cada forma variante, bem como a quantidade de artigos em que cada forma ocorre:

Tabela 1: Candidatos a UFEs designativas de potenciais subáreas do PFOL.

| LÍNGUA ADICIONAL | | |
|---------------------------------|-------------|---------|
| CANDIDATO A UFE | OCORRÊNCIAS | ARTIGOS |
| Português como Língua Adicional | 6 | 5 |

| LÍNGUA DE ACOLHIMENTO | | |
|--------------------------------------|-------------|---------|
| CANDIDATOS A UFE | OCORRÊNCIAS | ARTIGOS |
| Português como Língua de Acolhimento | 35 | 8 |
| Português Língua de Acolhimento | 11 | 3 |

| LÍNGUA DE HERANÇA | | |
|----------------------------------|-------------|---------|
| CANDIDATOS A UFE | OCORRÊNCIAS | ARTIGOS |
| Português como Língua de Herança | 16 | 7 |
| Português Língua de Herança | 3 | 2 |

| LÍNGUA ESTRANGEIRA | | |
|---|-------------|---------|
| CANDIDATOS A UFE | OCORRÊNCIAS | ARTIGOS |
| Português como Língua Estrangeira | 60 | 31 |
| Português Língua Estrangeira | 61 | 24 |
| Português como LE | 4 | 2 |
| Português LE | 3 | 2 |
| Língua Portuguesa como língua estrangeira | 3 | 2 |

| SEGUNDA LÍNGUA | | |
|-------------------------------|-------------|---------|
| CANDIDATOS A UFE | OCORRÊNCIAS | ARTIGOS |
| Português como Segunda Língua | 26 | 14 |
| Português Língua Segunda | 6 | 4 |
| Português como L2 | 2 | 2 |
| Português como Língua Segunda | 4 | 2 |
| Português L2 | 2 | 2 |

Fonte: Elaborada pelo autor.

A partir da Tabela 1, é possível observar que os candidatos a UFEs designadoras de algumas possíveis subáreas, como “Português como Língua Estrangeira” e “Português como Segunda Língua”, apresentam menor grau de padronização terminológica, neste *corpus* de estudo, em face a outros como “Português como Língua Adicional”. Essa quantificação de ocorrência dos candidatos a UFEs e dos artigos em que cada candidato foi utilizado auxiliou na eleição das formas que constariam na proposta de árvore de domínio para designar as possíveis subáreas de ensino e aprendizagem de PFOL, pois elegemos as formas mais frequentes e que foram utilizadas no maior número de artigos.

É relevante pontuar que, além dos candidatos a UFEs apresentados na Tabela 1, há outros que também ocorreram em mais de um artigo (com variações), mas que não foram considerados designadores de possíveis subáreas de atuação do PFOL. Esses candidatos a UFEs estão apresentados na Tabela 2, a seguir:

Tabela 2: Candidatos a UFEs não considerados como designadores de possíveis subáreas do PFOL.

| CANDIDATO A UFE | OCORRÊNCIAS | ARTIGOS |
|--------------------------------|-------------|---------|
| Português no Estrangeiro | 10 | 2 |
| Português para Hispanofalantes | 2 | 2 |
| Português para Indígenas | 2 | 2 |
| Português para Alunos Chineses | 2 | 2 |

Fonte: Elaborada pelo autor.

O candidato “Português no Estrangeiro” não foi considerado como designador de uma possível subárea de atuação do PFOL, nesta pesquisa, devido à sua imprecisão em relação aos contextos de ensino e aprendizagem. À título de exemplo, observemos as seguintes situações:

- (i) um estudante polonês, de família polonesa, aprendendo português em seu país de origem;
- (ii) um estudante australiano, filho de pais brasileiros, aprendendo português também em seu país de origem, mas como forma de preservar a língua e cultura constituintes da formação identitária de sua família.

Se considerarmos Português no Estrangeiro como uma subárea de ensino e aprendizagem do PFOL, o estudante polonês do primeiro exemplo estaria inserido simultaneamente em duas possíveis subáreas: Português no Estrangeiro e Português como Língua Estrangeira¹³. No que tange ao segundo exemplo, o estudante australiano estaria igualmente inserido em duas possíveis subáreas: Português no Estrangeiro e Português como Língua de Herança¹⁴.

13 Isto porque consideramos Português como Língua Estrangeira o ensino e aprendizagem de língua portuguesa que ocorre fora de países cuja língua majoritária é o português, ou seja, em contextos de não imersão linguística (cf. Seção 2), como é o caso do exemplo apresentado.

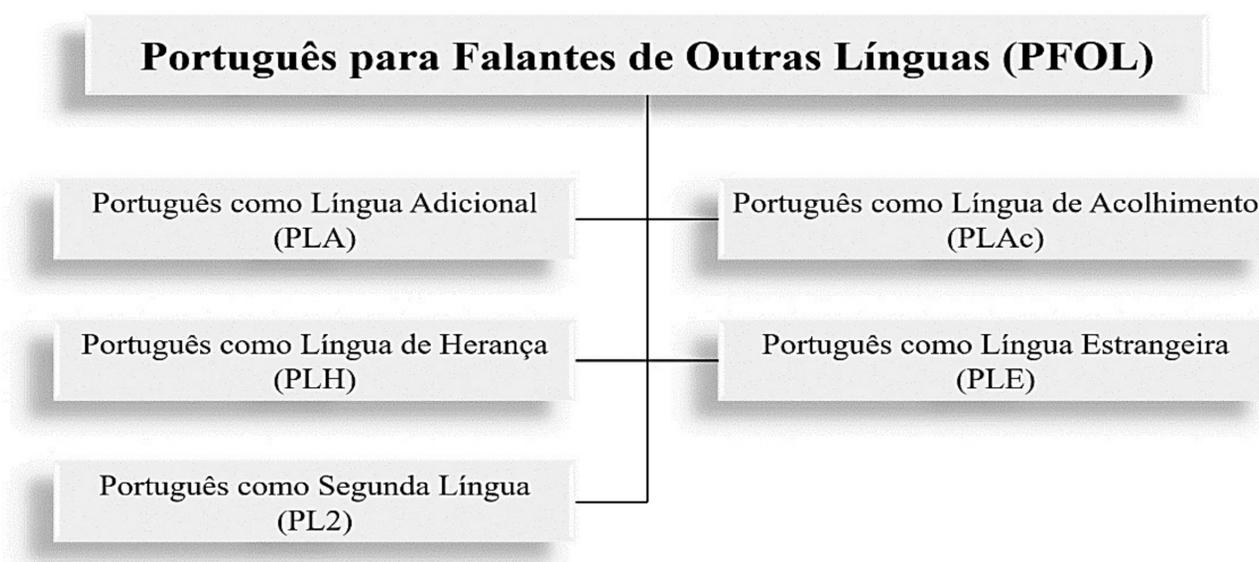
14 No caso ilustrado no segundo exemplo, poder-se-ia argumentar, também, que o estudante estaria inserido no ensino e aprendizagem de Português como Língua Estrangeira. Não obstante, a motivação de preservar um legado linguístico-cultural familiar é um traço caracterizador da subárea Português como Língua de Herança, traço este que não ocorre no primeiro exemplo. Ver mais sobre o Português como Língua de Herança em Mendes (2012), Cunha (2013), dentre outros.

A problemática não reside na possibilidade de aproximação de metodologias e abordagens de duas subáreas distintas em função do contexto de ensino e aprendizagem e das necessidades linguísticas do alunado, mas sim na imprecisão de “Português no Estrangeiro” como designação de uma possível subárea, pois nos dois exemplos apresentados, o ensino da língua ocorre em contextos estrangeiros em relação a contextos lusófonos. Essa imprecisão justifica a não consideração do candidato a UFE “Português no Estrangeiro” como designação de uma possível subárea de ensino e aprendizagem de PFOL.

Em relação aos outros três candidatos a UFEs apresentados na Tabela 2, “Português para Hispanofalantes”, “Português para Indígenas” e “Português para Alunos Chineses”, apesar de reconhecermos que há especificidades de ensino e aprendizagem em função da língua materna do público-alvo, entendemos que a UFE designativa “PFOL” já abarca o ensino e aprendizagem para falantes de línguas diversas, sejam elas línguas indígenas, chinesa ou espanhola.

Após a análise do *corpus* de estudo e as reflexões sobre as possíveis subáreas de ensino e aprendizagem de PFOL, foi construída a proposta inicial árvore de domínio inicial da área, que está apresentada na Figura 1:

Figura 1: Proposta inicial da árvore de domínio do PFOL.



Fonte: Elaborada pelo autor.

As siglas e acrônimos utilizados na proposta de árvore de domínio apresentada na Figura 1 foram retirados do *corpus*, contudo, algumas variações foram identificadas, por exemplo: para Português como Língua de Herança, foram observadas as siglas PLH e POLH, e para Português como Segunda Língua identificou-se PSL, PL2 e PLE. Em vista disso, para eleger uma única sigla ou acrônimo referente a cada possível subárea para compor a proposta de árvore de domínio, fizemos o levantamento quantitativo das ocorrências das siglas que vinham acompanhadas do candidato a UFE a que estavam se referindo. Portanto, as siglas e acrônimos que compõem a árvore de domínio foram os mais utilizados para designar cada uma das potenciais subáreas do PFOL apresentadas.

Por fim, é importante enfatizar que a árvore de domínio apresentada é apenas um esboço realizado a partir da análise terminológica exploratória de um *corpus* composto por artigos publicados em apenas um periódico especializado em PFOL. Para que seja possível realizar um estudo mais aprofundado sobre a terminologia designativa das subáreas do PFOL, que seja empiricamente fundamentado e que dê bases para afirmações mais precisas, é necessário partir de um *corpus* mais representativo da área, considerando os critérios de autenticidade, diversidade (de autores, de gêneros textuais e de origens dos textos), dentre outros critérios apontados em Berber Sardinha (2004), que são basilares para os estudos com *corpora*.

5 Considerações finais

Neste capítulo, apresentou-se os resultados oriundos de um estudo exploratório que visou realizar um mapeamento inicial das possíveis ramificações da área de PFOL, a partir da análise de um *corpus* composto por 66 artigos científicos publicados na *Revista SIPLE*. Outrossim, apresentou-se uma proposta inicial de árvore de domínio do PFOL, desenvolvida a partir da análise exploratória do *corpus* de estudo.

É relevante enfatizar a proficuidade da utilização de programas de análise lexical de base estatística, como o WST, bem como os princípios metodológicos da LC para guiar a realização de estudos desse tipo. Isto porque se o trabalho apresentado neste capítulo fosse realizado manualmente, precisaríamos de muito mais tempo para concluí-lo, teria sido relativamente mais difícil organizar todos os candidatos a UFEs designativas em seus contextos de ocorrência e não teríamos comprovações quantitativas satisfatoriamente precisas para corroborar nossas análises.

Por fim, destacamos que a experiência na realização do estudo exploratório apresentado neste capítulo nos instigou a desenvolver uma pesquisa de mestrado recém-defendida (LISBOA, 2021), cujo objetivo foi propor uma harmonização da terminologia designativa de área e subáreas do PFOL¹⁵, com base em um *corpus* de amostragem mais abrangente.

| Referências

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

CABRÉ, M. T. **Terminology: theory, methods and applications**. Amsterdam: John Benjamins Publishing Company, 1999.

¹⁵ Como este estudo exploratório foi realizado em 2019, optamos por manter a designação da área como PFOL neste capítulo, haja vista os dados obtidos no estudo em questão. Contudo, conforme a harmonização terminológica proposta em nossa pesquisa de mestrado, designamos a área como Português como Língua Não Materna (PLNM).

CUNHA, M. J. C. Língua de herança: estratégias na aquisição da língua dos pais. **Revista SIPLE**, Brasília, v. 4, n. 2, p. 19-28, 2013. Disponível em: https://docs.wixstatic.com/ugd/aa72b0_9aeb1cd15ab547df801c509e7978e193.pdf. Acesso em: 20 out. 2019.

KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à terminologia**: teoria e prática. São Paulo: Contexto, 2004.

LISBOA, J. V. R. **Proposta de harmonização da terminologia designativa de área e subáreas do Português como Língua Não Materna baseada em corpus**. 2021. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2021. 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>.

MENDES, E. Vidas em português: perspectivas culturais identitárias em contexto de português língua de herança (PLH). **Platô**, Praia, v. 1, n. 2, p. 20-31, 2012.

MOITA LOPES, L. P. (org.). **O português no século XXI**: cenário geopolítico e sociolinguístico. São Paulo: Parábola, 2013.

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

SANTOS, D. C. Análise de erros na produção escrita de estudantes de PL2: uma abordagem sociocultural. **Domínios de Lingu@gem**, Uberlândia, v. 12, n. 2, p. 803-838, 2018. DOI: <https://doi.org/10.14393/DL34-v12n2a2018-3>.

SCOTT, M. **WordSmith Tools version 6**. Stroud: Lexical Analysis Software, 2012. Disponível em: <https://www.lexically.net/wordsmith/downloads/>. Acesso em: 20 out. 2019.

O léxico *kimbundu* no português oralizado em Angola: uma análise da frequência em *Os contos de Ukamba Kimba*¹

Ivonete da Silva Santos²

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 da Linguagem pela Universidade Federal de Catalão, sob orientação da Profa. Dra. Maria Helena de Paula. Lattes: <http://lattes.cnpq.br/9475011369057638> E-mail: nete.silva.santos@hotmail.com

1 Introdução

Neste capítulo, são apresentados os resultados de um estudo exploratório sobre a existência de um léxico misto no português falado em Angola, a partir da análise quali-quantitativa da obra *Os contos de Ukamba Kimba* (VILANOVA, 2013), obra composta por vinte e quatro contos escritos por um escritor angolano, que retrata a realidade linguística do português oralizado em Angola. O estudo quali-quantitativo de palavras do léxico *kimbundu* no português angolano é pertinente para se justificar a relevância dessa língua nativa para a formação do conjunto vocabular do português falado em Angola.

Este estudo exploratório objetivou fazer um levantamento quantitativo semiautomático do tesouro vocabular do *kimbundu* na obra *Os contos de Ukamba Kimba* (VILANOVA, 2013), assim como verificar a frequência do uso de palavras do léxico *kimbundu* na obra analisada, sob a perspectiva da Linguística de *Corpus* (LC). Esses objetivos são relevantes para nortear os caminhos que responderão à pergunta: pensando na formação do português falado em Angola, qual é a representatividade do léxico do *kimbundu*, identificado na obra analisada?

Esta análise se faz necessária e fundamental para a promoção de material teórico e analítico voltado às variedades da língua portuguesa, em especial, à variedade angolana. O que será uma contribuição para novas descobertas e registros escritos de fenômenos linguísticos que, ainda, não foram evidenciados no âmbito da variedade angolana, principalmente no contexto da obra que constitui o *corpus* desta pesquisa. Esta obra é, ainda, pouco explorada no contexto literário, e menos conhecida ainda no âmbito dos estudos linguísticos, por isso a importância desta análise.

As análises foram empreendidas com o uso de métodos oriundos da LC, tendo como aporte teórico Berber Sardinha (2009), Novodvorski e Finatto (2014), Perini (2004), dentre outros necessários a este estudo. A seguir, serão brevemente apresentadas as noções teóricas sobre língua e linguagem, bem como a LC. Posteriormente, os caminhos percorridos para a compilação e preparação do *corpus* serão descritos e, por fim, um recorte da análise e discussão dos resultados será apresentado. A seção a seguir apresenta o aporte teórico que sustenta este estudo.

2 Fundamentação teórica

A língua é um instrumento de comunicação necessário à integração social do ser humano, o que justifica a flexibilidade que a língua tem de se adaptar às configurações culturais de uma determinada comunidade. Nesse sentido, a língua portuguesa (LP) é um instrumento de comunicação que motiva a interação dos membros das comunidades que integram a lusofonia

com o mundo em si. Ademais, a LP é um sistema de signos sensível à realidade sociocultural do espaço de atuação e, por isso, está sujeita a variações fonético-fonológicas, lexicais e semânticas.

Diante disso, é pertinente pensar na importância das interfaces da LP em consonância com as culturas vigentes em cada comunidade da lusofonia. Por isso, o estudo do léxico das variedades da LP é importante para que se entendam as configurações atuais do léxico do português falado em cada comunidade. Neste estudo exploratório, será considerado o léxico do português oralizado em Angola, bem como as influências do léxico da língua nativa *kimbundu* nessas configurações, ambos a partir da análise da obra que se constitui como nosso *corpus* de estudo.

A língua se configura, nesse sentido, como parte essencial da linguagem, existindo como um produto social depositado na memória coletiva que é herdada culturalmente dentro de um processo inteiramente coercitivo (SAUSSURE, 2006). A importância e o papel da língua materna como fator de determinação do modo como o ator social pensa e apreende o mundo é fundamental para sua sobrevivência em comunidade. A língua é um fenômeno social que, em si mesmo, é puramente abstrato, porque não apresenta uma existência física, mas em determinadas ocasiões é atualizado no comportamento linguístico. Dessa forma, viver em comunidade pressupõe uma interação recíproca entre língua/falante/cultura. Assim, Perini (2004, p. 52) afirma que “cada língua é um retrato do mundo, tomando de um ponto de vista diferente, e que revela algo não tanto sobre o próprio mundo, mas sobre a mente do ser humano”.

Diante disso, o léxico de uma língua se constitui como repertório linguístico geral utilizado pelos falantes para construir frases, períodos, parágrafos, textos e discursos, de modo a corresponder às necessidades comunicacionais dos interactantes. Isto é, o léxico de uma língua, segundo Biderman (2001), constitui uma forma de registrar o conhecimento do universo. O léxico se diferencia do vocabulário pelo fato de este último ser próprio do indivíduo (idioleto) e ser mais concreto, enquanto o léxico é mais abstrato e virtual (TIMBANE, 2013).

Para auxiliar os estudos do léxico das variedades da LP, acreditamos que os princípios metodológicos da LC para a geração de dados para análise são profícuos. Apesar de a LC ser considerada ora como área, ora como metodologia de pesquisa, ela tem sido cada vez mais útil a pesquisas em diferentes áreas, pois, como campo que se preocupa com a “criação e análise de *corpora*” (BERBER SARDINHA, 2009, p. 7), tem se mostrado inovadora no âmbito dos estudos que investigam a linguagem, pois possibilita ao pesquisador um “certo número de dados antes inacessíveis” (BERBER SARDINHA, 2009, p. 7), dados esses processados e organizados por meio de recursos computacionais. Assim, a LC se constitui “tanto como uma

metodologia quanto como uma abordagem teórica diferenciada dos Estudos da Linguagem” (NOVODVORSKI; FINATTO, 2014, p. 7).

Por tudo isso, a LC é pertinente para os estudos linguísticos, nesse caso, estudos do léxico, pois permite ao pesquisador um tratamento adequado dos dados. Isto posto, a LC é um caminho relevante que permite ao linguista uma análise profunda dos fatos da língua, até porque, segundo Perini (2006, p. 40), o bom linguista “estuda a língua como ela é, e não ideias a respeito de como ela deveria ser”. A esse respeito, é importante destacar que uma análise sistemática dos fatos de uma determinada língua é condicionada pelo modo como o linguista trabalha os dados, e a LC pode ser considerada como um caminho pertinente para tal.

Toda pesquisa carece de uma metodologia que justifique e sustente o desenvolvimento da análise proposta, por isso a seção a seguir trata especialmente dos passos que foram necessários na realização deste trabalho.

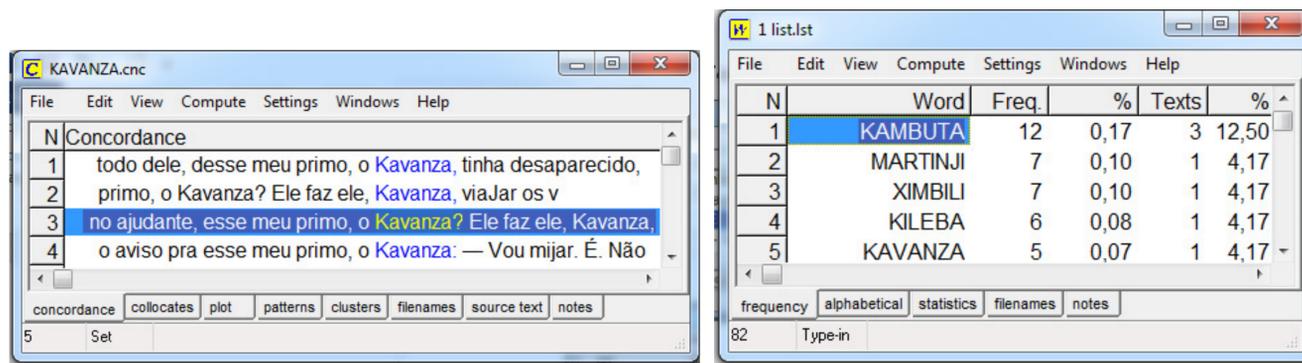
3 Metodologia

Os dados apresentados neste trabalho resultam da análise do *corpus* constituído pela obra *Os contos de Ukamba Kimba* (VILANOVA, 2013), um conjunto de vinte e quatro (24) contos que a integra. Esta obra apresenta uma linguagem que é reflexo do português angolano. Por serem textos com riqueza variacionista, são pertinentes para os estudos linguísticos.

Pois bem, a pesquisa foi desenvolvida em três fases: i) *Compilação do corpus*: essa fase compreende a seleção da obra (texto escrito) e a digitalização dos textos (24 contos); ii) *Preparação do corpus*: foi importante a essa fase a aplicação de OCR, a conversão do texto em formato .pdf para .doc., e posteriormente para .txt, a limpeza, a revisão, bem como o armazenamento dos textos em subpastas; iii) *Análise*: essa diz respeito à inserção do *corpus* no programa *WordSmith Tools* – doravante WST (SCOTT, 2004) e a utilização das ferramentas *WordList* e *Concord* para levantamento dos dados estatísticos e para a realização da análise quali-quantitativa.

O *software* WST, juntamente com suas ferramentas *WordList* e *Concord*, foram fundamentais para a análise do *corpus* desta pesquisa. Por meio da ferramenta *WordList*, foi possível gerar uma lista de palavras com todas as palavras existentes no *corpus*, tornando mais evidente a frequência de ocorrência de cada uma delas. Por sua vez, a ferramenta *Concord* nos possibilitou verificar as ocorrências das palavras do léxico *kimbundu* em seus contextos linguísticos. A Figura 1 ilustra a apresentação dos dados em cada ferramenta evidenciada.

Figura 1: Ferramentas *Concord* (à esquerda) e *WordList* (à direita).



Fonte: Programa WST.

Apenas as palavras do léxico *kimbundu* foram privilegiadas, tanto na lista de palavras quanto nas linhas de concordâncias. Após os resultados quantitativos semiautomáticos, foi empreendida a análise e discussão dos dados, o que pode ser verificado na seção a seguir.

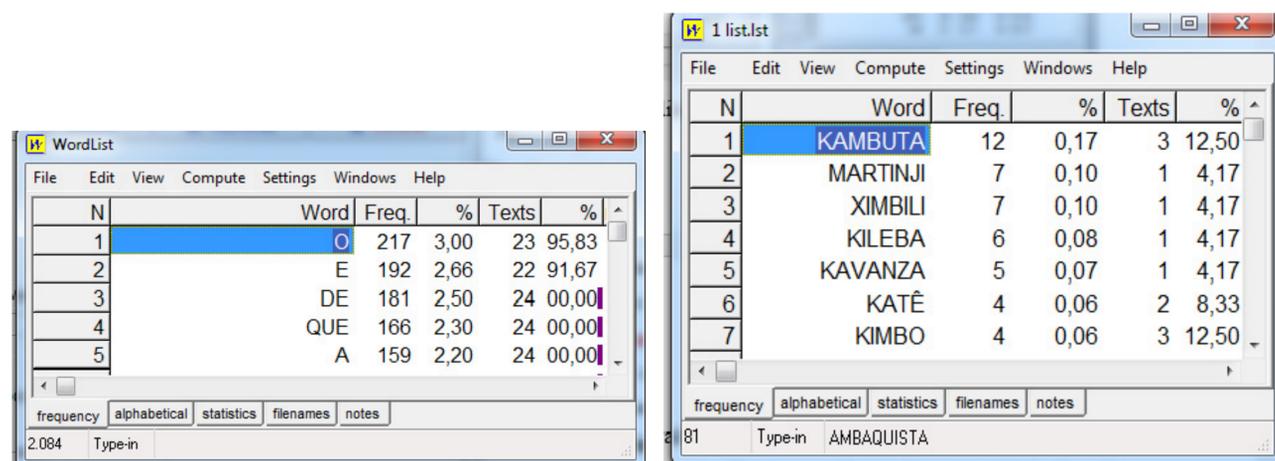
4 Análise e discussão dos resultados

Nesta seção, propõe-se a apresentação da análise e, posteriormente, a discussão dos resultados. Passos estes importantes para tornar credível a pesquisa, uma vez que os resultados comprovam a veracidade da análise em si.

4.1 A frequência de palavras do léxico *kimbundu* no léxico do português angolano

A análise da frequência de palavras *kimbundas* no *corpus* evidenciou a ocorrência de 81 vocábulos do *kimbundu* para um quantitativo de 2.084 *types*, que representam o total de palavras diferentes, no conjunto de 24 contos que compõem a obra sob análise, como é possível verificar na Figura 2, a seguir:

Figura 2: Quantitativo de palavras *kimbundas* (à esquerda) em relação ao total de *types* do *corpus* (à direita).



Fonte: Programa WST.

Com base nos dados evidenciados na Figura 2, é possível observar que o número de *types* de palavras *kimbundas* (81) é relativamente significativo se observada a quantidade de *types* no *corpus* como um todo (2.084). A partir disso, é pertinente dizer que o léxico do português angolano é marcado fortemente pela influência do *kimbundu*.

É importante ressaltar que, para além dessa constatação, a análise revelou que um conjunto de palavras do léxico *kimbundu* registrou certa recorrência, fato que leva a concluir que não se trata de ocorrências aleatórias. A Figura 3, a seguir, ilustra:

Figura 3: Alguns vocábulos do *kimbundu* de maior frequência no *corpus*.

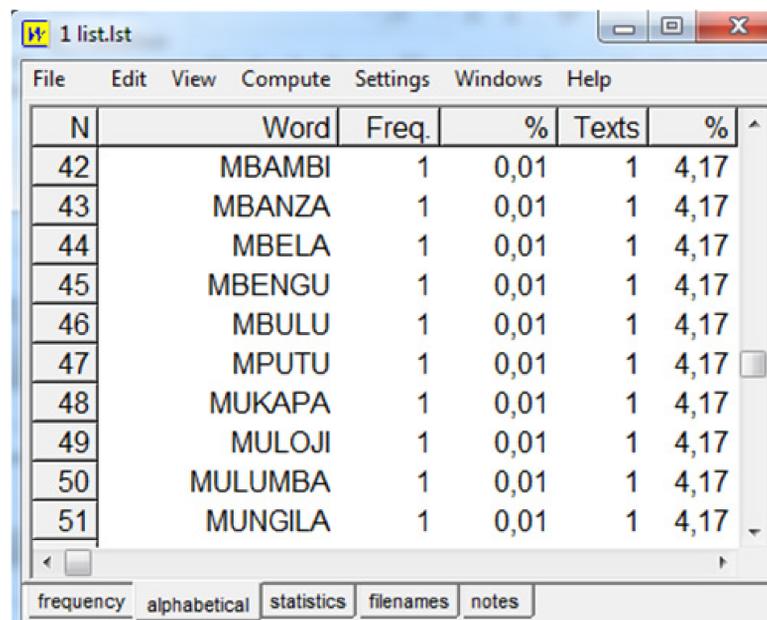
| N | Word | Freq | % | Texts | % |
|----|----------|------|------|-------|------|
| 9 | KAMUTUTA | 3 | 0,01 | 1 | 4,17 |
| 10 | KARIANGA | 3 | 0,04 | 1 | 4,17 |
| 11 | KATÉ | 2 | 0,03 | 2 | 8,33 |
| 12 | KATÊ | 4 | 0,06 | 2 | 8,33 |
| 13 | KAVANZA | 5 | 0,07 | 1 | 4,17 |

Fonte: Programa WST.

A Figura 3 apresenta as palavras que possuem um número maior de ocorrências no *corpus* sob análise, com indicativos de porcentagens e em quais textos essas palavras ocorreram. Por exemplo, a palavra *kambuta*, cujo significado é “de pequena estatura; anão”, apresenta uma frequência de 12 ocorrências, e aparece em três dos vinte e quatro contos que compõem o *corpus*. Um outro exemplo é o da palavra *kavanza*, que, por sua vez, significa “com desordem; com confusão”. Esta é a segunda palavra que ocorreu mais vezes no *corpus*, mas, por outro lado, ocorreu em apenas um dos textos. Outro fato curioso que importa mencionar é a ocorrência das palavras *katé* e *katê*, cuja diferença é marcada pela acentuação gráfica em que o fonema **-e** apresenta som fechado e aberto, mas o significado se mantém o mesmo para as duas versões: “até”.

Foram evidenciadas, também, palavras que ocorreram em menor grau, como se pode verificar na Figura 4, ou seja, essas apresentam uma frequência relativamente menor em relação às apresentadas na Figura 3.

Figura 4: Algumas das palavras *kimbundas* menos frequentes no *corpus*.



| N | Word | Freq. | % | Texts | % |
|----|---------|-------|------|-------|------|
| 42 | MBAMBI | 1 | 0,01 | 1 | 4,17 |
| 43 | MBANZA | 1 | 0,01 | 1 | 4,17 |
| 44 | MBELA | 1 | 0,01 | 1 | 4,17 |
| 45 | MBENGU | 1 | 0,01 | 1 | 4,17 |
| 46 | MBULU | 1 | 0,01 | 1 | 4,17 |
| 47 | MPUTU | 1 | 0,01 | 1 | 4,17 |
| 48 | MUKAPA | 1 | 0,01 | 1 | 4,17 |
| 49 | MULOJI | 1 | 0,01 | 1 | 4,17 |
| 50 | MULUMBA | 1 | 0,01 | 1 | 4,17 |
| 51 | MUNGILA | 1 | 0,01 | 1 | 4,17 |

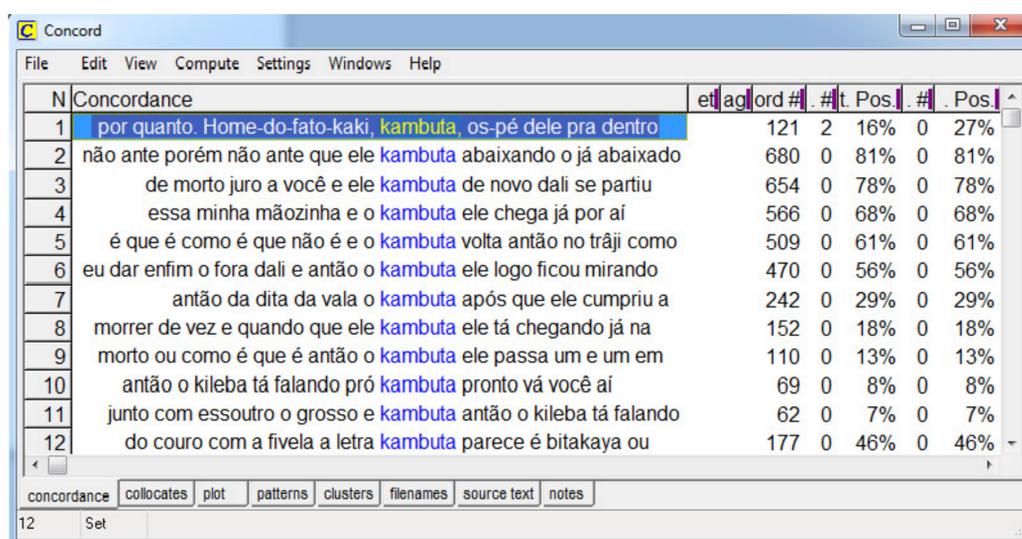
Fonte: Programa WST.

As palavras *mbanza* (aldeia), *mbengu* (rio), *mbela* (magreza), apresentadas na Figura 4, são exemplos de palavras do léxico *kimbundu* com menor grau de ocorrência no *corpus*, uma vez que os resultados apontam para uma baixa frequência tanto em relação à ocorrência das palavras em si, quanto em relação aos textos em que ocorrem. Os contextos gramaticais são uma constante no uso do léxico do *Kimbundu*, por isso a seguir dá-se uma atenção especial a este tema.

4.2 Contextos gramaticais de uso do léxico do *kimbundu*

Para entendermos as razões semânticas que justificam o uso dessas palavras no *corpus*, é preciso evidenciar alguns exemplos obtidos por meio das linhas de concordâncias. A Figura 5, a seguir, apresenta as linhas de concordâncias da palavra *kambuta*.

Figura 5: Linhas de concordância para a palavra *kambuta*.

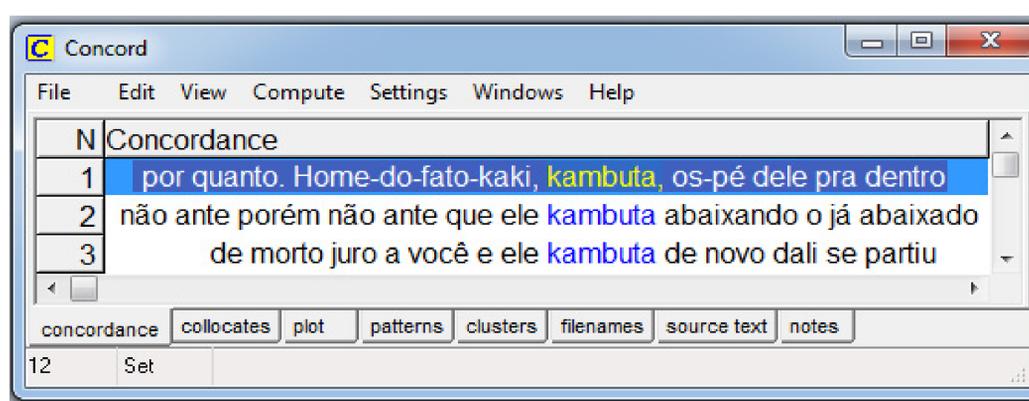


| N | Concordance | et | ag | ord # | # | t. Pos. | # | . Pos. |
|----|--|-----|----|-------|---|---------|---|--------|
| 1 | por quanto. Home-do-fato-kaki, kambuta , os-pé dele pra dentro | 121 | 2 | 16% | 0 | 27% | | |
| 2 | não ante porém não ante que ele kambuta abaixando o já abaixado | 680 | 0 | 81% | 0 | 81% | | |
| 3 | de morto juro a você e ele kambuta de novo dali se partiu | 654 | 0 | 78% | 0 | 78% | | |
| 4 | essa minha mãozinha e o kambuta ele chega já por aí | 566 | 0 | 68% | 0 | 68% | | |
| 5 | é que é como é que não é e o kambuta volta antão no trãji como | 509 | 0 | 61% | 0 | 61% | | |
| 6 | eu dar enfim o fora dali e antão o kambuta ele logo ficou mirando | 470 | 0 | 56% | 0 | 56% | | |
| 7 | antão da dita da vala o kambuta após que ele cumpriu a | 242 | 0 | 29% | 0 | 29% | | |
| 8 | morrer de vez e quando que ele kambuta ele tá chegando já na | 152 | 0 | 18% | 0 | 18% | | |
| 9 | morto ou como é que é antão o kambuta ele passa um e um em | 110 | 0 | 13% | 0 | 13% | | |
| 10 | antão o kileba tá falando pró kambuta pronto vá você aí | 69 | 0 | 8% | 0 | 8% | | |
| 11 | junto com essoutro o grosso e kambuta antão o kileba tá falando | 62 | 0 | 7% | 0 | 7% | | |
| 12 | do couro com a fivela a letra kambuta parece é bitakaya ou | 177 | 0 | 46% | 0 | 46% | | |

Fonte: Programa WST.

O curioso nessa análise é que, no caso da palavra cujas linhas de concordância foram apresentadas na Figura 5, os resultados indicam para um caminho em que a palavra mencionada é alvo do uso, cuja classificação gramatical não é a mesma ao longo das doze linhas de concordância. Isso quer dizer que, pela leitura dos resultados, a palavra *kambuta*, por significar “baixa estatura” ou “anão”, indica uma relação semântica de qualificador, adjetivo. Mas, ao aprofundar a análise, é evidente que, por vezes, funciona também como nome próprio, substantivo, e, portanto, assume um papel sintático de sujeito. As linhas de concordância dessa palavra, apresentadas na Figura 6, comprovam o uso dessa palavra com teor qualificativo, como podemos verificar a seguir:

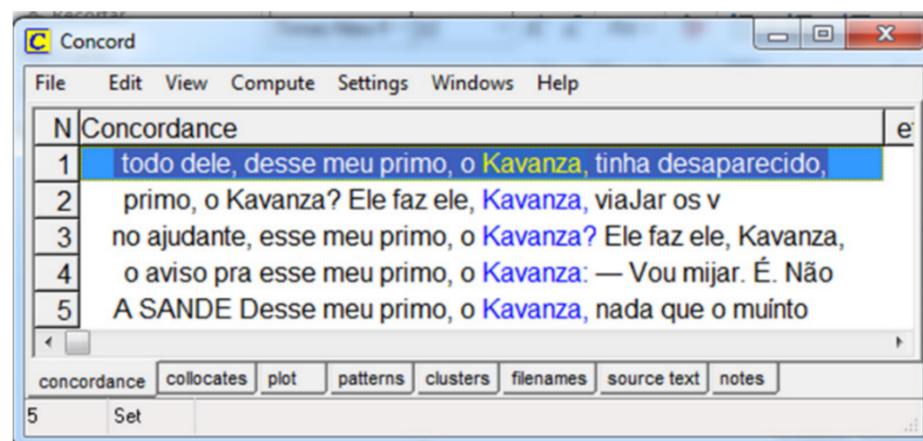
Figura 6: Linhas de concordância da palavra *kambuta* como qualificador, adjetivo.



Fonte: Programa WST.

Outro exemplo dessa natureza é o que diz respeito à palavra *kavanza*. A seguir, na Figura 7, é possível constatar a frequência desta palavra como qualificador, não descartando seu uso, também, como substantivo.

Figura 7: Linhas de concordância da palavra *kavanza* como qualificador, adjetivo.



Fonte: Programa WST.

Na Figura 7, temos um exemplo de uma palavra cuja relação simbólica é de qualificador, isto é, *kavanza* é o qualificar de primo, o que equivale ao “primo confuso, desordenado”. Dessa forma, é pertinente dizer que a qualificação é, para esse exemplo, de teor negativo, uma vez que uma pessoa confusa ou em desordem não expressa positividade. Tendo em atenção o que se desenvolveu ao longo deste trabalho, passa-se às considerações finais.

5 Palavras finais

Neste capítulo, apresentamos um estudo exploratório da obra *Os contos de Ukamba Kimba* (VILANOVA, 2013), que objetivou realizar um levantamento quali-quantitativo de palavras do *kimbundu* no português oralizado em Angola, representado na obra em questão. Os resultados apontam para a presença marcada do léxico do *kimbundu*, atestado pela frequência dessas palavras ao longo do *corpus*. Isso prova que o léxico do português angolano é constituído, também, por palavras do *kimbundu*.

A análise revelou, ainda, que não se pode negar o fato de que o falante angolano faz uso de um léxico misto, uma vez que é nítida, nos exemplos, a presença de palavras respectivas aos dois léxicos: português e *kimbundu*. Outra descoberta interessante é a questão da resignificação de palavras do *kimbundu* no léxico do português angolano, isso porque ora elas assumem o papel de qualificador, ora figuram como substantivo. O primeiro caso é o mais frequente em nosso *corpus* de estudo.

À guisa de conclusão, importa dizer que a LC é profícua para estudos desse tipo, visto que proporciona ao pesquisador princípios e métodos indispensáveis ao tratamento adequado de um *corpus*, facilitando, assim, a interpretação dos dados.

| Referências

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

BIDERMAN, M. T. C. As ciências do léxico. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. 2. ed. Campo Grande: EDUFMS, 2001. p. 13-22

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. Disponível em: <http://www.seer.ufu.br/index.php/letraseletras/article/view/28516/15799>. Acesso em: 16 maio 2020.

PERINI, M. A. **A língua do Brasil amanhã e outros mistérios**. São Paulo: Parábola, 2004.

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

SAUSSURE, F. **Curso de lingüística geral**. 27. ed. Tradução Antônio Chelini, José Paulo Paes e Izidoro Blikstein. São Paulo: Cultrix, 2006.

SCOTT, M. **WordSmith Tools version 4**. Oxford: Oxford University Press, 2004. Disponível em: <https://lexically.net/wordsmith/downloads/>. Acesso em: 19 maio 2020.

TIMBANE, A. A. A variação linguística e o ensino do português em Moçambique. **Confluências**, Rio de Janeiro, n. 33, p. 261-284, 2013. Disponível em: <http://lp.bibliopolis.info/confluencia/pdf/682.pdf>. Acesso em: 16 maio 2020.

VILANOVA, J.-M. **Os contos de Ukamba Kimba**. Vila Nova de Cerveira: Nós somos, 2013.

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 UFMG, com Pós-doutorado pela UFRGS (2020), 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).



Joel Victor Reis Lisboa é doutorando em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), Mestre em Estudos Linguísticos pela mesma universidade 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). Possui experiência docente em Inglês como Língua Estrangeira e Português como Língua de Acolhimento.



Yohanna Tamal Hernández Consoro possui graduação em Educação Menção Inglês pela Universidade do Caribe (UNICARIBE) (2014), mestrado em Educação e Novas Tecnologias pela Universidade a Distância de Madrid (UDIMA) e pelo Centro de Estudos Financeiros (CEF) em 2016. Atualmente, é doutoranda em Educação no Programa de Pós-Graduação em Educação (PPGED/UFU/CAPES). Bolsista pela Organização de Estados Americanos (OEA) e do Grupo Coimbra de Universidades Brasileiras (CGUB) pelo Programa de Alianças para a Educação e Capacitação (PAEC-2018). É membro do Grupo de Pesquisa em Educação Especial e Inclusão Educacional (GEPEPES) e participa do Grupo de Pesquisa em Análise do Discurso Crítica e Linguística Sistêmico-Funcional ADC&LSF (UFU/CNPq). Possui experiência como docente no curso de Graduação em Educação nas disciplinas de tecnologias aplicadas à educação no Ensino Superior na Universidade Domínico-Americana (UNICDA), assim como também tem atuado como docente no Nível Médio e facilitadora da Formação Técnico-Profissional na República Dominicana.



Bianca Mara Guedes de Souza é mestranda e bolsista CAPES no Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia (UFU). Bacharela em Comunicação Social com habilitação em Jornalismo pela UFU (2017). É membro fundador e agora colaboradora na Conexões – Agência de Notícias de Políticas, Ciências e Educação e no Luminar – Observatório de Mídia e Políticas Públicas. Durante a graduação foi bolsista do programa Jovens Talentos para a Ciência 2014 e bolsista da FAPEMIG no projeto de pesquisa “Políticas Sociais em Pauta - Mapeamento, monitoramento da cobertura e criação de Agência de Notícias para a mídia de Uberlândia/MG”. Membro do Grupo de Pesquisas e estudos em Análise de Discurso Crítica e Linguística Sistêmico Funcional (UFU/CNPq).



Conceição Maria Alves de Araújo Guisardi é graduada em Letras Português e Inglês pela Universidade Estadual de Goiás (UEG) e em Pedagogia pelo Instituto de Ensino Superior de Goiás. Especialista em Terapia familiar, pela Universidade Cândido Mendes (UCM). Mestre em Letras, pela Universidade Federal de Uberlândia (UFU). Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia, sob orientação da Dra. Maria Aparecida Resende Ottoni e do Dr. Teun van Dijk. Professora de Língua Portuguesa e de Língua Inglesa da Secretaria de Educação do Distrito Federal. Atuou em turmas do ensino superior, em cursos de Letras, Pedagogia, Administração e Jornalismo. Possui experiência com ensino superior, médio e ensino fundamental, aceleração da aprendizagem, educação integral, coordenação e supervisão pedagógica. Participou como delegada em GT para reformulação do Currículo da Educação Básica do Distrito Federal. Membro do Grupo de Pesquisas e Estudos em Análise de Discurso Crítica e Linguística Sistêmico Funcional (UFU/CNPq).



Laura Alejandra Guerrero Calderón é mestranda do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia (UFU) e bolsista CAPES. Sua pesquisa visa a entender como o fenômeno de assassinatos de líderes sociais na Colômbia se vê representado nas práticas discursivas das diferentes mídias digitais desse país. É membro do Grupo de Pesquisa e Estudo em Análise de Discurso Crítica e Linguística Sistêmico-Funcional e atualmente faz parte do projeto “Discursos; Identidades e Letramento: um olhar para diferentes práticas sociais”, liderado pelas professoras Dra. Maria Aparecida Resende Ottoni e Dra. Maria Cecília de Lima.



Layane Campos Soares é doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), mestra em Educação pela Universidade Federal dos Vales do Jequitinhonha e Mucuri (2017), graduada no curso de licenciatura em Letras Português/Inglês pela Universidade Federal dos Vales do Jequitinhonha e Mucuri (2015). É pesquisadora nas áreas de Análise de Discurso Crítica, Formação Docente e Ensino de Língua Portuguesa. É membro do Grupo de Pesquisa e Estudo em Análise de Discurso Crítica e Linguística Sistêmico-Funcional (UFU/CNPq).



Luana Duarte Silva é doutoranda pelo Programa de Pós-Graduação em Estudos da Linguagem, da Universidade Federal de Catalão (UFCAT), sob a orientação da Profa. Dra. Vanessa Regina Duarte Xavier; Mestre em Estudos da Linguagem pelo Programa de Pós-Graduação em Estudos da Linguagem da Universidade Federal de Goiás/Regional Catalão (UFG-RC), sob a orientação da professora Dra. Maria Helena de Paula, e beneficiada pela bolsa FAPEG; Especialista em Letras – Leitura e Ensino, pela Universidade Federal de Goiás/Câmpus Catalão, concluída em 2010; e Graduada em Letras Português pela mesma instituição (2007). Atualmente, é pesquisadora do Grupo de Estudos e Pesquisas em História do Português (GEPHPOR/CNPq) da Universidade Federal de Catalão e secretária executiva da mesma Instituição, atuando na Diretoria de Pós-Graduação. Tem experiência na área de Letras, atuando principalmente nos seguintes temas: cultura, léxico e história.



Maria del Rosario Mestanza é graduada em Letras Espanhol (2019) pela Universidade Federal de Uberlândia (UFU) e em Psicologia pela Universidade Inca Garcilaso de La Vega, (UIGV) Lima – Peru (1986), e possui especialização em Psicopedagogia (FACED/UFU). Atualmente, é mestranda em Estudos Linguísticos no Programa de Pós-Graduação em Estudos Linguísticos da UFU (PPGEL/UFU), sob a orientação do Prof. Dr. Ariel Novodvorski. Seus interesses de pesquisa incluem: Fraseologia, Estudos da Tradução e Linguística de *Corpus*. É membro do Grupo de Pesquisa em Estudos Contrastivos – GECon (UFU/CNPq). Tem experiência de mais de dez anos no ensino de espanhol no Ensino Médio, com ênfase na preparação para o ENEM, e é avaliadora para obtenção do diploma DELE do Instituto Cervantes.



Kássia Mariano de Souza é professora na Universidade Federal de Catalão (UFCAT), doutoranda no Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia (UFU), sob orientação do Prof. Dr. Ariel Novodvorski, Mestra em Estudos da Linguagem (2018) pelo Programa de Pós-Graduação em Estudos da Linguagem da Universidade Federal de Goiás – Regional Catalão (UFG-RC). Especialista em Libras pela Universidade Cândido Mendes (UCA) (2017). Graduada em Letras Português e Inglês pela UFG-

RC (2016). Aprovada pelo Ministério da Educação (MEC) no Exame Nacional de Certificação em Proficiência no Ensino de Libras (2015). É coordenadora do projeto de pesquisa “Língua Brasileira de Sinais: estudos do léxico, políticas linguísticas e aspectos socioculturais” (UFCAT) e pesquisadora dos Grupos de Estudos e Pesquisas em História do Português (GEPHPOR) e Grupo de Estudos Contrastivos (GECon). Desenvolve pesquisas linguísticas em Libras com ênfase na Lexicologia, Terminologia e Terminografia e Toponímia da Língua Brasileira de Sinais.



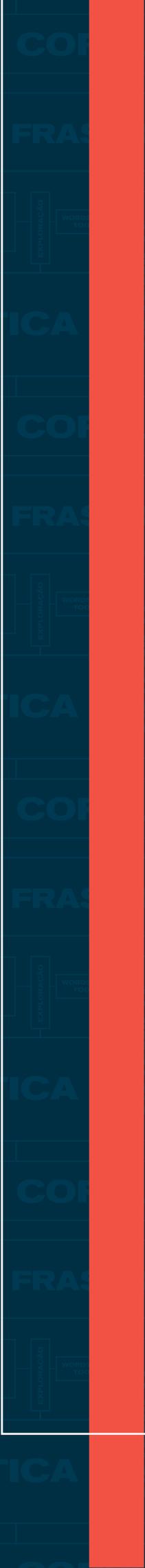
Ivonete da Silva Santos é graduada com dupla titulação em Letras Português pela Universidade Federal de Goiás-Regional Catalão (UFG-RC)/Universidade de Coimbra. Mestra em Estudos da Linguagem pela Universidade Federal de Goiás-Regional Catalão e doutoranda em Estudos da Linguagem pela Universidade Federal de Catalão (UFCAT), sob orientação da professora Dra. Maria Helena de Paula. Bolsista Capes. Pesquisadora na área de Linguística, Língua Portuguesa, Variedades do português, português na lusofonia, Políticas Linguísticas.

Professora efetiva de Língua Portuguesa na Prefeitura Municipal de Uberlândia, séries finais do Ensino Fundamental.

Publique seu e-book com a gente!

Letraria 





Letraria 