

Domínios de Lingu@gem

Revista do Instituto de Letras e Linguística
Universidade Federal de Uberlândia

Análise do Discurso Discurso
Fraseologia Funcionalismo
Gramática Gênero Leitura
Lexicografia Linguagem
Linguística Linguística
Aplicada Língua Histórica
Linguística Corpus Léxico
Língua Inglesa Sentença Sociolinguística Sujeito
Terminologia Tradução

Descrição de gêneros textuais/discursivos
com apoio computacional

Organização: Maria José Bocorny Finatto

EDUFU

V. 11, n. 2
Abril/Junho 2017
ISSN: 1980-5799



Domínios de Lingu@gem

Descrição de gêneros
textuais/discursivos com apoio
computacional

Organização: Maria José Bocorny
Finatto (UFRGS)

2º Trimestre 2017
Volume 11, número 2
ISSN: 1980-5799

Expediente

Universidade Federal de Uberlândia

Reitor

Prof. Valder Steffen Jr.

Vice-Reitor

Prof. Orlando Cesar Mantese

Diretor da EDUFU

Prof. Guilherme Fromm

Diretor do Instituto de Letras e Linguística

Prof. Ariel Novodvorski

EDUFU – Editora e Livraria da Universidade Federal de Uberlândia
Av. João Naves de Ávila, 2121 - Bloco 1S - Térreo - Campus Santa Mônica - CEP: 38.408-
144 - Uberlândia - MG
Telefax: (34) 3239-4293
E-mail: vendas@edufu.ufu.br | www.edufu.ufu.br

Editoração: Guilherme Fromm
Diagramação: Guilherme Fromm

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

Domínios de Lingu@gem, v. 11, n. 2, 2017, Uberlândia, Universidade
Federal de Uberlândia, Instituto de Letras e Linguística, 2007-

Trimestral.

Modo de acesso:

<http://www.seer.ufu.br/index.php/dominiosdelinguagem>.

Editoração: Guilherme Fromm.

Organização: Maria José Bocorny Finatto

ISSN: 1980-5799

1. Linguística - Periódicos. 2. Linguística aplicada - Periódicos.
I. Universidade Federal de Uberlândia. Instituto de Letras e Linguística.

CDU: 801(05)

*Todos os artigos desta revista são de inteira responsabilidade de seus autores, não cabendo qualquer
responsabilidade legal sobre seu conteúdo à Revista, ao Instituto de Letras e Linguística ou à Edufu.*

Diretor

Guilherme Fromm (UFU)

Conselho Editorial

Ariel Novodvorski (UFU)

Betina Ribeiro Rodrigues da Cunha (UFU)

Eliana Dias (UFU)

Fabio Izaltino Laura (UFU)

Cristiane Carvalho de Paula Brito (UFU)

Marleide Dias Esqueda (UFU)

Comissão Científica

Adriana Cristina Cristianini (UFU), Aldo Luiz Bizzocchi (FMU), Alice Cunha de Freitas (UFU), Andréia Guerini (UFSC), Ataliba T. de Castilho (USP/UNICAMP), Brett Hyde (Washington University in St. Louis – Estados Unidos), Carla Nunes Vieira Tavares (UFU), Cecilia Magalhães Mollica (UFRJ), Cintia Vianna (UFU), Cirineu Cecote Stein (UFPB), Claudia Zavaglia (UNESP/SJ Rio Preto), Cláudio Márcio do Carmo (UFOP), Cleci Regina Bevilacqua (UFRGS), Clecio dos Santos Bunzen (UNIFESP), Cristiane Brito (UFU), Deise Prina Dutra (UFMG), Dilys Karen Rees (UFG), Eduardo Batista da Silva (UEG), Elisa Battisti (UFRGS), Elisete Carvalho Mesquita (UFU), Ernesto Sérgio Bertoldo (UFU), Fabiana Vanessa Gonzalis (UFU), Fernanda Costa Ribas (UFU), Filomena Capucho (Universidade Católica Portuguesa – CECC - Portugal), Francine de Assis Silveira (UFU), Francis Henrik Aubert (USP), Gabriel Antunes Araujo (USP), Gabriel de Avila Othero (UFRGS), Giacomo Figueredo (UFOP), Hardarik Bluehdorn (Institut für Deutsche Sprache Mannheim – Alemanha), Heliana Mello (UFMG), Heloisa Mara Mendes (UFU), Igor Antônio Lourenço da Silva (UFU), Irenilde Pereira dos Santos (USP), Jacqueline de Fatima dos Santos Morais (UERJ), Janice Helena Chaves Marinho (UFMG), João Bôscio Cabral dos Santos (UFU), Jose Luiz Fiorin (USP), José Ribamar Lopes Batista Júnior (CAF/UFPI), José Sueli de Magalhães (UFU), Karylleila Santos Andrade (UFT), Krzysztof Migdalski (University of Wroclaw – Polônia), Luiz Carlos Travaglia (UFU), Liliane Santos (Université Charles-de-Gaulle - Lille 3 – França), Manoel Mourivaldo Santiago-Almeida (USP), Marcelo Módolo (USP), Márcia Mendonça (UNICAMP), Maria Angélica Furtado da Cunha (UFRN), Maria Aparecida Resende Ottoni (UFU), Maria Cecília de Lima (UFU), Maria Célia Lima-Hernandes (USP), Maria de Fátima Fonseca Guilherme (UFU), Maria do Perpétuo Socorro Cardoso da Silva (UEPA), Maria Helena de Paula (UFG), Maria José Bocorny Finatto (UFRGS), Maria Luisa Ortiz Alvarez (UnB), Maria Luiza Braga (UFRJ), Maria Suzana Moreira do Carmo (UFU), Marlúcia Maria Alves (UFU), Maurício Viana Araújo (UFU), Michael J. Ferreira (Georgetown University – Estados Unidos), Montserrat Souto (Universidade Santiago de Compostela – Espanha), Nilza Barrozo Dias (UFF), Patricia de Jesus Carvalhinhos (USP), Paulo Osório (Universidade da Beira Interior – Portugal), Paulo Rogério Stella (UFAL), Pedro Malard Monteiro (UFU), Pedro Perini-Santos (PUC-Minas), Raquel Meister Ko. Freitag (UFS), Roberta Rego Rodrigues (CLC/UFPEl), Rolf Kemmler (Universidade de Trás-os-Montes e Alto Douro – Portugal), Sebastião Carlos Leite Gonçalves (UNESP/S.J. Rio Preto), Silvana Maria de Jesus, (UFU), Silvia Melo-Pfeifer (Universidade de Hamburgo – Alemanha), Simone Floripi (UFU), Simone Tiemi Hashiguti (UFU), Sinara de Oliveira Branco (UFCG), Stella Esther Ortweiler Tagnin (USP), Teresa Maria Wlosowicz (University of Social Sciences - Polônia), Tommaso Raso (UFMG), Ubirajara Inácio Araújo (UFPR), Valeska Virgínia Soares Souza (IFTM), Vanessa Hagemeyer Burgo (UFMS), Vânia Cristina Casseb Galvão (UFG), Vera Lucia Menezes de Oliveira e Paiva (UFMG), Vitalina Maria Frosi (UCS), Walcir Cardoso (Concordia University – Canadá), Waldenor Barros Moraes Filho (UFU), Zelina Márcia Pereira Beato (UESC).

Participaram dessa edição como pareceristas *ad hoc*

Alena Ciulla (UFRGS)

Cibele Krause Lemke (UNICENTRO)

Emeli Borges Pereira Luz (UFU)

Fabíola Sartin (UFG)

Heberth Paulo Souza (IPTAN)

Lucivaldo da Silva Costa (UNIFESSPA)

Sumário

Expediente.....	278
Sumário	281
Apresentação	282
Descrição de gêneros textuais/discursivos com apoio computacional - Maria José Bocorny Finatto (UFRGS).....	282
Artigos.....	289
O gênero entrevista radiofônica em comunidades hispânicas: um aporte da Análise Textual Automática - Leandro Silveira de Araujo (UFU).....	289
Interação em rede social: das reações às características do gênero comentário - Roberlei Alves Bertucci (UTFPR), Paula Ávila Nunes (UTFPR).....	314
Um estudo sobre referência e a construção da opinião a partir de um corpus textual extraído do YouTube - Alexandre Ribeiro Afonso (UnB), Jordão Té (UFG).....	340
Gêneros textuais datasheet e artigo científico em aulas de ESP: levantamentos léxico-estatísticos para fins educacionais - Sabrina Bonqueves Fadanelli (UCS), Andrea Jessica Monzón (IFRS)	352
Português para Fins Acadêmicos sob o aporte da Linguística de Corpus e do Processamento de Linguagem Natural - Cristina Becker Lopes Perna (PUCRS), Lucelene Lopes (PUCRS), Lucas Zambrano Rollsing (PUCRS)	381
Ideologia em gêneros discursivos pela análise associada da Linguística Sistêmico-Funcional e Linguística de Corpus: cantigas galego-portuguesas - Ulisses Tadeu Vaz de Oliveira (UFMS)	396
Modelagem topológica do perfil sistêmico-funcional de manuais de instrução no espaço gramatical - Francieli Oliveira (UFMG), Thaís Campolina (UFOP), Giacomo Figueredo (UFOP).....	421

Descrição de gêneros textuais/discursivos com apoio computacional

Com muita satisfação, apresentamos o volume 11, número 2, de 2017 da nossa querida revista Domínios de Lingu@gem, que nos tem trazido importantes trabalhos das mais variadas áreas dos estudos linguísticos. Desta vez, o volume temático gira em torno da descrição de gêneros discursivos e/ou textuais com algum tipo de suporte computacional. Instaure-se, assim, desde a chamada, alguma interdisciplinaridade, seja teórica ou metodológica, entre os Estudos do Texto e do Discurso e a Informática. Ao propor este volume, imaginamos poder reunir trabalhos que trouxessem, partindo da descrição e da caracterização – ou mesmo da problematização do estatuto de gênero, diferentes tipos de aplicações. A descrição do nosso objeto linguístico, tornado concreto via texto e práticas de discurso, conforme entendemos, dá suporte à análise. Dessa análise, feita sob diferentes pontos de vista teóricos, os quais iluminam determinadas facetas do nosso objeto, poderemos partir para a explicação e, posteriormente, para as generalizações e previsibilidades. Feito esse percurso, pode-se retornar ao ponto de partida teórico e verificar o quanto ele, efetivamente, nos dá suporte.

Desde muito, o papel do processamento, via computador, da língua natural tem já se mostrado um empreendimento científico e multidisciplinar tanto relevante quanto desafiador. Se hoje testemunhamos muitos avanços para línguas como o inglês, para o português, o avanço é ainda desigual, embora reconhecido como necessidade. Para se ter uma ideia, em 1999, já líamos em uma notícia do jornal português *Público* a chamada para a seguinte notícia (<https://www.publico.pt/ciencias/jornal/a-lingua-de-camoes-na-informatica-132347>): “A língua de Camões na Informática”. Nessa notícia, entrevistava-se a nossa colega pesquisadora Diana Santos, e assinalava-se “A adaptação dos computadores à língua portuguesa é considerada pelo Ministério da Ciência e da Tecnologia uma área prioritária”.

Anos mais tarde, em 2007, em livro organizado por Diana Santos (SANTOS, 2007), encontramos um exemplo do avanço disso no trabalho de Quaresma e Rodrigues (2007), ambos cientistas da Computação. Nele vemos uma pesquisa dedicada ao tema da Recuperação da Informação que reúne e explora uma base de texto 7 mil (!) pareceres da Procuradoria-

Geral da República Portuguesa, produzidos tais documentos entre 1940, compondo um universo de cerca de dez milhões de palavras. A proposta dos autores era a de um sistema interativo para recuperação de informações nesse universo de documentos. Para tanto, desde a unidade palavra até a unidade “tipo de texto” precisou ser enfrentada, e bem sabemos o quanto a linguagem jurídica é complexa.

Atendendo à proposta desses tipos de estudos, estão aqui artigos que nos trazem um rico panorama: da descrição da linguagem – falada ou escrita - para fins diversos – como a arquitetura de sistemas computacionais que lidam com textos e conseguem, de algum modo, representar seu conteúdo – à obtenção de recursos para o ensino/aprendizagem de línguas. A intersecção, entre os Estudos da Linguagem, o Processamento da Linguagem Natural e a Linguística de *Corpus*, de certo modo esperável nesse cenário, destaca mais pontos de diálogo possíveis, instigantes e produtivos, como o aporte da Linguística Sistêmico-Funcional.

Enfim, muitos linguistas e cientistas da Computação – entre outros cientistas – têm lidado com gêneros textuais e discursivos sob diferentes enfoques. Não raro, encontramos belos trabalhos sobre esse tema em eventos de áreas de conhecimento cujo diálogo com a área de Letras não tem sido muito frequente: não é difícil imaginar, por exemplo, em um evento de Farmacologia, vários estudiosos de Ciências da Saúde debruçando-se, entre outras coisas, sobre os problemas conceituais, linguísticos e terminológicos das bulas de medicamentos. Em paralelo, sabemos, se precisar supor, que esse mesmo objeto, a bula, já rendeu vários estudos na área de Letras, de artigos a teses de doutorado. Entendemos, assim, que o tópico gênero – seja textual ou discursivo – é um potente agregador, uma ponte que pode facilitar nosso encontro, cada um com a sua ciência em relação com o outro e com a sua ciência. Além disso, essa noção de gênero parece que precisa sempre ser retomada entre nós que pesquisamos sobre a linguagem em funcionamento, a linguagem que instaura diálogos e cumpre funções.

Sucedendo à publicação deste volume, vale já anunciar aos interessados no tema aqui tratado o XIV Encontro de Linguística de *Corpus* (ELC) e a IX Escola Brasileira de Linguística Computacional (EBRALC). São dois eventos associados, que devem ocorrer em agosto de 2017. Na edição de 2017, O ELC, justamente, visa provocar reflexões sobre as contribuições da Linguística de *Corpus* para os estudos sobre gêneros textuais e discursivos. E a EBRALC quer mostrar como podemos nos aproximar das tecnologias computacionais para o estudo da linguagem. Fica, portanto, reforçada a temática dos gêneros como algo

importante, sem esquecer de mencionar aqui o próprio SIGET - IX Simpósio Internacional de Estudos dos Gêneros Textuais, maior evento da área, que ocorre também em agosto de 2017.

Feita essa breve introdução passamos agora a comentar um pouco sobre cada um dos trabalhos que perfazem este volume, esperando que esse comentário possa contribuir para a entusiasmar a leitura de cada um.

Inicia o volume o artigo intitulado “**O gênero entrevista radiofônica em comunidades hispânicas: um aporte da análise textual automática**” da autoria de Leandro Silveira de Araujo. Nesse trabalho, é descrito o gênero *entrevista radiofônica* em comunidades de língua espanhola. Para tanto, recorre-se a uma retomada dos principais aspectos envolvidos na caracterização dos gêneros discursivos conforme a abordagem sociointeracionista. O trabalho nos apresenta o *Tropes 7.23*, um *software* de processamento automático de textos, útil para a descrição dos âmbitos temático, estilístico, estrutural e funcional dos enunciados que compõem a *entrevista radiofônica*. Foram estudadas entrevistas produzidas em Madri, Buenos Aires e San Miguel de Tucumán – cidade da Argentina – difundidas por rádio e pela *web*. Esse trabalho nos traz ótimas ideias para novos estudos que se ocupem de questões relacionadas à hispanofonia, à variação diatópica e apresenta-nos um suporte metodológico muito atraente. A ferramenta computacional, a serviço da descrição desse gênero, conforme vemos no trabalho, permite perceber algumas nuances que talvez não fossem tão evidentes sem esse aporte tecnológico.

O segundo artigo é “**Interação em rede social: das reações às características do gênero comentário**” de Roberlei Alves Bertucci e Paula Ávila Nunes. O objetivo do trabalho é analisar a relação entre o gênero comentário em uma página no *Facebook* e reações de usuários nesse contexto. Nada mais atual, considerando o papel das redes sociais no cenário atual. Com o aplicativo *Netvizz*, os autores examinam o “odiar” manifestado pelos comentaristas. Conforme os autores mesmos apontam, estudar a linguagem em contextos da chamada era digital “é imperativo em nossa constante tentativa, como linguistas, de entender o modo pelo qual a linguagem se coloca como fundamento antropológico principal, sobretudo em sua relação com a tecnologia”. Outro ponto a destacar neste artigo é a verificação das condições e características de um gênero textual emergente. Esse tipo de objeto, o teor de comentário de usuários em redes sociais, bem sabemos, tem rendido toda uma série de pesquisas, do Marketing à prospecção de tendências políticas. Em Processamento de Linguagem Natural, esse campo tem sido denominado de *Análise de Sentimentos* e reúne

várias áreas, como Linguística, Psicologia e Computação. Para os linguistas interessados nessa inter- e multidisciplinaridade, sugerimos consultar a tese recente de Silva (2016), em Ciência da Computação, cujo título é, justamente, *Análise de sentimentos em textos curtos provenientes de redes sociais*. Veja a referência ao final deste texto.

O terceiro artigo, que também trata do tópico *Análise de Sentimentos*, é “**Um estudo sobre referenciação e a construção da opinião a partir de um corpus textual extraído do YouTube**”, de Alexandre Ribeiro e Afonso Jordão Té. Os dois autores não são linguistas; são formados e atuantes nas áreas da Ciência/Gestão da Informação e Computação e empreendem um belo trabalho de diálogo com os estudos da Linguagem, especialmente com a Linguística do Texto. Esse trabalho analisa um *corpus* de opiniões extraído das postagens do *YouTube*. Os autores verificam como o fenômeno discursivo da referenciação ocorre em tais postagens, isto é, buscam compreender como um dado objeto de interesse em discussão na mídia é renomeado. Foram examinadas as postagens opinativas favoráveis ao *impeachment* da então presidente do Brasil, Dilma Rousseff, e as postagens contrárias ao processo. Os dados foram gerados de três vídeos que relatam ou comentam sobre o tema. A ferramenta computacional para tratamento do texto foi o nosso conhecido *software* AntConc, recurso de acesso gratuito bastante utilizado por quem lida com Linguística de *Corpus* no Brasil.

O quarto trabalho do volume, “**Gêneros textuais *datasheet* e artigos científico em aulas de ESP: levantamentos léxico-estatísticos para fins educacionais**” de Sabrina Bonqueves Fadanelli e Andrea Jessica Borges Monzón, ambas professoras de inglês em cursos tecnológicos no Sul do Brasil, objetivou extrair dados relevantes, de um *corpus* textual composto por textos técnicos e científicos, para a preparação de material para o ensino de terminologia técnica e científica em aulas de ESP – *English for Specific Purposes*. As autoras examinaram artigos acadêmicos da área de Computação e um gênero textual bem específico da área de Eletrotécnica: os *datasheets*. As ferramentas computacionais empregadas foram o já antes citado AntConc e o extrator automático de candidatos a terminologias, também de acesso gratuito, o *TermoStat*. Interessaram às autoras especialmente as colocações mais frequentes em torno das terminologias empregadas nos textos.

O quinto artigo é “**Português para Fins Acadêmicos sob o Aporte da Linguística de Corpus e do Processamento de Linguagem Natural**”, que reúne, como autores Cristina Becker Lopes Perna, professora e pesquisadora da área de inglês e Linguística, Lucelene Lopes, graduada em Matemática e pesquisadora de Ciência da Computação/PLN, e Lucas

Zambrano Rollsing também da área de Letras/Linguística. O artigo traz resultados iniciais de estudo sendo desenvolvido entre dois programas de Pós-Graduação da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), o Programa de Pós-Graduação em Letras (PPGL), na sua área de concentração em Linguística, através do Grupo de Pesquisa sobre o Uso e Processamento de Língua Adicional (UPLA), e o Programa de Pós-Graduação em Ciência da Computação (PPGCC), através do Grupo de Processamento em Linguagem Natural. O tópico de interesse do trabalho conjunto é o Português para Fins Acadêmicos, representando por um *corpus* de 200 dissertações e teses sobre Linguística produzidas na PUCRS. Com apoio da ferramenta EXATO, esse acervo de pesquisas é tratado como *corpus* e são reconhecidos, automaticamente, os principais temas dos trabalhos, nas diferentes subáreas de Linguística abrigada pelo PPGL. Com o estudo, demonstra-se como se poderia, por exemplo, representar os principais temas de pesquisa ao longo de todo um quadriênio de um programa de pós-graduação. Em tempos de avaliação e prospecção da produtividade de nossos cursos, esse sistema poderia ser muito útil para gestores e para os próprios pesquisadores-produtores dos trabalhos.

O penúltimo artigo, **“Ideologia em gêneros discursivos pela análise associada da Linguística Sistêmico-Funcional e Linguística de *Corpus*: cantigas galego-portuguesas”** de Ulisses de Oliveira, logo de início, situa a recorrente discordância, entre linguistas, na definição de gêneros do discurso. Razão pela qual esse volume emprega, justamente, gêneros discursivos/textuais. Além disso, o artigo propõe tratar do tópico ideologia, uma dimensão contextual superior, que considera ser geralmente desprestigiada nas análises de gênero, dado seu caráter amplo e multinível, de complexa manipulação para o pesquisador. O trabalho traz uma proposta de análise cruzada de marcadores ideológicos postos nos textos, com aporte da Linguística Sistêmico-Funcional e da Linguística de *Corpus*. Enfim, temos um belo exemplo de enfrentamento de um tema bastante complexo, a Ideologia.

O último trabalho do volume intitula-se **“Modelagem topológica do perfil sistêmico-funcional de manuais de instrução no espaço gramatical”**. Esse trabalho, de autoria de Francieli Silvéria Oliveira, Giacomo Patrocínio Figueredo e Thaís Guerra Campolina, todos egressos da área de Letras/Linguística, investiga o modo pelo qual os sistemas gramaticais da oração contribuem para a produção de significados do texto, apoiando-se também em pressupostos da Linguística Sistêmico-Funcional e da Linguística de *Corpus*. O objeto sob estudo, tomado como *corpus*, é o gênero *manual de instrução* em português brasileiro. São

examinados diferentes tipos de manuais, que abrangem desde tema como a criação de peixes até procedimentos de Enfermagem. O estudo, apoiado pelo *software UAM Corpus Tool*, demonstra, com grande detalhamento, como a análise gramatical é capaz de revelar como se organizam aspectos fundamentais dos gêneros textuais. O tratamento quantitativo dos dados, fartamente ilustrado com gráficos multidimensionais, é um rico diferencial do trabalho. Conforme os autores, é possível prever, com confiança, o comportamento da gramática oracional do português brasileiro para quaisquer textos caracterizados como manual.

Assim, conforme o nosso leitor poderá conferir, com a leitura dos artigos deste volume, as pesquisas recentes com os *corpora* envolvidos em cada um dos estudos não trouxe mera contagens de palavras em textos com a ajuda de algum *software*. O apoio computacional não é um fim, ele é um meio, e ele não precisa nos intimidar.

Os artigos deste volume, no seu conjunto, mostram e validam métodos, sim, mas também apontam e reiteram correlações entre gêneros e tipologias textuais, cenários comunicativos, elementos lexicais, gramaticais, discursivos, semânticos e temáticos, sejam feitos por linguistas, por cientistas de Computação/Informação ou por ambos, associados. Esses trabalhos também se colocam epistemologicamente, filiam-se a teorias e a diferentes ciências. Esse é o diferencial desses trabalhos, ousar em direção ao diálogo interdisciplinar e instigar novas pesquisas, cada um a seu modo. Dos exemplos citados no início desta apresentação, da notícia sobre o processamento computacional do português de 1999, à pesquisa de 2007 com textos jurídicos, até o todo representado pela pequena amostra que vemos aqui reunida neste ano de 2017, pode ser percebido um percurso evolutivo em prol do tratamento da linguagem. Assim pensando, espero que cada vez mais o diálogo entre nós linguistas e os colegas de Computação e de áreas afins possa ser fortalecido. Desejo uma boa leitura a todos e que os trabalhos deste volume possam nos inspirar a prosseguir outros tantos.

Maria José Bocorny Finatto (UFRGS/CNPq/CAPES).
Abril de 2017.

Referências

QUARESMA, P.; RODRIGUES, I. Avaliação de sistemas interactivos de recuperação de informação em bases de texto jurídicas. p. 257-265. In: SANTOS, D. (Ed.) **Avaliação conjunta. Um novo paradigma no processamento computacional da língua portuguesa**. Ist Press: Lisboa, 2007. p. 257-265.

SANTOS, D. (Ed.) **Avaliação conjunta. Um novo paradigma no processamento computacional da língua portuguesa.** Ist Press: Lisboa, 2007.

SILVA, N. F. F. da. **Análise de sentimentos em textos curtos provenientes de redes sociais** [online]. Tese de Doutorado em Ciências de Computação e Matemática Computacional. São Carlos: Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2016. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-27092016-143947/> Acesso em: 7 de abril de 2017.

O gênero entrevista radiofônica em comunidades hispânicas: um aporte da Análise Textual Automática

The radio interview in Hispanic communities: a contribution of the Automatic Textual Analysis

Leandro Silveira de Araujo*

RESUMO: Neste trabalho descrevemos o gênero *entrevista radiofônica* em comunidades de língua espanhola. Para tanto, recorreremos a uma breve revisão dos principais aspectos envolvidos na caracterização dos gêneros discursivos segundo uma abordagem sociointeracionista. Com auxílio do *software Audacity 1.3*, efetivamos a gravação e transcrição de entrevistas radiofônicas veiculadas por rádios de transmissão online, as quais nos serviram como *corpus* para a análise do funcionamento desse gênero do discurso. Finalmente, valemo-nos do *Tropes 7.2.3, software* de processamento automático de textos, para auxiliar-nos na descrição dos âmbitos temático, estilístico, estrutural e funcional dos enunciados que compõem a *entrevista radiofônica*.

PALAVRAS-CHAVE: Entrevista Radiofônica. Gêneros do discurso. Análise textual automática. Linguística de *Corpus*. Língua Espanhola.

ABSTRACT: This paper describes the radio interview in Hispanic-speech communities. Therefore, we resort to a brief review of the main aspects involved in the characterization of discourse genres according to a sociointeractionist approach. With the help of *software Audacity 1.3*, we recorded and transcribed the radio interviews transmitted by online broadcast radio, which served as the *corpus* for the analysis of the functioning of this discourse genre. Finally, we used the *Tropes 7.2.3*, a word processing software, to assist us in the description of the thematic, stylistic, structural and functional domains of the radio interviews.

KEYWORDS: Radio interview. Discourse genres. Word processing software. *Corpus* linguistics. Spanish.

“[...] la radio se apoya en las normas de prestigio, nacionales y regionales. La radio permite el acceso instantáneo a las voces y a las ideas de todos los miembros de la sociedad, como oyentes o como participantes ocasionales” (LIPSKI, 2011, p.162).

* Professor do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia. Doutorando do programa de pós-graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras (FCL/Araraquara), da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP). E-mail: araujoleandrosilveira@gmail.com.

1. Introdução

Apresentaremos, neste trabalho, algumas contribuições da análise textual automática para a descrição dos gêneros discursivos, com especial atenção a enunciados pertencentes ao gênero **entrevista radiofônica**, produzidos em Madri, Buenos Aires e San Miguel de Tucumán – cidade ao norte da Argentina – e difundidos por rádios com transmissão também via *web*. Com esse objetivo, começamos pela apresentação do gênero em questão, descrevendo, de modo geral, as categorias próprias desse constructo sociodiscursivo, isso é, sua função, estilo, conteúdo temático e composicional. Para tanto, nos apoiamos numa discussão que articula e relaciona tanto o apoio teórico sobre gêneros do discurso (BAKHTIN, 1997; MARCUSCHI, 2008), como a descrição sobre as características da entrevista radiofônica segundo os profissionais que interagem cotidianamente com esse gênero (PÉREZ COTTEN; TELLO, 2004; PORTUGAL; YUDCHAK, 2008).

Uma vez construída essa fundamentação, apresentaremos os softwares a que recorreremos para compilar (*Audacity 1.3*) e processar automaticamente (*Tropes 7.2.3*) os textos que constituem o *corpus* de análise, explicando sucintamente a contribuição deles para cumprir os objetivos estabelecidos neste estudo. Em seguida, apresentaremos o material que compõe o *corpus* de análise, descrevendo o processo de compilação e algumas características linguísticas e extralinguísticas dessa base de dados. Finalmente, apresentaremos os dados provenientes da análise das entrevistas radiofônicas coletadas, expondo, desse modo, as contribuições aportadas pelo *Tropes 7.2.3* para a descrição desse gênero do discurso, no que se relaciona a suas categorias de estilo, conteúdo temático e composicional.

2. O gênero discursivo entrevista radiofônica

A fim de proceder ao estudo da entrevista radiofônica, refazemos o percurso epistemológico da conceitualização desse gênero do discurso orientando-nos por Pérez Cotten e Tello (2004, p. 28), autores que identificam um formato ou gênero que se denomina entrevista jornalística e que conforme sua difusão em suporte papel ou por meios eletrônicos define a entrevista jornalística como escrita, televisiva ou radiofônica. Em outras palavras, verificamos nessa asserção a definição do **domínio discursivo**¹ a que pertence o gênero em

¹ Segundo explica-nos Marcuschi (2008), parece que conceito de *domínio discursivo* é equiparável à “esfera da atividade humana”, presente em Bakhtin (1997), pois segundo afirma o teórico brasileiro, “os domínios

questão, bem como o **suporte** de sua circulação social. Isto é, podemos inferir que dentro da esfera jornalística há um modo de interação comunicativa estável que se chama **entrevista**, o qual, como sabemos, define-se pelas especificidades, exigências e necessidades do jornalismo².

Os traços que aporta o suporte (**rádio**) ao “formato” (**entrevista jornalística**) revelam-nos que apesar do suporte não alterar o valor dos textos que propaga, ele pode definir o gênero desses enunciados (MARCHUSCHI, 2008). No caso do rádio, um suporte virtual convencional, a ausência da imagem (presente na televisão) ou da possibilidade de releituras (possível no jornal impresso), entre outras características que descreveremos mais adiante, diferenciam a entrevista radiofônica das entrevistas difundidas por outros suportes do mesmo domínio discursivo.

Em complemento, como adverte Marcuschi (2008), a função desempenhada pelo gênero é um constituinte que deve ser considerado quando se deseja descrevê-lo. Visando encontrá-la, verificamos novamente nas palavras de Pérez Cotten e Tello (2004) o possível propósito da entrevista radiofônica:

Sirve para la construcción discursiva de diferentes relatos periodísticos [...] y, también, para escuchar directamente la voz del entrevistado [...] en ambos casos, el objetivo inicial (y por cierto, final) es el de producir conocimiento [...] (PÉREZ COTTEN; TELLO, 2004, p. 29)³.

La entrevista, por lo tanto, es el principal recurso periodístico para acceder a la información, ampliar una noticia, obtener la voz de algún personaje (PÉREZ COTTEN; TELLO, 2004, p. 34)⁴.

discursivos produzem modelos de ação comunicativa que se estabilizam e se transmitem de geração para geração com propósitos e efeitos definidos e claros. Além disso, acarretam formas de ação, reflexão e avaliação social que determinam formatos textuais que em última instância desembocam na estabilização de gêneros textuais” (MARCUSCHI, 2008, p. 294).

² São exemplos de outros gêneros presentes no domínio discursivo jornalístico: os editoriais, as notícias, as reportagens, a nota social, os artigos de opinião, a história em quadrinhos, a crônica policial, a crônica esportiva, os anúncios, os classificados, as cartas do/ao leitor, as entrevistas televisivas e radiofônicas, as notícias de TV/rádio; as reportagens ao vivo, as discussões, debates, etc. (MARCUSCHI, 2008, p. 195).

³<Tradução nossa> “Serve para a construção discursiva de diferentes relatos jornalísticos [...] e, também, para escutar diretamente a voz do entrevistado [...] em ambos os casos, o objetivo inicial (e certamente, final) é o de produzir conhecimento [...]”.

⁴<Tradução nossa> “A entrevista, portanto, é o principal recurso jornalístico para aceder à informação, ampliar uma notícia, obter a voz de algum personagem”.

A estruturação do gênero relaciona-se, portanto, à reconstrução de um evento por meio do discurso, à apreensão e à avaliação de opiniões envolvidas com acontecimentos das mais diversas esferas da sociedade. Ou seja, a entrevista radiofônica se organiza em função do “informar”, fazendo da informação, “notícia”⁵.

O interessante, no entanto, é que, se nos atemos apenas a esse traço, não avançamos muito no que corresponde ao conhecimento das características próprias apenas a esse gênero, isso porque veicular informações pertence à essência da esfera jornalística e, portanto, é de se esperar que verifiquemos o mesmo propósito em outros gêneros do discurso pertencentes à esfera jornalística.

Por isso a diferenciação da *entrevista radiofônica* vai se afirmando, de fato, à medida que se identificam novas características do gênero. Assim, como lemos nos dois fragmentos acima, o aporte de novas informações se dá, nesse gênero, pela recepção da voz direta do entrevistado (“voz de algum personagem”). **Voz** que, por sua vez, responde outra, geralmente indagadora. A atitude responsiva verificada nesse embate resgata um caráter dialógico, que na entrevista é visto como:

[...] un encuentro de absoluta formalidad⁶ donde los roles están bien definidos. Hay un actor que propicia el desarrollo del conocimiento de un tema a través de otro actor. Pero este actor conoce algo sobre el tema, tiene algunas ideas. Éste, a través de preguntas facilita que se produzca conocimiento nuevo sobre determinado tema (PÉREZ COTTEN; TELLO, 2004, p. 25)⁷.

Em outros termos, o diálogo, além de possuir uma forma bem definida e conhecida por seus participantes, é instaurado por dois atores, cujos papéis estão bem estabelecidos: um que se acredita dominar o conteúdo que está na pauta da entrevista e outro que, a partir do

⁵ Pérez Cotten e Tello (2004, p. 29) diferenciam “informação” de “notícia”. A primeira seria somente o relato de um evento ocorrido ou que está por acontecer, por sua vez, a *notícia* é a informação que se tornou difundida pelo interesse despertado em seu espectador. Logo, toda notícia implica necessariamente uma informação, o contrário não é verdadeiro.

⁶ Formalidade, aqui, não é entendida como nível de monitoramento linguístico, mas como estruturação (forma) bem definida e marcada.

⁷<Tradução nossa> “[...] um encontro de absoluta formalidade, onde os papéis estão bem definidos. Há um ator que propicia o desenvolvimento do conhecimento de um tema através de outro ator. Mas esse ator conhece algo sobre o tema, tem algumas ideias. Este, através de perguntas, facilita que se produza conhecimento novo sobre determinado tema”.

conhecimento superficial que adquiriu sobre o assunto, conduz o debate, tentando facilitar, por perguntas, o descobrimento de novas informações sobre o tema.

No gênero **entrevista radiofônica**, o diálogo formalizado com fins informativos assume um caráter eminentemente “público”, pois se constrói para difundir uma informação (transformando-a em notícia) e, logo, satisfazer o anseio que o ouvinte tem por determinado assunto. Não é por acaso que se caracteriza a entrevista nesse domínio discursivo como “a mais pública das conversações privadas” (PÉREZ COTTEN; TELLO, 2004. P.24).

A cena construída pelo entrevistador e pelo entrevistado (**Eu↔Tu**) debatendo um tema (Ele) aparentemente comporia uma situação suficiente para a concretização de enunciados. No entanto, isso não é assim no diálogo da entrevista jornalística. Graças ao traço “público”, há na entrevista, imprescindivelmente, a exigência de mais um personagem, o telespectador/ouvinte, quem motiva e, por fim, recebe toda a mensagem.

A fim de melhor avaliar o diálogo na entrevista radiofônica e, conseqüentemente, entender como esse conceito ajuda-nos a entender o gênero, recorreremos a Bakhtin (1997), que identifica a diferença existente entre o diálogo presente no gênero de discurso primário, denominado simples, e o diálogo próprio do gênero de discurso secundário, denominado complexo:

Os gêneros secundários do discurso — o romance, o teatro, o discurso científico, o discurso ideológico, etc. — aparecem em circunstâncias de uma comunicação cultural, mais complexa e relativamente mais evoluída, principalmente escrita: artística, científica, sociopolítica. Durante o processo de sua formação, esses gêneros secundários absorvem e transmitem os gêneros primários (simples) de todas as espécies, que se constituíram em circunstâncias de uma comunicação verbal espontânea. Os gêneros primários, ao se tornarem componentes dos gêneros secundários, transformam-se dentro destes e adquirem uma característica particular: perdem sua relação imediata com a realidade existente e com a realidade dos enunciados alheios [...] (BAKHTIN, 1997, p. 281).

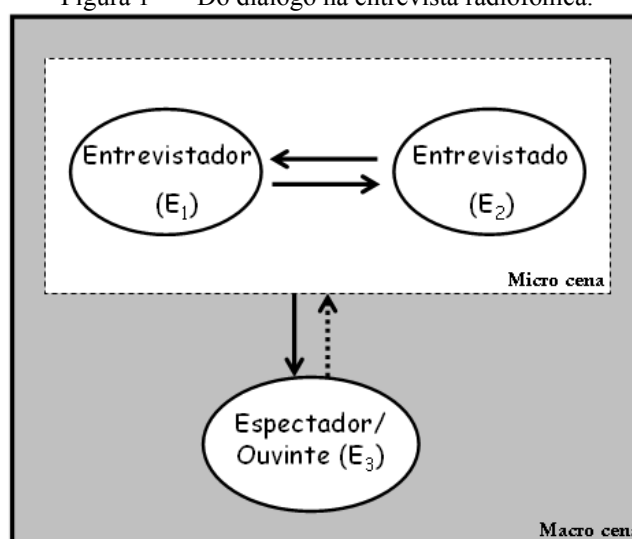
A simplicidade dos gêneros primários parece estar relacionada à espontaneidade observável na interação verbal do cotidiano, que se iguala ou se aproxima ao diálogo *stricto*. Por outro lado, se nota nos gêneros secundários uma espécie de apropriação dos gêneros primários, assimilação que é seguida por uma reelaboração que distancia os enunciados secundários do diálogo cotidiano. Esse processo de “reformulação”, por fim, culmina na constituição de gêneros mais elaborados, cuja “relação imediata com a realidade” dialógica do enunciado, presente nos gêneros primários, já não é evidenciada.

Em síntese, a categorização dos gêneros em **primários** e **secundários** leva em consideração o conceito de **diálogo**, *stricto* ou *lato*:

[...] A distinção entre gênero primário e gênero secundário [...] retoma, respectivamente, as duas maneiras de se considerar o diálogo, a que já fizemos menção: em *stricto sensu*, o diálogo cotidiano, espontâneo e, com base nele, o diálogo mais extenso e complexo que constitui todo e qualquer enunciado. (MARCHEZAN, 2006, p. 119).

Aplicando essa discussão à análise da relação dialógica instaurada na entrevista radiofônica (sintetizada pela figura 1), percebemos uma micro cena, na qual se instaura um diálogo entre entrevistador (E_1) e entrevistado (E_2) para satisfazer, unicamente, o desejo informativo que possui o espectador/ouvinte (E_3), na macro cena⁸. Por seu turno, a resposta do espectador (E_3) ao enunciado composto por E_1 e E_2 não é necessariamente imediata, posto que pode ser dada de diferentes formas: desde modos mais perceptíveis (como em comentários por chamadas telefônicas à rádio) até de maneiras menos conexas (por exemplo, com um diálogo traçado entre E_3 e outro interlocutor (E_n) com quem possa interagir, até mesmo, assincronicamente). Por isso, escolhemos representar essa interação por meio de uma flecha pontilhada.

Figura 1 — Do diálogo na entrevista radiofônica.



Fonte: elaborada pelo autor.

⁸ Não entender macro e micro, como sentidos *stricto* e *lato* do diálogo, verificados em Bakhtin (1997).

Frente ao cenário apresentado, definimos o gênero **entrevista radiofônica** como *secundário*, pois por se valer da esfera **jornalística** e do **rádio** (“circunstâncias de uma comunicação cultural mais complexa e relativamente mais evoluída”⁹ (BAKHTIN, 1997, p. 281), esse gênero apoia-se em uma estruturação do diálogo mais complexa, que, como vimos, vai além da simplicidade verificada no diálogo cotidiano (*stricto sensu*). Desse modo, constrói-se um diálogo (entre E₁ e E₂) mais próximo do cotidiano no primeiro plano (micro cena), no qual as relações são sincrônicas, mais espontâneas e onde, aparentemente, se configura a entrevista. Porém essa micro cena não tem valor por si mesma, posto que sua finalidade é informar E₃ (macro cena) e, portanto, estabelecer uma relação dialógica (*lato sensu*) com ele (E₃). Tanto é assim que os integrantes do primeiro plano têm consciência de que estão ali por causa do interlocutor (E₃) – o ouvinte da rádio. As interações nessa macro cena ([E₁↔E₂]↔E₃) são viabilizadas pelo entrevistador e não são necessariamente evidentes.

Em outros termos, o que aparentemente configuraria um gênero primário (E₁↔E₂), torna-se componente do gênero secundário ([E₁↔E₂]↔E₃), isto é, “[...] transforma-se dentro deste e adquire uma característica particular: perde sua relação imediata com a realidade existente [...]” (BAKHTIN, 1997, p. 281). Isso porque o diálogo aparente se estrutura para satisfazer as necessidades presentes na realidade que envolve o ouvinte (E₃). Encontramos nas palavras de Farneda (2007) a definição sintética do que tratamos nos parágrafos anteriores:

A entrevista radiofônica é um gênero jornalístico que diz respeito ao encontro de um entrevistador (jornalista) e um entrevistado (especialista em um assunto em particular), cujo interesse é fazer falar o *expert* a respeito dos diferentes aspectos de uma questão e, dessa forma, levar as informações fornecidas, por essa interação, a terceiros. Sendo contrária a uma conversa comum, a entrevista apresenta um caráter estruturado e formal, cujo objetivo é satisfazer as expectativas do destinatário (FARNEDA, 2007, p. 2).

Além de obter a atenção de seu espectador, o entrevistador deve buscar também a simpatia de seu entrevistado, que lhe poderá conceder maior informação à medida que se construa um vínculo de maior confiabilidade. Daí que se infere que:

⁹ Afastando-nos de qualquer mal-entendido que a citação possa gerar, ao se apropriar desses atributos para caracterizar os gêneros secundários, Bakhtin (1997) tinha em mente que as evoluções tecnológicas presenciadas em algumas comunidades poderiam servir como um agente catalisador da criação de novos gêneros. Isto porque, os gêneros **secundários**, tido como mais complexos, “inspiram-se nos gêneros primários e aparecem na sociedade quando esta atinge um certo grau de sofisticação cultural, permitido, principalmente pelo letramento” (SILVEIRA, 2005, p.63).

[...] la entrevista es un acto de seducción mutua. El entrevistador seduce para obtener más y mejores respuestas. El entrevistado busca convencer a su entrevistador, llevarlo a su juego, decir lo que quiere decir y evitar decir lo que no quiere decir (PÉREZ COTTEN; TELLO, 2004, p. 37).¹⁰

Outra característica da entrevista radiofônica que merece atenção refere-se à já mencionada presença da voz do entrevistado. O uso do discurso direto nesse gênero (i) aproxima o ouvinte do entrevistado, já que aquele sabe que este fala para ele; (ii) elimina a intermediação direta de um terceiro (jornalista), criando uma (iii) aparente objetividade; (iv) permite que o ouvinte tire suas próprias conclusões, protegendo, assim, o jornalista de eventuais asserções comprometedoras; (v) cria empatia ou antipatia entre ouvinte e entrevistado; além, é claro, de (vi) permitir o conhecimento imediato da voz e os testemunhos envolvidos nos episódios noticiados (PÉREZ COTTEN; TELLO, 2004, p. 33).

Haja vista que a presença direta de pelo menos duas vozes (entrevistador e entrevistado) é imprescindível nesse gênero, o estudo de sua construção composicional deverá considerar o modo como, formalmente, organiza-se o diálogo nessa interação verbal. Nesse sentido, Farneda (2007, p. 2) mostra-nos que o “entrevistador e o entrevistado ocupam posições assimétricas”, mas que este “deve respeitar a agenda de perguntas prevista pelo jornalista. Os turnos de pergunta devem sempre terminar com uma interrogativa” e por sua vez, os turnos do entrevistado revestem-se unicamente de respostas. Ao entrevistador cabe a abertura e o fechamento da entrevista e não lhe compete “formular expressões de ratificação, opinião ou comentário, abstendo-se de formar opinião contra ou a favor do entrevistado”.

Com o propósito de construir certa imparcialidade, o entrevistador controlará os turnos por meio de falas mais breves. O entrevistado, por sua vez, terá falas mais longas a fim de transmitir a informação que detém. A essas características composicionais, somam-se as instruções dadas por Portugal e Yudchak (2008, p. 76), para quem a condução da entrevista radiofônica deve ser feita por meio de perguntas individualizadas, de modo claro, curto e concreto. Na figura 2, verificamos como essas características composicionais se realizam em uma das entrevistas que compõem *corpus*.

¹⁰ <Tradução nossa> “[...] a entrevista é um ato de sedução mútua. O entrevistador seduz para obter mais e melhores respostas. O entrevistado procura convencer seu entrevistador, levá-lo a seu jogo, dizer o que quer dizer e evitar dizer o que não quer dizer”.

Portugal e Yudchak (2008, p. 84) destacam ainda a preferência por uma linguagem simples e, a partir de relatos de importantes jornalistas, mostram-nos o estilo mais empregado nesse gênero. Assim, Alfredo Leuco, jornalista cordobês, diz utilizar “uma linguagem mais simples, frases curtas, [apelar] à sabedoria popular, à linguagem da rua, à linguagem coloquial, [...] com o que se ganha maior familiaridade e proximidade”. Notemos como o uso de uma linguagem menos erudita possibilita a aproximação entre jornalista e sua audiência.

Figura 2 — Do modelo de composição estrutural da entrevista radiofônica.

KAR 0'00"	Diez de la mañana treinta y seis minutos. Como todos los lunes a la mañana, está con nosotros el doctor Guillermo Alonso, jefe del servicio de acupuntura del hospital de clínicas Presidente Nicolás Avellaneda. Le damos la [bienvenida] ¿Cómo le va, doctor?	} El inicia o diálogo e apresenta E2.
ALO 0'14"	Bien bien.	
KAR 0'14"	Bien venido.	
ALO 0'15"	Muchísimas gracias.	
KAR 0'16"	Bueno, hoy un tema, todo el mundo tiene... se hace esta pregunta. Eh... Sobre el te... sobre el tabaquismo, ¿la acupuntura ayuda a superarlo?	→ Turnos menores e interrogativos
ALO 0'30"	Bien, sí. Hay hay numerosos pacientes que concurren a nuestro servicio. No vamos a decir que todos tienen éxito, pero un porcentaje importante, sí, lo tienen. Eh... Porque muchas veces tienen la convicción de dejar de fumar. Eh... muchas veces co... conocen obviamente cuales son los efectos deletéreo que está haciendo sobre su organismo, sobre la parte respiratoria e sobre el resto de su organismo, conocen las limitaciones... sí sí, claro claro, hay que [####], a la está llegando por su problema de tabaquismo, pero, a pesar de eso y a pesar de su bueno propósito ellos no pueden dejar de fumar. Entonces, buscan esta ayuda.	} Entrevistado domina turnos maiores.
KAR 1'04"	¿Y cuáles serían los mecanismos a los que abunda la acupuntura?	→ Turnos menores e interrogativos
ALO 1'07"	Bien, en primer lugar, eh por vía refleja, o sea, con las agujas, lo que logramos, por un lado, disminuir el grado de ansiedad porque, de alguna manera, quien fuma tiene un beneficio de [####]. Un beneficio entre comillas, o sea, el cigarrillo le ayuda en este momento a a disminuir la ansiedad. Obviamente que después viene el conflicto: "¿Por qué <u>he fumado</u> ?", " <u>he sido débil</u> ", "no... no <u>he tenido convicciones</u> ", "no <u>he tenido voluntad</u> " y eso le genera de nuevo... le genera de nuevo la ansiedad y paradójicamente le hace fumar de nuevo mucho más. O sea que, por un lado, le disminuimos el tenor de la ansiedad a a esta persona. Por otra parte, el fumador tiene bloqueado el gus... el gusto y el olfato. O sea, se recuerdan los que fuman que los primeros cigarrillos eran duros, rígidos, fuertes, raspaban... y que después uno va logrando que todos sus... esas malas sensaciones se desaparezcán y hace una filtración de lo que pretende del cigarrillo y nada más. [...] Obviamente que... algunos pacientes necesitan más tiempo, otros menos tiempo, pero hay un porcentaje importante de paciente que lo logra.	} Entrevistado domina turnos maiores.
KAR 3'15"	Perfecto. Doctor, ¿cómo se lo... se encuentra? ¿Todavía sin teléfono?	→ Turnos menores e interrogativos
ALO 2'20"	Antes teníamos... tenía la comunidad, los los pacientes ya conocidos como los nuevos me llamaban por teléfono para pedir asesoramiento o turno. Hoy tienen que molestarse tanto a los nuevos como los que ya tienen eh... su historia clínica de servicio, hasta el servicio en el hospital Avellaneda, en los horarios que ustedes conocen, que es de siete y media a doce, porque lamentablemente todavía no tenemos teléfonos, como [todo un nada de este lugar] donde está a servicio de cirugía. Que [Javier] <u>ha hecho</u> una obra muy importante, se <u>ha llevado</u> por adelante el cabrero [#### #### #### ####] nuevamente	} Entrevistado domina turnos maiores
KAR 3'50"	O sea que la gente tiene que ir y pedir contacto.	
ALO 3'53"	Exactamente	
KAR 3'53'	Perfecto. Con nosotros, como todos los lunes, el doctor Guillermo Alonso, jefe del servicio de acupuntura del hospital de clínicas presidente Nicolás Avellaneda. Gracias, doctor, hasta lunes.	} El retoma o turno para terminar o diálogo.
ALO 4'03"	Gracias a usted	

Fonte: elaborada pelo autor.

Da mesma maneira, Graciela Mancuso diz que “hoje a linguagem foi liberada, fala-se como na vida” e Chiche Gelblung afirma que “a rádio tem que falar como o ouvinte”. Essa jornalista destaca ainda o caráter espontâneo que tem a linguagem nessa interação discursiva: “em realidade, a rádio tem que falar como você quer, nada substitui o pessoal”.

Parece-nos evidente, no entanto, que, diante de um meio de ampla difusão e expressividade, não podemos crer que, de fato, a seleção de elementos linguísticos nos

enunciados pertencentes ao gênero é exatamente igual à escolha verificável no uso comum, rotineiro e casual¹¹, no qual não se observa preocupação com a fala. Diante dessa situação de enunciação, o falante parece monitorar, mesmo que discretamente, sua fala a fim de alcançar um padrão linguístico que esteja de acordo com o que ele avalia como mais adequado a um meio de maior prestígio e desenvolvimento tecnológico, como é o caso da rádio.

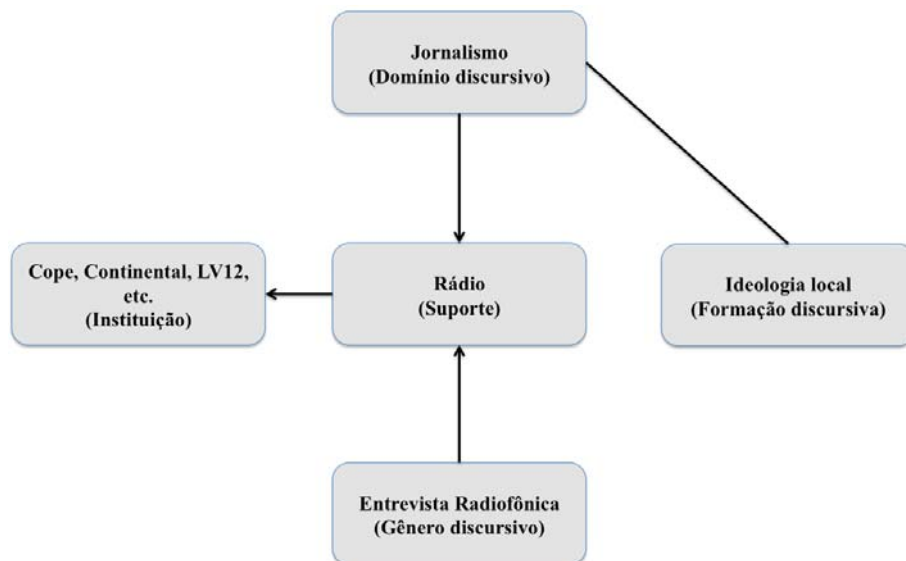
Por outro lado, a existência de uma maior espontaneidade nesse gênero pode estar relacionada, em partes, à modalidade de concepção e transmissão dos enunciados, ou seja, por pertencer ao domínio da oralidade, torna-se menos possível fazer releituras corretivas, pois tendo sido enunciada, a mensagem já não pode ser apagada – sobretudo quando se trata de entrevistas ao vivo.

Por fim, destacamos a recorrência de diferentes tipos textuais nesse gênero. Tanto é assim que **entrevista radiofônica** pode agregar quase que a totalidade deles, descrevendo eventos (descrição), ordenando-os temporalmente (narração), explicando e analisando determinadas situações ou ideias (exposição), bem como contrapondo ideias (argumentação).

De modo prático e sintético, se aplicarmos a prévia discussão teórica sobre o gênero **entrevista radiofônica** aos enunciados do *corpus*, chegaremos à composição da figura 3, na qual nos é revelado como as categorias teóricas vinculadas à noção de gênero do discurso se organizam na produção da entrevista radiofônica:

Figura 3 — Da relação do *corpus* com as categorias teóricas envolvidas no gênero discursivo.

¹¹ Segundo Labov (2008, p. 111) “por fala casual, em sentido estrito, entendemos a fala cotidiana usada em situações informais, em que nenhuma atenção é dirigida à linguagem”.



Fonte: elaborada pelo autor.

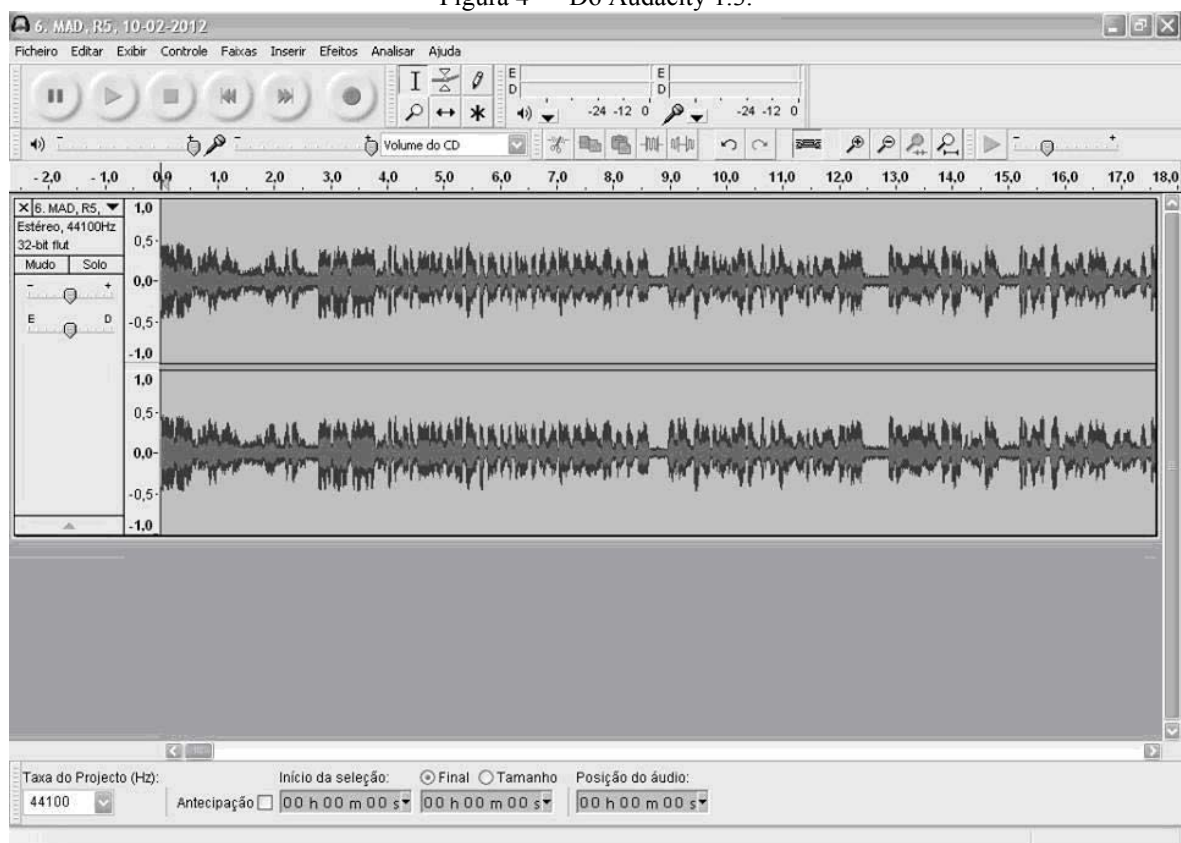
Uma vez construída esse panorama geral do funcionalmente e caracterização do gênero **entrevista radiofônica**, apresentaremos, a seguir, os softwares a que recorreremos para compilar (*Audacity 1.3*) e processar automaticamente (*Tropes 7.2.3*) os textos que constituem os *corpora* de análise, explicando sucintamente a contribuição deles para cumprir os objetivos estabelecidos neste estudo. Isso para que, em seguida, possamos explorar mais concretamente o material que compõe o *corpus* de análise, isto é, descrever os enunciados coletados e verificar como as características do gênero discursivo em foco realizam-se neles.

3. Ferramentas tecnológicas para o estudo da linguagem

3.1 Audacity 1.3

Trata-se de um *freeware* de gravação e edição de áudio em formato WAV, AIFF, MP3 e OGG. Por sua facilidade de manuseio, serviu-nos para a gravação das entrevistas radiofônicas transmitidas por internet, sua edição, além de nos auxiliar na transcrição dos enunciados. Algumas ferramentas como desaceleração do áudio, seleção de fragmentos, entre outras, viabilizaram a realização dessa última função. A informação do tempo de gravação é também um dado fornecido pelo programa por meio da linha do tempo (indicada pela flecha). A seguir, expomos uma imagem da interface do *Audacity 1.3*.

Figura 4 — Do Audacity 1.3.



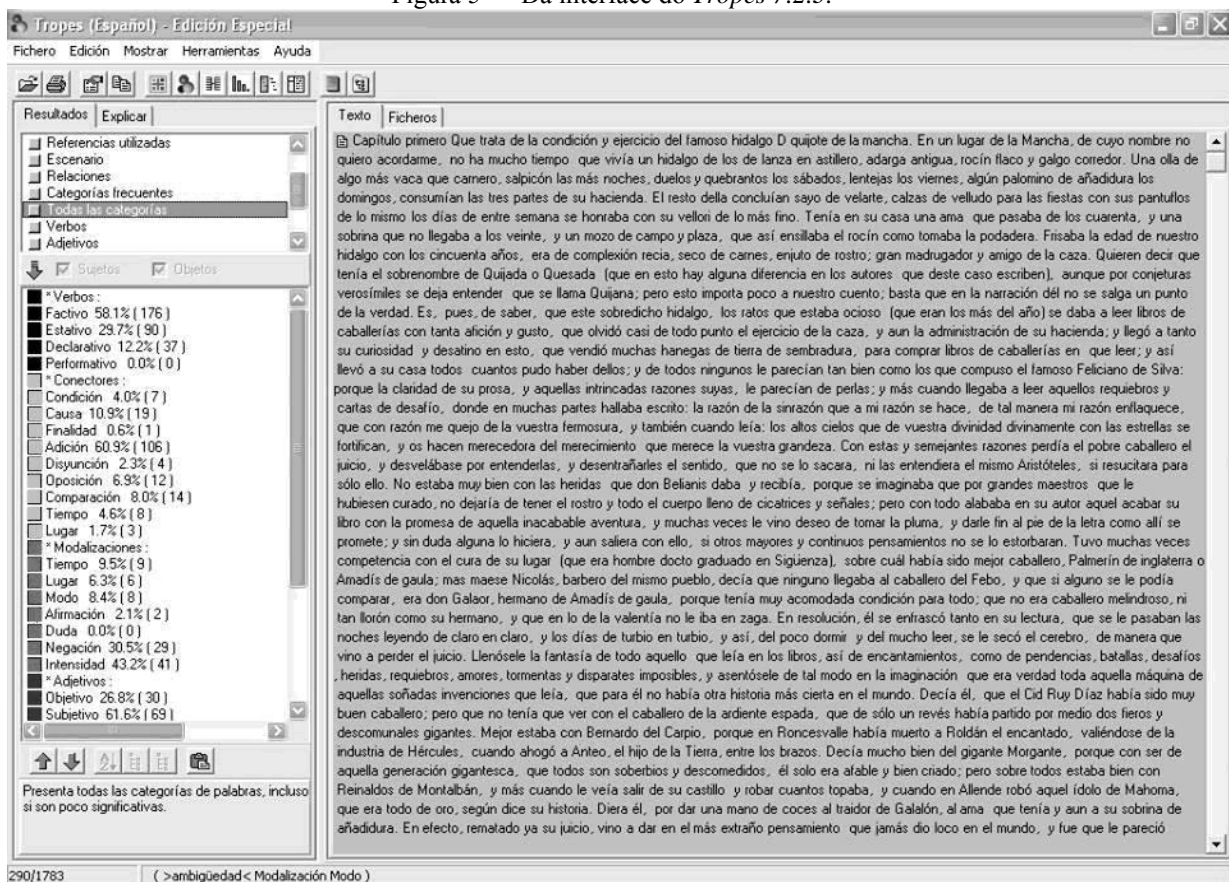
Fonte: própria.

3.2 Tropes 7.2.3

O segundo *software* destaca-se pelo processamento semântico de textos em línguas naturais. Para descrever as características dos enunciados em análise, o *Tropes 7.2.3* vale-se de critérios linguísticos pré-programados e os associa às estruturas linguísticas encontradas nos textos processados. Frente à desenvoltura metodológica tida pelo programa e nosso objeto de estudos, faz-se imprescindível a adoção da versão em língua espanhola.

Como se observa na figura 5, a interface do programa apresenta duas colunas, uma de *Resultados* (à esquerda), na qual verificamos as características textuais e suas proporções, e uma de *Texto* (à direita), na qual são apresentadas sequências textuais que ilustram os resultados da esquerda.

Figura 5 — Da interface do Tropes 7.2.3.



Fonte: própria.

O *software* baseia sua análise na observação de seis classes de palavras previamente definidas pelo próprio sistema do Tropes 7.2.3¹², as quais se estruturam nos grupos dos “verbos”, “conectores” (conjunções), “modalizações” (advérbios e locuções adverbiais), “adjetivos qualificadores”, “pronomes pessoais”, “substantivos e nomes próprios” – conforme se observa na coluna esquerda, da figura 5. Com o processamento automático do texto indexado ao *software* em formato.txt, o Tropes 7.2.3 classifica as palavras mais significativas presentes no material tendo em vista essas seis categorias de classe de palavras.

Dita classificação não apenas identifica as palavras em classe, mas também subclassifica cada um dos itens segundo a função semântica que apresenta dentro de cada classe de palavras. Essa subclassificação também é pré-definida pelo próprio *software* e se organiza em:

¹² Essas informações foram retiradas do site que divulga e vende o *software* em questão: <https://www.semantic-knowledge.com/doc/V70/text-analysis/index.html>. Acessado em 09.01.2017.

- **Verbos:** *factivos* (de ação: “fazer”, “trabalhar”, “andar”, etc.), *estativos* (estados ou *posses*: “ser”, “estar”, “possuir”, etc.), *declarativos* (declaração sobre circunstâncias, seres, objetos: “achar”, “julgar”, “crer”) ou *performativos* (ato de linguagem ou realizado por meio da linguagem: “prometer”, “dizer”, etc.).
- **Conectores:** *condicional* (“se”, “caso”, etc.), *causa* (“porque”, “portanto”, etc.), *finalidade* (“para que”, “afim de”, etc.), *adição* (“e”, “além disso”, etc.), *disjunção* (“ou”, “ou... ou”, etc.), *oposição* (“mas”, “contudo”, etc.), *comparação* (“como”, “bem como”, etc.), *tempo* (“quando”, “enquanto”, etc.), *lugar* (“onde”, “em que”, etc.).
- **Pronomes pessoais:** organizados em gênero, número e pessoa.
- **Modalidades:** *tempo* (“agora”, “ontem”, “amanhã”, etc.), *lugar* (“lá”, “aqui”, “abaixo”, etc.), *modo* (“diretamente”, “em conjunto”, etc.), *asserção* (“absolutamente”, “certamente”, etc.), *dúvida* (“talvez”, “provavelmente”, etc.), *negação* (“não”, “nunca”, “nada”, etc.), *intensidade* (“muito”, “fortemente”, etc.).
- **Adjetivos:** *objetivos* (caracterizam os seres ou objeto, independentemente do ponto de vista do enunciador. Adjetivos de cor, por exemplo), *subjetivos*: (indicam julgamento sobre algo ou sobre alguém, permitindo assim expressar o ponto de vista do enunciador – “bonito”, “pequeno”, “bom”, etc.), *numeral* (agrupando em numeração ordinal e cardinal).

Ao examinar a ordem de ocorrência das palavras tanto no interior dos orações como no texto, de modo geral, o *software* não apenas **categoriza as palavras** mais frequentes, mais também infere, a partir dos dados provenientes dessa classificação, a **tipologia textual** preponderante nos textos, comparando os resultados obtidos da análise específica do texto indexado com um “padrão” pré-estabelecido pelo programa – o qual deriva do estudo de um grande número de textos diferentes da língua analisada.

Assim, a classificação da tipologia de textos ocorre a partir da informação que as palavras mais significativas aportam, sendo possível a identificação de quatro tipos:

- **Argumentativo:** o enunciador envolve-se, argumenta, explica ou analisa, a fim de tentar convencer o interlocutor.

- **Narrativo:** um narrador afirma uma série de eventos, acontecendo em um determinado momento e em um determinado lugar.
- **Enunciativo:** o enunciador e o interlocutor estabelecem uma relação de influência mútua, compartilhando seus pontos de vista.
- **Descritivo:** um narrador descreve, identifica ou classifica algo/alguém.

Finalmente, além da **categorização das classes de palavras** e da **tipologia textual**, outra ferramenta oferecida pelo *Tropes 7.2.3* que nos será útil no cumprimento dos objetivos deste trabalho está relacionada à identificação dos **campos de referência semântica**. Função que aloca as palavras que aparecem com maior frequência ao longo do texto em um grupo de significados estreitamente relacionados. Assim, identificam-se as temáticas preponderantes do texto. Esse é o caso, por exemplo, da categorização de palavras como “pai”, “mãe”, “irmão”, “avó” e “tio”, pertencentes ao campo de referência familiar.

Será, portanto, a partir da combinação dessas três funções (**categorização das classes de palavras**, da **tipologia textual** e dos **campos de referência semântica**) que as análises oferecidas pelo *Tropes 7.2.3* irão nos auxiliar na descrição das entrevistas radiofônicas, oferecendo-nos, mais especificamente, informações sobre o estilo e o conteúdo temático do gênero. Isso posto, passemos, então, à exploração do *corpus* compilado.

4. Compilação do *corpus* de análise

Segundo Pérez Hernández (2002), a revolução tecnológica vivenciada pela humanidade permitiu-nos, entre outras coisas, o estudo linguístico fundamentado na análise conjunta de enunciados pertencentes às múltiplas situações de interação discursiva, afastando-nos, conseqüentemente, de análises impressionistas, baseadas apenas nas intuições linguísticas do estudioso. Valendo-nos de parte desses recursos, constituímos um *corpus* de análise a partir de materiais linguísticos que registram “a linguagem natural realmente utilizada por falantes e escritores da língua em situações reais” (BERBER SARDINHA, 2002, p. 352) e que se caracteriza por apresentar um

[...] conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y profundidad de manera que sean representativos del total del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador con

el fin de obtener resultados varios útiles para la descripción y el análisis (SÁNCHEZ, 1995, p. 8-9)¹³.

Note que, segundo o autor, um *corpus*, além de ser composto por textos efetivamente realizados na prática comunicativa, deverá seguir critérios que lhe assegurem representatividade de, pelo menos, algum âmbito da língua. Esses parâmetros, por sua vez, se caracterizam por condições de naturalidade, de autenticidade, de diversidade, de dimensões e a outros padrões que possam estar relacionados aos propósitos investigativos. A disponibilidade para o processamento digital é também uma das características que deve possuir o *corpus*, posto que com o auxílio dos recursos tecnológicos podemos analisar, de maneira mais precisa e ágil, uma grande quantidade de enunciados.

Considerando as limitações espaciais e temporais que enfrentamos e buscando respeitar as características apresentadas, podemos encontrar condição que satisfaça as exigências propostas por Sánchez (1995) em um *corpus* composto por **entrevistas radiofônicas**. Isto porque, além de encontrarmos esses enunciados disponíveis na rede mundial de computadores – em rádios das três regiões dialetais, que disponibilizam sua transmissão *on-line* –, esse gênero, como já mencionado, apresenta uma variedade linguística mais próxima ao vernáculo.

Uma vez que enunciados pertencentes a um único gênero discursivo e apenas à modalidade falada da língua não podem constituir um *corpus* representativo da totalidade de usos linguísticos de uma comunidade de fala, reconhecemos que as apreciações e conclusões provenientes deste estudo estão limitadas a um importante âmbito da língua empregada em três variedades diatópicas (Madri, Buenos Aires e San Miguel de Tucumán), no qual se observa o domínio da oralidade com pouco monitoramento. Além disso, o emprego desse gênero discursivo nos possibilita a observação da fala de uma gama diversificada de informantes das regiões, tornando mais viável um estudo diatópico que requeira, por exemplo, as informações referentes aos indivíduos que participam da construção dos enunciados e de seu entorno de enunciação, isso porque a opção por esse gênero e o apoio da internet nos

¹³[...] conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos e dispostos de tal modo que possam ser processados por meio de computador, com o fim de obter resultados vários e úteis para a descrição e a análise.

possibilitam o acesso a esse tipo de dados – ora por inferência na própria entrevista, ora por contato direto com as rádios ou, até mesmo, por meio de rede de relacionamentos (*Facebook*).

Sobre a obtenção dos textos, quando não disponibilizados para *download* pelo próprio site da rádio que difundiu a entrevista, o uso do *software Audacity 1.3*, como previamente assinalado, serviu-nos para gravação das entrevistas. Uma vez que a metodologia deste trabalho é uma continuidade da metodologia empregada em trabalhos anteriores (ARAUJO, 2012, 2013), parte do material que compõe o *corpus* desta pesquisa provém da primeira base de dados compilada. Assim, valemo-nos dos *corpora* preexistentes de Buenos Aires e San Miguel de Tucumán, ampliando-os, no entanto, em quantidade de tempo e palavras. Assim, o que antes se limitava a aproximadamente quarenta e sete minutos e oito mil palavras por área diatópica, nessa etapa do processo investigativo, foi ampliado para mais de duas horas, o que equivale a mais de vinte mil palavras por variedade.

Salientamos, contudo, que o *corpus* de Madri foi inteiramente compilado, haja vista que essa variedade não havia sido contemplada em nossas análises prévias. Dessa maneira, a primeira fase dos trabalhos de gravação e transcrição ocorreu em 2010, ao passo que a segunda fase se deu entre os anos de 2013 e 2014. Dispomos na tabela 1 as três regiões dialetais apreciadas neste estudo com algumas informações das entrevistas. Como se pode notar, as mais de 6 horas de gravações (Tempo de grav.), referentes às 28 entrevistas radiofônicas (Nº de entrev.), forneceram-nos quase 66 mil palavras¹⁴, sendo, em média, mais de vinte e um mil a quantidade de palavras provenientes de cada uma das três variedades diatópicas.

É interessante destacar que no grupo de 48 informantes (Nº de info.), encontram-se falantes que possuem desde 26 anos até 70 anos (Faixa etária). No entanto, os dois extremos não constituem parte do grupo etário preponderante nesse gênero. Tanto é assim que a média aritmética da idade dos 51 falantes mostra-nos que a idade média dos falantes é de aproximadamente 44 anos. Em relação ao gênero/sexo dos informantes, notamos uma clara preponderância de homens, posto que do total, apenas 12 são mulheres (23%). As informações referentes à idade e sexo levam-nos a refletir, mais uma vez, sobre o gênero

¹⁴ As informações de tempo de gravação e quantidade de palavras foram obtidas, respectivamente, por meio da linha do tempo exposta no *Audacity 1.3* e pela ferramenta contar palavras, do editor de texto *Microsoft Office Word*.

discursivo em questão e sua função social, posto que a entrevista radiofônica parece estar marcada pela presença de uma população masculina e adulta.

Tabela 1 — Da descrição das entrevistas radiofônicas que compõem o *corpus*.

Variedade diatópica	Rádio	Programa	Nº de entrev.	Tempo de grav.	Nº de palav.	Nº de infor.	Faixa etária	Mulher
Madri	Cope	El partido de las doce	11	2h03'58''	23.357	20	26 – 57	4
	Radio 5	Entrevista en R5						
Buenos Aires	Continental	La mañana	8	2h01'30''	21.124	16	28 – 70	4
	Palermo	Comunas en Plural						
		Entre nosotras						
	Cooperativa	El vermucito del domingo						
Los más grandes								
S. M. Tucumán	LV 12	Manyines en la radio	9	2h00'57''	21.221	12	30-59	4
	LV 7	La mañana de LV7						
		La tarde de LV7						
	Fish	Sin pescado concebido						
Total			28	6h06'25'	65.702	48	26 – 70	12

Fonte: elaborada pelo autor.

Seguindo a tipologia proposta pela Linguística de *Corpus* (BERBER SARDINHA, 2000), nosso material de análise se identifica com o “modo falado”, pois tanto em sua concepção como em sua propagação faz uso da oralidade. Quanto ao “tempo”, trata-se de um “*corpus* sincrônico e contemporâneo”, por abordar um único período: o corrente. É “dialetal” e “especializado”, por apresentar, respectivamente, um conteúdo que visa satisfazer um estudo dialetológico e por decorrer de um único domínio discursivo: o jornalístico. Por fim, é um *corpus* de “língua nativa”, já que os falantes possuem o espanhol como língua materna.

Quanto ao tamanho do *corpus* compilado e sua relativa representatividade, é importante sabermos que, de acordo com o que propõe Berber Sardinha (2000, p.346), manipulamos um *corpus* pequeno, por possuir menos de 80 mil palavras. No entanto, ainda orientados pelo autor, somos conscientes de que “a quantidade mínima de dados necessários para a formação de um *corpus* nunca foi estimada [...], sendo o critério de tamanho empregado subjetivamente na definição de *corpus*” (p.345). Desse modo, o *corpus* compilado é suficientemente significativo para o estudo das características de estilo, forma e conteúdo do gênero **entrevista radiofônica**.

5. Aportes da análise textual automática para a descrição do gênero entrevista radiofônica

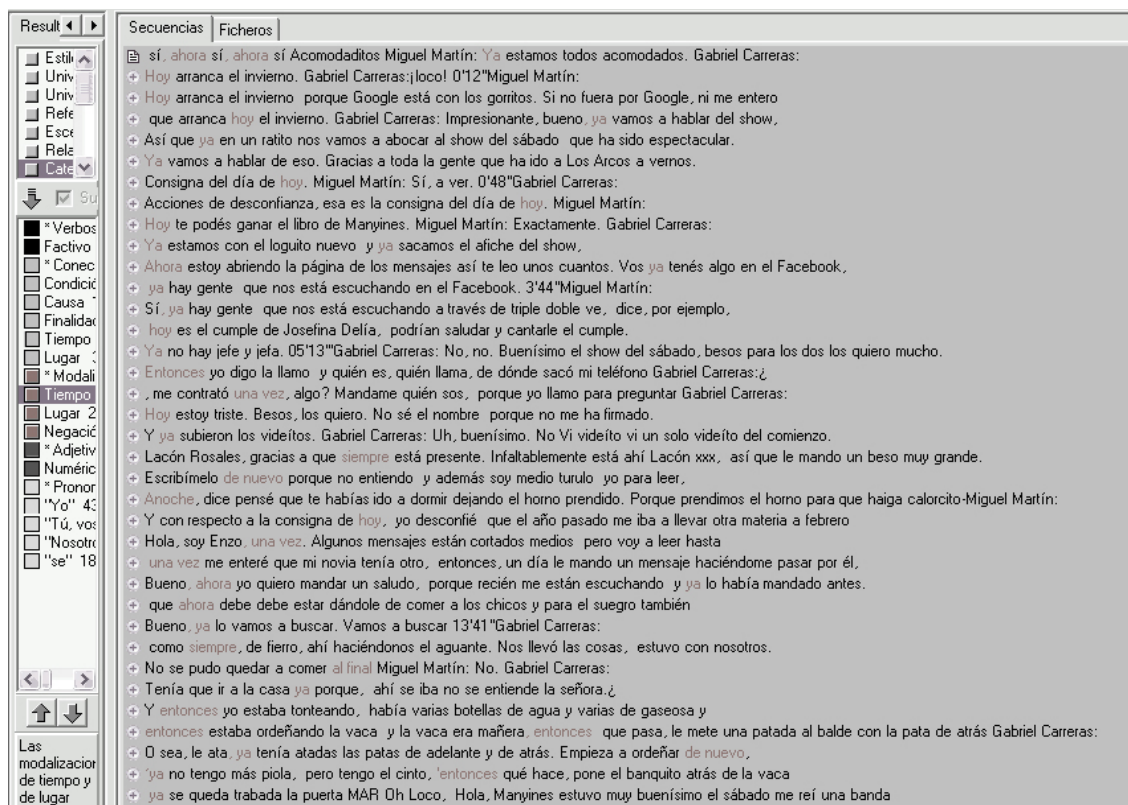
A fim de proceder à descrição de algumas das características relacionadas ao estilo, ao tema e à composição dos enunciados veiculados pela entrevista radiofônica, recorreremos ao auxílio do *Tropes 7.2.3*, *software* de processamento semântico. Segundo sua análise, o tipo textual preponderante na composição dos enunciados do *corpus* é o **argumentativo**, ainda que os demais tipos sejam encontrados com frequência igualmente significativa. Quanto à temática, foi salientada a preponderância de debates sobre questões sociais. No entanto, também estiveram presentes temas ligados à geografia e urbanismo, arte e cultura, política, ciências e tecnologia, economia e finanças, jogos e esportes, animais, alimentação, empresarial, saúde e doenças, direito e justiça, emprego e trabalho¹⁵, ou seja, temas que envolvem a vida cotidiana da sociedade e que, na esfera jornalística, são resgatados e transformados em notícias por serem de interesse do ouvinte.

Por fim, sobre o estilo, o *software* aponta a preponderância de verbos **factivos** e **estativos**¹⁶, que contribuem para a constituição de uma encenação dinâmica. Por meio da categoria **modalização temporal**, a mais preponderante das modalizações no *corpus*, podemos verificar uma grande recorrência de advérbios e locuções adverbiais de tempo que envolvem, fundamentalmente, o âmbito da enunciação (*ahora, hoy, ya*), que expressam frequência (*siempre, una vez, de nuevo*) e fragmentação do dia (*tarde, mañana, anoche*), além de alguns advérbios de posterioridade e anterioridade (*mañana, ayer, el año pasado, en aquel momento*), entre outros. A interface do *software* apresenta-nos essas informações na figura 6.

Figura 6 — Da exposição da modalização temporal no *Tropes 7.2.3*.

¹⁵ Apropriamo-nos da terminologia empregada pelo próprio *software*.

¹⁶ Verbos que, segundo o *software*, expressam, respectivamente, ação e estado.



Fonte: própria.

Destacamos também a preponderância de formas pronominais de primeira e segunda pessoas do singular – reflexo evidente da situação discursiva instaurada no gênero, na qual duas pessoas estão em diálogo direto –, e a tendência da variedade linguística usada nesse gênero aproximar-se do uso vernacular, isto é, menos monitorado e, logo, mais espontâneo. Isso é o que elucidam os seguintes fragmentos:

(01) CAR: [...] ahora sí, ahora sí, ahora sí. Acomodaditos. MAR: Ya estamos todos acomodados. CAR: Impresionante este feriado. MAR: ¿Qué onda? <TI>
 CAR: [...] agora sim, agora sim, agora sim. Acomodadinhos. MAR: Já estamos todos acomodados. CAR: Impresionante este feriado. MAR: ¿Qual é boa?

(02) Hola, Víctor Hugo. Tanto tiempo, ¡qué placer! Me estaba riendo con los chistes de tu compañero <BI>.
 Olá, Victor Hugo. Tanto tempo. Que prazer! Estava rindo com as piadas do seu companheiro.

(03) Antonio Adán, ¿qué tal? Buenas noches. ¿Qué tal estás? <MI>
 Antonio Adan, e aí? Boa noite. ¿Como está?

Nos fragmentos apresentados, além de expressões que indicam uma situação de aproximação e espontaneidade (“*¡qué placer!*”, “*me estaba riendo*” e “*chistes*”), encontramos

conjugação verbal (*estás*) e paradigma pronominal (*tu compañero*) referente à segunda pessoal do singular *tú/vos*, cujo uso, na Espanha e Argentina, respectivamente, também está relacionado a circunstâncias de menor formalidade e maior familiaridade. Destacam-se também **repetições com efeito cômico** (“*ahora sí, ahora sí, ahora sí*”), **uso de diminutivos** (“*Acomodaditos*”), **expressões de gosto pessoal** (“*¡qué placer!*”, “*Impresionante*”), **formas de questionamento e cumprimentos menos formais** (“*Hola*”, “*¿Cómo estás?*”, “*¿Qué onda?*” “*¿Qué tal?*”), entre tantos outros exemplos verificáveis na totalidade do *corpus*.

Por outro lado, conforme advertido anteriormente, sabe-se que por se tratar de um gênero midiático, no qual não raramente o falante vê-se especialmente atento à situação enunciativa em que se encontra, podemos identificar marcas que apontam, de algum modo, para um monitoramento linguístico. Tanto é assim que é comum constatar, entre outros, o uso recorrente dos seguintes tipos de marcadores conversacionais:

(04) **Repetições de palavras ou grupo de palavras:** [...] *pasearse con las desgracias de la gente y... y bueno queriendo mostrar que ellos tienen la verdad absoluta y realmente eso muchas veces confunde al electorado y... y da mucha bronca eh... que que haya tanta mala gente aprovechándose de de la... de lo... de los que por, como lo dije, cayeron en desgracia [...]* <T7>.

[...] *passar com a desgraça das pessoas e... e bom, querendo mostrar que eles detêm a verdade absoluta e realmente isso muitas vezes confunde o eleitor e... e dá muita raiva é... que que haja tanta gente má se aproveitando de de la... de lo... de los... que por, como disse, caíram na desgraça [...]*.

(05) **Pausas na fala:** *Eh... Madrid también ha mirado sus intereses y yo he mirado por los míos. Eh... ha habido propuestas eh... de otros clubs eh... Y bueno al final por unas circunstancias o por otras ninguna... ninguna se ha cerrado ¿no? Eh... yo creo que es una negociación amplia... [...]* <MI>

É... *Madri também olhou seus interesses e eu olhei pelos meus. É... houve propostas é... de outros clubes é... e bem, por fim, por uma ou outra circunstância, nenhuma... nenhuma foi concluída, né? É... eu acho que é uma negociação ampla... [...]*.

(06) **Marcadores conversacionais:** como uso excessivo de “eh”, “este” e “bueno”: “*Eh... bueno, una Eh... una sesión muy emotiva*”; “[...] *un romance que que se mantiene y, bueno, eh... [...]*”; “*Este... de estas personas que [avalaron] la desaparición de personas, la detención de hijos de desaparecido, la apertura, este... yo digo que este... nadie puede dormir con esta sensibilidad [...]*” <B3>

É... *bom, uma é... uma sessão muito emotiva; [...]* *uma novela que se mantém, e bom é... [...]*; *Este... destas pessoas que avaliaram a desapareição de pessoas, a prisão de filhos de desaparecidos, a abertura, este... eu digo que este... ninguém*

pode dormir com esta sensibilidade [...].

Diante desses dados, não negamos a efetiva aproximação entre o uso vernacular e a variedade linguística usada no gênero **entrevista radiofônica**, contudo, consideramos também que inserido nessa situação de interação humana, o falante apresenta uma sensibilidade maior à realidade linguística instaurada nesse gênero, de maneira que se atém um pouco mais ao modo como fará uso da linguagem. Assim, parece-nos ingênuo considerar as informações provenientes da observação dessa base de dados como válidas, necessariamente, para um contexto cotidiano de monitoramento próximo ao nulo, como se espera, por exemplo, do diálogo trivial de amigos.

6. Considerações finais

Propusemos, neste trabalho, descrever o gênero **entrevista radiofônica** em algumas comunidades hispânicas (Madri, Buenos Aires e San Miguel de Tucumán), avaliando as contribuições do *Tropes 7.2.3 – software* de processamento automático de textos – para os objetivos traçados. Com esse fim, vimos que os gêneros do discurso caracterizam-se como estruturas textuais relativamente estáveis que nos servem de base para alcançarmos, por meio da língua, determinados objetivos comunicativos (BAKHTIN, 1997). Ainda segundo o Bakhtin (1997), os gêneros se caracterizam por um estilo, tema e composição próprios. Dirigindo-nos às entrevistas radiofônicas, verificamos que esse gênero se define funcionalmente pelo interesse em reconstruir um evento por meio do discurso, apreendendo e avaliando as opiniões envolvidas nos acontecimentos das mais diversas esferas da sociedade. Ou seja, a entrevista radiofônica se organiza em função do “informar”, fazendo da informação, “notícia”.

Composicionalmente, a entrevista radiofônica apresenta uma estrutura clara de diálogo, que, como vimos, constrói-se em duas frentes. Na primeira, observa-se um diálogo mais próximo do cotidiano, no qual as relações são sincrônicas, mais espontâneas e, aparentemente, se configura a entrevista do enunciador com o enunciatário. Contudo, essa situação não tem valor por si só, já que visa informar um terceiro personagem (enunciatário-ouvinte/espectador) e, portanto, estabelecer uma relação dialógica (*lato sensu*) com ele. Tanto é assim que os integrantes do primeiro plano têm consciência de que estão ali por causa desse interlocutor – o ouvinte da rádio, presente na segunda frente.

Orientados por pressupostos teóricos e metodológicos da linguística de *corpus*, procedemos à coleta e à transcrição dos dados que compuseram o *corpus* analisado pelo *Tropes 7.2.3*. Esse material foi composto por mais de 65 mil palavras e 6 horas de gravação – aproximadamente 2 horas e 22 mil palavras por variedade diatópica investigada –, frutos da gravação de 28 entrevistas, com 48 informantes. Em síntese, a base de dados compilada, caracteriza-se pela modalidade falada, pelo tempo sincrônico e contemporâneo, pelo traço dialetal, especializado e por ser de língua materna (BERBER SARDINHA, 2000).

Sobre as contribuições pontuais do *Tropes 7.2.3* para proceder ao estudo do gênero entrevista radiofônica, observamos a identificação de algumas características composicionais, estilísticas e de conteúdo que, entre outras coisas, comprovaram uma relativa aproximação entre o uso vernacular e a variedade linguística usada nesse gênero discursivo. No nível **composicional**, o *software* identificou o tipo textual argumentativo como mais recorrente – sem excluir, contudo, a presença significativa dos demais tipos. No nível **temático**, foi possível elencar um conteúdo diversificado, versando sobre os muitos temas que envolvem a vida cotidiana da sociedade e que, na esfera jornalística, são resgatados e transformados em notícias por serem de interesse do ouvinte (por exemplo, questões sociais, geográficas, culturais, políticas, etc.).

Finalmente, no nível **estilístico**, foi possível observar as estruturas verbais e temporais mais recorrentes nos enunciados do gênero entrevista radiofônica. De maneira que os verbos mais frequentes – do subgrupo factivo e estativo – favorecem uma encenação mais dinâmica e os marcadores temporais assinalam a temporalidade instaurada nos textos, isso é, uma concepção marcada, como vimos, pela proximidade com o momento de enunciação, pela expressão de frequência, pela fragmentação do dia e pela indicação de posterioridade e anterioridade. Ainda foi possível identificar a recorrência de formas pronominais de primeira e segunda pessoas do singular – reflexo evidente da situação discursiva instaurada nesse gênero discursivo, na qual duas pessoas estão em diálogo direto.

Referências bibliográficas

ARAUJO, L. S. **Os valores atribuídos ao pretérito perfecto compuesto espanhol nas regiões dialetais da argentina**. 2012. 212 f. Dissertação (Mestrado em linguística e língua portuguesa) - Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara, 2012a.

ARAUJO, L. S. **O pretérito em espanhol**: usos e valores do perfecto compuesto nas regiões dialetais argentinas. São Paulo: Cultura Acadêmica, 2013.

BAKHTIN, M. **Estética da Criação Verbal**. São Paulo: Martins Fontes, 1997.

BERBER SARDINHA, T. Lingüística de *corpus*: histórico e problemática. **D.E.L.T.A.** São Paulo, v. 16, p.323-367, 2000. Disponível em: <http://www.scielo.br/pdf/delta/v16n2/a05v16n2.pdf>. Acesso em: 04 abr. 2016.

FARNEDA, E. S. Perguntas e Respostas na Entrevista Radiofônica. **Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Lingüística e Literatura Letra Magna**, São Paulo, v. 6, p.1-18, 2007. Disponível em: <http://www.letramagna.com/entrevistaradio.pdf>. Acesso em 01 abr. 2015.

LABOV, W. **Padrões Sociolinguísticos**. Trad. Marcos Bagno. São Paulo: Parábola, 2008.

LIPSKI, J. M. **El español de América**. 7 ed. Trad. Silvia Iglesias Recuero. Madrid: Cátedra, 2011.

MARCHEZAN, R. C. Diálogo. In: BRAIT, E. (Org.). **Bakhtin**: Outros conceitos-chave. São Paulo: Contexto. 2006. p.115-131.

MARCUSCHI, L. A. **Produção textual, análise de gêneros e compreensão**. São Paulo: Parábola, 2008.

PÉREZ COTTEN, M.; TELLO, N. **La entrevista radial**. Buenos Aires: La Crujía, 2004.

PÉREZ HERNÁNDEZ, M. C. Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. **Estudios de Lingüística del Español**. Barcelona, v. 18, p.1 –75, 2002. Disponível em: <http://elies.rediris.es/elies18/>. Acesso em 31 mar. 2016.

PORTUGAL, M.; YUDCHAK, H. **Hacer radio**: guía integral. Buenos Aires: Galerna, 2008.

SANCHEZ, A. (Org.). **CUMBRE – Corpus Lingüístico del Español Contemporáneo – Fundamentos, Metodología, y Aplicaciones**. Madrid: SGEL, 1995.

SILVEIRA, M. I. M. **Análise de Gênero Textual**: Concepção Sócio-Retórica. Maceió: EDUFAL, 2005.

Artigo recebido em: 30.10.2016

Artigo aprovado em: 30.01.2017

Interação em rede social: das reações às características do gênero comentário

Interaction on social networks: from reactions to the characteristics of comments

Roberlei Alves Bertucci*
Paula Ávila Nunes**

RESUMO: Este artigo tem como objetivo analisar a relação entre o gênero comentário em uma página no Facebook e reações de usuários nesse contexto. Consideramos que a navegação e a leitura que os usuários fazem na rede devem ser relacionadas (COSCARRELLI, 2016) e que o modo de interação nesse ambiente pode variar do mais básico (reação como “curtir”, “amar”, “rir” etc.), passando pelo compartilhamento até o comentário, possivelmente o nível mais profundo de interação dos usuários com uma publicação. Se estes participam da rede por questões de afinidade (RECUERO, 2012) e fazem uso de seu excedente cognitivo para criar uma cultura de participação nesses espaços (SHIRKY, 2011), pressupõe-se que determinados temas sejam mais convidativos à interação, o que se tenta verificar pelo conjunto das reações analisadas. Espera-se que, por meio delas e de sua relação com o texto a que se referem, seja possível prever as características do comentário realizado e do tipo de texto comentado. Assim, a partir da extração de dados pelo aplicativo Netvizz de uma página específica na rede, recortaremos nosso objeto às reações “odiar”, levantando a hipótese de que essa reação deve indicar um tópico polêmico ou de insatisfação do grupo.

PALAVRAS-CHAVE: Gênero textual. Facebook. Netvizz. Comentários. Reações.

ABSTRACT: This paper aims to analyze the relationship between users' reactions and comments on a Facebook page. We consider that actions performed by users, such as reading and surfing the Internet, must be related (COSCARRELLI, 2016), and that the way interaction takes place in this environment can vary from a basic level (in reactions such as “like”, “love”, “laugh” etc.), passing through sharing posts, to commenting, possibly the deepest level of interaction between users and posts. If users take part of these networks seeking affinity (RECUERO, 2012) and make use of their cognitive surplus to create a culture of participation in these spaces (SHIRKY, 2011), we expect that certain subjects are more inviting to interaction, an idea which we seek to verify using the group of reactions analyzed. We expect that these reactions, as well as their relationship with the original post they refer to, will allow us to predict certain characteristics of the comment made and of the type of text being commented. Thus, based on the data extraction made by the Netvizz app on a Facebook page, we focus on the “hate” (or “Grr”) reaction, hypothesizing that it may indicate either a controversial subject or group dissatisfaction.

KEYWORDS: Genre. Facebook. Netvizz. Comments. Reactions.

* Doutor em Linguística pela USP. Professor da UTFPR (Curitiba). bertucci@utfpr.edu.br

** Doutora em Linguística pela UFRGS. Professora da UTFPR (Curitiba). paulanunes@utfpr.edu.br

1. Introdução

Na história humana, compartilhar e avaliar experiências sempre foram ações fundamentais, inclusive para a sobrevivência da espécie. Hoje, em tempos em que tal sobrevivência já foi assegurada por nossos antepassados, elas, acima de tudo, contribuem para promover mudanças importantes na sociedade, tendo impacto nos tipos de tecnologia que, dialogicamente, também influenciam no compartilhamento e avaliação dos fenômenos da vida humana. É o caso, por exemplo, das tecnologias digitais, as quais, entre outras coisas, fomentaram uma participação ainda maior das pessoas na produção de conteúdo disponibilizado livremente na grande rede, verificada também na publicação de suas atividades e visões pessoais. Ainda que aparentemente trivial sob o olhar do século XXI, isso faz, como afirma Shirky (2011), com que “os usos sociais de nossos novos mecanismos de mídia [sejam] uma grande surpresa, em parte porque a possibilidade desses usos não estava implícita nos próprios mecanismos” (p. 18). Entretanto, como ainda ressalta o autor, “o uso de uma tecnologia é muito pouco determinado pelo próprio instrumento; quando usamos uma rede, a maior vantagem que temos é acessar uns aos outros” (idem). O “acesso” a outras pessoas, aparentemente tão comum quando falamos nos tempos atuais, em que redes sociais estão tão em voga, não pode, contudo, ser visto com ineditismo:

A atomização da vida social no século XX deixou-nos tão afastados da cultura participativa que, agora que ela voltou a existir, precisamos da expressão “cultura participativa” para descrevê-la. Antes do século XX, realmente não tínhamos nenhuma expressão para cultura participativa; na verdade, isso teria sido uma espécie de tautologia. Uma fatia expressiva da cultura era participativa – encontros locais, eventos, performances – porque de onde mais poderia vir a cultura? *O simples ato de criar algo com outras pessoas em mente e então compartilhá-lo com elas representa, no mínimo, um eco daquele antigo modelo de cultura, agora em roupagem tecnológica* (p. 23, grifos nossos).

Nesse contexto, é possível, pois, depreender dois aspectos essenciais: publicar, compartilhar e reagir a conteúdos não são práticas *mais* presentes em nossas vidas. Elas apenas se dão por meio de tecnologias tão pervasivas que nos dão essa sensação de ubiquidade. O segundo ponto, igualmente importante, é que, como se sabe, a linguagem exerce um papel decisivo na produção de cultura e, evidentemente, não é o simples artefato tecnológico que tiraria dela esse lugar central, até mesmo por ser a linguagem requisito fundamental para a criação das tecnologias produzidas pelo homem (VARGAS, 2009;

XAVIER, 2013) e condição *sine qua non* para seu aprimoramento, sendo ambas, linguagem e tecnologia, o que diferencia o homem como ser social do homem como ser biológico (PINTO, 2005).

Tratar de linguagem em contextos da chamada “era digital”, portanto, é imperativo em nossa constante tentativa, como linguistas, de entender o modo pelo qual a linguagem se coloca como fundamento antropológico principal, sobretudo em sua relação com a tecnologia, uma vez que “o verdadeiro cerne de nossa cultura [...] é sua ciência, arte e tecnologia, a soma total das conquistas, invenções e descobertas que definem nossa ideia de ‘civilização’” (GEERTZ, 2008). Ademais, a profícua relação entre linguagem e tecnologia – atemporal, mas tornada mais perceptível pelas atuais instâncias de interação mediada por computadores, sobretudo na massiva quantidade de interações proporcionadas pelas redes sociais – tem proporcionado mudanças significativas na sociedade, especialmente quando se considera a relação de mediação da segunda em relação à primeira, materializada em alguns problemas que os linguistas (e outros estudiosos) têm tratado de investigar: do tipo de texto ao modo de escrever; da circulação ao “choque” com o leitor (BARTON; LEE, 2015; JENKINS et al., 2014; LEMOS, 2015; LIMA, 2012; SOARES, 2002; XAVIER, 2013).

Nessa conjuntura, a tecnologia, ao criar um novo espaço para as ações humanas, o “ciberespaço”, formata uma era que tem contribuído para a circulação de um novo formato de texto: o hipertexto. Assumindo a proposta de Xavier (2013, p. 153), de que o hipertexto é “um dispositivo ‘textual’ semiolinguístico”, consideraremos uma postagem no Facebook como um tipo de hipertexto, no qual o sujeito navegador pode traçar um caminho próprio, a partir das possibilidades existentes entre leitura e navegação nesse ambiente (COSCARELLI, 2016).

Nesse sentido, é relevante também tratarmos da emergência de gêneros textuais no ambiente digital. Enquanto alguns deles parecem ser típicos do meio digital (memes e gifs, por exemplo), outros ocorrem em muitos outros ambientes, mas parecem ter ganhando mais frequência ali (tais como os comentários e as notícias), provocando uma ação discursiva igualmente mais intensa dos membros da rede (SANTOS; ALVES FILHO, 2014). Com isso, a análise dos gêneros midiáticos está intimamente ligada às formas de letramento, por oferecer uma reflexão sobre as práticas sociais efetivas dos indivíduos em uma sociedade (SOARES, 2002). Isso decorre do fato de toda a tecnologia provocar mudanças importantíssimas em nossas práticas comunicativas (POE, 2012), e não diferentemente ocorre

no meio digital, com implicações, por exemplo, na relação entre escritor, leitor e texto, conforme apontam diferentes autores (SOARES, 2002; XAVIER, 2013).

Dessa forma, o presente artigo tem o objetivo de investigar se algumas dessas práticas linguísticas, materializadas em gêneros específicos – ainda que não restritos ao ambiente da grande rede –, podem ser descritas por meio de uma ferramenta computacional. Mais especificamente, queremos investigar se as reações de usuários do Facebook podem indicar alguma característica da publicação à qual reagiram. Para isso, escolhemos, como recurso de análise, a ferramenta Netvizz, um aplicativo utilizado para coleta de dados especificamente nessa rede social, e a página Lítera Brasil, disponível na mesma rede, e cujos motivos de escolha são delineados mais adiante neste escrito. Nossa investigação, então, é basicamente a seguinte: as reações dos usuários a uma publicação podem indicar as características linguísticas (composicionais) do texto publicado? Mais do que isso: essas reações podem antecipar o tipo de comentário a ser feito nesse ambiente, de modo a se prever se ele será mais próximo de um comentário argumentativo ou não? A partir de uma análise preliminar, nossa hipótese é de que algumas reações da rede podem indicar pontos relevantes na análise de gêneros no ambiente hipertextual do Facebook. De forma mais pontual, nossa hipótese é de que comentários argumentativos e avaliativos¹ são mais frequentes em textos de alto índice de reação “odiar”.

Esse tipo de pesquisa torna-se relevante especialmente quando se trata de mapear as características dos gêneros textuais com o uso ferramentas computacionais. Se, por um lado, os pesquisadores sempre se questionam sobre o que torna um gênero textual específico e diferente de outros gêneros, na outra ponta surgem cada vez mais ferramentas e aplicativos que tendem a lidar com grande número de dados e que podem auxiliar nessa diferenciação. Assim, quanto mais conseguirmos cruzar os dados relativos a diferentes textos (e reações), melhor deverá ser nosso entendimento sobre as questões relativas ao gênero.

Para cumprir o objetivo aqui proposto, o artigo está organizado da seguinte maneira: na seção 2, apresentamos os pressupostos teóricos de nossa pesquisa; na terceira parte, os

¹ Vamos considerar, *grosso modo*, “argumentativo” como uma característica do texto que apresenta algum tipo de argumento a respeito de um tema, apresentando, assim, maior recorrência do tipo textual argumentativo, conforme tipologia de Marcuschi (2002); os avaliativos, por seu turno, são aqueles que contêm uma avaliação sobre um fato, mas sem necessariamente apresentar um argumento para isso. Essas duas categorias se distinguem daqueles comentários que apresentam, por exemplo, o nome (link) de um usuário, em que o comentário serve, na verdade, para indicar a publicação para a pessoa referida, mediante citação de seu nome nos comentários.

detalhes da metodologia; na parte 4, divulgamos os resultados a que chegamos e, na última seção, as considerações finais.

2. Pressupostos teóricos

2.1. Tecnologia

O desenvolvimento de técnicas incorporadas ao cotidiano do homem foi uma constante ao longo de nossa história. Além das necessidades da realidade imediata, que apareciam e exigiam novos instrumentos (como a agricultura, por exemplo), esse desenvolvimento parece ser parte da capacidade humana de sempre buscar aperfeiçoar aquilo que já existe, ou, em outros termos, são criações tributárias da inigualável capacidade humana de projetar (PINTO, 2005). Constatação semelhante também é feita por Vargas (2009, p. 9), ao afirmar que

(...) a técnica não se resume à invenção e uso de um instrumento. Ela tem a característica marcante de que, uma vez inventado o primeiro instrumento, desencadeia-se um processo de melhoria de suas formas e usos para satisfazer necessidades crescentes da humanidade.

Ainda que a linguagem, por si só, não possa ser considerada como uma tecnologia, dado que é natural e inata ao ser humano², ela sempre teve papel fundamental no desenvolvimento das técnicas pelo homem, por permitir que fossem feitas relações por diferentes sujeitos em diferentes tempos, sobretudo uma vez criado o aparato simbólico da escrita, nossa primeira máquina do tempo (GNANADESIKAN, 2009), a qual nos permitiu suspender a linearidade temporal e plotar uma língua em um papel, traduzindo o tempo (da fala) no espaço (da escrita)³. A capacidade de descrever e analisar o mundo – sempre existente, mas facultada especialmente pelo salto simbólico necessário à aquisição da escrita – , aliada a uma inerente criação imaginativa, dá ao homem o poder de “progredir” em suas

² Esse, na verdade, não é exatamente ponto pacífico entre os pesquisadores da filosofia da técnica. Entretanto, subscrevemos a ideias de autores como Pinto (2005), Vargas (2009) e Poe (2012), para quem a tecnologia é uma produção *deliberada* por parte do ser humano, sendo fruto de sua capacidade de criação e raciocínio. Portanto, a linguagem, concebida como a capacidade humana inata de simbolizar, não pode, acreditamos, ser tratada de tal forma, já que não é uma *criação*.

³ É interessante a análise realizada por Marshall Poe (2012) quanto à invenção da escrita e os novos atributos que tal criação deu à interação humana, especialmente quando a comparamos à fala, sendo talvez os mais expressivos a privacidade, o alcance e a persistência da informação.

técnicas de modo único na natureza. O desenvolvimento científico, por consequência, contribui para o avanço tecnológico. As inúmeras descobertas feitas possibilitaram o desenvolvimento de diferentes ferramentas em diferentes áreas, como a agricultura, a saúde e, claro, as comunicações. Marshall Poe (2012), inclusive, dedica todo seu livro, *A history of communications*, às grandes criações tecnológicas que afetaram a forma como nos comunicamos hoje, desde a escrita manuscrita até a Internet, sendo esta última bastante especial por parecer englobar todos os outros tipos de mídia precedentes (verbal – seja oral ou escrito – e audiovisual).

A relação do homem com a tecnologia, no entanto, sempre foi ambígua: se, de um lado, é inegável que a ela é capaz de melhorar a qualidade de vida dos seres humanos (GRINSPUN, 2009; MARTÍNEZ, 2015), de outro, é igualmente inegável que há consequências no uso dessas tecnologias que podem prejudicar a qualidade de vida das pessoas (MARTÍNEZ, 2015), como ocorre com a excessiva produção de lixo nos dias de hoje (além do lixo tóxico de equipamentos tecnológicos, mas, até mesmo, o chamado *lixo cultural*),⁴ ou mesmo com o aumento de níveis de ansiedade, provocado pelo excesso de interação possibilitado pelas mídias como um todo, mas potencializado pelos meios digitais⁵.

Ainda que haja uma série de pesquisas relacionadas à influência na qualidade de vida das pessoas que utilizam redes sociais,⁶ tanto estas quanto seus usuários parecem crescer a cada dia: em junho de 2016, eram 1,65 bilhões de pessoas no Facebook, sendo quase 100 milhões de usuários apenas no Brasil,⁷ contra 800 milhões em 2011 (RECUERO, 2012, p. 15). Isso faz com que não se possa ignorar o impacto que tais redes têm na vida das pessoas, nem em seus modos de produção e circulação de conteúdos. De modo mais específico, essas

⁴ Sobre isso, ver, por exemplo, o trabalho de Wolton (2012).

⁵ Ver, por exemplo, a asserção de Shirky (2011), que já notava tal fato relativamente à televisão: “[...] os economistas comportamentais [...] concluem não apenas que pessoas infelizes assistem consideravelmente mais televisão do que pessoas felizes, mas, além disso, que ver TV também afasta outras atividades que provocam menos interesse imediato, mas que podem produzir maior satisfação a longo prazo. Passar várias horas vendo TV, por outro lado, associa-se a maiores aspirações materiais e a um *aumento de ansiedade*” (p. 11, grifos nossos).

⁶ Ver, por exemplo, as matérias “Facebook entristece as pessoas, diz estudo”, publicada no jornal Bem Paraná, em 23 de novembro de 2016, disponível em: <http://www.bemparana.com.br/noticia/476008/facebook-entristece-as-pessoas-diz-estudo>, acessada em 23. nov. 2016 e “Se você quer ser feliz, saia do Facebook”, publicada em 25 de novembro pelo El País (Brasil), disponível em http://brasil.elpais.com/brasil/2016/11/23/tecnologia/1479897698_896068.html?id_externo_rsoc=Fb_BR_CM, acessada em 25. nov. 2016.

⁷ Conferir reportagem do portal G1: <http://g1.globo.com/tecnologia/noticia/2016/06/instagram-ultrapassa-os-500-milhoes-de-usuarios.html> Acesso em 23. nov. 2016.

redes modificam algumas relações sociais (criam grupos de interesse, por exemplo), modificam o modo como os seres humanos se apropriam das tecnologias (um telefone, meio de comunicação, pode gravar algo para uma denúncia ser feita em uma rede social, por exemplo), além de apresentar algumas especificidades linguísticas (o modo de escrita nas redes pode diferir de outros lugares de escrita, por exemplo).

Isso tudo justifica o trabalho que aqui se apresenta, uma vez que a forma como se dão as práticas languageiras nesse ambiente é diferente da conversa presencial. Assim, se, como afirma Recuero (2012, p. 28), “num diálogo, tudo é informação: elementos prosódicos (como o tom da voz, a entonação e as pausas da fala), elementos gestuais e, evidentemente, as palavras”, no ambiente virtual todos os outros sinais semiotizados precisam ser lidos, a fim de que o leitor construa os sentidos ali presentes. Nesse sentido, a navegação e interação no meio digital, pela complexidade da convergência de diferentes semioses, são diferentes também da interação puramente escrita. Essas questões abrem espaço para uma pesquisa ampla, que pode envolver as reações numa publicação no Facebook, a forma de pontuação que ocorre em textos ali escritos, a construção de identidades no espaço virtual, ou mesmo a utilização de uma ferramenta de apreciação de um comentário, tema de nosso estudo aqui. Tudo em busca de compreender melhor como se dão as práticas e trocas sociais ali estabelecidas por meio da linguagem.

Tendo isso em vista, analisamos, neste trabalho, características de algumas publicações da rede social Facebook por meio da ferramenta Netvizz. Por ela, tencionamos verificar se é possível prever o tipo de comentário feito em algumas publicações da rede, analisadas aqui em uma perspectiva hipertextual.

2.2. Texto e hipertexto

A partir da perspectiva sociointeracionista, assumimos que as atividades sociais ocorrem por meio da linguagem (XAVIER, 2013). Em outras palavras, entendemos, junto a outros autores, que a linguagem é elemento *sine qua non* para as atividades sociais cotidianas no ser humano. Por isso, tanto atividades simples, como fazer um pedido, quanto complexas, como escrever uma tese, são feitas por meio da linguagem, o que sempre produz uma ação.

Dessa forma, são igualmente essenciais os processos de semiotização dessas ações, sejam por signos linguísticos, imagens, ícones etc. Tais processos, integrados, constituem a totalidade textual e hipertextual, de onde se constroem os gêneros capazes de fazer circular os

discursos, ou seja, suas práticas sociais intermediadas pela linguagem. Seguindo Xavier (2013, p. 139), podemos entender textos como

resultados de cruzamentos entre um conjunto de matrizes: linguístico-cognitivas (capacidade para mobilizar e processar racionalmente recursos verbais), biofísicas (condições articulatórias e motoras para produzir fonemas e grafemas sistematizados e compreensíveis) e históricas (contexto sócio-político a partir do qual uma palavra, expressão ou longo trecho deve ser compreendido). Por causa da instabilidade de tais variáveis, o texto deve ser concebido como uma ação de um sujeito em processo e não como um produto finalizado.

Nesse sentido, cada prática social feita linguisticamente relaciona a ação do falante e do ouvinte: tudo contribui para a construção do processo, ainda que nunca se finalize o produto ao qual chamamos de “texto”. Nessa relação, outro texto pode surgir em “reação” ao primeiro, igualmente incompleto, mas que concretiza a noção de prática social da atividade linguística como uma troca entre indivíduos.

De forma complementar, o hipertexto tem ampliadas suas condições de produção e seus percursos de leitura. Assumindo a proposta de Xavier (2013, p. 153), definimos “o hipertexto como um dispositivo ‘textual’ digital semiolinguístico (dotado de elementos verbais, imagéticos e sonoros) on-line, isto é, indexado à Internet”. Em outras palavras, o hipertexto deverá ser realizado sempre em meio digital e aglutinar diferentes linguagens.

Ao integrar elementos multissemióticos, acaba por permitir ao leitor um caminho de leitura próprio, o que também deve influenciar no modo como as pessoas reagem a ele. Em outras palavras, se, no texto “padrão”, o autor delineava seu texto e esperava do leitor que construísse um caminho, no hipertexto, o autor oferece caminhos diversos ao leitor, que pode, por sua vez, se deparar com outros hipertextos e outros caminhos na navegação. Por isso, mais do que um texto, o hipertexto mostra-se como um ambiente específico de navegação.

Para Soares (2002), a necessidade de reflexão sobre essas novas práticas sociais está diretamente ligada às formas de letramento disponíveis na sociedade. Ainda que o presente trabalho não se debruce nos pressupostos de letramento, pode servir de base para análises desses fenômenos, já que, para a autora, a era atual é

um momento privilegiado para, na ocasião mesma em que essas novas práticas de leitura e de escrita estão sendo introduzidas, captar o estado ou condição que estão instituindo: um momento privilegiado para identificar se as práticas de leitura e de escrita digitais, o letramento na

cibercultura, conduzem a um estado ou condição diferente daquele a que conduzem as práticas de leitura e de escrita quirográficas e tipográficas, o letramento na cultura do papel (SOARES, 2002, p. 146).

Em outras palavras, se a escrita causou mudanças extremamente significativas na história humana (FISCHER, 2009; LIMA, 2012; XAVIER, 2013), qual não seriam as mudanças causadas pela tecnologia digital, que exige, cada vez mais, práticas de leitura e escrita diversificadas e aprofundadas? Acima de tudo, essas modificações parecem exigir novas formas de letramento.

Conforme Soares (2002), a cibercultura sugere que devemos falar em *letramentos*, no plural, exatamente pelas inúmeras questões colocadas nas práticas de leitura e escrita nesse ambiente, especialmente quando se observa que são delineadas por distintas tecnologias de escrita e leitura, o que, por sua vez, exige diferentes letramentos. Em outras palavras, devemos considerar, com a Linguística, que diferentes tecnologias de escrita mobilizam “diferentes espaços de escrita e diferentes mecanismos de produção, reprodução e difusão da escrita” o que resulta “em diferentes *letramentos*” (SOARES, 2002, p. 156).

Poderíamos exemplificar a identificação dessas novas práticas, apontando para uma diferença fundamental que a era digital tem provocado: a relação entre leitura e navegação. Para Coscarelli (2016), a leitura em ambiente virtual depende da perfeita relação entre esses dois pontos. De um lado, a navegação envolve o percurso a ser construído: “a navegação efetiva nos ambientes virtuais requer que os usuários saibam onde estão, onde precisam ir, como chegar lá e quando eles chegaram” (LAWLESS; SCHRADER, 2008, p. 269, apud COSCARELLI, 2016, p. 65). Isso significa, para a autora, que navegar não é algo banal.

De fato, navegar requer do usuário a capacidade de identificar e interpretar os elementos multissemióticos que permitem que ele responda aos requisitos acima. Nesse sentido, a navegação exige também um grau de leitura (e letramento) profundo do usuário: para cada elemento semiótico com o qual se depara, ele precisa interpretá-lo (lê-lo, portanto) para tomar uma decisão de navegação. De forma circular, cada navegação também exige do usuário uma leitura específica e assim por diante. Dessa forma, assumimos que é nesse movimento que os sentidos vão sendo produzidos e a coerência textual de um hipertexto vai sendo produzida. De igual forma, acreditamos que os elementos em análise neste texto, quais sejam, as reações a uma publicação, são formas não apenas de navegação, mas de leitura de um conteúdo produzido. Por envolver diferentes elementos semióticos, conforme mostramos

adiante, configura-se, portanto, como um ambiente privilegiado para investigação dos dois elementos estudados por Coscarelli. Propomos, dessa forma, que a análise da página Lítera Brasil, a partir de suas reações e comentários em suas publicações, pode contribuir para as pesquisas que levam em conta os pressupostos apontados acima.

2.3. Gêneros textuais

Marcuschi e Xavier (2010) sustentam que as formas de utilização de linguagem são modificadas à medida que se modificam as tecnologias associadas a elas, o que leva a um pressuposto básico de criação ou alteração de gêneros textuais à medida que novas tecnologias linguísticas surgem.⁸ Nesse sentido, a análise das interações que tomam lugar nos ambientes digitais é importante por promover uma reflexão sobre os gêneros ali presentes.

Assumimos, com Marcuschi (2010), que, bakhtinianamente, os gêneros são fenômenos sociais e históricos: primeiro, porque são fundamentais para as práticas sociais; depois, porque surgem em momentos determinados e com funções específicas. Em outras palavras, “são formas sociais de organização e expressões típicas da vida cultural”, ainda que não sejam “categorias taxionômicas para identificar realidades estanques” (MARCUSCHI, 2010, p. 19).

Se concebermos gênero como um instrumento complexo de ação social, somos levados a considerar os três pontos fundamentais da concepção de Bakhtin (1992): primeiro que ele sempre aborda um *tema* (conteúdo) num dado contexto; depois, tem uma *composição* característica (estrutura textual/gramatical); e, finalmente, apresenta o *estilo* (as opções) do autor para expressar e enunciar seu texto.

Para Marcuschi (2010), a relação íntima entre o gênero e a tecnologia pode ser vista tanto na noção bakhtiniana de gênero como algo “relativamente estável”, estilística e composicionalmente, quanto na função de instrumento comunicativo. Ora, se é assim, quaisquer interferências nas tecnologias linguísticas vão refletir diretamente tanto na “estabilidade” quanto na “instrumentalidade” dos gêneros, o que parece acontecer com o fenômeno hipertextual.

⁸ Chamamos de *tecnologias linguísticas* todas as criações humanas utilizadas para produção, compreensão e difusão de suas produções linguísticas, tais como dicionários, gramáticas ou mesmo tradutores automáticos.

Embora quase sempre multimodal, o hipertexto é essencialmente escrito, ou pelo menos exige as leituras para a construção de sentidos no percurso do leitor, conforme a relação já apresentada entre ler e navegar (COSCARELLI, 2016). Nesse sentido, consideramos que, numa página de Facebook e em suas postagens, tanto a escrita quanto as suas reações são essenciais para essa construção. Indo além da leitura, os itens que formam o aparato hipertextual e interacional de uma página nesse ambiente são passíveis de análises que podem ser essenciais para a compreensão das práticas de letramento na sociedade.

As publicações em si não devem ser analisadas como um gênero particular, já que podem ser compostas por diferentes deles: podem conter uma foto com ou sem um texto integrado (legenda ou mesmo um comentário); um vídeo com uma introdução textual sobre ele; um link entre outros. No entanto, a rede limita os tipos de materiais que podem ser publicados ali, sendo os mais comuns imagem, vídeo, link e documentos de texto. Um determinado usuário que navega pela rede faz escolhas importantes, que revelam o seu grau de envolvimento e letramento com tais publicações. Ao se deparar com uma publicação, o usuário a lê e decide como reagir a ela: pode simplesmente ignorá-la, ou utilizar umas das opções sugeridas pela rede (curtir, amar etc.). No entanto, pode ir além: compartilhar aquela publicação, por considerá-la relevante em algum grau, ou mesmo comentá-la.

Vamos considerar, aqui, que o comentário é o grau mais importante de envolvimento de um usuário, porque exige, além de uma navegação e leitura atentas, disposição e capacidade de contribuir para a ampliação da publicação. É um outro texto, que revela o percurso construído pelo leitor e que soma outros sentidos ao primeiro. Por isso, os comentários são o grau mais intenso de interação na tríade autor, texto, leitor numa rede como o Facebook, quando se considera a modalidade escrita.

De algum modo, portanto, neste artigo, cruzamos os dados da interação mais básica (reação) com os da interação mais profunda (comentários), a fim de verificar se isso nos dá previsões sobre o gênero da publicação e/ou sobre os tipos de comentários a serem encontrados ali. Na relação com os gêneros textuais, tentamos cruzar o índice de determinadas reações à composição dos possíveis comentários que uma postagem pode receber. Mais especificamente, queremos relacionar o índice de reações “odiar” a comentários de teor avaliativo ou argumentativo na postagem.

Os usuários da rede estão aos poucos utilizando outras reações além de “curtir”, a primeira e mais utilizada delas. Por isso, além de descrevermos a frequência de “odiar”,

parece interessante, igualmente, relacionar essas reações a possíveis publicações e comentários que possam ser feitos. Intuitivamente, tomamos a reação “odiar” como algo bastante relevante na rede porque, partindo do princípio de que as redes sociais são espaços de interação e formação de grupos de afinidade, bem como de participação de eventos (ou páginas) de interesse (RECUERO, 2012), podemos assumir que essa reação esteja associada a postagens que ou apresentam um fato que desagradava uma comunidade em geral (aqueles que curtem e/ou seguem a página), ou um fato polêmico. Dessa forma, esperamos encontrar, nos comentários da respectiva postagem, argumentos que revelem pelo menos uma das características da postagem apontadas acima (fato desagradável ou polêmico).

De forma semelhante a Santos e Alves Filho (2014), consideramos que o gênero comentário é um tipo de resposta a uma publicação, constituindo-se de uma “réplica” tanto à própria publicação como a outros comentários. Nesse sentido, sua composição deve ter uma referência àquilo que foi postado, direta ou indiretamente, com modos de referência, por exemplo, enquanto seu estilo segue a subjetividade dos sujeitos que ali interagem.

Como temos dito, tentamos verificar, basicamente, de que forma os comentários retomam o tema da publicação e trazem um argumento com relação a ele. Por isso, não vamos delinear as estratégias de referência e argumentação, que estão fora do escopo deste trabalho.

3. Metodologia

Barton e Lee (2015, p. 37) afirmam que o ambiente *online* apresenta novos métodos de pesquisa da linguagem: “investigar textos e práticas *online* proporciona novas possibilidades de metodologia de pesquisa linguística”. A quantidade de dados que a internet proporciona, e de forma gratuita, aliada a estudos multidisciplinares, como os de mídia, podem ser fundamentais para se investigar a linguagem como um todo (e não necessariamente apenas a *online*). Outro ponto importante é que, na rede, os dados são gerados fora de um ambiente de controle de pesquisa (ainda que o possa ser com relação à escrita), o que pode aparentar mais “naturalidade” em sua produção.

Lançado em 2004, nos Estados Unidos, o Facebook reúne uma série de funcionalidades que envolvem diferentes práticas sociais: marcação, convite e confirmação para eventos, formação de grupos específicos de interesse, atualizações de *status* (formato de *microblogging*), comentários sobre as postagens, transmissão ao vivo (mais recente), entre

outros. Com relação aos comentários, Barton e Lee (2015, p. 59) afirmam que “por vezes, [o recurso de comentário] age como um *site* para mini fóruns de discussão”.

Além dos comentários, uma postagem pode gerar uma série de reações, a saber: curtir (👍), amar (❤️), rir/haha (😂), admirar/wow (😲), chorar (😭) e odiar/grr (😡). Vamos assumir que, com exceção de “curtir” e “admirar”, as outras estão organizadas em pares de antônimos: amar *vs.* odiar e rir *vs.* chorar. Neste artigo, trabalhamos, pelos motivos já citados, com o segundo elemento do primeiro par: “odiar”.

Para restringir a quantidade de dados, vamos comparar as publicações com maiores índices de “odiar”, em relação às demais. Nossa intenção é dupla: primeiro, verificar o que esse tipo de reação pode indicar com relação ao tipo de texto publicado; depois, se os comentários feitos nessa publicação reforçam uma análise negativa por parte dos usuários.

Para a coleta de dados, utilizamos o aplicativo Netvizz, disponível na própria rede social Facebook.⁹ O aplicativo extrai dados de páginas e grupos que podem servir para pesquisas de diferentes áreas. Depois de extraídos, os dados precisam ser consolidados em uma planilha específica, ou mesmo analisados por softwares.¹⁰

Para a presente pesquisa, escolhemos a página Lítera Brasil, que divulga postagens sobre Literatura, Artes e Cultura pop no Facebook, atualmente com mais de 66 mil curtidas, e administrada por dois professores formados em Letras.¹¹ Em sua descrição, afirma-se o objetivo de “divulgar grandes nomes da Literatura nacional e universal, bem como auxiliar os vestibulandos de todo país na leitura e estudo das Obras e das Escolas Literárias”. Entretanto, na rolagem de suas publicações, encontram-se muitas publicações de cunho humorístico (alguns “memes”, por exemplo) e algumas notícias sobre a área que aborda. Exatamente por isso é que a página foi escolhida: por agregar um conteúdo leve e descontraído sobre obras, personagens e autores literários, o que poderia indicar uma baixa ocorrência de textos com polêmica ou de desagrado dos seguidores. Ou seja, esperamos, ali, encontrar um baixo índice de reações “odiar” em suas publicações.

A figura a seguir apresenta a página inicial do Lítera Brasil.

⁹ Disponível em: <<https://apps.facebook.com/Netvizz>> Acesso em 24. nov. 2016.

¹⁰ Por questões de espaço, vamos deixar o detalhamento da utilização do Netvizz para o uso em trabalhos de Linguística para outro artigo, o qual já está em preparação. No entanto, indicamos Freire (2015) para mais detalhes gerais sobre o aplicativo.

¹¹ Disponível em: <<https://www.facebook.com/LiteraBrasil/>> Acesso em 24. nov. 2016.

Figura 1 – Página inicial do Lítera Brasil.



Por razões de tempo e espaço, restringimos o período de extração de dados entre 05 e 11 de novembro de 2016, com um total de 601 publicações, gerando mais de 270 mil reações, mais de 20 mil comentários e mais de 110 mil compartilhamentos. As variáveis que tentamos investigar foram o número de reações “odiar” no total de publicações e em publicações específicas (aquelas de alta ocorrência). A outra variável foi a presença de indicativos de posição argumentativa nos comentários referentes às publicações com esse tipo de reação. Os resultados são analisados na seção a seguir.

4. Resultados

A extração de dados entre 05 a 11 de novembro da página Lítera Brasil está presente no Quadro 1, a seguir. Percebe-se que os números são bastante relevantes para o cruzamento de dados que desejamos fazer neste trabalho, ou seja, entre o tipo de reação e o teor argumentativo dos comentários.

Quadro 1 — Números gerais do Lítera Brasil (05 a 11 de novembro de 2016).

Item	Números
Publicações	601
Comentários	20.808
Reações	275.981
Compartilhamentos	113.223
Engajamentos	410.012

É importante dizer que o item “engajamento” refere-se ao envolvimento dos usuários nas publicações da página, seja por meio de reações, comentários ou compartilhamentos, por

exemplo.¹² Em outras palavras, é a soma dos níveis de interação do leitor com o texto: comentários, reações e compartilhamentos que houve no período analisado.

A tabela 2, a seguir, descreve o tipo e o número de publicações, relacionando isso ao tipo e ao número de comentários e reações. Acrescentamos, também, a porcentagem desses números em relação aos totais apresentados no Quadro 1.

Tabela 1 — Tipo e quantidade de publicações.







Publicação	Quantidade	Comentários	Reações	Compartilhamentos	Engajamentos
Vídeo	34/601 (5,6%)	2.222/20.808 (10,7%)	9.844/275.981 (3,6%)	8.322/113.223 (7,3%)	20.388/410.012 (5%)
Link	64/601 (10,6%)	727/20.808 (3,5%)	10.317/275.981 (3,7%)	4.467/113.223 (4%)	15.511/410.012 (3,8%)
Status	22/601 (3,7%)	78/20.808 (0,4%)	1.663/275.981 (0,6%)	241/113.223 (0,2%)	1.982/410.012 (0,5%)
Evento	1/601 (0,1%)	0/20.808 (0%)	1/275.981 (0%)	0/113.223 (0%)	1/410.012 (0%)
Foto	480/601 (80%)	17.781/20.808 (85,4%)	254.156/275.981 (92,1%)	100.193/113.223 (88,5%)	372.130/410.012 (90,7%)

Os dados revelam que o tipo de publicação de maior impacto nessa página, no período em questão, são as imagens (fotos): ocupam 80% do total de publicações da página, mas têm um índice de 90,7% de todo o engajamento que feito ali. Na outra ponta, está o evento, com apenas uma publicação e 0% de engajamento (houve apenas uma reação). Outro ponto que deve ser levado em conta é a diferença entre vídeos e links: enquanto estes têm mais de 10% do total de publicações, contra 5,6% daqueles, os comentários são muito mais frequentes nos vídeos do que nos links; também se nota diferença no índice de engajamento, sendo 5% do total para os vídeos, contra 3,8% dos links. Ou seja, mesmo com praticamente a metade de publicações, os vídeos geram mais impacto na página Lítera Brasil. No total, portanto, percebe-se que imagens e vídeos são aqueles que mais geram engajamentos na página, somando 85% do total de publicações e quase 96% do total de engajamentos. Esse dado revela a importância de uma análise que leve em conta a relação entre reações e comentários, o que coloca em perspectiva as hipóteses que temos lançado ao longo do texto.

¹² Mais informações em: < <http://trends.rmacomunicacao.com.br/visibilidade-engajamento-e-reputacao-como-definir-as-metricas-que-interessam-para-a-sua-marca-nas-redes-sociais>> Acesso em 24. nov. 2016.

O quadro seguinte compara os tipos de publicações aos tipos de reações. É daqui, basicamente, que recortamos os dados para a análise dos comentários. É importante destacar que há uma diferença na soma das reações do Quadro 3 com relação ao total no Quadro 2. Isso ocorre porque o Facebook permite que se desfça uma reação. Dessa forma, enquanto o Quadro em 2 apresenta o total de reações da publicação (feitas e desfeitas), o Quadro 3 apresenta apenas aquelas que foram mantidas.

Tabela 2 — Comparação entre os tipos de publicações e os tipos de reações.

Reação							Total
	Curtir	Amar	Rir	Admirar	Chorar	Odiar	
Publicação							
Vídeo	7.275	2.405	23	35	0	0	9.738
Link	8.009	1.309	211	207	142	309	10.187
Status	1.328	171	141	1	0	0	1.641
Evento	1	0	0	0	0	0	1
Foto	202.613	24.363	17.531	1.484	4.044	715	250.750
Total	219.226	28.248	17.906	1.727	4.186	1.024	272.317
% do total	80,5%	10,4%	6,6%	0,6%	1,5%	0,4%	100%

A Tabela 2 nos mostra que a reação mais utilizada nas publicações foi a “curtir”, com mais de 80% de cliques¹³. Por outro lado, a menos utilizada foi a reação “odiar”, foco de nosso trabalho.¹⁴ Nessa reação, o dado mais surpreendente é que, com exceção das fotos, cujos dados são majoritários em todas as reações, “odiar” ocorreu com bastante frequência com os links publicados na página, sendo a terceira reação mais clicada; em outras publicações, “rir” (ou “admirar”) ocupavam esse posto. Nossa intenção, agora, é relacionar algumas dessas publicações e alguns comentários para verificar se o que dissemos

¹³ Além do fato de que apenas recentemente o Facebook disponibilizou outras opções que não apenas “curtir”, muito em função de postagens relacionadas à divulgação de eventos tido como tristes, como a morte de alguém, por exemplo, também levantamos a hipótese de que “curtir” ainda é a reação mais realizada por exigir menos esforço por parte do usuário. Enquanto, para “curtir”, é necessário apenas clicar no botão correspondente, quem quiser realizar qualquer outro tipo de apreciação terá que mover o cursor por sobre o botão, esperar o aparecimento das outras reações e, só então, clicar em alguma delas.

¹⁴ Vale ressaltar que a reação com menor número foi “Agradecer”, com 11 cliques, mas, como não está mais presente nas opções do Facebook, não a contabilizamos.

anteriormente a respeito da relação direta entre essa reação e fatos desagradáveis ou polêmicos pode ser verificada.

Primeiramente, precisamos destacar que, na página analisada, 5 das 10 publicações mais “odiadas” estavam relacionadas a votações divulgadas pela página. A mais votada, com 360 “odiar”, por exemplo, fazia parte da enquete sobre o melhor personagem masculino da literatura brasileira. Cada personagem ali apresentado estava relacionado com um tipo de reação, que seria equivalente ao voto do internauta. Por isso, o número de 360 reações “odiar”, na verdade, é a quantidade de votos dados a Brás Cubas, que está em terceiro lugar nas “Olimpíadas” da página, ficando atrás de Bentinho (383 “amar”) e João Grilo (618 “chorar”).¹⁵

Sendo assim, consideramos pertinente retirar da análise as publicações com “odiar” que fizessem parte da enquete, já que era uma votação e não uma reação à publicação como em outros casos.

A segunda publicação com o maior número de reações “odiar” é um link da revista Exame, que trata da reescrita da obra de Machado de Assis. Sua entrada era: “Machado simplificado: a notícia é velha, mas nossa indignação, eterna.” A figura a seguir apresenta a referida publicação.¹⁶

¹⁵ Disponível em: < <https://www.facebook.com/349691378526332/posts/590917534403714>> Acesso em 24. nov. 2016.

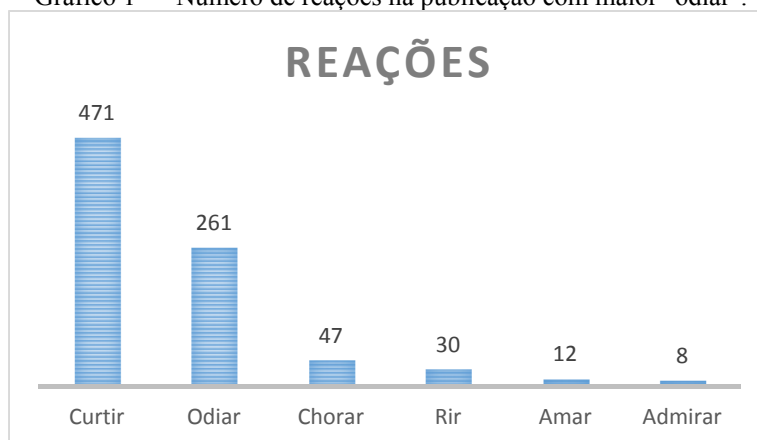
¹⁶ Disponível em: <<https://www.facebook.com/349691378526332/posts/630512320444235>> Acesso em 24. nov. 2016.

Figura 2 — Publicação com 261 reações "odiar".



Conforme a previsão, o fato relacionado ao alto índice de reações contrárias deve-se a uma notícia sobre a simplificação de uma obra literária, o que desagradou muitas pessoas. Nessa publicação, o número de reações para “odiar” foi bastante alto, conforme se vê no Gráfico 1, a seguir, sendo que, fora “curtir”, que é de longe a reação mais utilizada pelos usuários, “odiar” ficou bem à frente das outras.

Gráfico 1 — Número de reações na publicação com maior "odiar".



Tais dados indicam uma reação negativa dos usuários ao assunto tratado na publicação de 02 de novembro de 2016. A notícia, no link, dizia respeito a um projeto de simplificação de obras clássicas, como “O Alienista”, de Machado de Assis, e “A pata da gazela”, de José de Alencar, publicado pela revista Exame em 2014 (a publicação teria ocorrido naquele ano). Outros poderiam ser “simplificados”, caso a escritora conseguisse mais verbas. Por essa razão, Lítera Brasil introduziu o link com a expressão “notícia é velha”, a fim de “lembrar” que a indignação continua. A manchete, relacionando Machado, foi escolhida porque o referido livro fora o primeiro a ser lançado. A obra de Alencar estava prevista para ser lançada em junho. A própria reportagem da revista já deixava claro que os leitores se dividiam entre apoiar ou não a simplificação, havendo uma tendência maior para as opiniões contrárias. Em outras palavras, o tema é polêmico e, tal como prevíamos, isso pode ser mapeado pelo alto índice de reações “odiar” na publicação. O que devemos esperar, então, é que os comentários sejam compatíveis com esse sentimento.

De fato, dos 93 comentários, 76 se mostraram contrários à proposta de simplificação e apenas 17 se mostraram simpáticos à proposta, ou viam menos problemas, se garantido o conteúdo. A seguir, apresentamos alguns desses comentários.

Quadro 2 — Exemplos de comentários.

Favoráveis	Contrários
Contemporaneidade tá aí, aceita que doi menos.	uma decadência imensurável
O problema é alterar a obra. Mas segundo autora da ideia não vai se modificar, somente a linguagem.	Mds, como assim facilitar a leitura To indignada temos que ensinar os jovens a ler não readaptar a linguagem deles
Vamo torcer pra ser o primeiro de muitos. Vontade de ler os Lusíadas mas n entendo nada do q ta la	Se durante tantos anos a obra foi trabalhada do jeito que realmente é, e muitos jovens ficaram apaixonados por ela (assim como eu) qual é a necessidade de mudar agora? Pobre Machado, deve estar se revirando no caixão

Os comentários mostram argumentos de diferentes fontes: para os favoráveis, deve-se aceitar isso como consequência da contemporaneidade; como algo natural, desde que não se modifique a obra; ou como algo que contribui para o entendimento de obras mais complexas. Já para os contrários, a obra literária tem seu valor se preservada, graças às escolhas do autor; é preciso “letrar” os mais jovens (e não facilitar leituras mais complexas); ou a obra pode fascinar as pessoas, sem que seja simplificada.

A outra matéria com número alto de “odiar” está apresentada a seguir. Como a anterior, a publicação tem um link que leva a uma notícia da revista Veja, e foi apresentada assim, em 29 de setembro de 2016: “Às vezes a ladeira parece não ter fim”.

Figura 3 — Publicação com 34 reações "odiar".



Novamente, o alto índice de reação “odiar” está relacionado a um fato desagradável ou polêmico, conforme previmos. Nesse caso, os leitores do Lítera Brasil, geralmente pessoas que têm interesse e gosto pela leitura, reagiram raivosamente ao fato de alguém publicar um livro sem gostar de ler. Ao ler a matéria, percebemos que o ator diz não ter necessariamente interesse em literatura; no entanto, sua publicação também não foi nessa área, o que poderia “amenizar” a relação entre seu pouco gosto pela leitura e o lançamento de um livro. A seguir, os números relativos às reações a essa publicação.

Gráfico 2 — Número de reações na publicação com alto índice de "odiar".



Aqui, ao contrário da anterior, temos em segundo lugar a reação “rir”, que deve estar associada ao fato de os leitores classificarem tal fato como uma espécie de “brincadeira”. No

entanto, assim como a postagem anteriormente analisada, a reação “curtir” é a que fica em primeiro lugar.

Com relação aos comentários, dos 23 feitos na publicação, nenhum aceitou a postura do ator de colocar-se como um escritor que não gosta de ler. Alguns escreveram que isso é algum muito ruim, e outro relacionou a falta de hábito de leitura do ator à baixa qualidade de suas performances como ator. Cinco deles, contudo, fizeram ressalvas à matéria por ter sido veiculada pela revista *Veja*, colocando em dúvida a credibilidade da notícia, por associação a fatos anteriores de teor parecido, publicados pelo periódico. Para esses usuários, é preciso verificar com cuidado a entrevista e sua repercussão, para garantir que não houve distorção dos fatos. Um exemplo é este: “Eu não confiaria na *Veja*, eles distorceram tudo que a Mandy Candy disse na entrevista do lançamento do livro dela”.

Se a nossa previsão de que “odiar” está relacionado a publicações polêmicas, fato que gera comentários com argumentos para cada um dos lados do debate, devemos verificar se mesmo as publicações com baixo índice dessa reação corroboram a hipótese. A postagem que contabiliza 7 reações “odiar” é apresentada na Figura 4, a seguir.

Figura 4 — Publicação com 7 reações "odiar".



A publicação é de um link do G1 acerca de uma pesquisa realizada sobre leitura no Brasil. Obviamente, a pesquisa é mais complexa do que a manchete parece sugerir: de que Kéfera (youtuber) possa estar “duelando” com Machado por uma posição de escritora de referência. Como estratégia de “chamariz” para a leitura do texto, a manchete contribui para uma reação imediata (e contrária) dos leitores. Nesse caso, vale dizer que, além das 7 reações “odiar”, houve 38 “chorar”, fato que parece indicar um lamento por parte dos leitores. Dos 23 comentários, alguns apontam para possibilidade da leitura de Kéfera levar os jovens a outros livros, comentários que chamamos de “favoráveis”. Por outro lado, quem é contra utiliza

argumentos de que isso reflete a sociedade e deve seguir acontecendo algo similar em outras artes, conforme indicamos no quadro a seguir.

Quadro 3 — Exemplos de comentários.

Contrários	Favoráveis
É só um reflexo do que a sociedade está se tornando!	Toda entrada na literatura é válida
Próxima notícia do G1 vai ser: Catra empata com Beatles. Taqueopariu!	Meu primeiro pensamento: "Oi?...aaaaff" Mas aí lembrei que quando comecei a ler Harry Potter, 15 anos atrás, uma conhecida me disse que aquilo era porcaria, que eu tinha que ler Machado de Assis (vejam só!). Hoje é indiscutível o quanto Harry Potter influenciou minha geração a ingressar no universo literário. Não vamos entrar no mérito de ser bom ou ruim (isso é muito subjetivo, muito complexo). Sendo assim, não tem problema nenhum a Kefera vender milhões de exemplares. Quero mais é que essa juventude leia.. começa com a Kefera e um dia, com muita sorte, eles encontrarão Machado, Érico, Eça, Jorge, Clarice, Drummond...

Para finalizar nossa descrição e análise, apresentaremos uma das postagens com 2 reações “odiar”, que faz referência a algo semelhante (Kéfera como escritora), conforme Figura 5.

Figura 5 — Publicação com 2 reações “odiar”.



A publicação, que brinca com o fato de alguém preferir uma youtuber escritora a Machado de Assis, teve um número alto de reações (2.837) e comentários (292). Apesar do baixo número de reações “odiar”, nossa previsão é de que possa incluir comentários com traços de argumentação.

Quadro 4 — Exemplos de comentários.

Contrários	Favoráveis
faço das palavras de eduardo as minhas, minha cara pense um pouco, qual livro traz correntes historicas e criticas humanitarias/politicas? o da kefera? acho q n	Também prefiro Machado. Mas tudo bem ler Kéfera.
Temo pelo o que há de vir nas próximas gerações. Não é a toa que a sociedade atualmente anda tão frustrada	Não acho errado, ela ler um livro de Youtuber. Se as pessoas parassem de tratar o clássico como uma obrigação, mas sim como prazer, talvez os jovens leriam livros clássicos. O erro é você que ler é uma obrigação tirando o prazer de ler. Se ler um tal livro, é um bom começo.

Os comentários da publicação apresentam debates bastante acalorados sobre a validade desse tipo de leitura, em relação à leitura dos clássicos. Resumimos, no Quadro 4, alguns dos argumentos principais: quem é contrário à leitura de livros dessa natureza alega que há uma decadência cultural da juventude de hoje e que Machado de Assis consegue aglutinar muita cultura em suas obras. Por outro lado, quem é favorável defende que é melhor esse tipo de leitura do que nada, que se pode começar por ali e expandir o conhecimento literário e, por fim, que é preciso respeitar o gosto dos mais jovens.

Portanto, todos os dados aqui apresentados corroboram nossa hipótese de que há uma relação direta entre o tipo de reação e a composição dos comentários. Nesse sentido, nossa análise parece demonstrar que a relação entre dados de navegação e dados de leitura (letramento) deve ser mais explorada no ambiente das redes sociais, a fim de tentarmos descrever melhor a características dos gêneros que ali circulam.

Não menos importante é a conclusão de que, no percurso de interação com o texto (da reação ao comentário), as escolhas dos leitores ajudam os pesquisadores na descrição e análise de sua língua. Tudo isso, claro, tem sido permitido pelo desenvolvimento de aplicativos computacionais que contribuem para a extração e análise de dados nesses ambientes. Sem o Netvizz, nesse caso, nossa tarefa aqui seria praticamente impossível, já que teríamos que verificar uma a uma as mais de 600 ocorrências extraídas pelo aplicativo no período em questão.

5. Considerações finais

Neste artigo, procuramos demonstrar que parece existir uma relação entre o tipo de reação (interação mais básica) e as características do gênero comentário, em publicações no Facebook. Ao focarmos na reação “odiar”, mostramos que essa reação indica que o assunto

tratado (a publicação) tem ou um grau de polêmica ou é algo em desacordo com a maioria dos seguidores da página, o que pode ser sustentado a partir da descrição dos comentários.

Nesse sentido, o trabalho contribuiu para mostrar que um aplicativo de coleta de dados em redes sociais pode contribuir bastante para a descrição e análise de alguns gêneros que ocorrem nesse ambiente digital, o que é cada vez mais relevante nessa era do hipertexto.

Ao concluirmos este artigo, queremos colocar algumas questões que podem contribuir para pesquisas futuras em perspectivas semelhantes. A primeira delas é: como tratar o alto índice de “curtir” presente nas publicações da rede? O que ele poderia revelar sobre a interação realizada pelos usuários? Essa pergunta tem relevância, uma vez que se constata, por exemplo, que “curtir” é usado inclusive para realizar outras ações, como simplesmente indicar ao elaborador da postagem que ela foi lida.

Outro ponto importante seria a tentativa de se descrever o alto índice de outras reações a outras características de comentários. Em outras palavras, poderíamos perguntar, por exemplo, o que um alto índice de “chorar” revelaria sobre os comentários.

Sobre a diferença com relação aos tipos de publicações, ou seja, entre o número de imagens e vídeos, frente aos outros tipos de publicações, ainda que não seja propósito deste trabalho, levantamos algumas hipóteses para tal ocorrência, a serem testadas em outros trabalhos: primeiro, que as publicações que exigem menos rolagem ou cliques devem ser as que mais provocam o engajamento; depois, que as publicações desse tipo têm um caráter mais lúdico, algo muito procurado em páginas desse tipo, especialmente com a proliferação de memes e gêneros similares.

Nossa contribuição para a área dos estudos em gêneros, língua portuguesa e ferramentas para análise linguística deverá continuar, a fim de verificar o seguinte ponto: esse padrão de característica dos comentários argumentativos em publicações com alto índice de “odiar” é verificado em outras páginas, como de jornais, por exemplo? Esperamos responder essa pergunta em análises futuras.

Agradecimentos

Agradecemos a Fernanda Cavassana (UFPR/UTFPR) pela instrução no uso com o Netvizz e aos dois pareceristas anônimos da Domínios de Lingu@gem pelas revisões e comentários. Os problemas persistentes são de nossa responsabilidade.

Referências

- BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 1992.
- BARTON, D.; LEE, C. **Linguagem online**: textos e práticas digitais. Trad. Milton Camargo Mota. 1. ed. São Paulo: Parábola Editorial, 2015.
- COSCARELLI, C. V. Navegar e ler na rota do aprender. In: COSCARELLI, C. V. (Org.) **Tecnologias para aprender**. São Paulo: Parábola, 2016.
- FREIRE, C. P. **Método de monitoramento de redes sociais**: epistemologia, técnicas e propostas de mineração de banco de dados para conteúdos gerados por fãs de telenovela em redes sociais. Tese de doutorado. Escola de Comunicação e Artes. Ciências da Comunicação. Universidade de São Paulo, 2015.
- FISCHER, S. R. **História da escrita**. São Paulo: Editora UNESP, 2009.
- GNANADESIKAN, A. E. **The writing revolution**: cuneiform to the Internet. Chichester: Willy-Blackwell, 2009.
- GRINSPUN, M. P. S. Z. (Org.). **Educação tecnológica**: desafios e perspectivas. 3. ed. São Paulo: Cortez, 2009.
- JENKINS, H.; GREEN, J.; FORD, S. **Cultura da conexão**: criando valor e significado por meio da mídia propagável. São Paulo: Editora ALEPH, 2014.
- LE MOS, A. **Cibercultura**: tecnologia e vida social na cultura contemporânea. 7. ed. Porto Alegre: Sulina, 2015.
- LIMA, C. **Cibercultura, ciberlinguagem e cibereducação**. São Paulo: Editora Biblioteca24horas, 2012.
- MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. In: DIONISIO, A. P. et al. (org.) **Gêneros textuais & ensino**. Rio de Janeiro: Lucerna, 2002, p. 19-36.
- MARCUSCHI, L. A. Gêneros textuais emergentes no contexto da tecnologia digital. In: MARCUSCHI, L. A.; XAVIER, A. C. **Hipertexto e gêneros digitais**: novas formas de construção de sentido. 3. ed. São Paulo, Cortez, 2010.
- MARCUSCHI, L. A.; XAVIER, A. C. **Hipertexto e gêneros digitais**: novas formas de construção de sentido. 3. ed. São Paulo, Cortez, 2010.
- MARTÍNEZ, F. S. **Impacto socio-cultural de las tecnologías**. Murcia (Espanha): Umtv (Universidad de Murcia), 2015. (8 min.). Disponível em: <https://www.youtube.com/watch?v=eOu9exw-Pv0>. Acesso em: 22 maio 2016.
- PINTO, Á. V. **O conceito de tecnologia**. Rio de Janeiro, Contraponto: 2005.

POE, M. T. **A history of communications**: media and society from the evolution of speech to the Internet. New York: Cambridge University Press, 2012.

RECUERO, R. **A conversação em rede**: comunicação mediada pelo computador e redes sociais na Internet. Porto Alegre: Sulina, 2012.

SANTOS, E. P.; ALVES FILHO, F. O plurilinguismo no gênero comentário online: encontro e confronto entre muitas vozes sociais. **Revista FSA**, Teresina, v. 11, n. 2, abr./jun. 2014, p. 301-317. Disponível em: <http://www4.fsnet.com.br/revista/index.php/fsa/article/view/502> Acesso em 24. nov. 2016.

SHIRKY, C. **A cultura da participação**: criatividade e generosidade no mundo conectado. Tradução: Celina Portocarrero. Rio de Janeiro: Zahar, 2011.

SOARES, M. **Novas práticas de leitura e escrita**: letramento na cibercultura. Educ. Soc., Campinas, vol. 23, n. 81, p. 143-160, dez. 2002.

VARGAS, M. Prefácio. In: GRINSPUN, M. P. S. Z. **Educação tecnológica**: desafios e perspectivas. 3. ed. São Paulo: Cortez, 2009. p. 7-19.

WAGNER, R. **A invenção da cultura**. São Paulo: Cosac Naify, 2008.

WOLTON, D. **Internet, e depois? Uma teoria crítica das novas mídias**. 3ª edição. Porto Alegre: Editora Sulina, 2012.

XAVIER, A. C. **A era do hipertexto**: linguagem e tecnologia. Recife: Pipa Comunicação, 2013.

Artigo recebido em: 14.12.2016

Artigo aprovado em: 31.01.2017

Um estudo sobre referenciação e a construção da opinião a partir de um *corpus* textual extraído do *YouTube*

A study about reference and the construction of the opinion from a textual *corpus* extracted from YouTube

Alexandre Ribeiro Afonso*
Jordão Té**

RESUMO: Este artigo descreve o fenômeno da referenciação a partir de um *corpus* contendo postagens opinativas coletadas do *YouTube*. Especificamente, os vídeos selecionados descrevem ou comentam sobre o processo de *impeachment* da presidente do Brasil Dilma Rousseff, iniciado no ano de 2015. Procurou-se descrever como o objeto discursivo *impeachment* é colocado em formas ou expressões nominais pela composição de unidades morfossintáticas variadas. Este estudo estabelece relações com a informática, no que diz respeito à análise automatizada de sentimentos, partindo-se dos dados em mídias sociais. Por isso, busca-se também a relação do referente e suas formas nominais com a positividade e a negatividade expressas nas opiniões postadas.

PALAVRAS-CHAVE: Referenciação. Mídias Sociais. Análise de Sentimentos.

ABSTRACT: This article describes the phenomenon of reference from a *corpus* containing textual opinions collected from *YouTube*. Specifically, the selected videos describe or comment about the *impeachment* of the President of Brazil Dilma Rousseff that started in 2015. It was noted the discursive object *impeachment* changes in nominal forms having different morphosyntactic units. Our study establishes a relationship between the reference phenomenon in social media data and automated sentiment analysis. Therefore, the aim is to establish a relationship between the reference phenomenon, its nominal forms, and the positivity and negativity expressed through the opinion texts.

KEYWORDS: Reference. Social Medias. Sentiment Analysis.

1. Introdução

Com o crescente desenvolvimento e utilização das mídias sociais e dos dispositivos móveis, abriram-se diversos canais de comunicação em rede, capazes de conectar milhares de pessoas ao redor do mundo, através de ferramentas de baixo custo, portáteis e facilmente utilizáveis.

Nesse cenário, de troca de informações e interação discursiva, a exposição e recepção de opiniões sobre pessoas, entidades e eventos de interesse da população ocorrem de forma

* Doutor em Ciência da Informação pela Universidade de Brasília - UnB.

** Graduando do curso de Bacharelado em Gestão da Informação da Universidade Federal de Goiás - UFG.

quase instantânea, gerando mensagens textuais opinativas em grande quantidade, originárias de localidades diversas, e estimulando o ato de conversação em rede. Dentre as investigações relacionadas às mídias sociais, é possível a análise do código linguístico utilizado pelos internautas, com métodos diversificados. Há, por exemplo, a possibilidade de descobrir e analisar o que um grande número de internautas, de forma coletiva, sente ou pensa sobre um produto comercializado, um processo de eleição ou alguma celebridade, quando se expressam por suas mensagens opinativas. Segundo Schiessl (2007), a essa atividade investigativa tem-se chamado de descoberta de conhecimento em textos; ela objetiva automatizar o processo de transformar dados textuais em informação para possibilitar a aquisição de conhecimento. Tal banco de dados de análise é construído a partir de textos diversos, desde artigos científicos até escolares e, inclusive, os das mídias sociais em questão.

O conhecimento descoberto permitiria às organizações a mudança de estratégias de marketing pessoal ou comercial, de maneira a prover maior índice de vendas, levar à mudança de estratégia em campanhas políticas e publicitárias, ou mesmo contribuir para elaboração de políticas públicas baseando-se nas opiniões da população. Os dados textuais das mídias sociais poderiam ser analisados utilizando-se softwares capazes de construir medidas estatísticas a partir de grandes quadros de opinião em rede.

Em outra vertente, de inovação tecnológica, ainda seria possível não só analisar dados provenientes das mídias sociais, mas construir aplicativos inteligentes, capazes de, automaticamente, sugerir um discernimento entre mensagens com teor positivo, negativo ou neutro sobre uma pessoa, evento ou entidade específica. Esses aplicativos procuram simular a inteligência humana e executam a análise automática de sentimentos. Os métodos de julgamento automático da expressão textual em polaridades positivas ou negativas são diversificados, e não são a única opção para tal inteligência computacional, sentimentos de surpresa, medo, tristeza, entre outros, também têm sido alvo de pesquisa na área (ARAÚJO *et al.*, 2013).

Seja no viés da análise de dados ou da inteligência computacional citada, deve-se considerar que uma das etapas iniciais é o entendimento do processo de comunicação. Sendo possível levantar como as pessoas utilizam a língua ao interagirem, certamente, tais observações serão fundamentais, como parâmetro, na análise de opiniões ou na construção de aplicativos inteligentes.

Sob tal perspectiva, do estudo do texto opinativo on-line, este artigo relata alguns experimentos, analisando um *corpus* de opiniões extraído das postagens do *YouTube*. Verifica-se como o fenômeno discursivo da referenciação ocorre em tais postagens, ou seja, busca-se compreender de que maneira o objeto de interesse em discussão na mídia é renomeado. São procuradas as características das formas nominais construídas e verifica-se a maneira pela qual o ponto de vista do internauta é inserido nessas formas.

As formas nominais de um objeto do discurso em interesse, segundo estudos realizados no âmbito da Linguística, por autores diversos, podem relatar o ponto de vista do enunciador, e neste trabalho, buscam-se relações desta afirmativa em ambientes computacionais interativos.

2. Trabalhos relacionados

Os estudos que analisam o uso da língua e a interação discursiva em mídias sociais ainda são discretos para o português do Brasil, se consideradas aplicações computacionais. Sobre o fenômeno discursivo da referenciação, aspectos cognitivos e linguísticos, e a elaboração do ponto de vista no texto, os trabalhos que se enquadram na especialidade da Linguística Textual estão próximos a este estudo.

Para Cortez e Koch (2013), por exemplo, os indivíduos retrabalham as formas sociais e culturais no discurso. A construção dos objetos de discurso homologa traços de um diálogo interior do sujeito enunciador consigo mesmo e com os outros. O ponto de vista, segundo as autoras, não se limita à expressão de uma percepção e integra julgamentos e conhecimentos que o locutor e/ou enunciador projetam sobre o referente. O ponto de vista se apresenta abertamente ou de forma mascarada, manifestando-se em todo o discurso.

A discussão referencial pode ser colocada ainda em um nível mais específico, onde os referentes não são apenas entidades materializadas no discurso, mas são reconstruídas. Lima e Feltes (2013) afirmam que os referentes, inclusive, podem não estar presentes na superfície do texto, estando sua reconstrução sujeita à recorrência a mecanismos inferenciais mais complexos, ancorados no nível das estruturas e do funcionamento cognitivo, mas sempre guiados pelo sinal linguístico.

A ideia de referenciação na descrição de Maruci (2010) é de uma atividade discursiva em que o sujeito opera com o material linguístico que tem à disposição, por ocasião da interação verbal. A referenciação consiste na construção e reconstrução de objetos do

discurso, ou seja, objetos designados, representados a partir de conhecimentos socialmente compartilhados e discursivamente (re)construídos e é, portanto, caracterizada como um processo em que o sujeito faz escolhas, baseadas em função de um querer-dizer.

Neste estudo, tem-se por base o conceito apresentado em Koch (2008) quanto ao ato do enunciador de referenciar. Procura-se verificar na mídia como o referente *impeachment* é categorizado e recategorizado pelos internautas em suas postagens, através de formas nominais que expressam seus pontos de vista. Segundo Koch (2008), o primeiro passo na construção de um texto é a introdução de um objeto de discurso na memória textual (em geral, por meio de um nome próprio ou forma nominal). Isto é, um novo objeto de discurso é construído e introjetado na memória, onde vai preencher um nódulo, ou seja, passar a ter um endereço cognitivo, de modo a ficar em foco e disponível para retomadas ou remissões. Quando a introdução se faz por meio de um nome próprio, tem-se apenas a nomeação do objeto. Já no caso de se tratar de uma expressão nominal, opera-se uma primeira categorização do objeto de discurso, o qual, a cada retomada, pode ser mantido como tal ou, então, recategorizado por outras expressões nominais.

O desenvolvimento de um sistema computacional que analise e julgue o sentimento de positividade e negatividade sobre o referente exige a compreensão do uso da língua na mídia social, além do entendimento sobre as estratégias de interação e as características de expressão dos usuários no ambiente midiático. Neste sentido, Foschiera (2012) contribui ao realizar um estudo apoiado na semântica de frames para verbos e adjetivos que descrevem emoções, considerando um *corpus* coletado do *Twitter*. Fernandes *et al.* (2015) conceitua e explica os processos de formações de neologismos, a partir de dados coletados do *Facebook* e *Twitter*. Cunha (2012) apresenta um estudo de inspiração sociolinguística acerca da utilização de *hashtags* pelos membros do *Twitter*. Sylvestre (2013) investiga como as relações discursivas evidenciam a construção de identidades nas Redes Sociais, em especial no *Twitter* e no *Facebook*, e como relações assimétricas de poder são mantidas ou modificadas nas práticas sociais nesses espaços.

No trabalho de DeAquino e Dioguardi (2013) realiza-se um estudo que se aproxima do enfoque desta pesquisa, e contribui para o conhecimento de uma concepção específica de orientação argumentativa que se dá a partir de elementos linguísticos e discursivos que, utilizados de modo coesivo particular, vão caracterizando o gênero *tweet*, para conferir-lhe coerência. O trabalho descrito por Scopim (2011) inspirou esta pesquisa, no sentido que

também descreve algumas características lexicais de textos opinativos para o português do Brasil, especificamente para um *corpus* de textos jornalísticos, e identifica expressões que denotam avaliação. Feito isso, a autora estabelece dois léxicos avaliativos: um positivo e um negativo.

A partir destes entendimentos comunicacionais e interativos, os sistemas computacionais de análise de sentimentos se tornarão mais robustos, pois terão esses parâmetros e heurísticas em seu código, permitindo analisar sentimentos com maior índice de acertos. Este estudo contribui no mesmo sentido, considerando como tópico de investigação o fenômeno da referenciação.

3. Questões de pesquisa e objetivos

No final do ano de 2015, vídeos diversos surgiram no *YouTube* relatando ou comentando sobre o processo de *impeachment* da presidente do Brasil Dilma Rousseff. Os acontecimentos vêm sendo acompanhados on-line pela população brasileira, em diversos tipos de mídias sociais digitais, e especificamente, no *YouTube*, os vídeos têm permitido a interação social sobre o assunto.

Nesta pesquisa, em análise de textos opinativos, diversas opiniões em texto foram coletadas de três vídeos provenientes do *YouTube*, sobre o processo de *impeachment*, levando à montagem de um *corpus* de opiniões.

A pesquisa objetivou analisar dois conjuntos distintos de postagens do *corpus*: as postagens opinativas favoráveis ao *impeachment* da presidente (positivas) e as postagens contrárias ao processo (negativas). Buscaram-se as características dos dois conjuntos analisados, especificamente, as formas nominais para o objeto discursivo *impeachment*, ou como o texto representante do referente é modificado pelos enunciadores quando as opiniões são lançadas.

Considera-se que o resultado da análise destas transformações sobre o referente poderia ser útil na análise de dados de postagens para se chegar ao conhecimento do sentimento coletivo, ou ainda, tais resultados sobre as formas nominais poriam ser implementados, num formato apropriado, em programas de computador que agrupem as mensagens automaticamente.

Procura-se responder a duas questões na pesquisa realizada:

- a) Existem mudanças na forma do objeto de interesse em discussão, no caso, o evento *impeachment*, e quais as características das formas nominais que o representam?
- b) Os conjuntos de formas nominais para o referente *impeachment* são diferentes para as opiniões positivas e negativas?

4. Materiais e Métodos

Para os experimentos realizados foram coletadas 215 (duzentas e quinze) postagens sobre o processo de *impeachment* da presidente Dilma Rousseff, a partir de 3 (três) vídeos que relatam ou comentam sobre o tema. A coleta das postagens foi uma ação aleatória, a verificação no ato foi notar se o texto colhido opinava sobre o processo de *impeachment*. Evitou-se também a coleta de postagens que eram respostas direcionadas a postagens prévias, pois estas poderiam levar em consideração um comentário pessoal prévio e não somente o tema em questão.

Após a coleta houve a filtragem, pois era o objetivo reter apenas postagens cujo conteúdo fosse de teor positivo ou negativo em relação ao processo de *impeachment*. Algumas mensagens não guardavam polaridade definida, ou por algum motivo não foi possível detectar a polaridade da opinião. Tais postagens foram descartadas, sendo que o *corpus* final gerou uma base de dados de 158 (cento e cinquenta e oito) postagens, destas, contabilizou-se 91 (noventa e uma) de teor positivo, ou favoráveis ao processo de *impeachment*, e 67 (sessenta e sete) de teor contrário, ou negativo.

Neste primeiro conjunto de experimentos, foram coletadas somente formas nominais que renomeassem o objeto *impeachment*, seja com formação simples contendo somente um substantivo (por exemplo, *processo*) ou composta, contendo substantivos seguidos de outras categorias morfossintáticas que modificam ou complementam o substantivo central: adjetivos, preposições, advérbios ou outro substantivo (por exemplo, *processo de impedimento* ou *ato de ruptura inconstitucional*). Não foram coletadas formas verbalizadas, porém, indiretamente, formas verbalizadas podem também renomear o referente, como no caso do verbo *golpear*, que remete a *golpe* substituindo *impeachment*, ou ainda a oração: *Dilma não deve estar no poder* que também remete a *impeachment*.

Após tais identificações, partiu-se para a localização das mudanças sobre o referente, o evento *impeachment*, localizando as formas nominais que o substituíssem. Lista-se então, as formas para o referente dentro do *corpus* de 158 postagens. A localização foi possível

utilizando os sistemas computacionais *Microsoft Excel* e o software livre para análise de *corpora AntConc* descrito por Karder e Richter (2013).

5. Experimentos e Análises

5.1 Experimento 1

Este experimento busca responder a primeira questão de pesquisa:

a) Existem mudanças na forma do objeto de interesse em discussão, no caso, o evento *impeachment*, e quais as características das formas nominais que o representam?

A lista a seguir descreve as formas encontradas nas postagens do *corpus*.

Lista 1— Lista de formas nominais encontradas para o evento *impeachment* no *corpus* de testes.

- | | |
|------------------------------------|--|
| 1. instrumento | 13. retrocesso |
| 2. solução | 14. mudança |
| 3. processo | 15. tapetão (<i>no sentido de "ganhar no tapetão"</i>) |
| 4. processo de impedimento | 16. machismo |
| 5. processo de impeachment | 17. espetáculo de ópera bufa |
| 6. julgamento | 18. pato (<i>no sentido de "pagar o pato"</i>) |
| 7. julgamento político | 19. golpe |
| 8. ato de ruptura inconstitucional | 20. injustiça |
| 9. saída de Dilma | 21. quadro |
| 10. justiça | 22. impedimento da presidenta |
| 11. pouca vergonha | 23. impedimento do atual mandato |
| 12. decisão | 24. punição |

Fonte: elaborada pelo autor.

As 24 (vinte e quatro) formas nominais relatam que o referente pode se camuflar sob o ponto de vista do enunciador. Como este estudo tem foco na caracterização das postagens como positivas e negativas, procurou-se esses pontos de vista a partir das formas nominais. Por exemplo, as formas listadas: 2 (*solução*) e 10 (*justiça*), isoladamente expressam concordância com o processo de *impeachment*, já as formas: 8 (*ato de ruptura inconstitucional*), 19 (*golpe*) e 13 (*retrocesso*) marcam contrariedade ao processo de *impeachment*. Obviamente, tais formas podem ter partículas de negação associadas no decorrer do texto e poderiam levar a um sentido contrário ao que é exposto isoladamente. As formas 3 (*processo*), 12 (*decisão*) e 21 (*quadro*) quando lidas isoladamente expressam neutralidade e necessitam de outros elementos textuais para a atribuição de positividade ou

negatividade à mensagem. As formas 14 (*mudança*), 16 (*machismo*) e 24 (*punição*) exibem pontos de vista, mas não é possível atribuir positividade ou negatividade em relação ao *impeachment* se lidas isoladamente.

Observa-se que as formas nominais também variam na estrutura, vão desde um melhor detalhamento do nome *impeachment*, como em 5 (*processo de impeachment*), ou mesmo a simples tradução, como ocorre em 4 (*processo de impedimento*), indo até como constituinte em uma expressão cristalizada, como no caso de 15 (*ganhar a presidência no tapetão*) ou 18 (*a classe trabalhadores pagará o pato*).

5.2 Experimento 2

Este experimento busca responder a segunda questão de pesquisa.

b) Os conjuntos de formas nominais para o referente *impeachment* são diferentes para as opiniões positivas e negativas?

Durante o experimento, dividiu-se o *corpus* original com 158 (cento e cinquenta e oito) postagens em duas partes: postagens com opiniões a favor do *impeachment*, que somam 91 (noventa e uma) opiniões e postagens contra o *impeachment*, que somam 67 (sessenta e sete) opiniões.

Após esta etapa, verificaram-se quais formas descritas na lista 1, anteriormente descrita, ocorriam no conjunto das postagens positivas (favoráveis) e quais ocorriam no conjunto das postagens negativas (contrárias). Os resultados estão listados a seguir:

Lista 2 — Formas nominais para o evento *impeachment* que ocorrem no conjunto de opiniões positivas (a favor do *impeachment*) à esquerda, e negativas (contra o *impeachment*) à direita.

1. instrumento	1. instrumento
2. solução	2. processo
3. processo	3. processo de impedimento
4. processo de impeachment	4. processo de impeachment
5. julgamento político	5. julgamento
6. ato de ruptura inconstitucional	6. julgamento político
7. saída de Dilma	7. pouca vergonha
8. justiça	8. decisão
9. decisão	9. retrocesso
10. golpe	10. mudança
	11. tapetão
	12. machismo

13. espetáculo de ópera bufa
14. pato
15. golpe
16. quadro
17. impedimento da presidenta
18. impedimento do atual mandato

Fonte: elaborada pelo autor.

As colunas da esquerda e direita mostram que as formas: *instrumento*, *processo*, *processo de impeachment*, *julgamento político*, *golpe* e *impedimento da presidenta* ocorrem nos dois vieses, nas postagens a favor e contra. Com exceção de *golpe*, estas formas têm caráter neutro, e não geram qualificação ou ponto de vista quando lidas e interpretadas isoladamente. A forma *golpe* traz o ponto de vista negativo, contrário ao processo de *impeachment*, mas esta interpretação necessita, além do contexto, o conhecimento do estado político brasileiro atual e de conhecimento enciclopédico para ser compreendida.

As formas que aparecem somente na lista à direita advindas de postagens contra o *impeachment*: *pouca vergonha*, *retrocesso*, *tapetão*, *machismo*, *pato*, *espetáculo de ópera bufa* e *punição* colocam uma qualificação de contrariedade sobre o referente *impeachment*. As formas *machismo*, *espetáculo de ópera bufa*, *tapetão* e *pato* necessitam de uma interpretação que vai além do conhecimento cedido no próprio texto e, neste caso, o conhecimento enciclopédico também é necessário para sua interpretação.

Nos itens que ocorrem somente na lista à esquerda advindas de postagens a favor do *impeachment*, observam-se as formas: *solução* e *justiça*, que geram a noção de resolução do caso, porém, suas interpretações necessitam do conhecimento do caso.

Especificamente, a forma nominal *golpe*, que marca uma contrariedade ao processo de *impeachment*, ocorre nos dois conjuntos e isso indica que algumas formas nominais, de alguma maneira, são negadas visando marcar o efeito de positividade na mensagem.

Para uma noção quantitativa da aparição da forma *golpe* nos dois conjuntos, uma análise por frequência foi realizada para os conjuntos de postagens de teor positivo e negativo, e verificou-se sobre as formas nominais que:

golpe ocorreu 23 vezes nas postagens a favor.
impeachment ocorreu 38 vezes nas postagens a favor.
golpe ocorreu 63 vezes nas postagens contrárias.
impeachment ocorreu 42 vezes nas postagens contrárias.

Ao analisar tal contagem, observa-se que a forma *golpe* obteve frequência consideravelmente maior (sessenta e três ocorrências) que a forma original *impeachment* (quarenta e duas ocorrências) no conjunto de postagens negativas, ou contrárias ao processo de *impeachment*.

É observado ainda, que a forma *golpe* também foi utilizada com notável frequência nas postagens a favor do *impeachment* (vinte e três ocorrências), neste caso, acredita-se, que a forma é utilizada junto a algumas marcas discursivas de negação ou ironia, conjectura que deve ser verificada apropriadamente em análises futuras.

6. Considerações finais

A pesquisa aqui descrita desenvolveu-se orientada à busca de alguns padrões linguísticos, especificamente, em relação às mudanças formais sobre os referentes, alterando-se ao longo do texto opinativo on-line. O diferencial maior apontado são verificações ocorrentes nas postagens de uma mídia social on-line.

Com o auxílio dos softwares para análise de *corpora*, verificou-se que o fenômeno da mudança das formas nominais para o referente ocorre com frequência, inclusive reproduzindo o ponto de vista do internauta ao comentar os vídeos que descrevem o processo de *impeachment*.

Foi observada a existência de formas nominais que ocorrem com alta frequência tanto no conjunto das postagens a favor, quanto no conjunto das postagens contrárias, ou seja, formas ocorrentes nos dois conjuntos ao mesmo tempo.

Também foi observado que as formas nominais têm estrutura de formação simples ou composta, onde os substantivos nucleares sofrem alterações, sendo as formas nominais utilizadas também em expressões cristalizadas. Percebeu-se ainda, que as formas que camuflam o referente podem ser verbalizadas, e não só nominais, como as que foram descritas neste trabalho.

Sobre a interpretação das formas nominais listadas e a definição de suas polaridades, observou-se que determinadas formas não trazem, quando lidas isoladamente, a carga de positividade ou negatividade em relação ao processo de *impeachment*, sendo que algumas, inclusive, não revelam positividade ou negatividade, mas outro tipo de sentido. As formas ainda podem necessitar da interpretação do contexto, de conhecimento enciclopédico ou do

conhecimento da situação política brasileira atual para a identificação da polaridade que indicam.

Continuando este trabalho inicial, a análise quantitativa mais elaborada possivelmente traga novos resultados sobre como estas formas nominais se modificam. No caso das expressões cristalizadas, por exemplo, que foram identificadas nos experimentos, poder-se-ia verificar se a presença destas expressões, contendo o objeto discursivo de interesse em seu corpo, seria algo recorrente ou casual em postagens. Ou mesmo, se a frequência de determinadas formas nominais é mais alta do que outras, como foi o caso da forma *golpe* que se destacou em preferência de uso.

Todas estas observações levam a crer que as variações de forma de um objeto em discussão devem ser consideradas no trabalho de análise de dados ou na construção de aplicativos que processam e julgam os sentimentos de tais registros opinativos. Acreditamos que seja importante o estudo da opinião em outras discussões, com outros tópicos discursivos de interesse, para conclusões mais elaboradas.

Referências Bibliográficas

ARAÚJO, M. et al. Métodos para análise de sentimentos no Twitter. In: **Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)**, 2013.

CORTEZ, S. L.; KOCH, I. G. V. A construção do ponto de vista por meio de formas referenciais. In: CAVALCANTE, M. M.; LIMA, S. M. C. (Org.). **Referenciação: teoria e prática**. São Paulo: Cortez, 2013. p. 9-29.

CUNHA, E. L. T. P. **Etiquetagem de micromensagens no Twitter: uma abordagem linguística**. Dissertação (Mestrado em Ciência da Computação). Belo Horizonte, MG: UFMG, 2012.

DE AQUINO, Z. G. O.; DIOGUARDI, G. Argumentação nas redes sociais: o tweet-caracterização e funcionamento. **ContraPonto**, v. 3, n. 3, p. 70-92, 2014.

FERNANDES, P. D. *et al.* Os Processos de Formação de Neologismos: uma análise nas redes sociais. **Revista Philologus**, v. 21, n. 61, jan./abr. 2015.

FOSCHIERA, S. M. P. **A Semântica da Emoção: um estudo contrastivo a partir da FrameNet e da Roda de Emoções**. Tese de Doutorado (Programa de Pós-Graduação em Linguística Aplicada). São Leopoldo, RS: UNISINOS, 2012.

KADER, C. C. C.; RICHTER, M. G. Linguística de *corpus*: possibilidades e avanços. **Instrumento**, v. 15, n. 1, p. 13-23, jan./jun. 2013.

KOCH, I. V. Como se constroem e se reconstroem os objetos-de-discurso. **Investigações**. Recife, v. 21, n.2, p. 99-114, 2008.

LIMA, S. M. C.; FELTES, H. P. M. A construção de referentes no texto/discurso: um processo de múltiplas âncoras. In: CAVALCANTE, M. M.; LIMA, S. M. C. (Org.). **Referenciação: teoria e prática**. São Paulo: Cortez, 2013. p. 30-58.

MARUCI, F. A. D. Desvendando o processo de referenciação no gênero artigo de opinião produzido por alunos da educação básica. In: Congresso Nacional de Linguística e Filologia, 14., 2010, Rio de Janeiro. **Cadernos do CNLF**. Rio de Janeiro: UERJ, 2010. p. 2551-2563, vol. XIV, t.3.

SCHIESSL, J. M. **Descoberta de conhecimento em texto aplicada a um sistema de atendimento ao consumidor**. Dissertação (Mestrado em Ciência da Informação). Brasília, DF: UnB, 2007.

SCOPIIM, D. **Estudo de Padrões Lexicais em Textos Opinativos**. Dissertação (Mestrado em Linguística). São Carlos, SP: Universidade Federal de São Carlos, 2011.

SYLVESTRE, A. P. M. **O eu e o outro online: discurso, poder e identidade nas redes sociais**. Dissertação (Programa de Pós-Graduação em Linguística). Brasília, DF: UnB, 2013.

Artigo recebido em: 27.07.2016

Artigo aprovado em: 15.01.2017

**Gêneros textuais *datasheet* e artigo científico em aulas de ESP:
levantamentos léxico-estatísticos para fins educacionais**
**Datasheet and research article genres in ESP classes: lexical-statistic surveys for
educational purposes**

Sabrina Bonqueves Fadanelli*
Andrea Jessica Monzón**

RESUMO: A linha Sociorretórica de estudos de gêneros textuais (SWALES, 1990; BAZERMAN, 1997) tem apresentado resultados interessantes para o ensino de Inglês para Fins Específicos (ESP), muitos destes voltados para gêneros acadêmicos (BIBER et al., 2002). O presente artigo objetivou extrair dados relevantes para a preparação de material para o ensino de terminologia técnica em aulas de ESP no ambiente acadêmico. Assim sendo, artigos acadêmicos da área de Computação foram comparados a um gênero específico da área de Eletrotécnica: os *datasheets*. Utilizando ferramentas computacionais como o AntConc e o TermoStat e *corpora* destes dois gêneros, foram observadas as características apresentadas pelas ocorrências mais frequentes de colocações especializadas (HAUSMANN, 1990). Os resultados apontam similaridades e diferenças entre as colocações especializadas nos gêneros artigo acadêmico e *datasheet*, evidenciando a importância da aplicação da Linguística de *Corpus* no planejamento do professor de ESP.

PALAVRAS-CHAVE: Gêneros acadêmicos. Léxico especializado. Inglês para fins específicos. *Corpora*.

ABSTRACT: The social-rhetorical line of studies of text genres (SWALES, 1990; BAZERMAN, 1997) has presented interesting results in the teaching of English for Specific Purposes (ESP). A considerable amount of this research is related to academic genres (BIBER et al., 2002). This paper aims at extracting relevant data for the preparation of material for the teaching of technical terminology in ESP in the academic environment. Research articles in the area of Computer Science have been compared to a very specific genre in the area of Power Engineering: *datasheets*. Through the use of computing tools such as AntConc and TermoStat and *corpora* of both genres, features of the most frequent occurrences of specialized collocations (HAUSMANN, 1990) have been observed. The results point to similarities and differences between specialized collocations in the academic articles and the *datasheets*, evidencing the importance of applying *Corpus* Linguistics in the planning of ESP teaching.

KEYWORDS: Academic genres. Specialized lexicon. English for specific purposes. *Corpora*.

* Professora na área de Letras da Universidade de Caxias do Sul (UCS). Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: sbfadane@ucs.br

** Professora de Português e Inglês no Instituto Federal do Rio Grande do Sul (IFRS) - Campus Feliz. Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: andrea.monzon@feliz.ifs.edu.br

1. Introdução

Nos últimos anos, a ciência tem assumido cada vez mais um caráter de força produtiva, haja vista a sua inserção no cotidiano das pessoas comuns. A atualização constante da tecnologia e o desaparecimento das fronteiras de comunicação por meio do advento da internet e das mídias em geral têm como uma das consequências a necessidade de se obter conhecimento sobre Língua Inglesa em praticamente todas as áreas do conhecimento. Isso porque essa língua é a língua franca (JENKINS, 2012) do conhecimento e das trocas de comunicação.

A Língua Inglesa é a principal linguagem da ciência mundialmente (CRYSTAL, 2010; GRADDOL, 2000), sendo pré-requisito na maioria dos cursos de pós-graduação e também em publicações de pesquisa em diversas revistas acadêmicas (PAIVA; PAGANO, 2001). Muitos cursos técnicos e de graduação também disponibilizam disciplinas de inglês em seus currículos. E, em paralelo, a linguagem técnica, mesmo em inglês, aparece cada vez mais em nosso dia a dia (FINATTO, 2011) no Brasil. Logo, tornou-se comum encontrar termos anglófonos em meio a textos técnico-científicos escritos em português, o que mostra a relevância de tal idioma nesse cenário.

O inglês voltado à compreensão da linguagem técnica de uma determinada área se insere no campo do ensino de ESP - *English for Specific Purposes* (Inglês para Propósitos ou Fins Específicos), uma abordagem em que o aluno recebe instrução para necessidades pessoais, acadêmicas e/ou profissionais específicas (HUTCHINSON; WATERS, 1987). Muitas vezes a disciplina nos institutos, faculdades e universidades também é chamada de Inglês Instrumental ou Inglês Técnico. Dois domínios técnicos que se enquadram neste cenário de necessidade de leitura em Língua Inglesa são a Ciência da Computação e a Eletrotécnica.

Na área de Ciência da Computação/Informática, para que alunos, docentes e profissionais se informem, capacitem e atualizem, os artigos científicos (SWALES, 1990) são os textos especializados mais respeitados. A maioria deles, entretanto, é publicada em inglês, o que demanda a leitura e compreensão de tal idioma, especialmente das terminologias de domínio.

A Eletrotécnica é a ramificação da Engenharia Elétrica responsável por instalações de redes elétricas residenciais e industriais. A principal necessidade de estudantes e profissionais para a realização de seu trabalho nesse domínio é a leitura de documentos técnicos chamados

datasheets (DEWEY, 1998), os quais contêm especificações sobre o funcionamento de dispositivos elétricos. Esses documentos se encontram na internet, praticamente apenas em Língua Inglesa e demandam um certo nível de proficiência na leitura (pelo menos um A2, de acordo com o *Common European Framework*¹).

Em se tratando de ensino-aprendizagem de ESP, segundo Bhatia, Anthony e Noguchi (2011), há 3 aspectos sobre os quais os professores e pesquisadores envolvidos com tal abordagem deveriam refletir: análise das necessidades, desenvolvimento de materiais (didáticos) e as práticas docentes. Algo que permeia esses aspectos é, tanto para discentes quanto docentes, conhecer diversos gêneros textuais com o intuito de participar de distintas comunidades discursivas (SWALES, 1990).

Desse modo, o presente trabalho tem o objetivo de identificar e analisar as características léxico-estatísticas dos gêneros textuais *datasheet* e artigo científico da Computação. Para tanto, aqui observamos as convencionalidades de tais textos sob duas perspectivas: 1) descrição linguística quantitativa e qualitativa através da compilação de *corpora* e do uso de duas ferramentas computacionais para as análises léxico-estatísticas, e 2) contribuições para as aulas de ESP em contexto de Ensino Técnico e Superior. Trata-se, aqui, de levantamentos e análises provenientes de duas teses de doutorado em fase de conclusão, as quais tratam da leitura em situação de ESP em cursos de Educação Profissional nas áreas de Computação e Eletrotécnica, sob as perspectivas teórico-metodológicas da Linguística Aplicada, Linguística de *Corpus* e Terminologia. Para o presente artigo, entretanto, não apresentamos e discutimos todos os resultados destas pesquisas, uma vez que aqui buscamos uma concentração nas análises lexicais dos *corpora* e suas possíveis contribuições linguístico-pedagógicas. Frisamos também que o gênero textual *datasheet* receberá uma descrição mais específica, não por ser mais relevante academicamente do que o gênero artigo acadêmico, mas justamente por ainda não ser tão estudado quanto o artigo acadêmico.

Assim, na seção 2, buscamos discutir aspectos teórico-metodológicos dos gêneros textuais e suas implicações para este trabalho. Já na seção 3, trazemos características e etapas metodológicas para o levantamento léxico-estatístico e, na seção 4, discutimos os resultados provenientes do mesmo, identificando as idiossincrasias da cada área do conhecimento aqui

¹ Disponível em: http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp

observada. Finalmente, na seção 5 apontamos o que aprendemos com este estudo, bem como indicamos potenciais caminhos a serem percorridos no futuro.

2. Os Gêneros Textuais e suas potencialidades teóricas, linguísticas e educacionais

Os gêneros textuais têm sido objeto de interesse por parte de estudiosos preocupados com uma prática que privilegie a interação verbal e as diversas situações de interlocução. Entretanto, nas palavras de Marcuschi (2008, p. 147), "o estudo dos gêneros textuais não é novo e, no Ocidente, já tem pelo menos vinte e cinco séculos, se considerarmos que sua observação sistemática se iniciou com Platão". Isso indica que, devido ao longo período de observações, o tema apresenta diversos enfoques teóricos ainda hoje.

A partir do século XIX, o estudo dos gêneros assumiu novas e importantes perspectivas com o surgimento da ciência da Linguagem, a Linguística. Por todo o século XX e início do XXI, houve o desenvolvimento de linhas de pensamento que enfatizam pontos de vista diferentes. As linhas de pensamento mais conhecidas, juntamente com seus principais expoentes, são (MEURER; BONINI; MOTTA-ROTH, 2005²):

- a) Sociodiscursiva (Bakhtin, Adam, Bronckart, Marcuschi, entre outros);
- b) Sociossemiótica (Hasan, Martin, Kress, Fowler entre outros);
- c) Sociorretórica (Swales, Bazerman, Bhatia, entre outros).

Mikhail Bakhtin formulou a teoria sobre gêneros que viria a influenciar o pensamento ocidental, com uma concepção de linguagem que incorporava às atividades de linguagem a visão de um sujeito que a utilizava como um instrumento de mudança social (BAKHTIN, 2003). Ele partiu do pressuposto de que a comunicação só se faz possível através de textos e/ou gêneros textuais. Destacando o lado social da noção de gênero, ele afirma que qualquer enunciado considerado isoladamente é individual, porém cada esfera de utilização da língua elabora seus tipos relativamente estáveis de enunciados, sendo por isso denominados gêneros do discurso.

Marcuschi (2003) transporta a linha de Bakhtin para a noção de gênero textual, afirmando que esses são fenômenos históricos profundamente vinculados à vida cultural e

² Os autores discorrem sobre as diferentes linhas teóricas referentes ao estudo de Gêneros ao longo de toda a obra.

social. São, portanto, entidades sociodiscursivas e formas de ação social em qualquer situação comunicativa.

Nas abordagens de ordem Sociodiscursiva e Sociossemiótica, percebe-se que o gênero textual é abordado como um resultado de eventos e atividades humanas. Nesta pesquisa, preferimos adotar a abordagem Sociorretórica dos estudos de texto, seguindo principalmente as linhas de Swales (1990) e Bazerman (2011). Isso se deve à característica impressa nesta abordagem de que o gênero textual é um agente que determina e modifica eventos e convenções sociais, ao invés de ser apenas visto como resultante dos mesmos. Bazerman (2011, p. 32) define gênero textual como:

fenômenos de reconhecimento psicossocial, que são parte de processos de atividades socialmente organizadas. Gêneros são tão-somente os tipos que as pessoas reconhecem como sendo usados por elas próprias e pelos outros. (...) Gêneros emergem nos processos sociais em que as pessoas tentam compreender umas às outras suficientemente bem para coordenar atividades e compartilhar significados com vistas a seus propósitos práticos.

Charles Bazerman (2005) entende o gênero como uma construção acionada pelo sujeito para reconhecimento e desenvolvimento de ações tipificadas em situações retóricas recorrentes, não como uma categoria linguística modelizada por certas características textuais. Esse direcionamento nos leva a perceber o gênero como um fato social, entendido como aquilo em que se acredita como sendo verdadeiro, afetando o modo como se definem as situações sociais. Cada texto bem-sucedido cria para seus leitores um fato social, consistindo em uma ação social significativa realizada pela linguagem.

Para Swales (1990), o gênero textual é, ainda, um conjunto de propósitos comunicativos formado pelo próprio discurso e seus participantes (comunidade discursiva). O propósito comunicativo para Swales é central em sua noção de gênero, visto que esse é a razão do uso da linguagem, fornecendo a justificativa por trás do gênero e contribuindo para a estruturação do discurso e das convenções de uso.

Há três conceitos intimamente relacionados: comunidade discursiva, propósito comunicativo e gênero. Esses são importantes na caracterização dos *datasheets* e dos artigos científicos da Computação, pois: (a) identificando quem os usa, (b) para que finalidade e (c) como os documentos se organizam, podemos mais adiante delinear contribuições mais eficientes para nossos contextos educacionais de ESP e às necessidades de nossos alunos.

Sobre os participantes da comunicação e seu posicionamento em relação às áreas de Elétrica e Computação, Swales afirma que a análise de um gênero textual deve iniciar identificando o seu propósito comunicativo dentro de uma comunidade discursiva. A comunidade discursiva que escreve os *datasheets* e Artigos Científicos da Computação (doravante ACCs) é, em sua maioria, altamente especializada. A princípio, o público-alvo de leitura dos *datasheets* e ACCs é de especialistas, contudo levamos aqui em consideração aqueles que são especialistas em formação, ou seja, pesquisadores iniciantes e/ou estudantes durante a sua formação acadêmico-profissional. Esse cenário se adapta bem à noção expressa por Swales e Marcuschi de que é a comunidade discursiva que produz o gênero de acordo com suas necessidades. Assim sendo, no caso dos *datasheets* e ACCs, o reconhecimento do gênero dentro de sua comunidade discursiva pode estar ligado, dentre outros aspectos, ao nível de especialização que os autores de ambos os gêneros textuais empregam nas terminologias dos documentos. Hoffmann (1998; KILLIAN; FINATTO, 2015)³, em sua descrição estrutural interna de textos especializados, aponta como uma característica fundamental do texto especializado o nível de especialidade dos participantes da comunicação. Cabré (2008) e Freixa (2002) determinam que a ‘quantificação’ do nível de especialização se dá principalmente através da relação entre parâmetros comunicativos e textuais, principalmente a relação entre os interlocutores. Ciapuscio e Kuguel (2002) também evidenciam a importância do nível de especialização dos participantes da comunicação para uma tipologização do texto especializado. Tanto Ciapuscio e Kuguel como Hoffmann determinam que o texto especializado pode ter participantes mais ou menos especialistas na área.

Neste trabalho, observamos os gêneros textuais não somente pelas suas características, construção e níveis de especialidade, como também entendemos suas contribuições para o ensino-aprendizagem de línguas estrangeiras e ESP.

³ In: FINATTO, M. J.; ZILIO, L. **Textos e termos por Lothar Hoffmann** - um convite para o estudo das linguagens técnico-científicas. Porto Alegre: Palotti, 2015.

2.1 Gêneros textuais e ESP

Inglês é o idioma que tem o lugar de língua franca (JENKINS, 2012) na comunidade internacional acadêmica e científica, ocupando, ainda, a posição de *global language* (CRYSTAL, 2003) no cenário mundial dos negócios, tecnologia e do trabalho. É principalmente através de tal idioma que pesquisas são divulgadas e compartilhadas com estudantes, professores e profissionais de todo o mundo.

No caso específico do alunado dos cursos da área de Informática/Computação, onde se concentra a maioria das matrículas na educação profissional federal, para que sejam iniciados e inseridos em seu domínio de formação acadêmico-profissional, é preciso ler e compreender textos especializados em inglês. Sendo essa a língua predominantemente utilizada em publicações científicas, livros, manuais e linguagens de computador, ela tem o status de ferramenta primordial para a assimilação e construção do “saber e [do] saber fazer específicos daquela ciência” (BARBOSA, 2009, p. 43).

Já no caso da área de Engenharia Elétrica, o artigo científico possui a sua devida importância, contudo no cotidiano acadêmico-profissional os alunos necessitam, em primeira instância, da leitura e compreensão dos *datasheets*. Nesse gênero textual são encontradas as terminologias em seu *habitat* natural (CABRE, 2009; 2002) para que os alunos, como futuros especialistas, possam vivenciar o uso e funcionamento de equipamentos elétricos recorrentes.

Para este contexto acadêmico, o da Educação Profissional, outra característica relevante dos gêneros textuais, apontada por Biber et al. (2004; 1995) é a sua composicionalidade. Existem padrões associativos que formam pacotes lexicais, os quais são “sequências recorrentes de palavras, que são importantes blocos constitutivos do discurso [acadêmico] nos registros falado e escrito”⁴ (BIBER, 2007, p. 1). O entendimento desses blocos de informações (SCARAMUCCI, 2007; 1998) só constrói seus sentidos através da leitura atenta, que não se atém a palavras isoladas, já que o significado de uma palavra “é sempre contextual” (FIRTH, 1957, p. 7).

No que tange à leitura, enquanto processamento, Koch (2005) defende a ideia de que os indivíduos desenvolvem uma competência metagenérica que lhes possibilita interagir com as línguas de forma conveniente. É “essa competência que possibilita a produção e compreensão de gêneros textuais” (KOCH; ELIAS, 2014, p. 102).

⁴ Tradução nossa.

2.2 Colocações especializadas: blocos lexicais idiossincráticos

As colocações são fenômenos linguísticos arbitrários que são consagrados pelo uso ou pela prática, ou seja, a convencionalidade estabelece a sua combinabilidade. Esse conceito, por sua vez, remete “à faculdade que os elementos linguísticos têm de se combinar. Existem palavras que se associam de forma tão natural que a única explicação possível é de que essa associação tenha sido consagrada pelo uso, isto é, de que ela seja convencional” (TAGNIN, 2013, p. 25-26).

O estudioso que introduziu o termo *colocação* para as convencionalidades léxico-sintáticas foi J. R. Firth (1957). Propõe-se, aqui, estudar as colocações sob a ótica de que as palavras se constituem de acordo com a companhia que elas têm. As coocorrências podem ser arbitrárias, mas “se um de seus componentes for alterado, poderá haver ruído na comunicação” (SANTOS, 2010, p. 100).

Aqui optamos por empregar o termo colocações especializadas - CEs (ORENHA-OTTAIANO, 2012; L’HOMME; BERTRAND, 2000), devido à relevância terminológica e conceitual desses padrões lexicais dentro do domínio da Computação, como algo que remete às metalinguagens que devem ser inferidas pelos alunos em formação. O contato com essas terminologias ocorre através da leitura do inglês especializado.

As colocações, por sua vez, distinguem-se das combinatórias livres, nas quais as palavras se combinam por mera afinidade. Nas colocações, um usuário da língua pode entender as palavras que as compõem, mas não saberá usá-las e/ou reproduzi-las, não sendo essa tarefa algo automático. Dentro de uma colocação, o status de cada uma de suas partes não é igual. Hausmann (1990) denomina *base*⁵ (*keyword*) o componente que não modifica a identificação do que é caracterizado; já *collocatif* (colocado) é o componente que só recebe seu significado semântico através da construção da colocação em si. Por exemplo, na Computação o termo *display server* tem o primeiro vocábulo como um colocado (*display*), que com sua base (*server*) constrói um significado diverso do termo *server* por si só.

De acordo com L’Homme e Bertrand (2000, p. 498), “Colocações são convenções dentro de uma determinada comunidade linguística; as combinatórias lexicais especializadas

⁵ Texto original em francês.

são convenções dentro de um grupo de especialistas”.⁶ Essas autoras afirmam, ainda, que tanto dizer colocações especializadas quanto combinatórias léxicas especializadas atende a uma mesma comunidade de especialistas. Para não nativos, a importância de se conhecer e estudar colocações implica entender melhor algo que não pode ser inventado ou parafraseado (SANTOS, 2010). Essas convenções não podem ser vistas como algo com regularidade sintática ou semântica, uma vez que os aprendizes e/ou não nativos devem adquiri-las partindo do pressuposto de que elas são imprevisíveis (HAUSMMAN, 1990; L’HOMME; BERTRAND, 2000). Nisso reside a relevância de se produzir compilações de CEs de todas as áreas e subáreas do conhecimento. Pesquisadores e estudantes dos mais diversos níveis de proficiência em inglês devem se comunicar de maneira produtiva em seus campos de estudo, divulgando e recebendo informações.

Para quem é aprendiz, portanto, a reprodução das colocações não ocorre automaticamente, ou seja, ele deve aprendê-las, “pois as línguas, dentro de sua totalidade de combinações lógicas possíveis, fazem escolhas idiossincráticas” (HAUSMMAN, 1990, p. 1010). Logo, a colocação é uma unidade, não da fala, mas da língua em si.

Seja em forma de repositórios, glossários, dicionários ou ferramentas pedagógicas, a arbitrariedade das CEs deve ser contemplada em produtos terminológicos que atendam a usuários reais com as dificuldades já discutidas.

Segundo Tagnin (2013), as colocações podem ser: nominais, adjetivas, verbais ou adverbiais. Sendo as CEs combinatórias que têm essencialmente cunho terminológico e que se configuram por uma combinabilidade convencionalizada por seus usuários especialistas, L’Homme (2000) menciona que sua base é majoritariamente nominal, havendo as seguintes possibilidades: verbo + substantivo, substantivo + adjetivo e substantivo + substantivo. Orenha-Ottaiano (2012), entretanto, discorda de que haja somente essas três possibilidades para CEs, uma vez que a pesquisadora verificou, em seus estudos colocacionais acerca de jornalismo de negócios e contratos sociais em textos anglófonos, que a taxonomia de Hausmann (Quadro 1) pode ser empregada tanto para analisar colocações da língua comum quanto da de especialidade. Assim sendo, essa foi a classificação aqui aplicada.

⁶ Tradução nossa do trecho “*Collocations are conventional within a given linguistic community; specialized lexical combinations are conventional a group of specialists*”.

Quadro 1 — Taxonomia de Hausmann para as colocações.

colocações nominais	substantivo + substantivo substantivo + preposição + substantivo
colocações verbais	verbo + substantivo substantivo + verbo verbo + preposição + substantivo verbo + partícula adverbial + substantivo verbo + adjetivo
colocações adjetivas	adjetivo + substantivo
colocações adverbiais	advérbio + adjetivo verbo + advérbio advérbio + verbo

Fonte: Hausmann, 1985 (apud ORENHA-OTTAIANO, 2012, p. 153).

Percebemos a relevância das CEs, enquanto blocos de informações convencionalizados, ao longo dos dois gêneros textuais aqui analisados - *datasheet* e artigo científico da Computação. Isso demonstra a potencialidade de se utilizar estudos linguístico-descritivos, quantitativos e qualitativos, no âmbito do léxico técnico-profissional como ferramentas pedagógicas e de aporte teórico-metodológico para aulas de ESP, as quais trabalham fundamentalmente com gêneros especializados.

2.3 Gênero textual *datasheet* e suas peculiaridades

Apresentamos a seguir algumas características macro e microestruturais de *datasheets* de Eletrotécnica. O Quadro 2, de nossa autoria juntamente com um especialista engenheiro elétrico, mostra os aspectos macroestruturais que os *datasheets* em geral apresentam. Um exemplo de *datasheet* pode ser visualizado na Figura 1.

Quadro 2 — Características em comum dos *datasheets*.

característica	lâmpada led	inversor fotovoltaico	capacitor
nome da empresa	X	X	X
endereço, site e contato		X	X
modelo/código do produto	X	X	X

aplicações	X	X	X
desenhos, dimensões, diagramas	X	X	X
gráficos	X	X	
condições de funcionamento (tensão, consumo, temperaturas, etc.)	X	X	X
características de padrão, elétricas e ópticas	X	X	X
símbolos, referências numéricas, abreviações	X	X	X

Fonte: elaborado pelas autoras.

A microestrutura dos *datasheets* pode mesclar características descritivas/argumentativas e expositivas: os dados técnicos e informações de atributos do produto são colocados objetivamente em tabelas ou introduzidas por marcadores. Aparecem fotos ou desenhos do produto e as informações de contato da empresa fabricante, além de gráficos. O documento pode apresentar um *layout* colorido ou preto-e-branco. *Datasheets* podem ser usados como ferramentas de venda do produto, como também podem acompanhar o mesmo no processo de pós-venda, a fim de fornecer dados para sua instalação e manutenção.

Figura 1 — Exemplo de *datasheet*

World Components Network Service Ltd.

Model No.	WCN-411HY1-110N
Doc. No.	LED-411HY1-110N

Applications:

- Advertising Signs
- Indicators
- LCD Back Light
- Illuminations

Absolute Maximum Ratings at Ta = 25°C

Items	Symbol	Absolute maximum Rating	Unit
Forward Current	I_F	50	mA
Peak Forward Current*	I_{FP}	200	mA
Reverse Voltage	V_R	5	V
Power Dissipation	P_D	130	mW
Operation Temperature	T_{opr}	-40 ~ + 95	°C
Storage Temperature	T_{sta}	-40 ~ + 100	°C
Lead Soldering Temperature	T_{sol}	Max.260°C for 3 sec Max. (3mm from the base of the epoxy bulb)	

pulse width $\leq 0.1\text{msec}$ duty $\leq 1/10$

Typical Electrical & Optical Characteristics (Ta = 25°C)

Items	Symbol	Condition	Min.	Typ.	Max.	Unit
Forward Voltage	V_F	$I_F = 20\text{mA}$	---	2.1	2.6	V
Reverse Current	I_R	$V_R = 5\text{V}$	---	---	100	μA
Luminous Intensity	I_V	$I_F = 20\text{mA}$	---	700	---	mcd
Peak Emission Wavelength	• • •	$I_F = 20\text{mA}$	---	---	---	nm
Dominant Wavelength	• • •	$I_F = 20\text{mA}$	585	591	595	nm
Spectral Line Half-Width	• • • •	$I_F = 20\text{mA}$	---	26	---	nm
50% Power Angle	$2\theta_{\pm}$	$I_F = 20\text{mA}$	---	100/40	---	deg

Important Notes:

- 1) All ranks will be included per delivery, rank ratio will be determined by WCN.
- 2) Tolerance of measurement of luminous intensity is $\pm 15\%$.
- 3) Tolerance of measurement of Vf is $\pm 0.05\text{V}$.
- 4) Packaging methods are available for selection, please refer to PACKAGING STANDARD.
- 5) Please refer to LED LAMP RELIABILITY TEST STANDARD for reliability test conditions.

Dimension Drawing

NOTES:

1. ALL DIMENSIONS ARE IN mm TOLERANCE IS ± 0.25 , UNLESS OTHERWISE NOTED.
2. BURR AROUND BOTTOM OF EPOXY MAY BE 0.5mm MAX.

3. Metodologia

A Linguística de *Corpus* (LC) é uma abordagem que possui como característica principal a análise empírica, a partir de padrões existentes em textos naturais, fazendo uso de ferramentas computadorizadas e dependendo de enfoques qualitativos e quantitativos (BERBER SARDINHA, 2004). Diferentemente das abordagens mais tradicionais, que determinam que a linguagem deve ser analisada de acordo com princípios determinados anteriormente, a LC primeiro observa o comportamento da linguagem em seu *habitat* natural (o texto) e depois teoriza sobre ela. Para esta abordagem da linguagem, existe uma correlação

entre as características linguísticas e contextuais, podendo-se observar mais facilmente padrões na língua sob análise⁷.

3.1 Compilação de *corpora*

O *corpus* de estudo para observação dos *datasheets* foi coletado a partir de documentos de componentes elétricos e áreas da Eletrotécnica sugeridas por dois engenheiros elétricos, de acordo com as necessidades mais proeminentes de leitura que os alunos desta área apresentam. Os *datasheets* podem ser facilmente encontrados na internet e podem ser baixados em formato eletrônico. Assim, separamos os textos em 11 componentes/áreas pertinentes à Eletrotécnica divididas em 11 arquivos txt, os quais limpamos conforme os passos propostos por Aluísio e Almeida (2006) removendo letras maiúsculas, números, pontuação e símbolos. Não removemos, contudo, as tabelas e figuras, uma vez que essas contêm informações extremamente relevantes e idiossincráticas dos *datasheets*. O total de *tokens* foi 29.718. Os componentes e áreas dos *datasheets* são os seguintes: capacitores, conversores, diodos, disjuntores, indutores, inversores fotovoltaicos, osciladores de cristal, resistores, transistores, transformadores e segurança.

Para a observação dos artigos científicos da Computação, foi compilado um *corpus* de cerca de 400.000 *tokens*, contendo 54 artigos científicos da Computação provenientes de seis periódicos da Association of Computer Machinery (ACM), datados de 2010 a 2014. Os títulos dos periódicos foram selecionados por dois professores especialistas da área de Informática, os quais lecionam em cursos técnicos e tecnológicos (Ensino Superior) de tal domínio.

3.2 Ferramentas de PLN para estudos léxico-estatísticos

As ferramentas extratoras utilizadas no estudo foram de natureza estatística e linguística, com o intuito de promover mais variedade nos resultados (VIEIRA; LOPES, 2010); dificilmente um extrator de candidatos a termos obtém resultados 100% confiáveis. Como exemplo podemos citar Teixeira (2010), que em seu estudo comparou quatro ferramentas de extração de termos diferentes e constatou que nenhum obteve mais do que

7 Reiteramos que para que as escolhas metodológicas fossem constituídas, foram realizados estudos-piloto (MONZÓN; FADANELLI, 2016) com *corpora* e com experimentos de leitura com os públicos-alvo discentes destas teses de doutorado.

20% de índice de acerto. As seguintes ferramentas computacionais foram escolhidas por serem de fácil uso e obtidas de forma gratuita na Internet: o *AntConc 3.2.4*, recurso que se utiliza mais de estatística de frequência; e o *TermoStat*, recurso que usa estatística linguística.

O AntConc é um *kit* de ferramentas de análise de *corpus* desenhado por Laurence Anthony (ANTHONY, 2005; 2004). É equipado com um concordanciador⁸, geradores de listas de frequência de palavras e a ferramenta utilizada para a extração, a função *Keyword* (oferece na tela a lista da candidatas a palavras-chave de um *corpus*, comparando o *corpus* de estudo com um outro *corpus* no mínimo cinco vezes maior - chamado de *corpus* de referência); o *corpus* de referência foi composto de textos extraídos do *corpus* COCA (*Contemporary Corpus of American English*) e do BNC (*British National Corpus*), pesquisando a ocorrência das 30 palavras mais frequentes na wordlist do *Corpus* COCA. O *corpus* de referência possui aproximadamente 211 mil *tokens*, mais do que cinco vezes o tamanho do *corpus* de estudo (SARDINHA, 2000). Os textos têm origens de diversos domínios, acadêmico, noticiários, literatura, internet, e passaram pela mesma limpeza de pontuação, etc. que o *corpus* de pesquisa.

O TermoStat (DROUIN, 2003) é um extrator de termos acessado gratuitamente na *web*, que primeiramente se baseia em outro programa, o *TreeTagger* (SCHMID, 1994) para *tokenizar*, lematizar e etiquetar gramaticalmente o *corpus*. Depois deste processo, o programa junta candidatos a termos (unigramas e bigramas) em uma lista de acordo com padrões gramaticais. O TermoStat fornece outras informações além da frequência, como por exemplo a classificação morfossintática dos termos, relações de combinabilidade e grafos de relações semânticas.

Essas duas ferramentas se complementam quando se trata de identificar e analisar padrões lexicais nos *corpora*, especialmente sob a ótica terminológica, a qual se sobressai nestes textos produzidos por especialistas. Salientamos, entretanto, que a seleção automática é complementada pela análise manual.

⁸ Um concordanciador, de acordo com Sardinha (2009, p. 9), é “uma ferramenta que: realiza concordâncias, ou listagens de uma palavra específica (ou nóculo) juntamente com parte do texto onde ocorreu. Oferece também listas de colocados, isto é, palavras que ocorreram perto do nóculo”.

4. Resultados e discussão

Apesar das particularidades e convencionais da Eletrotécnica e da Computação, no que diz respeito à composicionalidade lexical observada em gêneros textuais relevantes para as comunidades de cada um desses domínios, existem aspectos semelhantes. Desse modo, nesta seção analisaremos os padrões léxico-estatísticos neste contexto, atentando para suas potenciais contribuições para o ensino-aprendizagem de ESP na Educação Profissional.

4.1 Padrões léxico-estatísticos em artigos científicos da Computação

Desse modo, ao se levantarem semiautomaticamente as palavras gramaticais mais frequentes, utilizou-se o AntConc em parceria com o TermoStat, para então se fazer uma lista de termos que são candidatos a serem bases (HAUSMANN, 1990) para CEs; obtiveram-se resultados um pouco diferentes através dessas ferramentas (Tabelas 1 e 2). É importante salientar, ainda, que como o TermoStat calcula o *score* de especificidade de cada termo (*score spécificité*), sua lista acaba trazendo informações complementares em relação ao AntConc. Mesmo assim, a análise é semiautomática⁹, pois, neste estudo, sempre houve o julgamento humano ao percorrer as listas linha por linha e “limpá-las” de forma a garantir a confiabilidade dos termos. Isso porque o levantamento e a análise exclusivamente automáticos trazem alguns equívocos bastante visíveis como não sendo de linguagem natural.

Tabela 1 — Lista dos 10 candidatos a termos (unigramas) mais recorrentes na *wordlist* do AntConc.

termo	frequência
time	965
set	754
students	749
algorithm	701
problem	692
user	584
use	578
data	539
number	539
learning	526

⁹ “A Linguística de *Corpus* tem se ocupado, dentre outras coisas, de criar procedimentos automatizados para o levantamento da frequência lexical, mas atualmente ainda é necessária a intervenção humana para apurar os resultados obtidos pelos programas de consulta aos *corpora*” (DURAN, XATARA, 2006, p. 44).

Tabela 2 — Lista dos 10 termos (unigramas) mais recorrentes na lista do TermoStat.

termo	frequência	score spécificité
algorithm	886	123.21
use	962	114.28
graph	561	94.04
tag	499	86.16
participant	539	82.79
vertices	284	70.23
set	685	69.13
edge	507	68.11
node	298	67.74
model	628	66.81

Outro julgamento e análise necessários à seleção de CEs foi verificar se cada uma das candidatas (Quadro 3) ocorria em pelo menos três artigos científicos e gerar uma lista mais apurada (Quadro 4). No Quadro 3 estão apresentadas as 17 CEs mais frequentes geradas pelo TermoStat. A primeira delas é *display server*, a qual foi possível verificar, através do AntConc, que está presente em somente um dos textos do *corpus* (Figura 2). Através da funcionalidade denominada *concordance plot* visualiza-se a distribuição de uma CE, ou qualquer outro tipo de *cluster*, ao longo dos textos em que a ocorrência é observada (Figura 3). Dessa forma, a primeira CE do Quadro 3, ao ser analisada, é encontrada em apenas um artigo do *corpus*. Consequentemente, essa CE foi extraída da lista final. O mesmo ocorreu com *information goal*, sendo que *active learning* foi a primeira CE da lista que ocorre em vários textos. Nesse caso, o termo ocorre em três textos, sendo que é bem mais frequente em um deles (Figura 3). Finalmente, chegou-se a uma listagem de 7 CEs (Quadro 4), eliminando-se 10 delas da lista inicial.

Quadro 3 — Lista de 17 CEs geradas pelo TermoStat.

CEs
display server
information goal
fairness property
surrogate competition
display space
active learning

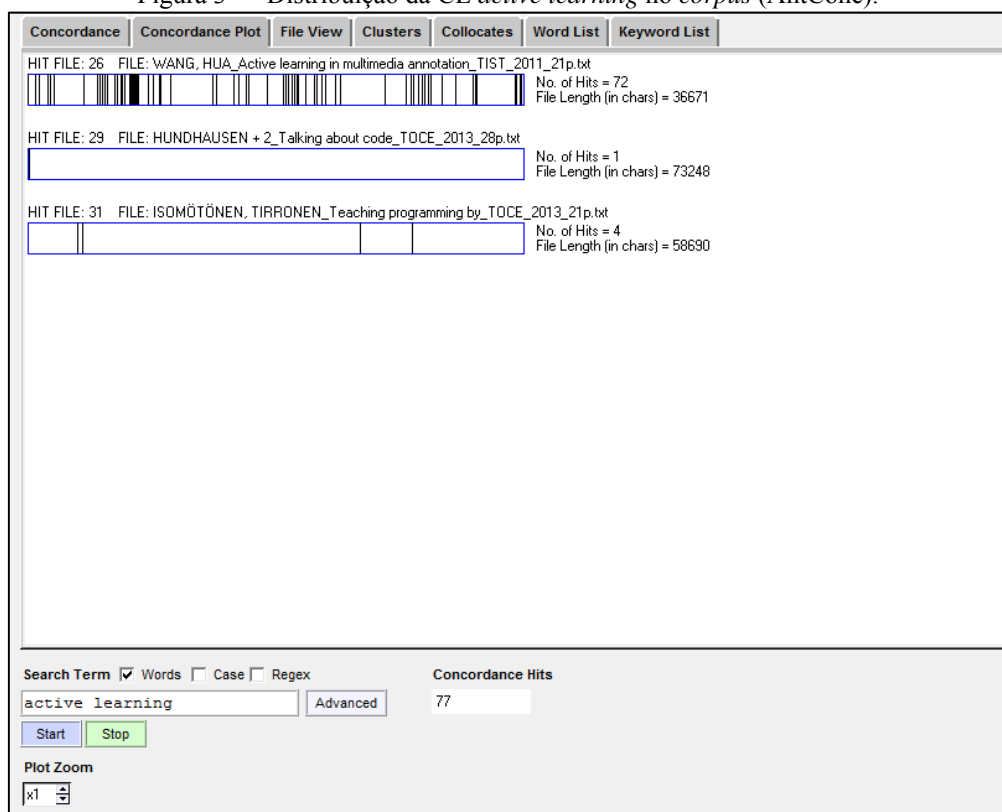
window broker
social network
user interface
nominal group
utility function
datum structure
relational structure
user interest
tag choice
broker software
high probability
social influence

Figura 2 — Distribuição da CE *display server* no *corpus* (AntConc).

The screenshot shows the AntConc interface with the following elements:

- Menu Bar:** Concordance, Concordance Plot, File View, Clusters, Collocates, Word List, Keyword List.
- File Information:** HIT FILE: 37 FILE: ARTHUR, OLSEN_Window broker_TOCHI_2012_21p.txt
- Statistics:** No. of Hits = 94, File Length (in chars) = 54922
- Search Settings:** Search Term: display server, Words checked, Case unchecked, Regex unchecked.
- Concordance Hits:** 94
- Plot Zoom:** x1

Figura 3 — Distribuição da CE *active learning* no corpus (AntConc).



Quadro 4 — Lista de 7 CEs depois da seleção manual.

CEs selecionadas
active learning (adj. + substantivo)
social network (adj. + substantivo)
user interface (substantivo + substantivo)
high propability (adj. + substantivo)
planar graph (adj. + substantivo)
polynomial time (adj. + substantivo)
data structure (substantivo + substantivo)

Levantou-se de forma automática que 31% das CEs são colocações adjetivas (substantivo + adjetivo)¹⁰ e 27% são nominais (substantivo + substantivo). Ademais, atestou-se que 24% dos termos são unigramas compostos por um substantivo. Logo, totaliza-se nessas três composições 82% dos termos listados pelo TermoStat (Tabela 3).

¹⁰ Reitera-se que Hausmann (1990), em sua tipologia de colocações, determina que uma colocação adjetiva é formada por substantivo + adjetivo, porém na Língua Inglesa essas posições se invertem.

Tabela 3 — Composição dos termos listados pelo TermoStat.

composição dos termos	frequência	porcentagem
adjetivo + substantivo	1.843	31%
substantivo + substantivo	1.586	27%
substantivo	1.417	24%
adjetivo + substantivo + substantivo	180	3%
substantivo + substantivo + substantivo	129	2%
adjetivo + adjetivo + substantivo	76	1%
	5.231	88%

Para melhor compreender as CEs de forma mais aprofundada, bem como o universo das terminologias da Computação, foi importante observar atentamente o comportamento dos termos que as constituíam como base (*base*, *node*). Para demonstrar como esses levantamentos são obtidos ao se submeter este *corpus* ao TermoStat, organizou-se a Tabela 4.

Tabela 4 — Características dos 10 termos mais frequentes como base de CEs (TermoStat).

Lema	frequência	score <i>spécificité</i>	variantes ortográficas	classe
algorithm	886	123.21	algorithm algorithms	substantivo
use	962	114.28	user users	substantivo
graph	561	94.04	graph graphs	substantivo
tag	499	86.16	tag tags	substantivo
participant	539	82.79	participant participants	substantivo
vertices	284	70.23	vertices	substantivo
set	685	69.13	set sets	substantivo
edge	507	68.11	edge edges	substantivo
node	298	67.74	node nodes	substantivo
model	628	66.81	model models	substantivo

Além disso, analisar as possibilidades combinatórias de um termo para gerar CEs é algo que agregou a este estudo, posto que as terminologias estabelecem as relações conceituais de uma determinada ciência (BARBOSA, 2009). Desse modo, quando se examinou o comportamento do termo (unigrama) mais frequente - *algorithm* -, notou-se uma gama bem interessante de relações de combinabilidade levantadas e quantificadas

estatisticamente pelo TermoStat (Figura 4). Salientamos, portanto, que somente as análises automáticas não nos trazem todas as ocorrências relevantes nos *corpora* pensando-se especialmente nas aulas de ESP. Desse modo, muito embora o termo *algorithm* componha uma CE em 30^a posição na listagem automática e 10^a posição na lista limpa manualmente, acreditamos que se trata de um conteúdo léxico-conceitual relevante para o contexto deste estudo. Observou-se, então, que o adjetivo *deterministic*, por exemplo, é o que apresenta o maior índice de especificidade ao coocorrer com *algorithm*. Investigando-se a aproximação desses vocábulos ao constituir o termo *deterministic algorithm*, observa-se que eles mantêm uma relação semântica relevante, uma vez que se trata, de acordo com o especialista consultado, de um tipo específico de algoritmo, o que pode ser observado na Figura 4.

Figura 4 — Combinatórias envolvendo o termo *algorithm*.

Termes en relation	deterministic (191.84) approximation (79.86) polynomial-time (63.68) linear-time (63.68) efficient (57.84) dynamic (36.52) greedy (34.7) randomize (28.91) multiplier (24.17) latter (23.13) state-of-the-art (23.13) whole (23.13) layout (21.41) k-stroll (9.21) first (7.66)
Inclus dans	approximation algorithm broker algorithm deterministic algorithm dynamic algorithm efficient algorithm first algorithm greedy algorithm k-stroll algorithm latter algorithm layout algorithm linear-time algorithm multiplier algorithm polynomial-time algorithm randomized algorithm state-of-the-art algorithm time algorithm

4.2 Padrões léxico-estatísticos em *datasheets*

Para obter as classes morfossintáticas que estão mais presentes nos *datasheets* usou-se o etiquetador *Claws Free Tagger* da Universidade de Lancaster, Reino Unido. O etiquetador aceita arquivos com a extensão txt e classifica cada palavra com códigos como NN1 para substantivos, por exemplo. Foram inseridos no etiquetador os termos resultantes dos extratores AntConc. Os termos apontados pelo TermoStat não precisaram ser inseridos, pois o

software já classifica morfologicamente os resultados. Como sabemos que este tipo de ferramenta não é 100% confiável por si só, um exame da etiquetagem foi feito e algumas classificações foram corrigidas manualmente. Após uma contagem das ocorrências de cada classificação morfossintática, as categorias que mais se destacaram, assim como nos ACCs, foram os substantivos, seguidos pelos adjetivos.

Segundo estudos previamente realizados em relação às características morfossintáticas de conceitos especializados, sabe-se que a grande maioria dos termos pertence à categoria dos substantivos (KRIEGER; FINATTO, 2004, p. 81). A alta quantidade de termos classificados como substantivos combina com o tipo de texto expositivo dos *datasheets*, cujo objetivo é apresentar peças e equipamentos elétricos e seus componentes.

A segunda categoria mais frequente foi a dos adjetivos, e aqui podemos notar a concordância com a característica de muitos *datasheets* que possuem textos descritivos e argumentativos em sua estrutura, onde os adjetivos desempenham um papel importante na linha textual.

O mesmo processo de julgamento para a escolha de CEs feito com os artigos científicos da área da computação foi aplicado para resultar tanto nos termos mais frequentes (Tabelas 5 e 6), como na composição morfossintática dos CEs.

Tabela 5 — Lista dos 10 termos mais recorrentes no AntConc.

Termo	Frequência
output	245
power	227
voltage	220
current	157
temperature	151
circuit	147
data	124
capacitors	123
input	115
high	102

Tabela 6 — Lista dos 10 termos mais recorrentes no TermoStat.

Candidato	Frequência	Variantes
output	260	output__outputs
voltage	249	voltage__voltages
capacitor	230	capacitor__capacitors
power	214	power
circuit	163	circuit__circuits
temperature	154	temperature__temperatures
input	127	input__inputs
converter	110	converter__converters
operation	110	operation__operations
product	105	product__products

Na extração da composição morfossintática dos CEs através do AntConc, os resultados excluíram os CEs que constituíssem o nome fantasia do aparelho ou dispositivo elétrico descrito pelos *datasheets*. A Tabela 11 mostra a listagem dos 17 CEs mais frequentes e que ocorrem em ao menos três textos.

Tabela 7 — Listagem dos 17 CEs mais frequentes no AntConc.

CE	Frequência	Em quantos datasheets aparece
Output voltage (noun + noun)	54	7
Single output (adj + noun)	26	4
Input voltage (noun + noun)	25	7
Short circuit (adj + noun)	24	7
Power factor (noun + noun)	23	4
Output power (noun + noun)	22	7
Voltage range (noun + noun)	22	4
Power supply (noun + noun)	21	6
Output current (noun + noun)	18	5
Supply voltage (noun + noun)	17	7
Phase noise (noun + noun)	16	4
Rated voltage (adj + noun)	15	6
Ac dc (noun + noun)	12	3
Frequency stability (noun + noun)	12	3

Silicon power (noun + noun)	10	3
Circuit current (noun + noun)	9	6
Power dissipation (noun + noun)	7	4

No TermoStat, as CEs geradas são parecidas (seguindo os mesmos critérios de exclusão de nomes de aparelhos ou dispositivos), com pequenas diferenças provavelmente devido à natureza sintático-estatística do TermoStat (Tabela 8).

Tabela 8 — Listagem dos 17 CEs mais frequentes no TermoStat.

CE	Frequência
output voltage (noun + noun)	59
single output (adj + noun)	26
short circuit (adj + noun)	24
input voltage (noun + noun)	21
power factor (noun + noun)	19
output power (noun + noun)	17
phase noise (noun + noun)	16
power supply (noun + noun)	13
voltage range (noun + noun)	14
supply voltage (noun + noun)	14
rated voltage (adj + noun)	12
frequency stability (noun + noun)	12
output current (noun + noun)	11
ac dc (noun + noun)	10
silicon power (noun + noun)	8
circuit current (noun + noun)	8
power dissipation (noun + noun)	5

As CEs sintaticamente classificadas pelos extratores apresentam nos *datasheets* uma maioria de ocorrências do padrão substantivo + substantivo, embora haja ocorrências de adjetivo + substantivo. As implicações destes resultados e dos resultados mostrados na seção sobre a análise dos artigos acadêmicos serão comentadas a seguir.

4.3 Lições aprendidas com a comparação dos dois gêneros textuais

Se olharmos para a composição morfossintática dos CEs mais frequentes nos *datasheets* apontados por ambos extratores, veremos que quase todos pertencem à combinação substantivo + substantivo. Essa característica se relaciona com a microestrutura dos *datasheets*, que apresenta uma descrição e exposição de um produto, a fim de informar sobre seus componentes; porém, existe uma representação muito relevante do padrão adjetivo + substantivo na composição dos CEs, evidenciando uma outra possibilidade de uso dos *datasheets*: usados para fazer propaganda e vender o dispositivo.

Outro ponto importante a ser considerado diz respeito ao significado e relevância das CEs mais frequentes. Assim como no caso dos artigos científicos de Computação (rever exemplo sobre o termo *algorithm* em 4.1), as colocações não parecem ser aleatórias, e sim determinam conceitos bem específicos que fazem parte dos componentes elétricos. Os termos *output voltage/input voltage*, por exemplo, são dois tipos diferentes de tensão elétrica (saída e entrada). Outra característica interessante é que muitos dos mesmos termos individuais se combinam entre si para formar conceitos diferentes (*output voltage; output power; output current*). E muitos termos individuais se posicionam de forma diferente na CE, e, embora sejam substantivos, acabam por assumir uma função adjetiva no significado da CE. Por exemplo: *supply voltage*, onde *voltage* é um substantivo; e *voltage range*, em que *voltage* assume uma função mais adjetivada.

A implicação mais evidente destes resultados é de que, embora sintaticamente a constituição dos CEs presentes em artigos acadêmicos e *datasheets* pareça se apresentar diferente, quando funcionalmente inseridas no gênero textual, os padrões substantivo + substantivo e adjetivo + adjetivo parecem se equilibrar.

Outra implicação muito importante, mais especificamente sobre os *datasheets*, é de que na verdade o vocabulário a ser trabalhado pelo professor não precisa ser muito vasto, já que as CEs mais frequentes se combinam entre si para formar conceitos diferentes. O significado isolado dos termos principais deve ser trabalhado como uma introdução para depois se enfatizar as combinatórias e os conceitos que as CEs expressam.

Já no caso dos ACCs, as CEs apresentam uma gama de vocabulário mais variado, significando que neste domínio científico o trabalho com os termos individuais e com as combinatórias deve receber a mesma atenção.

5. Conclusão

Em vista dos dados apresentados nas seções anteriores, foi possível verificar que a constituição de CEs entre os artigos acadêmicos de Computação e os *datasheets* de Eletrotécnica apresentam-se diferentes à primeira vista: nos artigos acadêmicos da área da Computação as combinações adjetivo + substantivo aparecem com maior frequência na escolha dos CEs que apareciam em três ou mais artigos acadêmicos. Já nos *datasheets*, a combinação substantivo + substantivo é mais recorrente nos CEs mais frequentes, sendo que a ocorrência do padrão adjetivo + substantivo também aparece, porém em menor escala.

No entanto, quando olhamos para as CEs e seu funcionamento dentro do gênero textual a que pertencem, percebemos que muitas das CEs que nos *datasheets* haviam sido classificadas como substantivos, se posicionam funcionalmente como adjetivos.

Todas estas descobertas são importantes para o professor de ESP que deseja produzir material e preparar aulas que realmente atendam às necessidades reais de seus alunos. Certamente, um destaque especial deve ser dado ao tratamento dos CEs nas diferentes áreas, sempre chamando a atenção do aprendiz para a conexão que se estabelece entre os padrões linguísticos e o gênero textual com a qual se está lidando.

Através do levantamento automático, utilizando-se as ferramentas AntConc (ANTHONY, 2005; 2004) e TermoStat (DROUIN, 2003), verificou-se que as CEs desempenham um papel importante no cenário terminológico dos textos de ambos *corpora* pesquisados. As CEs mais frequentes de ambos gêneros textuais mencionados neste trabalho representam um exemplo muito importante de elementos a serem incluídos em atividades de aprendizagem e trabalho com uma língua especializada: os exemplos citados, com suas respectivas combinatórias, foram apontados por especialistas das áreas de Eletrotécnica e Computação como conceitos essenciais para o estudo das áreas em questão.

Por fim, percebemos que o escopo deste estudo ainda apresenta inúmeras possibilidades de expansão, através da análise de diferentes gêneros textuais de diversos domínios científicos, com o objetivo de profissionalizar a montagem de ferramentas, aulas e material didático para o ensino da Língua Inglesa com propósitos específicos (ESP).

Referências bibliográficas

ALUÍSIO, S.; ALMEIDA, G. O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, São Leopoldo/RS, v. 4, n. 3, p. 155-177, 2006.

ANTHONY, L. AntConc: a learner *corpus* and a classroom friendly multi-platform *corpus* analysis toolkit. **Proceedings of Interactive Workshop on Language E-Learning**, Tokyo/Japan, p. 7-14, 2004.

ANTHONY, L. AntConc: design and development of a freeware *corpus* analysis toolkit for the technical writing classroom. **Proceedings of International Professional Communication Conference IPCC**, pp. 729-737, 2005.
<https://doi.org/10.1109/ipcc.2005.1494244>

BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 2003.

BARBOSA, M. A. Terminologia aplicada: percursos interdisciplinares. **Polifonia** (UFMT), v. 4, p. 29-44, 2009.

BAZERMAN, C. **Gêneros textuais, tipificação e interação**. São Paulo: Cortez, 2005.

BIBER, D. The comprehensive analysis of register variation. In: **Dimensions of register variation: a cross-linguistic comparison**. Cambridge University Press, 1995.
<https://doi.org/10.1017/CBO9780511519871.002>

BIBER, D.; CONRAD, S.; REPPEN, R.; BYRD, P.; HELT, M. Speaking and writing in the University: A multidimensional comparison. **TESOL Quarterly**, v.36, p. 9-48, 2002.
<https://doi.org/10.2307/3588359>

BIBER, D.; CONRAD, S.; CORTES, V. If you look at...: lexical bundles in university teaching and textbooks. **Applied Linguistics**, v. 25, n. 3, p. 371-405, 2004.
<https://doi.org/10.1093/applin/25.3.371>

CABRÉ, M. T. La Teoría Comunicativa de la Terminología: una aproximación lingüística a los términos. **Revue Française de Linguistique Appliquée**, v. 14, n. 2, p. 9-15, 2009.

CABRÉ, M. T. Textos especializados y unidades de conocimiento: metodología y tipologización. In: GARCÍA PALACIOS, J. and FUENTES, M. T. (Eds.). **Texto, terminología y traducción**. Salamanca: Ediciones Almar, 2002.

CIASPUCIO, G.; KUGUEL, I. Hacia una tipología del discurso especializado: aspectos teóricos y aplicados. In: PALACIO, J.G. e FUENTES, M.T. (eds.). **Entre la terminología, el texto y la traducción**. Salamanca, Almar, p. 37-73, 2002.

DEWEY, F. R. A Complete Guide to Datasheets. **Sensors Magazine**, 1998. Disponível em: <http://www.allegromicro.com/~media/Files/Technical-Documents/pub26000-Complete-Guide-To-Datasheets.ashx>. Acesso: 01 dez. 2014.

DROUIN, P. Term extraction using non-technical *corpora* as a point of leverage. **Terminology**, v. 9, n. 1, p. 99-117, 2003. <https://doi.org/10.1075/term.9.1.06dro>

DURAN, M.; XATARA, C. A metalexicografia pedagógica. **Cadernos de Tradução (UFSC)**, v.2, n. 18, p. 41-66, 2006.

FADANELLI, S. Arquitetura de um glossário terminológico Inglês-Português na área de Eletrotécnica: rumo a uma Terminografia Didático-Pedagógica. **Anais do III Congresso Internacional Linguagem e Interação**, São Leopoldo/RS, 2015, p.165-166.

FINATTO, M. J.; ZILIO, L. **Textos e termos por Lothar Hoffmann** - um convite para o estudo das linguagens técnico-científicas. Porto Alegre: Palotti, 2015.

FIRTH, J. R. **Papers in Linguistics 1934-1951**. London: Oxford University Press, 1957.

HAUSMANN, F. J. Le dictionnaire de collocations. In: HAUSSMAN, F. J. [et al.]. **An international encyclopedia of lexicography**. V. 1. Berlin, New York: Walter de Gruyter, 1990, p. 1010-1019.

HOFFNAGEL, J. C.; DIONÍSIO, Â. P. (Orgs.). **Gêneros textuais, tipificação e interação**. São Paulo: Cortez, 2011.

HUTCHINSON, T.; WATERS, A. **English for Specific Purposes**. Cambridge University Press, 1987. <https://doi.org/10.1017/CBO9780511733031>

JENKINS, J. English as a *lingua franca*: from the classroom to the classroom. **ELT Journal**, Special Issue, p. 486-494, 2012. <https://doi.org/10.1093/elt/ccs040>

KOCH, I.; ELIAS, V. M. **Ler e compreender**: os sentidos do texto. São Paulo: Contexto, 2014.

KOCH, I. **O texto e a construção dos sentidos**. São Paulo: Contexto, 2005.

L'HOMME, M. C. Understanding Specialized Lexical Combinations. **Terminology**, v. 6, n. 1, p. 89-110, 2000. <https://doi.org/10.1075/term.6.1.06hom>

L'HOMME, M. C.; BERTRAND, C. Specialized Lexical Combinations: Should they be described as Collocations or in Terms of Selectional Restrictions? **Proceedings of Euralex 2000** - European Association of Lexicography, Stuttgart-Germany August 8th-12th.

MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, A. P.; MACHADO, A.R.; BEZERRA, M.A. (orgs.). **Gêneros Textuais e Ensino**. Rio de Janeiro: Lucena, 2003.

MARCUSCHI, L. A. **Produção textual, análise de gêneros e compreensão**. São Paulo: Parábola Editorial, 2008.

MEURER, J. L.; BONINI, A.; MOTTA-ROTH, D. **Gêneros. Teorias, métodos, debates**. São Paulo: Parábola Editorial, 2005.

MONZÓN, A.; FADANELLI, S. Leitura de textos especializados anglófonos no Ensino Técnico: idiossincrasias terminológicas e pedagógicas sob a perspectiva da Linguística de *Corpus*. **LínguaTec**, v. 1, n. 1, 20165, p. 1-28.

PALTRIDGE, B.; STARFIELD, S. **Thesis and dissertation writing in a second language**. London: Routledge, 2007.

ORENHA-OTTAIANO, A. Semelhanças e diferenças entre colocações e colocações especializadas. In: ORTIZ-ALVAREZ, M. L. (Org.). **Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia**. 1 ed. Campinas: Pontes, 2012, p. 147-163.

PAIVA, V. L. M. O; PAGANO, A. S. English in Brazil with an outlook on its function as a language of science. In: AMMON, U. (Ed.). **The dominance of English as a language of science**. Mouton de Gruyter, 2001, p. 425-445. <https://doi.org/10.1515/9783110869484.425>

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

SARDINHA, T. B. **Linguística de Corpus: Histórico e Problemática**. D.E.L.T.A., v. 16, n. 2, p. 323-367, 2000.

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

SCHMID, H. Probabilistic part-of-speech tagging using decision trees. **Proceedings of the International Conference on New Methods in Language Processing**. Manchester, United Kingdom, p. 44-49, 1994.

SANTOS, A. Análise de colocações adverbiais em inglês para negócios. In: **Corpora no Ensino de Línguas Estrangeiras**. VIANA, V.; TAGNIN, S. E. O. São Paulo: Hub Editorial, 2010, p. 97-136.

SWALES, J. M. **Genre Analysis - English in Academic and Research Settings**. Cambridge University Press, 1990.

SWALES, J. M.; FEAK, C. B. **Academic Writing for Graduate Students: Essential Skills and Tasks**. 2nd edition. Ann Arbor, MI: University of Michigan Press, 2004.

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

VIEIRA, R.; LOPES, L. Processamento de Linguagem Natural e o tratamento computacional de linguagens científicas. In: PERNA, C. B. L.; DELGADO, H. O. K.; FINATTO, M. J. B. (Org.). **Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa**. Porto Alegre: EDIPUCRS, 2010, p. 128-151.

Artigo recebido em: 13.12.2016

Artigo aprovado em: 06.04.2017

Português para Fins Acadêmicos sob o aporte da Linguística de *Corpus* e do Processamento de Linguagem Natural

Portuguese for Academic Purposes with the contribution of Corpus Linguistics and Natural Language Processing

Cristina Becker Lopes Perna*

Lucelene Lopes**

Lucas Zambrano Rollsing***

RESUMO: O presente artigo intenta apresentar um projeto de interface entre Linguística de *Corpus* e Processamento de Linguagem Natural (PLN) em andamento na Pontifícia Universidade Católica do RS. Tal projeto consiste na exploração, via *corpus* escrito, de teses e dissertações em Linguística oriundas do Programa de Pós-Graduação em Letras da referida instituição de ensino superior, por meio de um software chamado ExATO (LOPES, 2012). A partir dessa ferramenta produzida em PLN, podemos ter em mãos uma série de recursos linguísticos que permitem a prossecução da nossa análise, tais como, a detecção de Hierarquia de Conceitos, Listas de Termos Extraídos, Concordanciador de Termos e Nuvens de Conceitos a fim de embasarmos uma proposta de ensino de Português como Língua Adicional, com vistas à proficiência dentro do gênero acadêmico por alunos não-falantes de Língua Portuguesa. Trazemos uma série de resultados já disponibilizados pelo ExATO, os quais fomentam a discussão acerca dos temas visualizados a partir dos dados obtidos, bem como fomentam a interdisciplinaridade necessária entre Linguística e Ciência da Computação, no que concerne à descrição e explicação do viés pragmático da LP em grande escala dentro dos limites da esfera discursiva universitária.

PALAVRAS-CHAVE: Linguística de

ABSTRACT: The article hereby intends to present an interface project between Corpus Linguistics and Natural Language Processing (NLP) in progress at the Pontifical Catholic University of RS. This project consists in the exploration, through a written corpus, of theses and dissertations in Linguistics from the Postgraduate Program in Letters of said institution of Higher Education, through software called ExATO (LOPES, 2012). From this tool produced in NLP, we can have in hand a series of linguistic resources that allow the continuation of our analysis, such as the detection of Concept Hierarchy, Extracted Term Lists, Concordance of Terms and Concepts Clouds in order to base a proposal of teaching of Portuguese as an Additional Language, with a view to the proficiency within the academic genre by non-Portuguese speakers. We bring a series of results already available by ExATO, which foment the discussion about the subjects visualized from the obtained data, as well as foment the necessary interdisciplinarity between Linguistics and Computer Science, as far as the description and explanation of the pragmatic bias of the LP on a large scale, within the limits of the university discursive register.

KEYWORDS: Corpus Linguistics. Natural

*Doutora em Linguística, Escola de Humanidades – Programa de Pós-Graduação em Letras, PUCRS.

**Doutora em Ciência da Computação, PUCRS.

***Mestrando em Linguística, Escola de Humanidades – Programa de Pós-Graduação em Letras, PUCRS.

<i>Corpus</i> . Processamento de Linguagem Natural. Português para Fins Acadêmicos.	Language Processing. Portuguese for Academic Purposes.
---	--

1. Considerações iniciais

O presente artigo visa apresentar os resultados iniciais de um estudo sendo desenvolvido na interface entre dois programas de Pós-Graduação da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), a saber, o Programa de Pós-Graduação em Letras (PPGL), na sua área de concentração em Linguística, através do Grupo de Pesquisa sobre o Uso e Processamento de Língua Adicional (UPLA)¹, e o Programa de Pós-Graduação em Ciência da Computação (PPGCC), através do Grupo de Processamento em Linguagem Natural². Cabe mencionar que o grupo UPLA opera sob a perspectiva pragmática para descrição e análise de suas pesquisas, utilizando o aporte da Linguística de *Corpus* (LC). Essa união epistemológica assumida pelo grupo proporciona condições de compilar e descrever *corpora* de Língua Portuguesa, seja em registro acadêmico, como veremos ao longo do presente trabalho, seja em qualquer esfera discursiva.

Essa profícua relação é bem apontada por Finatto, Lopes e Ciulla (2015, p. 43), ao atestarem que o diferencial da LC: “[...] é o estudo da língua em uso, verificado o uso sempre em grande escala, com apoio informatizado e tratamento estatístico”, e acrescentam que: “Seu principal objetivo é a descrição de usos”. Logo, através da metodologia de análise de *corpora* proposta pela LC em conjunto com a Pragmática, obtemos dados reais sobre o registro acadêmico, em condições de extrairmos deles considerações que possam nos levar a um entendimento mais profundo sobre os gêneros do discurso.

Nosso objeto de estudo, portanto, são textos acadêmicos produzidos no âmbito do PPGL, dentro da área de concentração em Linguística, conforme descreveremos na Metodologia. As pesquisas sobre o Português para Fins Acadêmicos (PFA) vêm crescendo nos últimos anos e, diante dessa nova demanda, é necessário que o ensino e a aprendizagem da língua portuguesa se tornem mais especializados, de acordo com os propósitos específicos para os quais a língua será utilizada.

Ressaltamos que este artigo está sob a égide das demais pesquisas já iniciadas no PPGL-PUCRS na área de PFA, vinculadas ao Grupo UPLA, o qual almeja ampliar os estudos

¹<http://dgp.cnpq.br/dgp/espelhogrupo/0705653235733182>

²<http://dgp.cnpq.br/dgp/espelhogrupo/9398077368852737>

sobre Português para Fins Acadêmicos (PFA), ao desenvolver *corpora* de estudo e metodologias de análise, para futuramente serem empregados como uma pedagogia de ensino de Português como Língua Adicional (PLA) para alunos estrangeiros em mobilidade acadêmica que necessitem de proficiência em PFA.

Com isso em mente, Molsing e Perna (2015, p.3), extraem e postulam as diretrizes para o trabalho em PFA, quando atestam que o Grupo UPLA: “[...] têm adotado, por enquanto, construtos teóricos similares àqueles que guiam a pesquisa na área de ESP (*English for Specific Purposes*), fazendo as adaptações específicas para a língua portuguesa, para a cultura brasileira e para a realidade da educação superior brasileira³”.

Tendo delineado nossa motivação, organizamos nosso artigo a partir da seguinte estrutura: na seção 2 apresentaremos os construtos da Linguística de *Corpus* e do Processamento de Linguagem Natural, que dão suporte à análise dos textos da produção acadêmica do PPGL-PUCRS; na seção 3 apresentaremos a metodologia empregada na compilação do *corpus* de estudo, denominado CorpAcad, e como se deu o processamento pelo software ExATO; na seção 4 faremos a discussão dos dados obtidos através do software em questão e, por fim, na seção 5 apresentaremos as considerações finais.

2. A Linguística de *Corpus* e o Processamento de Linguagem Natural

As definições das áreas de Linguística de *Corpus* (LC) e de Processamento de Linguagem Natural (PLN) são sujeitas a certas discrepâncias segundo a fonte consultada. No entanto, é um consenso entre as principais fontes da área (e.g., BIBER et al., 1998; MANNING; SCHATZ, 1999; DALE et al., 2000; KENNEDY, 1998; MITKOV, 2003; TEUBERT; CERMÁKOVÁ, 2007) que se trata de duas áreas de pesquisa de origens distintas, mas que colaboram estreitamente em diversas iniciativas práticas e teóricas. Logo, nos parágrafos que se seguem a este, conceituaremos ambas as áreas e as caracterizaremos, a fim de contextualizarmos a presente pesquisa.

O PLN é “uma área de Ciência da Computação, mais especificamente, da área de Inteligência Artificial, que estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais”

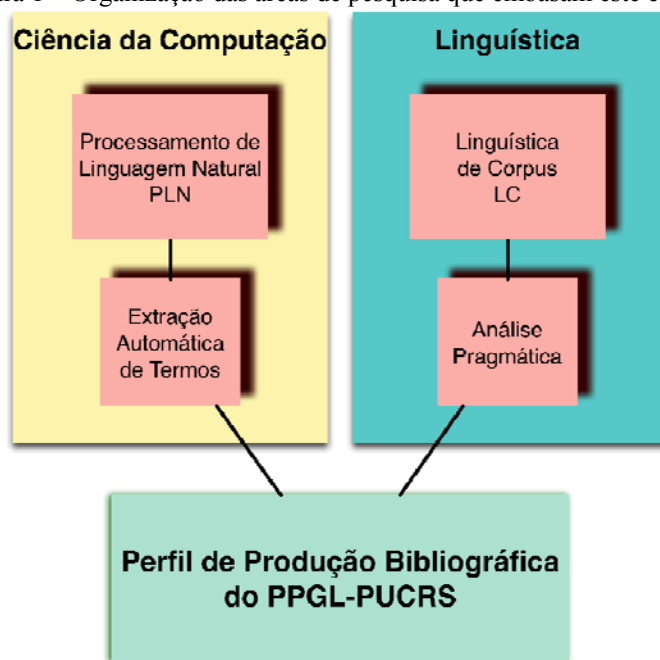
³ Texto original: “[...] we have adopted, for the time being, similar theoretical constructs as those that guide the research in the area of ESP, making specific adaptations for the Portuguese language, Brazilian culture and the reality of Brazilian higher education.”

(LOPES; VIEIRA, 2010, p. 184). A partir disso, podemos afirmar que o PLN tem por propósito buscar o desenvolvimento de técnicas, algoritmos e, conseqüentemente, programas de computador que possam tratar informações complexas expressas por linguagens que possuem todas as características que são encontradas na expressão humana, seja ela escrita ou falada. O PLN se destaca de outras áreas usuais da Ciência da Computação que tratam linguagens ditas formais, onde os conceitos não se prestam à interpretação subjetiva ou à existência de ambigüidade, como é o caso das linguagens de programação ou mesmo das expressões matemáticas ou de lógica de primeira ordem.

Já a LC é uma área da Linguística, mais especificamente, da área de Linguística Aplicada, que não pretende desenvolver técnicas, algoritmos ou programas capazes de analisar a linguagem natural dos textos acadêmicos por nós reunidos, e sim valer-se destas ferramentas teóricas e práticas para promover um estudo sistemático sobre grandes volumes de expressões humanas. Conceitualmente, essas expressões também podem ser faladas, porém é mais frequente limitar-se a textos escritos. Esses conjuntos de textos são denominados *corpora* (no singular: *corpus*) e representam amostras de expressões relativas a entidades específicas. Dessa forma, um *corpus* é uma maneira prática bastante eficaz e usualmente livre de preconceitos de descrever, através de exemplos de utilização, uma determinada entidade, seja ela uma língua, um domínio, um grupo ou uma pessoa. Dessa forma, o que usualmente define um *corpus* é o critério com o qual ele é construído. Caso construa-se um *corpus* com os textos de um determinado autor, teremos uma amostra do tipo de discurso utilizado por este autor. Conseqüentemente, a análise deste *corpus* permitirá chegar a conclusões sobre este autor de forma sistemática e sem o viés de uma análise tradicional.

Apresentadas as duas áreas, a Figura 1 apresenta esquematicamente a relação entre as áreas envolvidas (LC e PLN). Especificamente, detalha-se que dentro de PLN utiliza-se como suporte a Extração Automática de Termos, enquanto que dentro de LC utiliza-se a análise pragmática de *corpora*.

Figura 1 – Organização das áreas de pesquisa que embasam este estudo.



Fonte: elaborada pelos autores.

De acordo com o esquema da Figura 1, podemos ter um resumo sobre a organização teórica do trabalho em questão. Essa estruturação nos conduzirá a análise e descrição do perfil da produção intelectual do Programa de Pós-Graduação em Letras da PUCRS, a fim de dar continuidade aos estudos que se debruçam sobre o português acadêmico e, por consequência disto, sobre o Português como Língua Adicional.

Na seção seguinte, explicitaremos como se procedeu com a extração automática de termos relevantes do *corpus* do PPGL, a partir da ferramenta de software ExATO.

3. Metodologia

Tomamos a Linguística de *Corpus* (LC) como a abordagem necessária para compilarmos o CorpAcad, bem como para a sua descrição e análise pragmática; conjuntamente a isso, utilizamos o software elaborado em PLN, a saber, o Extrator Automático de Ontologias⁴ (ExATO), para o processamento dos dados. Escolhemos o

⁴A ferramenta ExATO na sua versão atual é uma evolução que inclui diversas otimizações do ponto de vista de informática, bem como a capacidade de estender o seu escopo de aplicação a corpora com textos em Inglês, enquanto que a versão inicial, denominada ExATO_{LP} (Extrator Automático de Ontologias em Língua Portuguesa), só era capaz de extrair termos relevantes de corpora com textos em Português (LOPES et al., 2009).

software ExATO, desenvolvido por Lopes desde sua tese de doutorado (2012), no PPGCC da Pontifícia Universidade Católica do Rio Grande do Sul, pois se trata de um software robusto e moderno, capaz de processar grandes quantidades de dados em linguagem natural. O ExATO é um extrator automático de termos relevantes (sintagmas nominais/verbais) que permite traçar um panorama do *corpus* de estudo, através da detecção de Hierarquia de Conceitos e outros recursos linguísticos, tais como Listas de Termos Extraídos, Concordanciador de Termos e Nuvens de Conceitos. Ele processa *corpora* em Português e Inglês, utilizando métodos linguísticos e estatísticos da área de PLN (LOPES; FERNANDES; VIEIRA, 2016).

De acordo com Lopes (2012, p. 33), conceito “é uma generalização associada a uma ideia, podendo ter várias manifestações textuais”. Devido a um conjunto de regras heurísticas, o ExATO nos fornece com precisão sintagmas nominais representativos para o domínio em questão, sob a forma de conceitos. Esses conceitos são gerados em listas ordenadas por grau de relevância, e podem ser visualizados de diferentes formas, como por exemplo, nuvens de conceitos ou árvores hiperbólicas. No contexto deste trabalho serão utilizadas as listas textuais e as nuvens de conceitos, cada qual com seu propósito, uma vez que a lista de conceitos abarca um número muito mais expressivo de sintagmas nominais, enquanto que a nuvem nos fornece somente os 150 sintagmas mais relevantes, mas os disponibiliza graficamente.

De um ponto de vista metodológico, o processo de extração realizado pelo ExATO pode ser detalhado em cinco etapas bem distintas:

- Anotação sintática e estrutural do texto, através de um *PoS-tagger* que anota, dentre outras informações, a classe gramatical de cada palavra do texto (substantivo, preposição, pronome, etc.) e um *parser* que anota as funções gramaticais desempenhadas por conjuntos de palavras (sujeito, verbo, objeto, etc.);
- Detecção e refinamento de sintagmas nominais através de heurísticas de base linguística que permitem fornecer termos de alto valor terminológico;
- Computação de um índice de relevância para cada sintagma nominal detectado através de um método de forte base estatística;
- Aplicação de um ponto de corte às listas de termos extraídos e organizados segundo seu índice de relevância de forma a escolher dentre todos os termos

aqueles que são suficientemente relevantes para serem considerados conceitos do *corpus* alvo;

- Geração de recursos linguísticos sofisticados que vão desde uma simples lista de conceitos até sofisticadas visualizações como é o caso, neste artigo, das nuvens de conceitos.

A anotação sintática e estrutural dos textos foi feita através do *parser* PALAVRAS⁵ (BICK, 2000), que executa tanto a tarefa de *PoS-tagging* quanto à estruturação gramatical das frases. A detecção de sintagmas nominais, por sua vez, segue um processo proposto por Lopes e Vieira (2012) e engloba várias heurísticas, que vai desde a eliminação de símbolos até a detecção de sintagmas implícitos. O cálculo do índice de relevância está baseado na utilização de *corpora* contrastantes com o uso do índice *tf-dcf* (*term-frequency, disjoint corpora frequency*), uma métrica com resultados mais efetivos que as atuais alternativas existentes (LOPES; FERNANDES; VIEIRA, 2016). Uma vez calculados os índices de relevância dos termos, estes podem ser ordenados pelo índice e então se aplica uma política de ponto de corte que escolhe quais termos possuem os valores de *tf-dcf* altos o suficiente para serem considerados relevantes para o domínio, ou seja, conceitos do *corpus* alvo (LOPES; VIEIRA, 2015). Finalmente, a geração de recursos linguísticos inclui um concordanciador, além das listas, nuvens e árvores citadas anteriormente. Considerando o material utilizado nesse artigo, a seguir trazemos exemplos das formas de saída do ExATO, a fim de ilustrarmos os resultados iniciais já obtidos.

Para a construção do CorpAcad, selecionamos o período de produção bibliográfica de 2006 a 2016 do PPGL da PUCRS. Já contamos com mais de 200 textos, dentre dissertações e teses que versam sobre Linguística, finalizando assim a etapa de compilação. Podemos conferir na Tabela 1 abaixo, a quantia de teses e dissertações, o número de sentenças, *tokens*⁶ e termos (sintagmas nominais) de acordo com cada ano, a quantia de conceitos segundo o ponto de corte estabelecido pelo ExATO e a representação percentual desses conceitos no CorpAcad como um todo. Cabe salientar que o período selecionado

⁵Mais detalhes em: <http://visl.sdu.dk/~eckhard/pdf/PLP20-amilo.ps.pdf>.

⁶“Uma única unidade linguística, na maioria das vezes uma palavra, embora dependente do sistema de codificação utilizado, uma única palavra pode ser dividida em mais de um token, por exemplo, em inglês “*he’s*” (*he + ’s*)”. Traduzido e adaptado de Baker, Hardie e McEnery (2006, p. 159).

contempla o início do arquivamento digital das teses e produções dos programas de pós-graduação da PUCRS. A partir do ano de 2006, podemos acessar as produções intelectuais da universidade, ainda que nesse mesmo ano, nem todas as produções tivessem sido armazenadas virtualmente. No entanto, conseguimos coletar o que estava disponível para cada ano, sendo crescente o número de trabalhos disponíveis até os dias de hoje.

Tabela 1 – Dados Gerais do CorpAcad.

Ano	Teses	Dissert.	Sentenças	<i>Tokens</i>	Termos/SN	Conceitos Ponto de Corte	Porcent.
2006	5	3	13483	347225	18803	1640	9%
2007	5	13	30758	700224	35971	3275	9%
2008	9	13	44695	1017660	49479	4796	10%
2009	10	13	45936	992365	48288	4636	10%
2010	6	12	33619	738295	37509	3680	10%
2011	9	13	44181	954984	48884	4534	9%
2012	5	13	35246	828666	46699	4096	9%
2013	8	12	32390	817770	43070	3777	9%
2014	11	13	43825	959624	50284	4460	9%
2015	14	16	75614	1722420	84942	5439	6%
2016	0	3	4106	96355	5985	309	5%
Geral	82	124	403853	9175588	333429	8056	2%

Fonte: elaborada pelos autores.

O CorpAcad já processado através do ExATO pôde nos propiciar, para cada ano de produção bibliográfica do PPGL, as nuvens e listas de conceitos relevantes para a área da Linguística, assim como uma nuvem e uma lista geral para o período total observado. Trazemos, como exemplo, as nuvens de conceitos (Figura 2) e a lista dos 20 termos mais relevantes (Quadro 1) referentes respectivamente a cada um dos anos de 2006 a 2016 e ao período de onze anos de produção.

Figura 2 – Nuvens de conceitos por ano.



Fonte: extraído da análise no programa ExATO.

Quadro 1 – Lista dos 20 termos mais relevantes por ano.

2006	2007	2008	2009
Discurso	Verbo	Criança	Criança
Enunciados	Discurso	falante	Oração
oração	Crianças	Consciência Fonológica	Verbos
Verbos	Aprendizes	Verbo	Adjetivos
Alocutário	sinonímia	Expressões Idiomáticas	Consciência fonológica
enunciação	falante	Implicatura	Sintagma
Encadeamentos	Processamento Auditivo	Orações	Discurso
Pronomes	mulher	aprendiz	Síndrome de Down
oração relativa	enunciado	Substantivos	falante
interlocutor	oclusiva dental	Discurso	Vogais
criança	Pragmática	palavra gramaticais	Advérbios
Enunciador	item lexical	enunciado	enunciado
Ducrot	palatalização de oclusiva dental	Pragmática	Brasil
Espanhol	Implicatura	Editoriais	João
r forte	UG	Informante	dialeto
Onset	Grupo Experimental	Grupo Experimental	Guarda chuva
r fraco	orelha	Epêntese	Encadeamentos
sentido de enunciado	GJTS	léxico	postônica
Retórica	paciente	GC	Flor
encadeamento argumentativo	oração	Português Brasileiro	pronomes

2010	2011	2012	2013
Consciência Fonológica	Criança	Discurso	Crianças
Discurso	Discurso	Enunciados	Discurso
enunciado	Verbo	Emoção	Enunciados
criança	Narrativa	Metáforas	HR
encadeamento	Mãe	falante	Tempos Verbais
falante	enunciado	interlocutor	Surdos
Informante	Latim	Encadeamento	pesquisadora
Vogais	nomeação	signo	crianças bilíngues
aprendiz	traço distintivo	corpus	Enunciação
conjunção	DFE	ato	Signo
Negação	menino	Diálogo	falante
enunciação	Vogal	input	Compreensão de textos
paciente	Enunciação	Serviços Sexuais	Acentos
verificação acústica	Adultos	Implicatura	Informante
Sílaba	Sílaba	mulher	Elisão
Ducrot	Menina	enunciação	interlocutor
narrativa	inteligibilidade	ANL	encadeamento
Redução vocálica	gravidade de DFE	Assibilação	Zona rural
Francês	pesquisadora	aluna	ANL
verificação perceptual	alocutário	Reescrita	Brasil

2014	2015	2016	2006-2016
Discurso	Discurso	Discurso	Discurso
Criança	Criança	ANL	Criança
Enunciado	Enunciado	Ducrot	Enunciado
Enunciação	interlocutor	diálogo	Verbo
Adjetivo	implicatura	asno	Falantes
Falantes	Signo	transposição didática	Consciência Fonológica
ANL	falante	língua adicional	Encadeamento
Encadeamento	Bush	produção oral	interlocutor
implicatura	encadeamento	competência interacional	Enunciação
AP	Enunciação	ensino de leitura	Oração
signo	TR	Alunos	Implicatura
Provérbio	diálogo	Alteridade	Signo
Teoria dos Topoi	Narrativas	entidade linguística	Aprendizes
Linguística	Charge	CLG	Brasil
Ducrot	ANL	Semântica Argumentativa	ANL
Sílaba	Rima	Saber Científico	Vogais
Vogais	Poema	Bloco Semântico	Narrativa
Aprendizes	Brasil	encadeamento argumentativo	Ducrot
Oração	Igreja	Ensino Fundamental	Diálogo
Acento	Atos	noção de valor	Ato

Fonte: elaborado pelos os autores.

4. Discussão dos resultados iniciais

Nas nuvens da Figura 2, como dito anteriormente, constam os 150 conceitos mais relevantes para o CorpAcad. Embora não seja um número muito expressivo, a partir das nuvens juntamente com as listas de conceitos, podemos notar que existe um padrão de produção intelectual no PPGL-PUCRS naquele dado período ao observarmos atentamente cada conceito extraído. Nossa análise inicial já pôde distinguir três grandes áreas de trabalho ao longo desses 11 anos, a saber, 1) Fonética/Fonologia, 2) Análise do Discurso e 3) Pragmática. Essas foram as três grandes áreas nas quais houve maior intensidade de produção devido à alta presença de conceitos pertinentes as suas terminologias.

Essa terminologia foi comprovada posteriormente por profissionais das respectivas áreas, pois repassamos as listas de conceitos do Quadro 1 a doutores e doutorandos em Linguística para que os termos ali expostos fossem analisados como pertencentes ou não ao domínio de sua expertise. Os especialistas puderam corroborar alguns conceitos como específicos da terminologia de sua área, assim como puderam acrescentar outros, cuja inserção não estava prevista na nossa análise inicial. Com isso chegamos à seguinte triagem de conceitos por área, demonstrada na Tabela 2, na qual temos os termos classificados em 5 categorias:

- Termos exclusivos da Fonética/Fonologia estão marcados com fundo roxo;
- Termos exclusivos da Análise do Discurso estão marcados com fundo amarelo;
- Termos exclusivos da Pragmática estão marcados com fundo verde;
- Termos comuns à Análise do Discurso e Pragmática estão marcados com fundo azul;
- Termos de classificação da Linguística em geral.

Existe, portanto, uma terminologia básica de cada área, que ocorre naturalmente nos textos do *corpus* de estudo, e acaba por caracterizar o seu aparato teórico-metodológico. A constatação desse perfil de especialidade também nos indica outras áreas não ali contempladas, as quais podem incrementar os estudos linguísticos do PPGL, impulsionando novas pesquisas em diferentes campos até então não explorados, ou não bem estabelecidos dentro da tradição linguística, como o já apresentado PFA. A constatação acima, através dos conceitos extraídos, das áreas da Linguística nos incentiva a continuarmos explorando o PFA a fim de estabelecê-lo como uma pedagogia de ensino de LP, assentada em metodologias

oriundas de pesquisa científica que nos forneçam informações confiáveis e valiosas sobre o registro acadêmico.

Somente definir as áreas de trabalho do PPGL durante os últimos 11 anos não é algo que nos traga uma real novidade. No entanto, temos, a partir das nuvens e outros recursos gerados pelo ExATO, a possibilidade de realizar uma análise mais profunda de cada área, apontando para questões eminentemente práticas, como a produção de glossários, dicionários terminológicos, ementas entre outras possíveis aplicações. Não só restrito às áreas em si, podemos refletir sobre os conceitos que as perpassam e tipificam a própria ciência linguística e, por extensão, conceitos que são necessários para uma pedagogia de ensino de PLA para alunos estrangeiros em mobilidade acadêmica que necessitem de proficiência em PFA.

Tomemos, por exemplo, os resultados referentes ao ano de 2015. Ao nos debruçarmos mais atentamente sobre os 150 conceitos encontrados, podemos observar, em ordem decrescente de relevância, os seguintes termos:

Tabela 2 – Alguns conceitos extraídos do ano de 2015.

posição	termo	frequência	posição	termo	frequência
3	enunciado	666.5	83	gêneros discursivos	90
4	interlocutor	644	88	substantivo	88
9	encadeamento	364	92	anáfora	84
12	diálogo	279	93	ensino de leitura	83
20	atos	224	94	estratégias de leitura	82
23	verbos	214	97	aprendizagem de leitura	81
33	negação	174	98	compreensão de texto	80
45	compreensão leitora	128	111	bloco semântico	75
57	pronomes	117	122	sintagma	71
66	oração	107	124	sentido de enunciado	70
70	preposição	101	125	encadeamento argumentativo	70
72	léxico	98			

Fonte: elaborada pelos autores.

À frente de cada sintagma nominal está sua colocação na lista de 1 até 150, e sucedendo cada termo está o seu índice de relevância indicado pelo ExATO, na etapa de computação desse índice através de um método de forte base estatística, como dito anteriormente.

A partir desses sintagmas, podemos intuir a seguinte consideração: tendo em vista que o ensino de PLA, assumido nesse artigo sob a égide da Pragmática, parte, portanto, do pressuposto de que a língua deve ser ensinada e adequada à real necessidade de ser utilizada em determinado contexto, devemos abarcar, então, esses conceitos em atividades desenhadas que visem à proficiência do aluno no contexto acadêmico.

Tomemos, por exemplo, os conceitos encontrados na Tabela 3:

“verbos / pronomes / oração / preposição / léxico / substantivo / anáfora / sintagma / sentido de enunciado / encadeamento argumentativo”.

Ao refletirmos sobre eles e sobre o ensino de PLA, podemos intuir que tais são essenciais para a escrita, por exemplo, de um texto acadêmico, no qual a tipologia textual requer um domínio maior das habilidades de escrita do aluno, a fim de se obter um texto coeso e coerente. Por isso, é valioso que acrescentemos às atividades de ensino de PLA tópicos como esses, cujo objetivo final é o desenvolvimento da proficiência linguística do aprendente de LP - ou mesmo do falante nativo - cuja proficiência no contexto acadêmico também deve ser alcançada para obter aprovação em suas atividades de estudo.

Da mesma forma, ao longo de uma tese ou dissertação, caso seja um aluno de pós-graduação, é fundamental que se encontre no seu texto “encadeamento argumentativo”, pois esses gêneros acadêmicos em questão são baseados em discussões científicas, nas quais existem argumentos que corroboram outros, assim como argumentos conflitantes sobre o mesmo assunto.

5. Considerações finais

O Português para Fins Acadêmicos está se solidificando no PPGL-PUCRS através das pesquisas desenvolvidas dentro do Grupo UPLA, o qual está crescendo e refinando sua abordagem pragmática. Esse refinamento se alcança devido à congregação epistemológica da LC e PLN. Essa interface bem estabelecida se estabelece para que o trabalho, tanto em PFA quanto em PLA, seja mais eficaz no que concerne a uma descrição e análise mais empírica sobre a LP no registro acadêmico. Ainda que a contribuição do artigo seja um pouco sucinta devido ao fato dos resultados de análise serem iniciais, já temos em mãos dados pertinentes que nos propiciam a descrever concretamente o registro acadêmico em seus gêneros tese e dissertação de forma mais eficiente.

Além disso, é importante salientar que trabalhos futuros aprofundarão a análise dos resultados obtidos, abrindo novos horizontes para a exploração do português dentro da academia. Com foco no ensino e aprendizagem de PLA, de maneira mais prática, já podemos vislumbrar um glossário de termos acadêmicos, um dicionário terminológico de Linguística, uma ementa que enfoque nas habilidades linguísticas necessárias para o aluno que utilize a LP como meio de estudo, quanto para aquele que pesquisa a LP, entre outras possíveis aplicações de extrema importância para a pesquisa, referência e consulta por alunos nativos e não nativos de LP.

Referências Bibliográficas

BAKER, P.; HARDIE, A.; MCENERY, T. **A Glossary of Corpus Linguistics**. Edinburgh: Edinburgh University Press, 2006. 187p.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus Linguistics: Investigating Language Structure and Use**. Cambridge: Cambridge University Press, 1998. <https://doi.org/10.1017/CBO9780511804489>

BICK, E. **The parsing system PALAVRAS: automatic grammatical analysis of Portuguese in constraint grammar framework**. 2000. Ph.D. (Doctoral Thesis). Arhus University, Arhus, 2000.

DALE, R.; MOISL, H.; SOMERS, H. **Handbook of Natural Language Processing** (first edition). New York: Marcel Dekker, 2000.

FINATTO, M. J. B.; LOPES, L.; CIULLA, A. Processamento de Linguagem Natural, Linguística de *Corpus* e Estudos Linguísticos: uma parceria bem-sucedida. In: **Domínios de Lingu@gem**. v. 9, n. 5 (dez. 2015).

KENNEDY, G. **An introduction to Corpus Linguistics**. London & New York: Longman, 1998.

LOPES, L. **Extração Automática de Conceitos a partir de Textos em Língua Portuguesa**. 2012. Tese (Doutorado em Ciência da Computação) – Programa de Pós-Graduação em Ciência da Computação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2012.

LOPES, L.; VIEIRA, R. Processamento de Linguagem Natural e o Tratamento Computacional de Linguagens Científicas. In: **Linguagens Especializadas em Corpora: modos de dizer e interfaces de pesquisa**. PERNA, C.; DELGADO, H.; e FINATTO, M. (orgs.). Porto Alegre: EdiPucrs, 2010. p.183-201.

LOPES, L.; FERNANDES, P.; VIEIRA, R. ExATO - High Quality Term Extraction for Portuguese and English. In: **2016 IEEE/WIC/ACM International Conference on Web Intelligence**, 2016, Omaha. Proceedings of International Conference on Web Intelligence. Omaha - Nebraska - USA, 2016. p. 1-6. <https://doi.org/10.1109/WI.2016.0092>

LOPES, L.; VIEIRA, R. Evaluation of cutoff policies for term extraction. **Journal of the Brazilian Computer Society**, v. 21(1), p. 1-9, Elsevier, 2015.

LOPES, L.; FERNANDES, P.; VIEIRA, R. Estimating term domain relevance through term frequency, disjoint *corpora* frequency - tf-dcf. **Knowledge-Based Systems**, v. 97: p. 237-249, Elsevier, 2016.

LOPES, L.; VIEIRA, R. Improving Portuguese Term Extraction. In: **International Conference on Computational Processing of the Portuguese Language - PROPOR**, 2012, Coimbra. Lecture Notes in Computer Science - Proceedings of PROPOR 2012. Heidelberg: Springer, 2012. v. 7243. p. 85-92. https://doi.org/10.1007/978-3-642-28885-2_9

LOPES, L.; FERNANDES, P.; VIEIRA, R.; FEDRIZZI, G. ExATOlP - An Automatic Tool for Term Extraction from Portuguese Language *Corpora*. In: **LTC'09 - 4th Language and Technology Conference**, 2009, Poznan, 2009, Poznan. Proceedings of the Fourth Language and Technology Conference. Poznan: Adam Mickiewicz University, 2009. p. 427-431.

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Cambridge: The MIT Press, 1999.

MITKOV, R. (ed.). **The Oxford Handbook of Computational Linguistics**. Oxford: Oxford University Press, 2003.

MOLSING, K. V.; PERNA, C. B. L.-P. Research and Teaching in Portuguese for Specific Purposes. **BELT-Brazilian English Language Teaching Journal** 5.2 (2015): 1-7. <https://doi.org/10.15448/2178-3640.2014.2.19701>

TEUBERT, W.; CERMÁKOVÁ, A. **Corpus Linguistics. A short introduction**. London: Continuum, 2007.

Artigo recebido em: 15.12.2016

Artigo aprovado em: 21.03.2017

Ideologia em gêneros discursivos pela análise associada da Linguística Sistêmico-Funcional e Linguística de *Corpus*: cantigas galego-portuguesas

Ideology in discursive genres by the associated analysis of the Systemic-Functional Linguistics and Corpus Linguistics: Galician-Portuguese chants

Ulisses Tadeu Vaz de Oliveira*

RESUMO: Recentes pesquisas linguísticas no escopo da Linguística Sistêmico-Funcional e Linguística de *Corpus* permitiram grandes avanços no entendimento da microestrutura e macroestrutura textual, devido à crescente preocupação de qualificar dados quantitativos, em especial as escolhas lexicogramaticais e os padrões de ocorrências em gêneros discursivos. Assim, a associação das ferramentas metodológicas fornecidas pelo arcabouço teórico de ambas as teorias permite um estudo abrangente do texto, uma vez que a análise cruzada de dados pode evidenciar importantes padrões de uso da língua para o reconhecimento e uso de gêneros discursivos. De fato, a recorrente discordância na definição de gêneros do discurso ocorre, em grande medida, pois a estabilidade relativa de gênero (cf. Bakhtin, 1992) ocorre em diferentes níveis linguísticos. A ideologia, considerada uma dimensão contextual superior, fica geralmente desprestigiada nas análises de gênero, justamente por seu caráter amplo e multinível, o que dificulta a manipulação de dados pelo pesquisador. Portanto, o artigo apresenta uma proposta de análise cruzada, ou seja, uma associação teórico-metodológica envolvendo a Linguística Sistêmico-Funcional e a Linguística de *Corpus*, para a delimitação de complexos ideológicos visualizados nos padrões de ocorrência de participantes-chave no discurso e o significado logogenético obtido da representação da experiência (metafunção

ABSTRACT: Recent linguistic research in the scope of Systemic-Functional Linguistics and Corpus Linguistics allowed major advances in understanding textual micro and macrostructure due to the growing concern to qualify quantitative data, especially lexicogrammatical choices and patterns of occurrence in genres. Therefore, the combination of methodological tools provided by the theoretical framework of both theories allows a comprehensive study of the text, since the cross-analysis of data might show significant standards of use of languages in order to recognize and use a given genre. In fact, the recurring disagreement about the definition of discursive genres is largely due to the relative stability of genres (cf. Bakhtin, 1992) that materialize at different linguistic levels. Ideology – assumed as a detectable and noticeable higher contextual dimension – is generally discredited in genres analysis, precisely because of its broad and multilevel characteristic, which makes data manipulation harder for the researcher. Therefore, this article proposes a cross-analysis, there is a theoretical-methodological association between the Systemic-Functional Linguistics and Corpus Linguistics in order to unveil the ideological complexes obtained from the patterns of occurrence of key participants (w-participants) in the discourse and the logogenetical meaning resulted from the representation of the experience (ideational

* Doutor em Linguística Aplicada e Estudos da Linguagem pela PUC-SP. Professor da Universidade Federal de Mato Grosso do Sul. E-mail: ulisvaz@gmail.com

ideacional) em cantigas de amor galego-portuguesas.

PALAVRAS-CHAVE: Gêneros discursivos. Linguística Sistêmico-Funcional. Linguística de *Corpus*. Ideologia.

metafunction) in Galician-Portuguese chants of love.

KEYWORDS: Discursive genres. Systemic-Functional Linguistics. *Corpus* Linguistics. Ideology.

1. Introdução

A escolha do tema deste estudo, isto é, o papel da ideologia em gêneros discursivos, deve-se prioritariamente às escassas pesquisas sobre gênero que estabeleçam correlação de aspectos ideológicos macro e microestruturais na delimitação de gêneros discursivos. Outrossim, a análise cruzada, isto é, uma associação teórico-metodológica entre a Linguística Sistêmico-Funcional e Linguística de *Corpus* revela-se absolutamente valiosa pela exequibilidade de um rigor científico, o qual é alcançado pela sobreposição de pesquisas quantitativas e qualitativas.

Para a Linguística Sistêmico-Funcional (doravante LSF) é importante o fato de que escolhas realizadas na microestrutura textual revelam significados nem sempre explícitos na macroestrutura do discurso. Pesquisas complementares (FAIRCLOUGH, 1992; FAIRCLOUGH; WODAK, 1997; FOWLER, 1996a) evidenciam de que forma noções como o alinhamento (*frame*) e a intersubjetividade participam para estabelecer uma relação fundamental entre os interlocutores (GOFFMAN, 1974) e sua interação com o texto.

A Linguística de *Corpus* (doravante LC), por sua vez, ocupa-se da coleta e cotejo de *corpora*, ou conjuntos de dados linguísticos textuais, os quais são obtidos criteriosamente em bancos de dados com vistas a alicerçarem pesquisas de uma língua ou variedade linguística (BERBER-SARDINHA, 1999). Segundo Berber-Sardinha (2000, p. 1) a LC “dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador” e sua influência para pesquisa linguística é grande.

A associação da LSF e a LC está enquadrada na proposta da Linguística Crítica, uma abordagem desenvolvida desde a década de 1970 por um grupo da Universidade de East Anglia (FOWLER, 1996b; KRESS, 1982). A Linguística Crítica tenta conciliar um método de análise linguística textual com uma teoria social do funcionamento da linguagem em processos políticos e ideológicos (OLIVEIRA, 2008).

2. Gêneros discursivos

O corolário sobre gênero baseia-se amplamente nas ideias difundidas por Bakhtin

(1986, [1952-1953], 1992), para quem a unidade real da comunicação verbal é o enunciado. O enunciado reproduz em seu conteúdo (temático) e seu estilo (verbal) as condições específicas e as finalidades de uma meta comunicativa, no enquadre de um determinado contexto. Ou seja, a construção composicional reflete seleções conscientes operadas através de recursos da língua: lexicais, fraseológicos e gramaticais.

Para Bakhtin (1992, p.281), os gêneros do discurso são “tipos relativamente estáveis de enunciados elaborados por cada esfera de utilização da língua”. Portanto, a língua em uso deve ser o objeto do linguista, uma vez que ignorar a natureza do enunciado e as particularidades de gênero que moldam a variedade do discurso levam ao formalismo e à abstração, desvirtuando a historicidade do discurso e enfraquecendo o vínculo existente entre a língua e a vida (BAKHTIN, 1986, [1952-1953], p. 282).

A importância dos estudos bakhtinianos é tamanha, que uma série de pesquisadores desenvolveu pesquisas e teorias no esteio da ideia geral do filósofo russo. Bathia (2004) destaca três enquadres desenvolvidos e usados na análise de gênero:

(a) a escola americana, representada por trabalhos como de Miller (1994), Bazerman (1994) e Berkenkotter; Huckin (1995);

(b) a escola australiana, com uma abordagem funcionalista de gênero, representada por Martin (1984, 2005) e Martin; Christie; Rothery (apud HYLAND, 2003).

(c) a escola britânica, representada pelos estudos de Swales (1990) e Bathia (1997, 2004, 2014).

Trabalhos como os de Bathia (1997, 2014), Hyon (1996) e Yunick (1997) apresentam diversas informações para o pesquisador interessado em aprofundar os conhecimentos sobre os enquadres supracitados. No que circunscreve este artigo, aproximamo-nos da visão sistêmico-funcional sobre gênero. Alguns estudos merecem especial destaque, conforme apresentaremos adiante.

Martin (1984) define gênero como uma atividade organizada em estágios, orientada para uma finalidade, na qual os falantes se envolvem como membros de uma cultura¹. Dessa forma, para o autor, o gênero pode ser visto tanto como um produto cultural determinado pelo contexto, quanto pela estrutura esquemática. Em 1992, Martin aborda a noção de macro

¹ Do original: “a staged, goal-oriented, purposeful activity in which speakers engage as members of our culture.” (MARTIN, 1984, p. 25).

gênero – um conceito guarda-chuva que abarcaria subcategorias – por exemplo, um manual universitário contém seções expositivas, descritivas ou regulativas. Menos tecnicamente, gêneros são como as coisas são feitas, quando a linguagem é usada para efetivá-las (MARTIN, 1984, p. 248). Para o pesquisador australiano, grande parte do choque cultural de um aluno num dado contexto é, de fato, um choque de gênero.

Halliday e Hasan identificaram padrões linguísticos nos textos por meio de um conjunto específico de valores realizados no discurso. Nesta configuração contextual, obtém-se dos textos uma estrutura de estágios, perguntando: (1) que elementos devem ocorrer; (2) que elementos podem ocorrer; (3) onde eles devem ocorrer; (4) onde eles podem ocorrer e (5) com que frequência eles podem ocorrer (HALLIDAY; HASAN, 1989, p. 55-56). A partir dessa orientação, Hasan (1992) desenvolve o conceito de Estrutura Potencial Genérica (EPG), ou ainda Estrutura Genérica Potencial (EGP). A EPG assemelha-se a um esboço que especifica “que elementos devem ocorrer em cada e toda instância de um discurso particular; e como esses vários elementos se relacionam” (HASAN, 1992, p. 128). “Elemento” é definido como um estágio com alguma consequência para a progressão de um texto, de tal forma que “cada estágio em uma atividade social é um elemento da estrutura dessa atividade” (HASAN, 1992, p. 142). Segundo a autora (HASAN, 2000, p. 53), os estágios podem ser:

- (a) obrigatórios, que especificam todos os elementos da estrutura na qual a presença é obrigatória;
- (b) opcionais, que devem enumerar todos os elementos que são considerados opcionais, de modo que sua presença ou ausência não deve afetar o texto;
- (c) sequenciais, que visam especificar a ordenação dos elementos obrigatórios e opcionais em comparação um com outro;
- (d) recursivos ou iterativos, que são aqueles que podem se repetir.

Admitindo o raciocínio de Hasan, a Linguística de *Corpus* oferece valiosa metodologia para determinar os estágios, uma vez que fornece dados quantitativos manipuláveis, obtidos por índices de ocorrência e posicionamento na oração.

Tendo apresentado em 1985 um modelo derivado de um “esquema convencional de notícias” na análise de reportagem de notícias, o que, de certa forma, é o mesmo que determinar o gênero de reportagem, Van Dijk (1997, 1999) avançou em seus estudos sobre gênero e discurso ao relacionar a noção macro da ideologia às noções micro dos discursos e

das práticas sociais de membros de um determinado grupo.

Para Swales, gênero compreende uma classe de eventos comunicativos, cujos membros compartilham algum conjunto de metas comunicativas, as quais são reconhecidas pelos membros da comunidade discursiva (1990, p.58).

Biber (1992) e Paltridge (2002) se interessam pela distinção entre gênero e tipo de texto, estabelecendo entre eles diferenças pragmáticas importantes. Já Biber e Conrad (1999) observam que textos dentro de certos gêneros podem diferir bastante em suas características linguísticas. Por exemplo, editoriais podem apresentar desde uma estrutura linguística bastante formal e elaborada, até um caráter narrativo e coloquial. Por outro lado, é possível verificar textos de gêneros diferentes bastante parecidos linguisticamente. Por exemplo, cartuns e tirinhas são frequentemente idênticos na forma linguística.

Kress aplica os estudos de Biber e cunha os conceitos de extensões, misturas e híbridos² presentes no discurso multimodal da mídia. Por analogia, extensões, misturas e híbridos são encontrados em textos mais tradicionais do discurso da mídia (p. ext. jornais, revistas, blogs). Reynolds (2000), por exemplo, ao analisar editoriais de jornais, verifica como o argumento predomina numa mistura de narrativa, descrição e argumento. O autor utiliza os conceitos de fusão linear e escalada para demonstrar como ocorre essa organização textual. Anos antes, Fairclough (1997) já apresentara as noções de “hibridismo intertextual” e “heterogeneidade linguística”, este último realizando o primeiro no nível textual.

Dentre os inúmeros pesquisadores que se debruçaram sobre o controverso tema dos gêneros discursivos, chamamos a atenção para os estudos de Douglas Biber. O pesquisador da Northern Arizona University considera o enquadre descritivo como não sendo “suficientemente explícito para ser usado em taxonomias situacionais” (1993, p. 38). Desta forma, e também em parceria com notáveis linguistas, o autor avançou em pesquisas que empregavam *corpora*³ para análise de gênero e no desenvolvimento da pesquisa em “Processamento de Linguagem Natural” (PLN) (BIBER, 1993; BIBER; CONRAD 1999; BIBER; CONNOR; UPTON, 2007). A seguir apresentaremos algumas considerações importantes sobre a Linguística de *Corpus* na supracitada interface.

² Tradução livre de “*extensions, mixes and hybrids*” (BIBER, 2004, p. 443).

³ *Corpora* (plural de *corpus*) linguísticos são os conjuntos de textos escritos e registros orais em uma determinada língua e que serve como base de análise.

3. A Linguística de *Corpus* como mudança epistemológica para a análise linguística

A Linguística de *Corpus*, de certa forma, renovou a empiria nos estudos linguísticos no momento que permitiu, através da análise de grande volume de dados, a aferição científica pela observação de evidências linguísticas em *corpora* autênticos e de linguagem em uso. Para Fillmore (1992, p. 35), o *corpus* proporciona a descoberta de fenômenos linguísticos, como de nenhum outro modo.

Dentre diversos autores, Sanchez apresenta uma definição abrangente de *corpus*:

Un corpus lingüístico es un conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y profundidad de manera que sean representativos del total del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador con el fin de obtener resultados varios y útiles para la descripción y el análisis. (SANCHEZ, 1995, p. 5)

Ao longo do século XX, a Linguística de *Corpus* floresceu gradativamente e teve um salto a partir da criação de *corpora* digitais. Inicialmente, as análises de *corpus* ocorriam manualmente, como nos trabalhos de Thorndike (1921), Sinclair (1966), Boas (*apud* JAKOBSON; BOAS, 1944). Depois do louvável trabalho de Leech (1966), a necessidade de *corpora* digitais passou a ser o foco das pesquisas da geração seguinte⁴ (SINCLAIR, 1991), porém, nesse segundo momento, houve um direcionamento geral para a aplicação da LC para orientar o ensino de línguas (SVARTVIK, 1992). Ao longo da década de 90, a LC passou a ser aplicada amplamente na descrição linguística da língua inglesa⁵ e de outras línguas internacionais (ZHOU; YU, 1997; CERMAK, 1997, BERBER-SARDINHA, 2014). Recentemente, estudos como os de Granger (1998) mostram uma tendência à retomada do uso de *corpus* com aplicação pedagógica. Para Berber-Sardinha (2000, p. 5),

a história da Linguística de Corpus está condicionada à tecnologia, que permite não somente o armazenamento de *corpora*, mas também a sua exploração. Por isso, a história da área está relacionada à disponibilidade de ferramentas computacionais para análise de corpus.

4 A obra “Corpus, Concordance, Collocation” (SINCLAIR, 1991) é considerada a precursora dos estudos com *corpus* digital.

5 A maior parte dos estudos no escopo da Linguística de *Corpus* é sobre a língua inglesa, uma vez que a disponibilidade de *corpora* eletrônicos nessa língua é muito superior às demais.

Sendo assim, ferramentas como o *Text Converter*, que faz parte do *WordSmith Tools* (SCOTT, 1996), ganha destaque por ainda ser um dos mais completos e operacionais recursos de cotejo de dados. Ainda, a popularização da LC tem sido viabilizada por ferramentas como o Ngram (fornecida pelo *website Google*), mesmo que consideradas suas limitações.

Biber (1998, p.4) e Leech (2004) alinham-se no destaque da Linguística de *Corpus* como uma das ferramentas mais efetivas para a análise linguística por reunir grande volume de dados, em textos autênticos e uso real; e, sobretudo, a importância no foco: (a) no desempenho linguístico; (b) na descrição linguística; e (c) numa visão empírica da pesquisa científica.

Muitos autores associam a Linguística de *Corpus* à teoria funcionalista da linguagem desenvolvida por Halliday (1994), a Linguística Sistêmico-Funcional. Isso ocorre porque a LSF encara a linguagem como um sistema probabilístico de escolhas, pelo qual traços linguísticos são determinados. Portanto, a associação de ambas as teorias se torna reveladora no sentido de fornecer dados qualiquantitativos complementares na análise do fenômeno linguístico.

4. A Linguística Sistêmico-Funcional e a construção do significado no sistema de escolhas

Para a Linguística Sistêmico-Funcional, a língua possui um nível intermediário de codificação: a lexicogramática, que possibilita à língua construir três significados concomitantes, que entram no texto através das orações, mediante escolhas feitas no sistema linguístico. Sendo assim, a noção de escolhas é muito importante para a LSF. Halliday (1994) argumenta que, quando se faz uma escolha no sistema linguístico, o que se escreve ou o que se diz adquire significado contra um fundo em que se encontram as escolhas que poderiam ter sido feitas, mas não foram (escolhas potenciais) – fato importante na análise do discurso. Em suma, a LSF procura desenvolver uma teoria sobre a língua como um processo social e, para tanto, fornece ferramentas metodológicas que permitem uma descrição detalhada e sistemática dos padrões linguísticos (HALLIDAY, 1994).

A LSF entende a língua estruturada para construir três tipos de significados concomitantes: ideacional, interpessoal e textual. A metafunção ideacional representa os eventos das orações em termos da representação da experiência por meio do sistema da transitividade; a metafunção interpessoal envolve as relações sociais com respeito à função da

oração no diálogo, e referem-se a dar/pedir informação ou bens & serviços; a metafunção textual organiza os significados ideacionais e interpessoais de uma oração, considerando os significados advindos da ordem das palavras na oração.

O estudo apresentado neste artigo enfoca a metafunção ideacional, na interface entre a Linguística Sistêmico-Funcional e a Linguística de *Corpus*.

4.1 A Metafunção Ideacional

De acordo com Halliday (1994), a metafunção ideacional tem a função de representar padrões de experiência. Nossa impressão mais poderosa da experiência é de que ela consiste de “eventos” (i.e., acontecer, fazer, sentir, significar, ser e tornar-se). O sistema da transitividade é o modo pelo qual o falante constrói o mundo da experiência em um conjunto de tipos de processos semânticos representados na oração.

Para Halliday (1994), os componentes da transitividade são os processos participantes e circunstâncias. Os processos são expressos pelo grupo verbal da oração; os participantes envolvidos no processo (realizados pelos grupos nominais da oração) e as circunstâncias (realizadas por grupos adverbiais ou preposicionais). O autor, e mais tarde estudos complementares (THOMPSON, 1998), propõe uma classificação dos processos de seis tipos, referindo-se a ações ou eventos do mundo (Quadro 1).

Quadro 1 – Componentes da Transitividade.

Processos	Participantes ligados ao Processo
Material	João <u>quebrou</u> a mesa <u>com um soco</u> Ator Material Meta Circunstância
Comportamental	Ele <u>sofreu</u> <u>com a derrota</u> Comportante Comportamental Alcance
Mental	Eu <u>entendi</u> <u>o seu sofrimento</u> Experienciador Mental Fenômeno
Verbal	O rapaz <u>contou</u> à mãe <u>sobre a difícil situação</u> Dizente Verbal Receptor Verbiagem
Relacional	João <u>continua</u> <u>deprimido</u> Portador Relacional Atributo
Existencial	Houve <u>motivos</u> <u>com certeza</u> Existencial Existente Circunstância

Fonte: Adaptação de Halliday (1994).

A Transitividade tem-se provado extremamente útil na análise textual de perspectiva crítica por também participar da construção do significado interpessoal do texto. Isso ocorre devido a um efeito prosódico obtido ao longo do fluxo discursivo pelas escolhas lexicogramaticais, é o que podemos chamar de Expansão Logogenética do Significado, a qual pode ser extremamente efetiva na análise da ideologia na microestrutura textual, por propiciar

a “construção dinâmica do significado conforme o texto se desenvolve” (OLIVEIRA, 2008, p. 56).

Martin (1992, p. 553-559) e outros autores notaram, com destaque a Bakhtin e a noção de expansão dialógica, que as realizações dos significados interpessoais tendem a ser mais prosódicas, espalhando-se através da oração ou da oração complexa. Halliday empregou o termo logogênese para identificar a construção dinâmica do significado, conforme o texto se desenvolve (HALLIDAY; MARTIN, 1993; HALLIDAY; MATTHIESSEN, 2004). Tratando especificamente da metafunção ideacional, Thompson (1998) denomina ressonância a harmonia de significados que é produto de uma combinação de escolhas; não identificáveis com qualquer outra escolha, se consideradas isoladamente. Tanto a logogênese quanto a ressonância são, do ponto de vista da construção do significado discursivo, instâncias da ideologia.

5. A Ideologia e os complexos ideológicos na formação dos gêneros discursivos

A ideologia é um fator que não pode ser descartado em análises de textos reais. Embora tenha grande importância, apenas alguns pesquisadores de áreas do conhecimento humano e do estudo da linguagem tentam analisar aspectos ideológicos refletidos nos textos (e.g. VAN DIJK, 1999). No particular da análise de gênero, a ideologia parece não receber a atenção que merece, sendo em menor número estudos que avaliam “as formas nas quais indivíduos respondem à ideologia incorporada nos gêneros” (ARTEMEVA, 2007).

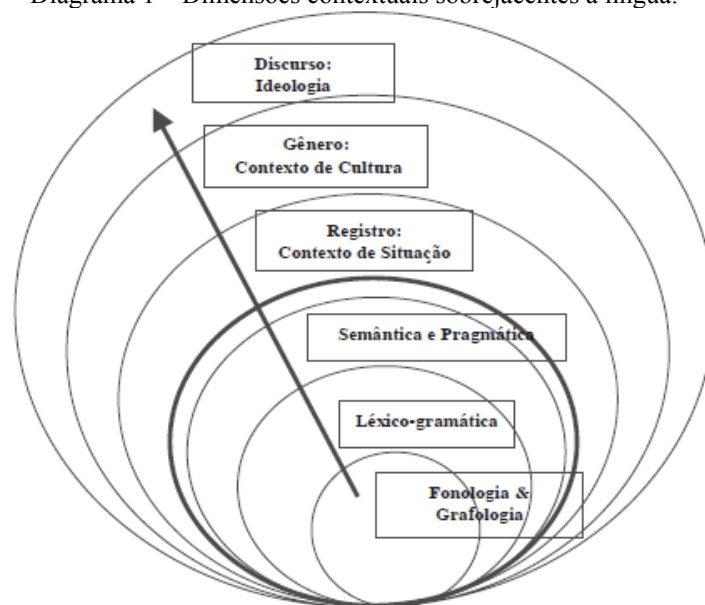
Segundo Fairclough (2001), há uma relação dialética entre os elementos das práticas sociais, o que significa a “interiorização” de uns por outros, sem que haja redução entre eles. As relações sociais, as identidades sociais, os valores e a consciência cultural apresentam uma faceta discursiva (semiótica), muito embora sejam comumente teorizados e pesquisados por teorias adjacentes às teorias linguísticas. Portanto, resta indagar como a construção de valores da consciência cultural e da identidade social (i.e., ideologia) pode se materializar na elaboração de textos? Para responder a essa questão, Bakhtin (cf. CLARK; HOLQUIST, 1984) considera um *self* individual, um eu, mas que incorpora também aspectos compartilhados com os outros. O filósofo russo usa o termo *elocução* para designar essa atividade capaz de compreender simultaneamente energias tão díspares. Os sistemicistas, por sua vez, lançam mão de dois conceitos que incidem diretamente sobre a *elocução*: Registro e Gênero, referentes, respectivamente, ao contexto de situação e ao contexto de cultura

(MARTIN, 1992).

Além das variáveis determinadas por situação e cultura, Fowler (1991) e Fairclough (1992) consideram a ideologia como um tipo de dimensão contextual superior. Para Fowler (*Idem*), qualquer aspecto da estrutura linguística carrega significação ideológica - seleção lexical, opção sintática, etc. – todos têm sua razão de ser. Há sempre modos diferentes de dizer a mesma coisa, e esses modos não são alternativas acidentais. A ideologia refere-se a posições de poder, a vieses políticos e a suposições sobre valores, tendências e perspectivas que os interlocutores trazem para os seus textos, e tem chamado a atenção dos sistemicistas, na medida em que, em qualquer registro, em qualquer gênero, o uso da língua será sempre influenciado pela nossa posição ideológica (OLIVEIRA, 2008).

Em 1985, Martin propõe a “Teoria de Gênero e Registro” (TGR) que investiga o modo como as diferenças de contexto constroem textos distintos atendendo a padrões linguísticos. Halliday e Martin (1993, p. 132) compreendem a dimensão contextual da ideologia como supra e inter em relação às camadas sobrejacentes à língua (Diagrama 1).

Diagrama 1 – Dimensões contextuais sobrejacentes à língua.



Fonte: Adaptação de Halliday e Martin (1993, p. 132).

Nesse diagrama, os autores reconhecem a força da ideologia, a qual incide sobre todas as camadas contextuais e linguísticas: macro e microestrutura textual. A análise do diagrama permite reconhecer que há especificações explícitas (idealmente quantificáveis de padrões gramaticais) semânticas do texto recebendo e fornecendo força à dimensão ideológica. Assim,

a explicação da variação linguística dar-se-á pelos elos explícitos entre traços do discurso e variáveis críticas do contexto social e cultural empregados para realizar os significados de um dado texto. Nesse sentido, apoiamo-nos em estudo de Li (2010), no qual o autor, guiado por propostas de análise crítica e com o apoio do contexto analítico da LSF, investiga as relações entre escolhas de certas formas linguísticas e ideologias e as relações de poder que subjazem a essas formas.

6. As Ideologias sociopolíticas e socioculturais na língua na proposta de Li

Em estudo intitulado “*Transitivity and lexical cohesion: Press representations of a political disaster and its actors*”, Li (2010) analisa aspectos da gramática da oração relacionados à transitividade e à coesão lexical em textos de reportagens de orientações ideológicas diferentes. Os resultados indicam que as interpretações e os papéis sociais dos participantes envolvidos no evento são construídos pelas escolhas específicas que cada veículo de comunicação faz na representação da experiência, com valor aparentemente informativo, mas com vistas a persuadir o leitor.

Li (2010) emprega as ferramentas metodológicas da LSF para analisar como as escolhas linguísticas expressam significados ideológicos. Dessa forma, o exame da seleção, categorização e ordenação do significado nas microestruturas no nível oracional é especialmente útil para uma análise sistemática do discurso, com enfoque nos traços linguísticos e na organização dos significados do texto. Hoey (1997, p. 3) sugere quais ocorrências devem ser observáveis nos padrões de ocorrência lexical:

- Quais são os padrões lexicais dos quais a palavra faz parte?
- A palavra se associa constantemente com outros sentidos específicos?
- Em quais estruturas ela aparece?
- Há correlação entre o uso/sentido do léxico e as estruturas das quais participa?
- A palavra está associada com alguma posição na organização textual?

As ideologias sociopolíticas ou socioculturais estão introjetadas na língua e no discurso (LI, 2010), portanto, o autor apoia-se na abordagem de Van Dijk (1997), para quem há um elo entre o social e o individual, o macro e o micro, o social e o cognitivo. Esta é a mesma abordagem de Fowler, que considera que há sempre valores implicados no uso da língua, pois “não há representação neutra da realidade” (1987, p.67).

Fundamentado nesses conceitos, o estudo ora apresentado investiga como complexos ideológicos estão presentes e são construídos na e pela representação (metafunção ideacional – LSF). A Linguística de *Corpus* tratará de fornecer as ferramentas metodológicas para a observação empírica dos padrões de ocorrência que deverão validar a hipótese ideológica que, por sua vez, é particular e constante (Estrutura Potencial Genérica) de determinados gêneros discursivos. Em outras palavras, o mapeamento de padrões de ocorrências linguísticas de motivação ideológica constrói um estágio genérico que, de outra forma, seria de difícil observação, justamente por construir-se logogeneticamente no fluxo discursivo.

7. Estudo de caso: análise cruzada da LC e LSF

Esta pesquisa é baseada em dados quantitativos e qualitativos e investiga cantigas de amor galego-portuguesas empregando a metafunção ideacional (LSF) e a Linguística de *Corpus*. Através das variáveis de ocorrências em *corpus* significativo, o que se aborda é o sistema da transitividade (HALLIDAY, 1994; THOMPSON, 1996, 2004; EGGINS, 1996) e a recorrência (THOMPSON, 2004), enfocando participantes e processos para caracterização de gênero pela Estrutura Potencial Genérica (cf. Teoria de Gênero e Registro – EGGINS; MARTIN, 1997).

7.1 *Corpus* e procedimentos de análise

O *corpus* foi extraído da base de dados online do “Projeto Littera” (FCSH-UNL)⁶, organizado pelo pesquisador Graça Videira Lopes. O objetivo desta base de dados é o de “colocar à disposição do público em geral, e também do público mais especializado, a totalidade dos textos que nos foram transmitidos pelos cancioneiros medievais galego-portugueses, de uma forma rigorosa mas acessível” (LOPES et al., 2011). A equipe do projeto reúne a totalidade das cantigas medievais galego-portuguesas profanas (cerca de 1680) que nos foram transmitidas pelos três grandes cancioneiros trovadorescos, o Cancioneiro da Ajuda (CA), o Cancioneiro da Biblioteca Nacional (CBN) e o Cancioneiro da Biblioteca Vaticana (CV). Além disso, múltiplas pesquisas podem ser feitas na base de dados, inclusive ocorrências lexicais e variações linguísticas.

As seguintes etapas foram adotadas:

⁶ Disponível em: < <http://cantigas.fcsh.unl.pt> >

- I) Formulação de hipótese: traços ideológicos e gênero discursivo.
- II) seleção e análise na análise do sistema da transitividade de cinquenta e duas (52) cantigas de amor galego-portuguesas de um trovador de destaque (o rei D. Dinis⁷);
- III) correlação do sistema da transitividade, com os complexos ideológicos e gêneros discursivos. O objetivo foi a formulação de hipótese de padrão de ocorrência (processos, participantes e circunstâncias – LSF) para cada um dos gêneros (EPG);
- IV) aplicação da LC para busca na base de dados online do “Projeto Littera” de processos-chave e participantes-chave para confirmação empírica das hipóteses;
- V) depois da verificação da incidência descrita em (IV) e categorizações, tecemos comentários sobre como isso se relaciona ao gênero da cantiga e ao posicionamento ideológico admitido pelo trovador. Assim, foi possível estabelecer estruturas estáveis no gênero no nível representacional (cf. princípio da ressonância, THOMPSON, 1998) e a motivação ideológica para essas ocorrências.

As etapas analíticas evidenciam a associação entre LC e LSF, considerando a possibilidade de observar um efeito macro discursivo ideológico em padrões de ocorrências na microestrutura do sistema da língua.

7.2 Análise

Devido ao constricto espaço, nos propusemos a mostrar, de forma muito sucinta, aspectos importantes e resultados obtidos em cada uma das etapas, no que diz respeito ao gênero em questão.

ETAPA I

A formulação de hipótese de traços ideológicos construindo e construídos pelo gênero discursivo escolhido, a saber, cantigas de amor galego-portuguesas medievais, buscou atender à formulação de complexos ideológicos no ideário literário medieval galego-português elaborada pelo professor e reconhecido medievalista Osvaldo Ceschin, em texto intitulado “A

⁷ D. Dinis foi o Rei de Portugal e Algarve de 1279 até sua morte. Grande amante das artes e letras, D. Dinis tornou-se um famoso trovador. O “Rei Poeta” cultivou as Cantigas de Amigo, de Amor e a sátira, contribuindo para o desenvolvimento da poesia trovadoresca na península Ibérica. Pensa-se ter sido o primeiro monarca português verdadeiramente alfabetizado, tendo assinado sempre com o nome completo. (GUERREIRO, 1978)

educação do cavaleiro e a expressão do amor no trovadorismo” (1998), e aplicada por Oliveira (2008).

Segundo o autor (Idem), na Europa medieval, o grupo dos trovadores obedecia a um rígido código de conduta inspirado, basicamente, na educação, ética, estética e ideal de cavalaria – podemos entender como um “código de cortesia”. Entre os princípios desse preceituário, constam a vassalagem paciente e humilde que compreendia quatro graus no seu aprendizado, de acordo com a intimidade entre o trovador e a donzela, todos extraídos do léxico provençal (com o correspondente em português, nos parênteses): *fenhedor* (ou suspirante), *precador* (ou suplicante), *entendedor* (ou namorado) e *drudo* (ou amante) (SPINA, 1996, p. 363).

Ceschin (1998) observa que os posicionamentos de *fenhedor*, *precador*, *entendedor* e *drudo* eram empregados pelos trovadores de maneira rigorosa, de acordo com o gênero (e.g. cantiga de amor – *fenhedor*); porém, no decorrer dos séculos, as regras de uso se afrouxaram, permitindo exceções. Este trabalho enfocará a persona (posicionamento ideológico) do *fenhedor*, a qual corresponde ao gênero das cantigas de amor. Ideologicamente, tal posicionamento é fruto da soma da ética clássica com a doutrina cristã, às quais se adicionaram o sistema feudal, a atmosfera cortês, a educação do trívio e do quadrívio e a formação do cavaleiro (CESCHIN, 1998).

O *fenhedor*, como grau primário na formação do trovador, surge no desenvolvimento de gostos e emoções criados sob a tutela da figura feminina. Nesse sentido, o *fenhedor* ama e suspira, mas não se atreve a demonstrar seu sentimento, ou sequer cortejar a dama (para os provençais) ou donzela (para os galego-portugueses). A condição do jovem nobre em relação à superioridade da donzela obriga-o a conter seus gestos, gostos e reações (OLIVEIRA, 2008).

De acordo com Spina, os trovadores observavam rigorosamente os preceitos da cortesia e retórica consagrados pela tradição. Para o autor, “o amor consistia, então, nessa vassalagem rigorosa do trovador à dama de sua eleição, e tal arte previa a observância escrupulosa de cânones preestabelecidos; em síntese, o preceituário amoroso...” (SPINA, [s.d.] *apud* CESCHIN, 1998).

Nosso objetivo foi identificar, operando as ferramentas metodológicas da LSF e LC, quais são implicações linguísticas deste “preceituário amoroso” e as fronteiras ideológicas destes “cânones preestabelecidos” para o gênero discursivo em questão. Dessa forma, a LC

viabiliza uma comprovação empírica da hipótese ideológica por meio de banco de dados digitais que, ao fornecer um padrão de ocorrências linguísticas, evidencia o posicionamento da figura do fenhedor.

ETAPA II

A análise do sistema da transitividade das cinquenta e duas (52) cantigas galego-portuguesas de D. Dinis buscou determinar a frequência dos tipos de processo (HALLIDAY, 1994), verificando quais são os mais e menos frequentes e o porquê disso. A categorização das ocorrências levou em conta a posição dos principais participantes das cantigas (participantes-chave) e o significado no contexto, relacionado com a estrutura genérica das cantigas, a cortesia e os posicionamentos ideológicos admitidos pelo trovador. Quanto à ocorrência dos processos, nota-se que os processos materiais, mentais e relacionais são massivamente superiores em relação aos demais, à medida que a representação da experiência é geralmente enquadrada nos domínios do “fazer”, “pensar” e “ser”. Portanto, neste artigo, nos focaremos em analisá-los. O Quadro 2 resume os resultados de ocorrências dos processos (metafunção ideacional) em números totais da coletânea de D. Dinis.

Quadro 2 – Processos da transitividade: cantigas de amor.

Cantigas de amor		
Processo	Nº de ocorrências	%
Mental	414	35,41
Relacional	331	28,31
Material	229	19,58
Verbal	97	8,29
Existencial	56	4,79
Comportamental	42	3,59
Total	1169	100

Fonte: Oliveira, 2008

Rememorando os estágios na expressão do cavaleiro descritos por Ceschin (1998) e compreendendo-os como os estágios na expressão do trovador, quando falamos em cantigas de amor, devemos ter em mente que a investidura no papel de fenhedor desencadeia uma rede ampla de associações prototipicamente presentes no uso da língua, bem como todas as implicações ideológicas das escolhas na lexicogramática. Dom Dinis, em suas cantigas de amor, incorpora a figura do fenhedor, ou seja, “ama”, “sofre”, “suspira”, mas não revela seu sentimento à donzela. Essa contenção do sentimento explica a falta de ação do trovador, o

que, por conseguinte, explica a maior incidência dos processos mentais e relacionais nas cantigas.

O eu-lírico do trovador se vê preso às amarras da cortesia, impelido a:

- 1) Descrever seu estado de alma – processos relacionais
- 2) Utilizar os processos mentais no sentido de manifestar: (a) o estado mental de prostração e dor – processos mentais emotivos; (b) o açoitado de ver e ouvir (ou ter visto e ouvido) a donzela – processos mentais perceptivos; (c) o pensar na donzela e seus conflitos internos – processos mentais cognitivos; (d) o desejar, o querer e não-querer da amada – processos mentais desiderativos.
- 3) Os processos materiais, verbais, existenciais e comportamentais, em menor número, também auxiliam na construção da imagem do fenhedor. Os materiais no “fazer mal” da donzela e o “fazer bem” do trovador, os verbais no “falar” que ofende o trovador e elogia a donzela, os existenciais que denotam a “existência” do sofrer para o trovador e do bem para a donzela e o comportamental na expressão do “sofrer” somente para o trovador (OLIVEIRA, 2008, p.98).

A análise qualitativa de duas estrofes da cantiga número 522 do Cancioneiro da Biblioteca Nacional (Quadro 3) exemplifica como ocorre a construção do mundo textual a partir da representação da experiência nas cantigas de amor. É um exemplo de como os processos mentais e relacionais se destacam superiormente para construir um significado ideológico subjacente ao texto (OLIVEIRA, 2008).

Quadro 3 – Excerto de análise da cantiga 522 do Cancioneiro da Biblioteca Nacional.

Cantiga – CBN 522 bis – edição de Machado (1970)	Metafunção ideacional Tipo de processo	Significados construídos
Senhor fremosa e de mui loução		
coraçõ, e queredes-vos doer	Mental: emotivo	pedido do amor
de mi, pecador, que vos sei querer	Mental: desiderativos	querer do trovador
melhor ca mi; pero sõo certão	Relacional: atributivo	trovador: estar certo
que mi queredes peor d’outra ren;	Mental: desiderativo	donzela quer mal
pero, senhor, quero -vos eu tal ben	Mental: desiderativo	trovador quer bem
Qual maior poss’ , e o mais encoberto	Relacional: atributivo	incapacidade do trovador
que eu poss’ ; e sei de Brancafrol	Relacional: atributivo/ Mental: cognitivo	incapacidade/ saber do trovador
que lhi non ouve Flores tal amor	Mental: emotivo	metáfora: ela não ama
qual vos eu ei ; e pero sõo certo	Mental: emotivo/ Relacional: atributivo	Metáfora: ele ama/ está certo disso
que mi queredes peor d’outra ren;	Mental: desiderativo	donzela quer mal

pero, senhor, quero -vos eu tal ben	Mental: desiderativo	trovador quer bem
pero, senhor, quero -vos eu tal ben	Mental: desiderativo	trovador quer bem

Fonte: Oliveira, 2008.

Paráfrase: Senhora formosa e de muito vistoso/coração, e queira vos doer/de mim, pecador, que vos quero/melhor que eu; mas estou certo/que mi quer pior que (qualquer) outra coisa/mas, senhor, eu vos quero tal bem/Qual maior (eu) possa, e o mais fingido/ que eu puder; e sei de Branca flor/ que tal amor não teve flores para você/ como por você eu tenho/ mas estou certo/que mi quer pior que (qualquer) outra coisa/mas, senhor, eu vos quero tal bem.

*O refrão desta cantiga contém certo veneno. “Tal” poderia significar “tão grande”, porém, no contexto em que está (“tal... qual”), geralmente, equivaleria ao comparativo de igualdade. Sendo assim, interpretar-se-ia o trecho como: “Ela o quer pior do que qualquer outra coisa” e ele a quer “tal bem qual ela o quer, o maior que puder”, ou seja, lhe devolveria o sentimento. (OLIVEIRA, 2008).

ETAPA III

A correlação do sistema da transitividade, com os complexos ideológicos e gêneros discursivos nos leva a formulação de hipótese de padrão de ocorrência (processos, participantes e circunstâncias – LSF) para cada um dos gêneros.

As análises comparadas ratificam o que já vinha sido dito sobre as cantigas de amor, mas que, por falta de metodologia adequada, carecia de comprovação empírica na microestrutura textual. Assim, é possível identificar estágios obrigatórios e opcionais nas cantigas de amor (cf. EPG), também no nível ideológico pois estão circunscritos à figura do fenhedor. Assim, os estágios obrigatórios são: (i) a “coita amorosa”; e (ii) a ênfase no sofrimento do trovador, pois, o sofrimento do trovador (eu-lírico) é fato, causa ou consequência em todas as cantigas do *corpus*.

A ‘coita’ é um símbolo de dupla face, à medida que o bem (estar e parecer – processos relacionais) da donzela é também o mal do trovador, simbolizado em vários momentos nas cantigas através dos diferentes processos e dos participantes-chave (a donzela e o trovador):

- materiais: o “fazer mal” da donzela e o “fazer bem” do trovador;
- relacionais: “estar” mal e não “ser” merecedor. A donzela “está” bem;
- mentais: “sentir” mal, “pensar” mal em si e bem na donzela; “querer” mal a si e bem a donzela.

Os estágios opcionais identificados são: (a) a retratação da beleza e formosura da donzela, (b) o malquerer da donzela e de Deus para com o trovador; e (c) o gatilho para a coita amorosa (ter visto ou ouvido a donzela).

ETAPA IV

A aplicação da LC para busca na base de dados online do “Projeto Littera” de processos-chave e participantes-chave confirma empiricamente as hipóteses. Considerando que a análise da metafunção ideacional das cinquenta e duas cantigas do *corpus* inicial (análise qualitativa) mostrou que, como fenhedor, o trovador se vê impelido a cumprir os estágios obrigatórios discursivos do preceituário amoroso, o uso da Linguística de *Corpus* confirmou o padrão de ocorrência de cada processo.

Valendo-se de processos mentais (35,41%), tais como: “saber”, “merecer”, “ver”, etc., o trovador recorre à voz de outro – Deus – para dizer o que ele próprio sente. Com os emotivos revela o seu “sentir” dor e amor pela donzela que o repudia, os cognitivos trazem seu “pensar” e “saber” de sua situação de incapacidade, os perceptivos assinalam basicamente o “ver” a donzela (início do seu mal) e os desiderativos manifestam, ainda que timidamente, seu “querer” amar. O trovador é quase sempre o Experienciador e, diferente do que se possa hipotetizar, o que “vê”, “sente”, “pensa” e “quer” não é essencialmente a donzela. O eu-lírico usa a grande maioria dos mentais para apresentar o seu ponto de vista e revelar seu “estado” de alma (OLIVEIRA, 2008). Na busca na base de dados online do “Projeto Littera”, apenas 4% dos processos apresentavam o trovador e a donzela como participantes na mesma oração (respectivamente Experienciador e Fenômeno), o que demonstra um dado importante, a orientação básica egocêntrica como traço característico do gênero. A Figura 1, abaixo, mostra como a busca do radical do verbo “sofrer” (processo mental) na base de dados – 369 ocorrências – confirma o posicionamento ideológico do fenhedor como estágio obrigatório do gênero (coita de amor).

Figura 1 – Busca na base de dados do processo mental “sofrer”.

369 ocorrências encontradas.

A *bôa dona por que eu trovava*, João Garcia de Guilhade
sofrendo coita sempre a servi; (linha 4)
pero, por coita grande que *sofri*, (linha 17)

A *dona fremosa do Soveral*, Lopo Lias
nom lho *sofrerei*; (linha 17)

A *dona que eu vi por meu*, Airas Veaz
afã e *sofrer* tanto mal, (linha 8)

A *la fé, Deus, se nom por Vossa Madre*, Gil Peres Conde
porque vos eu a vós esto *sofresse* (linha 7)

A *maior coita que eu vi sofrer*, Fernão Velho
A maior coita que eu vi *sofrer* (linha 1)
eu mi a *sofro*; e já que est assi, (linha 3)

A *mia senhor atanto lhe farei*;, Fernão Gonçalves de Seabra
quero-lh'eu já *sofrer* tod'outro mal (linha 2)
de pram, aquesto lhe nom *sofrerei*; (linha 4)
sofrer quero de nunca lhe dizer (linha 6)
Mais, de pram, esto nom posso *sofrer*; (linha 9)
E *sofrer*-lh'-ei quanta coita me dá, (linha 11)
mas, de pram, esto nom *sofrerei* já; (linha 14)

A *mia senhor já lh'eu muito neguei*, João Garcia de Guilhade
ousei na coita que *sofri* e no mal (linha 8)
a gram coita que por ela *sofri*; (linha 14)

A *mia senhor que eu por mal de mi*, D. Dinis
pero mi tod'este mal faz *sofrer*, (linha 10)

- *Abri! Pérez, mult'hei eu gram pesar*, Bernal de Bonaval
da gram coita que vos vejo *sofrer*, (linha 2)
tanto me faz maior coita *sofrer*. (linha 14)

Agora *m'hei eu a partir*, João Soares Somesso
nom lhe quero já *sofrer* rem, (linha 9)

Agora *me quer'eu já espedir*, Nuno Anes Cerzeo
quanto *sofri* (linha 26)

Agora *viv'eu como querria*, Rui Queimado
de já mais aquesta coita *sofrer* (linha 20)
do que *sofri*; e desejaria (linha 21)

Ai *amiga, tenh'eu por de bom sem*, Rodrigo Anes de Alvares
nem por senhor tam gram coita *sofrer* (linha 17)
com'el *sofre*, há mui longa sazom, (linha 18)

Ai Deus! *que coita de sofrer*, Airas Carpancho
Ai Deus! que coita de *sofrer* (linha 1)

Ai Deus! *Que grave coita de sofrer*;, Pero Garcia Buralgés
Ai Deus! Que grave coita de *sofrer*; (linha 1)

Ai *eu! que mal dia naci*, Pero Garcia Buralgés
que me mia senhor faz *sofrer*, (linha 23)

Fonte: Projeto Littera - <http://cantigas.fcsh.unl.pt>.

Os processos relacionais totalizavam 28,31% do total de ocorrências do *corpus* da análise inicial. Assim como os mentais, acompanham a orientação ideológica, pois grassam as orações com a exultação de um “estado” e um “ser” negativo para o eu-lírico, principalmente nos processos relacionais atributivos intensivos (a maioria das ocorrências), os quais apresentam atributos negativos ao trovador – aquele que, novamente, é quase supremo na primeira posição de participante (Portador e Identificado), e quando não, aparece como vítima na posição de Beneficiário (OLIVEIRA, 2008). Contrariamente, a donzela surge na posição de Portador com atributos positivos direcionados a ela; é o que comprova a busca na base de dados. Através das ocorrências do léxico “bem” (3400 ocorrências) obtém-se a confirmação ideológica do posicionamento de fenhedor no maniqueísmo (bem x mal) dos participantes-chave, exemplificado na Figura 2.

Figura 2 – Busca na base de dados do léxico “bem” (processo relacional).

Pesquisar em Cantigas
(Palavras nos versos:)

BEM

Pesquisa avançada

Filtrar Cantigas

Por cancionero

Por género

Por tipo

Por tema

Todos os temas e subtemas

A dona que eu quero bem, Rui Fernandes de Santiago
A dona que eu quero BEM (linha 1)

A dona que eu vi por meu, Airas Veaz
A que eu quero mui gram BEM (linha 13)
parar-me de lhe BEM querer, (linha 16)

A dona que home, Paio Gomes Charinho
todo seu BEM, "senhor" a chamaria. (linha 7)
Ca senhor é de muito BEM; e vi-a (linha 8)
ca BEM soub'eu quanto m'end'averria: (linha 11)
no seu BEM, que me nom soube guardar; (linha 17)
nem cuidava que tam BEM parecia (linha 18)
É por esto que BEM conselharía (linha 22)
quantos oirem no seu BEM falar: (linha 23)

A Far[õ] um dia irei, madre, se vos prouguer, João de Requeixo
rogar se verria meu amigo, que mi BEM quer, (linha 2)
e se cedo veer e o vir eu, BEM talhada, (linha 10)

Fonte: Projeto Littera - <http://cantigas.fcsh.unl.pt>.

Os processos materiais acompanham o posicionamento ideológico, mostram que o trovador “serviu” a donzela; da parte da donzela, ela o “maltrata”; e por fim, o trovador deixa a causa nas mãos de Nosso Senhor, que a “fez” valer mais do que as outras e deve “ajudá-lo”, mas, de fato, é seu algoz e o “faz” sofrer por ter o poder e não agir a seu favor.

Valendo-se do princípio da Ressonância (THOMPSON, 1998), as cantigas de amor apresentam um significado ideológico condensado na representação da experiência

(metafunção ideacional), o que se configura em estágios do gênero. Ao final das leituras das cantigas o leitor percebe o tom auto piedoso e dramático que o trovador pretende mostrar para criar a figura do fenhedor, este que “chora” e “balbucia”, tal qual uma criança (CESCHIN, 1998). Os processos, participantes e suas respectivas posições na oração criam, portanto, um efeito de ressonância nas cantigas de amor, a qual ecoa uma ideologia relativamente estável no discurso.

8. Considerações Finais

O presente artigo procurou revelar como a Linguística Sistêmico-Funcional e a Linguística de *Corpus* podem ser associadas para análises sobre gêneros discursivos. Nesse caso, o destaque foi mostrar como os complexos ideológicos que permeiam a fala do fenhedor nas cantigas de amor podem se consubstanciar em estágios estáveis e característicos do gênero. O trovador, enquanto fenhedor no enquadre do gênero em questão, não podia se manifestar diretamente ou pedir o amor (nem o poder amar) da sua donzela. A análise do sistema da transitividade (metafunção ideacional – LSF) revelou como os complexos ideológicos estão introjetados nas cantigas de amor no uso dos processos (representação da experiência). Os processos mentais (do “sentir”) são majoritários e criam o efeito da “coita amorosa” no mundo textual. Já os processos relacionais (“ser” sofredor) exultam a donzela e diminuem o trovador, criando uma hierarquização social característica do posicionamento do fenhedor e, por conseguinte, do gênero. Por outro lado, o “fazer” (processos materiais) está adstrito à incapacidade do eu-lírico para agir, em contrapartida, Deus surge como participante-chave no papel de ator e torna-se responsável por fazer o mal ao trovador.

Ao detectar os estágios obrigatórios e opcionais do gênero, a análise pela perspectiva cruzada da Linguística de *Corpus* e Linguística Sistêmico-Funcional avança nos pressupostos e doutrinas estabelecidas por análises literárias tradicionais relacionadas às cantigas de amor. Primeiramente, porque a análise linguística comprova como aspectos perceptíveis e sensíveis ao leitor incorporam uma dimensão ideológica superior ao texto. Além disso, pesquisadores, professores e alunos podem realizar análises mais robustas ao notar como os estágios do gênero e a ideologia estão presentes no texto linguisticamente, neste caso, pela representação da experiência. É necessário dizer que os mesmos padrões discursivos e outros traços característicos deverão também ser observáveis por análises sintático-semânticas análogas.

A Linguística de *Corpus* possibilita uma comprovação empírica da hipótese através do

auxílio de banco de dados digitais, para essa análise, o banco de dados do “Projeto Littera”. A pesquisa mostrou como uma sucessão de orações com um padrão de ocorrência de processos, participantes e circunstâncias pode realizar linguisticamente uma macro-argumentação ideológica e logogenética essencial e que não deve ser desconsiderada em análises abrangentes sobre gênero discursivo.

Referências

- ARTEMEVA, N. Approaches to learning genres: A bibliographical essay. In: ARTEMEVA, N.; FREEDMAN, A. **Rhetorical genre studies and beyond**. Winnipeg, Manitoba, Canada: Inkshed. p. 9-99, 2007.
- BAKHTIN, M. **Speech genres and other late essays**. Austin: University of Texas Press, 1986.
- _____. Os gêneros do discurso. **Estética da criação verbal**, v. 4, p. 261-306, 1992.
- BAZERMAN, C. Systems of genres and the enactment of social intentions. **Genre and the new rhetoric**, v. 79101, 1994.
- BHATIA, V. K. Introduction: Genre analysis and world Englishes. **World Englishes**, v. 16, n. 3, p. 313-319, 1997. <https://doi.org/10.1111/1467-971X.00066>
- _____. **Worlds of written discourse: A genre-based view**. A&C Black, 2004.
- _____. **Analysing genre: Language use in professional settings**. Routledge, 2014.
- BERKENKOTTER, C.; HUCKIN, T. **Genre knowledge in disciplinary communities**. Hillsdale, HJ: Lawrence Erlbaum, 1995.
- BERBER-SARDINHA, T. Lingüística de corpus: histórico e problemática. **Delta**, v. 16, n. 2, p. 323-367, 1999.
- _____. O que é um corpus representativo. **Direct Papers**, v. 44, 2000.
- BIBER, D. On the complexity of discourse complexity: A multidimensional analysis. **Discourse Processes**, v. 15, n. 2, p. 133-163, 1992. <https://doi.org/10.1080/01638539209544806>
- BIBER, D; CONRAD, S. Lexical bundles in conversation and academic prose. **Language and Computers**, v. 26, p. 181-190, 1999.
- BIBER, D.; CONNOR, U.; UPTON, T.A. **Discourse on the move: Using corpus analysis to describe discourse structure**. John Benjamins Publishing, 2007. <https://doi.org/10.1075/sci.28>

BIBER, D. et. al. **Corpus linguistics – Investigating language structure and use**. Cambridge: Cambridge University Press, 1998. <https://doi.org/10.1017/CBO9780511804489>

CERMAK, F. Czech National Corpus: A case in many contexts. **International Journal of Corpus Linguistics**, 2.2: 181-198, 1997. <https://doi.org/10.1075/ijcl.2.2.03cer>

CESCHIN, H. L. A educação do cavaleiro e a expressão do amor no trovadorismo In: **Quaderni Nuova Serie** - nº 10. v. 10, São Paulo: [s.n.], 1998. p. 139-173.

CLARK, K.; HOLQUIST, M. **Mikhail Bakhtin**. Boston: Harvard University Press, 1984.

EGGINS, S.; MARTIN, J. R. Genres and registers of discourse. **Discourse as structure and process**, v. 1, p. 230-256, 1997. <https://doi.org/10.4135/9781446221884.n9>

FAIRCLOUGH, N. **Language and power**. London; New York: Longman, 1992.

FAIRCLOUGH, N.; WODAK, R. Critical discourse analysis. In: VAN DIJK (T.). **Discourse as Social Interaction**. London: Sage, 1997.

FILLMORE, C. Corpus linguistics or computer corpus linguistics. **Directions in corpus linguistics. Proceedings of Nobel symposium 82**, Stockholm, Ed. Jan Svartvik, 35-60. Berlin/New York, De Gruyter, 1992.

FOWLER, R. The intervention of the media in the reproduction of power. **Approaches to discourse, poetic and psychiatry**, p. 67-80, 1987.

_____. **Language in the news: Discourse and ideology in the press**. London: Routledge, 1991.

_____. **Linguistic criticism**. (2ª ed.). Oxford: Oxford UP, 1996a.

_____. On critical linguistics. **Texts and practices: Readings in critical discourse analysis**, p. 1, 1996b.

GUERREIRO, M. V. **Para a história da literatura popular portuguesa**. Inst. de Cultura e Língua Portuguesa, Min. da Educação e Cultura, 1978.

GOFFMAN, E. **Frame analysis**. New York: Harper Colophon Books, 1974.

GRANGER, S. (Org.) **Learner English on Computer**. New York: Longman, 1998.

HALLIDAY, M. A. K. **An Introduction to Functional Grammar**. (2ªed.). London: Edward Arnold, 1994.

HALLIDAY, M. A. K.; HASAN, R. Language, context and text. **Aspects of language in a social-semiotic perspective**. 2ªed. London: Oxford University Press, 1989.

HALLIDAY, M. A. K.; MARTIN, J. R. **Writing Science: literacy and discursive power**. London/Washington: The Falmer Press, 1993.

HALLIDAY, M. A. K.; MATHIESSEN, M. I. M. **An introduction to functional grammar**. (3^{ed.}). London: Arnold, 2004.

HASAN, R. Speech genre, semiotic mediation and the development of higher mental functions. **Language Sciences**, v. 14, n. 4, p. 489-528, 1992. [https://doi.org/10.1016/0388-0001\(92\)90027-C](https://doi.org/10.1016/0388-0001(92)90027-C)

_____. The uses of talk. **Discourse and social life**, p. 28-47, 2000.

HYLAND, K. Genre-based pedagogies: A social response to process. **Journal of second language writing**, v. 12, n. 1, p. 17-29, 2003. [https://doi.org/10.1016/S1060-3743\(02\)00124-8](https://doi.org/10.1016/S1060-3743(02)00124-8)

HYON, S. Genre in three traditions: Implications for ESL. **TESOL quarterly**, v. 30, n. 4, p. 693-722, 1996. <https://doi.org/10.2307/3587930>

JAKOBSON, R.; BOAS, F. Franz Boas' approach to language. **International Journal of American Linguistics**, v. 10, n. 4, p. 188-195, 1944. <https://doi.org/10.1086/463841>

KRESS, G. **Learning to write**. London, Routledge & Kegan, 1982.

_____. Multimodality, Multimedia, and Genre.?. **Visual rhetoric in a digital world**, p. 38-54, 2004.

LEECH, G. **English in advertising: a linguistic study of advertising in Great Britain**. London: Longman, 1966.

_____. Reading images: Multimodality, representation and new media. **Information Design Journal**, v. 12, n. 2, p. 110-119, 2004. <https://doi.org/10.1075/idjdd.12.2.03kre>

LI, J. Transitivity and lexical cohesion: Press representations of a political disaster and its actors. **Journal of Pragmatics**, v. 42.12. 2010. p. 3444-3458. <https://doi.org/10.1016/j.pragma.2010.04.028>

LOPES, G.V et al. **Cantigas Medievais Galego Portuguesas** [base de dados online]. Lisboa: Instituto de Estudos Medievais, FCSH/NOVA. Disponível em: <http://cantigas.fesh.unl.pt>. Consulta em: 01 nov. 2016.

MARTIN, J. R. et al. Language, register and genre. **Children writing: reader**, v. 1, p. 984, 1984.

_____. Process and text: two aspects of human semiosis. **Systemic perspectives on discourse**, v. 1, n. 15, p. 248-274, 1985.

_____. Genre and literacy-modeling context in educational linguistics. **Annual review of applied linguistics**, v. 13, p. 141-172, 1992. <https://doi.org/10.1017/S0267190500002440>

_____. Analysing genre: functional parameters. In: JR Martin. **Genre and institutions: Social processes in the workplace and school**, p. 3, 2005.

MILLER, C. R. The cultural basis of genre. In: **Genre and the new rhetoric**. 1994.

OLIVEIRA, Ulisses Tadeu Vaz de et al. **O fenhedor e o precador nas cantigas líricas galego-portuguesas de D. Dinis**: uma perspectiva sistêmico-funcional. 2008. 162 f. Dissertação (Mestrado). LAEL, Pontifícia Universidade Católica, São Paulo, 2008.

PALTRIDGE, B. Genre, text type, and the English for Academic Purposes (EAP) classroom. **Genre in the classroom: Multiple perspectives**, p. 73-90, 2002.

REYNOLDS, M. The blending of narrative and argument in the generic texture of newspaper editorials. **International Journal of Applied Linguistics**, 10.1, 2000, p. 25-40.

SÁNCHEZ, A. **Cumbre**: corpus lingüístico del español contemporáneo: fundamentos, metodología y aplicaciones. Sociedad General Española de Librería, 1995.

SCOTT, M. **WordSmith tools**. Oxford: Oxford University Press, 1996.

SINCLAIR, J. McH. Beginning the study of lexis. In: C. E. BAZELL (org.). **In Memory of J. R. Firth**. London: Longman, 1966.

_____. Collocation: a progress report. In: R. STEELE & T. THREADGOLD (org.). **Language topics – Essays in honour of Michael Halliday** (Vol. 2). Amsterdam/Philadelphia: John Benjamins, 1987.

SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

SVARTVIK, J. (Org.) **Directions in Corpus Linguistics**: Proceedings of Nobel Symposium 82 – Stockholm, 4-8 August 1991 (Trends in Linguistics – Studies and Monographs: 65). Berlin, New York: Mouton De Gruyter, 1992.

SPINA, S. **A lírica trovadoresca**. 3ª ed. São Paulo: Edusp, 1996.

SWALES, J. **Genre analysis**: English in academic and research settings. Cambridge University Press, 1990.

THOMPSON, G. **Introducing Functional Grammar**. London: Arnold, 1996.

_____. Resonance in text. In: SANCHEZ-MACARRO & CARTER, R. (eds.) **Linguistic choice across genres**: variation in spoken and written English. London: John Benjamins, 1998. <https://doi.org/10.1075/cilt.158.05tho>

THORNDIKE, E. L. **Teacher's Wordbook**. New York: Columbia Teachers College, 1921.

VAN DIJK, T. A. Structures of news in the press. **Discourse and communication**, p. 69-93, 1985.

_____. **Discourse studies**: a multidisciplinary introduction. v.1 London: Sage, 1997. <https://doi.org/10.4135/9781446221884.n1>

_____. **Cognição, discurso e interação**. São Paulo: Contexto, 1999.

YUNICK, S. Genres, registers and sociolinguistics. **World Englishes**, v. 16, n. 3, p. 321-336, 1997. <https://doi.org/10.1111/1467-971X.00067>

ZHOU, Q.; YU S. Annotating the Contemporary Chinese Corpus. **International Journal of Corpus Linguistics**, 2.2: 199-238, 1997. <https://doi.org/10.1075/ijcl.2.2.05qia>

Artigo recebido em: 15.11.2016

Artigo aprovado em: 02.04.2017

Modelagem topológica do perfil sistêmico-funcional de manuais de instrução no espaço gramatical

Modeling topologically systemic-functional profiling of handbooks in the grammar space

Francieli Oliveira*

Thaís Campolina**

Giacomo Figueredo***

RESUMO: Embasado nos pressupostos da Linguística Sistêmico-Funcional e da Linguística de *Corpus*, o presente artigo investiga o modo pelo qual os sistemas gramaticais da oração contribuem para a produção de significados do texto. Mais especificamente, objetiva construir um modelo dinâmico da gramática que descreva a produção textual no manual de instrução em português brasileiro. Uma vez que a língua é constituída de probabilidades de ocorrência de funções que compõem subconjuntos, aos quais se denominam aqui, é possível descrevê-los à medida que as funções linguísticas são descritas. Para a modelagem, foi compilado um *corpus* de manuais de instrução e analisado segundo os sistemas principais da oração. Os resultados mostram como foi possível produzir um modelo, o qual aponta que as funções que realizam o manual prototipicamente são processo material, relacional e verbal, modo declarativo e jussivo e tema default e perspectiva. Além disso, configura-se como uma das formas de generalização das probabilidades maiores de escolhas dos sistemas gramaticais para esse gênero textual.

PALAVRAS-CHAVE: Modelagem sistêmico-funcional. Perfilação sistêmica. Linguística Computacional. Linguística de *Corpus*. Manual de instrução.

ABSTRACT: This paper adopts Systemic Functional Linguistics and Corpus Linguistics to investigate clause grammar systems and their contributions to text meaning-making. More specifically, it aims at building a dynamic model of grammar describing text production in handbooks in Brazilian Portuguese. Language is a probabilistic system of functions and these probabilities form subsets of textual genres. As the language functions are described, genres can be described and modelled in order to understand how functions build their meaning-making. A handbook corpus was compiled and profiled for the main clause grammar functions. The model shows that the prototypical handbook functions are material, relational and verbal process, declarative and jussive mood, and default and perspective theme. Accordingly, it generalizes the most frequent probabilities for choices in grammar systems for the handbook.

KEYWORDS: Systemic functional modeling. Systemic profiling. Computational Linguistics. Corpus Linguistics. Handbooks.

* Mestranda em Linguística Aplicada. Universidade Federal de Minas Gerais (UFMG).

** Graduada em Letras. Universidade Federal de Ouro Preto (UFOP).

*** Doutor em Linguística Aplicada. Departamento de Letras. Universidade Federal de Ouro Preto (UFOP).

1. Introdução

Campos disciplinares vinculados à Linguística e à Ciência da Computação, dentre eles a Linguística Computacional e a Inteligência Artificial, vêm crescentemente demandando contribuições mais efetivas dos estudos da língua para o Processamento de Linguagem Natural (PLN) (MATTHIESSEN; BATEMAN, 1991). Esse fato tem impulsionado a Linguística Aplicada a produzir conhecimento para oferecer modelos confiáveis da produção linguística, que não se baseiem apenas em regras formais de gramáticas livres de contexto; mas, de outra forma, tomem como parâmetro de processamento a Compreensão das Línguas Naturais (*Natural Languages Understanding*) (WINOGRAD, 1972; HALLIDAY, 2005).

De modo complementar, tem sido incrementada, no âmbito da Linguística Aplicada, a necessidade de implementação do processamento de línguas naturais na interação entre o humano e o computador (WINOGRAD, 2006), incorporando a máquina às suas práticas de pesquisa e aplicação. Dentre essas demandas, destaca-se a busca por modelos que extrapolem a ordem da oração, em particular voltando-se para a estrutura de textos e, por conseguinte, de gêneros textuais (MATTHIESSEN, 2015).

Situando-se nesse contexto, este artigo visa contribuir com soluções para as demandas do PLN relativamente à Linguística Aplicada. Dessa forma, tem por objetivo oferecer ao PLN um modelo linguístico dos sistemas gramaticais que realizam o gênero textual “manual de instrução”, pautado pelos pressupostos da Linguística de *Corpus* (BERBER SARDINHA, 2000; VIANA, 2011) e pelo arcabouço sistêmico-funcional (HALLIDAY, 2002; HALLIDAY; MATTHIESSEN, 2014).

Por conseguinte, este trabalho analisa as formas pelas quais a gramática da oração constrói a produção de significado do manual de instrução. Mais especificamente, produz um modelo capaz de descrever a maneira pela qual se dão as seleções sucessivas de opções nas redes dos sistemas (*system networks*) e, conseqüentemente, como essa dinâmica explica a organização do texto. A partir desse modelo, visa-se apresentar uma descrição gramatical do manual de instrução contribuindo para a descrição de gêneros textuais.

A Linguística de *Corpus* compreende que o estudo do funcionamento da língua deve ser realizado de forma empírica através de um *corpus* composto por textos falados e/ou escritos de forma espontânea (VIANA, 2011). Biber (2010) propõe que o estudo dos registros de uma língua deve ser realizado com base em um *corpus* composto por textos pertencentes ao gênero, para que seja possível investigar os padrões que o definem. Como a língua é um

sistema de probabilidades (HALLIDAY, 1991; BERBER SARDINHA, 2000), esses padrões apresentam um subconjunto de probabilidades de ocorrência de funções que fazem parte da construção de significados da língua. Isso significa que um gênero textual pode ser modelado exatamente desta maneira: um subconjunto de probabilidades de funções (HALLIDAY, 2002).

A Linguística Sistêmico-Funcional (LSF) (cf. HALLIDAY, 2002; HALLIDAY; MATTHIESSEN, 2014; MARTIN, 1992; MATTHIESSEN, 1995, dentre outros) compõe a base teórica e analítica deste artigo. Em primeiro lugar, essa escolha se deu pelo fato de essa teoria conceber a língua como um sistema de recursos que permite a representação, encenação e construção da experiência de mundo e das relações sociais dos falantes por meio de unidades de significado organizadas como configurações de funções, ou ‘textos’. Somando-se a isso, na qualidade de teoria abrangente, a LSF proporciona um arcabouço de descrição e análise do sistema linguístico como um todo, em seus diferentes estratos, de forma que a gramática pode ser examinada tanto em relação aos elementos que a expressam quanto em relação ao texto que organiza (HALLIDAY, 2010). Dessa maneira, a análise gramatical é capaz de revelar como se organizam os aspectos fundamentais dos gêneros textuais.

Sendo a produção de significado do manual de instrução o objeto de estudo deste artigo, cabe apontar que, do ponto de vista sistêmico-funcional, o significado é a propriedade da língua de relacionar funções organizadas sistemicamente, de forma que ao contraste entre elas se atribua um *valeur*¹ (SAUSSURE, 2006). Assim, o conceito de ‘função’ depende tanto dos recursos de produção de significado quanto da forma como a língua se organiza (HALLIDAY 2008). O modelo da língua ou o modelo de um gênero textual consiste em um conjunto de sistemas, a partir dos quais as relações entre os recursos geram as funções. Assim, toda e qualquer análise, descrição e aplicação sistêmico-funcionais são derivadas, em última instância, da dimensão sistêmica do sistema linguístico – denominada ‘dimensão do eixo’ (*dimension of axis*) (MARTIN, 2013).

A metodologia e a análise presentes neste artigo, portanto, procuram derivar seus protocolos de realização igualmente da dimensão do eixo. Para tanto, utilizam o *software* UAM CorpusTool (O’DONNELL, 2008) para análise semiautomática do *corpus* e, em seguida, a metodologia da perfilação sistêmica (FIGUEREDO, 2014), a qual estabelece

¹ O termo *valeur* é adotado pela LSF e se conserva a palavra original em francês, mesmo quando empregado em outras línguas (cf. MARTIN, 2013).

distâncias topológicas entre as funções sistêmicas e seu movimento no espaço gramatical. A perfilação gramatical – ou a aplicação da perfilação sistêmica à gramática – possibilita criar um modelo da produção de significado nos textos, de modo a descrever e prever seu funcionamento (OLIVEIRA, 2015).

Neste artigo, a definição de ‘modelagem’ parte da LSF e implica a criação de um modelo linguístico para futura aplicação, que pode variar desde áreas como o ensino e a análise do discurso até a Linguística Computacional, em estudos que envolvem o Processamento de Linguagem Natural (PLN). Dessa maneira, a modelagem proposta neste artigo se configura como uma resposta positiva à demanda da Linguística Aplicada por aplicações sistêmicas derivadas da dimensão do eixo, as quais envolvam o estudo de sistemas, escolhas e funções na constituição de textos/tipos de texto (MAHBOOB; KNIGHT, 2010) incluindo sua variação de frequências e probabilidades de conformação ao gênero textual (HALLIDAY, 1991; LEMKE, 1993).

Especificamente neste artigo, a modelagem sistêmico-funcional de perfilação é realizada em um *corpus* de manuais de instrução, tendo também como meta construir um modelo dinâmico da gramática que descreva a produção textual no manual de instrução em português brasileiro. Com isso, este trabalho apresenta, com base na perfilação gramatical, os recursos linguísticos utilizados para compreender os significados construídos no manual de instrução.

Ao final, este artigo demonstra como o arcabouço da LSF e Linguística de *Corpus* permite não somente analisar e descrever o gênero textual manual de instrução, como também, criar um modelo que aponte as probabilidades linguísticas que caracterizam esse texto. Uma vez que o modelo possui potencial de implementação computacional, ele pode contribuir para a interface entre o PLN e a Linguística Aplicada.

2. Pressupostos teóricos

Halliday (2002) aponta que, na tentativa de compreender melhor a língua, os falantes tentam o tempo todo estabelecer padrões, os quais são somente criados mediante as diferenças existentes na língua. Sendo assim, os significados são compreendidos por meio do contraste entre funções, ou *valeur* (GEASON, 1965; SAUSSURE, 2006). Pensando nisso, a descrição linguística consiste em conhecer a língua a fim de se estabelecer uma explicação para o seu funcionamento, de forma que essa explicação represente a modelagem linguística. Nesse

sentido, Halliday (2002) afirma que um modelo da língua deve contemplar seus diferentes contextos para diferentes propósitos, pois não é somente a base para a compreensão da língua, pelo modelo também é possível enxergar as informações relevantes sobre outros aspectos linguísticos.

Para modelar a língua, primeiramente, é necessária a sua descrição. Nesse sentido, a LSF oferece o arcabouço para a investigação de sistemas que trazem as possibilidades de escolhas de uma língua (HALLIDAY et al., 1964), assim como o seu significado e sua diferença com relação às outras escolhas (SAUSSURE, 2006). Por meio da organização sistêmica, na dimensão do eixo, é possível descrever, contribuir com a teoria linguística e modelar a língua.

Para a construção de um modelo da produção textual, toma-se a LSF como teoria para especificamente abordar três questões importantes (HALLIDAY, 2002; HALLIDAY, 2008): (i) a generalização da descrição de um *corpus* específico para o gênero textual; (ii) o grau de variação entre textos distintos e a acomodação de sua variabilidade no modelo; e (iii) a aplicabilidade do modelo para necessidades reais do uso da língua.

Como forma de atender a essas três questões, adota-se a modelagem topológica da gramática. O mapa topológico é a representação espacial das funções de um determinado sistema gramatical (MARTIN; MATTHIESSEN, 1991; FIGUEREDO, 2014). As funções são derivadas das opções (*features*) em um sistema. Por exemplo, o sistema de MODO possui em um primeiro nível duas opções, indicativo e imperativo. Destas opções se derivam as funções gramaticais de Indicativo e Imperativo.

Além disso, quando duas opções são derivadas de um mesmo sistema são denominadas formas agnatas. Neste caso as opções são, por um lado, semelhantes por serem derivadas de um mesmo sistema; mas, por outro lado, são opostas pelo fato da seleção de uma opção implicar na exclusão da outra opção. No sistema de MODO, por exemplo, indicativo e imperativo são agnatas. Com isto, são semelhantes por ambas serem derivadas do sistema de MODO e, concomitantemente, são opostas por que a seleção de indicativo implica na exclusão de imperativo. Esta relação de semelhança-e-oposição se denomina agnação (GLEASON, 1965, p. 199).

A partir deste pensamento sistêmico, é possível compreender que o significado não está exatamente em nenhuma opção sistêmica, mas na relação entre opções. Por conseguinte, nenhuma função linguística possui valor “em si”; o valor se encontra na relação de

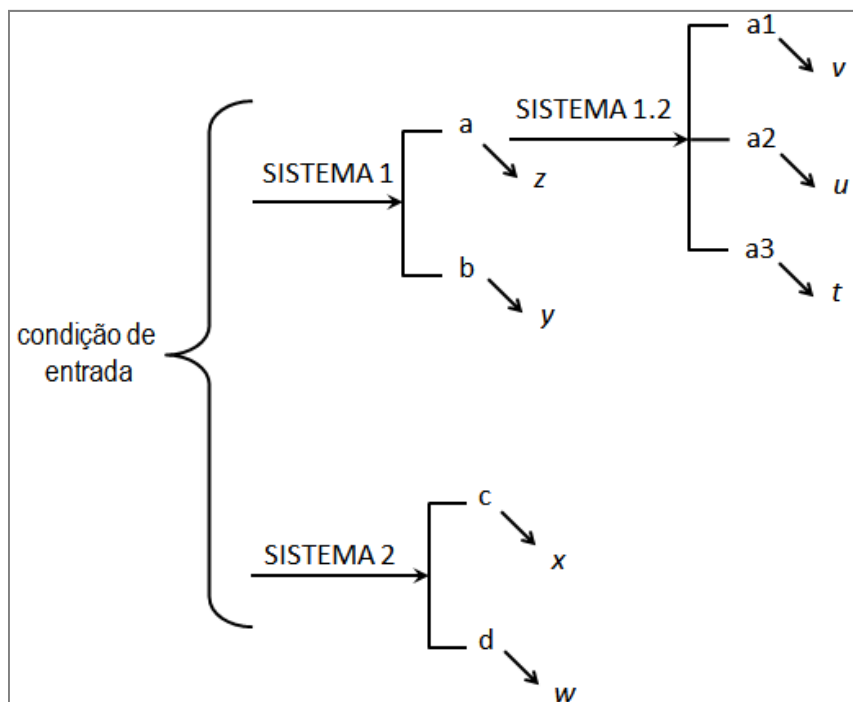
semelhança-e-oposição sistêmica. Este pensamento sistêmico, introduzido nos estudos linguísticos por Saussure (2006), foi caracterizado da seguinte forma.

Verifiquemos inicialmente que, mesmo fora da língua, todos os valores parecem estar regidos por esse princípio paradoxal. Eles são sempre constituídos: (1) por uma coisa dessemelhante, suscetível de ser trocada por outra cujo valor resta determinar; (2) por coisas semelhantes que se podem comparar com aquela cujo valor está em causa (p. 134).

O significado relacional das funções linguísticas se denomina valor, ou *valeur*. No mapa topológico, o *valeur*, que é estabelecido pela agnação, é representado pela variação em função da distância – *i.e.*, a distância gramatical, que é medida em *valeur* e suas reactâncias – assim, quanto maior o *valeur* entre duas funções, maior a distância entre elas. Assim, a distância entre as funções Indicativo e Imperativo é pequena, pois o *valeur* que as relaciona é pequeno. Por outro lado, a distância entre as funções Indicativo: Interrogativo: Polar e Imperativo: Jussivo: Ordem é maior, pois o *valeur* que as relaciona é relativamente grande.

O primeiro passo para a produção de modelos está no entendimento de como a LSF concebe a organização da língua como sistema. Uma vez que o objetivo da organização sistêmica é apresentar o funcionamento da língua, cada sistema corresponde a uma classe; assim, uma rede de sistemas corresponderá a uma parte da língua que se quer descrever para, posteriormente, ser modelada. Dessa forma, o sistema constitui-se como a base para a modelagem linguística. A Figura 1 apresenta a disposição do sistema em rede.

Figura 1 – Rede do sistema.



Fonte: os autores.

O sistema possui relações que são determinantes para a sua criação e interpretação. Como mostra a Figura 1, o primeiro elemento do sistema é a condição de entrada, o qual determina a unidade que será descrita. Em seguida apresentam-se os princípios de classificação, que são chamados de sistemas – escritos em caixa alta (MARTIN, 2013). Cada sistema possui opções de classes (‘a’, ‘b’, ‘c’, ‘d’, ‘a1’, ‘a2’, ‘a3’) e cada classe é apresentada com sua realização (‘z’, ‘y’, ‘x’, ‘w’, ‘v’, ‘u’, ‘t’).

Existe uma relação de oposição no sistema, como por exemplo, entre a classe ‘a’ e a classe ‘b’. Apesar de serem geradas pelo mesmo sistema, são opostas. Essa relação de semelhança-e-oposição, como visto anteriormente, é a agnação (GLEASON, 1965). Nesse sentido, a classe ‘a’ e a classe ‘b’ são agnatas. Além disso, quanto mais à direita estiver o sistema, maior o nível de distinção entre os itens, sendo esse nível de distinção chamado de delicadeza (MARTIN, 1992). Após a descrição de uma unidade, é possível visualizar as relações entre as categorias e o significado de cada uma dentro da língua em funcionamento.

Nesse sentido, devido à possibilidade de construção de sistemas de categorias linguísticas (HALLIDAY, 2002), o texto passa a ser compreendido como o produto de escolhas que geram significados nos diversos contextos de situação. Assim, a língua passa a ser um conjunto de textos diversos, organizados, primeiramente, pela dimensão do eixo (URE, 1989; MARTIN, 2013).

A organização da língua conforme a dimensão do eixo abre a possibilidade para a cosseleção de subsistemas para um mesmo fenômeno. Esta cosseleção, por sua vez, permite que os subsistemas se agrupem de forma consistente na categorização de determinado fenômeno. Assim, o estrato gramatical se organiza não só pelos subsistemas, mas, igualmente, pelo agrupamento desses na forma de cosseleções. A cada um desses agrupamentos dá-se o nome de metafunção (ou a função que um grupo de subsistemas desempenha no próprio sistema). Na gramática, as metafunções são três: ideacional, interpessoal e textual.

A metafunção ideacional é responsável pela representação da experiência do mundo; para isso, ela utiliza o sistema da TRANSITIVIDADE, que gera as funções dos cinco Tipos de Processo do português brasileiro² (FIGUEREDO, 2011): Material, Mental, Verbal, Relacional e Existencial. A metafunção interpessoal é responsável pela interação existente entre os falantes; para isso, ela utiliza o sistema de MODO (FIGUEREDO, 2011), que gera cinco funções de interesse para a presente pesquisa, a saber, Imperativo: Jussivo, Imperativo: Sugestivo, Indicativo: Declarativo, Indicativo: Interrogativo: Polar e Indicativo: Interrogativo: Elemental. A metafunção textual é responsável por semiotizar os significados da metafunção ideacional e interpessoal em texto. Ela gera, por meio do sistema de TEMA, que é o “ponto de partida”³ (MARTIN; ROSE, 2007, p. 107) da mensagem, as funções: Default, Tema Elemental, Ângulo Fonte, Ângulo Ponto de Vista, Perspectiva, Intensivo (FIGUEREDO, 2011).

A posição relativa de uma opção sistêmica reflete o seu valor funcional, e o mapa topológico opera como uma “descrição dinâmica” da língua (LEMKE, 1993). Como consequência, as relações entre as opções sistêmicas da produção textual podem ser quantificadas, o que abre a possibilidade para seu comportamento (*i.e.*, variação) ser modelado.

A partir da distância topológica, é possível projetar como determinado texto se desenvolve no tempo, relativamente às funções que emprega. Em outras palavras, é possível descrever as funções na primeira oração do texto (tempo = 1), em seguida na segunda oração

² Ao contrário da gramática do inglês, a gramática do PB não apresenta motivação sistêmica para o Tipo de Processo Comportamental. Em PB, o comportamento emprega ora a gramática material, ora a gramática mental/verbal.

³ “Starting point” (MARTIN; ROSE, 2007, p. 107).

(tempo = 2), na terceira (tempo = 3) e assim por diante, obtendo-se uma “descrição instantânea” da língua.

Assim, um texto que esteja ainda em produção pode ir sendo descrito à medida que vai sendo produzido a cada instante. A descrição instantânea se define como uma descrição gramatical do texto em qualquer momento de sua produção, e não apenas do produto final; é a descrição do texto como processo.

Para tanto, tomam-se a oração como unidade de análise, as opções dos sistemas gramaticais fundamentais da oração TRANSITIVIDADE (trs.), MODO (mod.) e TEMA (tem.) como eixos de coordenadas para a variação de distância e a dinâmica textual para o emprego de cada oração na variação de tempo, em que o tempo tem como quantidade uma oração. Isso leva conseqüentemente à produção de um “vetor-oração” que descreve o movimento do texto entre as orações. A “distância gramatical” e a “sequência temporal” mostram como os sistemas gramaticais “se movem” no espaço, o que se denomina ‘dinâmica de sistemas’.

Como forma de aplicar essa concepção de modelo ao *corpus* da pesquisa, apresenta-se a seguir a metodologia utilizada nesta pesquisa.

3. Metodologia

A metodologia para a perfilação sistêmica se divide em duas partes. Na primeira, apresenta-se o *corpus* investigado; na segunda, os procedimentos que levam à análise, os quais se dividem em (1) estudo do *corpus* e (2) anotação sistêmica.

O presente trabalho utilizou os dados de um *corpus* em português brasileiro (PB) composto por dez⁴ textos caracterizados comumente como *manual*, como aponta Biber (2010) que a descrição das funções gramaticais para um gênero textual não precisa, de fato, ser realizada em textos completos, mas pode ser realizada por amostras. Por conseguinte, os textos foram compilados a partir da tipologia da língua no contexto de cultura (URE, 1989), a qual classifica os textos em oito processos sociossemióticos (explicar, relatar, recriar, compartilhar, fazer, recomendar, capacitar e explorar) e quatro modos de produção (escrito/monólogo, escrito/diálogo, falado/monólogo e falado/diálogo) (HALLIDAY;

⁴ A quantidade de textos empregados no estudo se baseou na capacidade de análise manual, uma vez que ainda não existem estudos indicando um número de textos que seja significativo para a modelagem da dinâmica de sistemas.

MATTHIESSEN, 2014). Mais especificamente, para a análise desta pesquisa, foi compilado um *corpus* de textos classificados como capacitar, escrito/monólogo.

Por conseguinte, o critério da tipologia da língua no contexto de cultura (URE, 1989) foi utilizado na compilação do *corpus* de manuais de acordo com as variáveis do contexto de situação, definidas pela LSF (HALLIDAY, 1978; PAGANO, 2015):

Campo:

- Explicação de um procedimento (MATTHIESSEN et al., 2008);
- Exposição das orientações para o leitor (TEICH, 1999).

Sintonia⁵:

- Grau de especialização da linguagem: especializado para não especializado (HALLIDAY, 1978; MATTHIESSEN et al., 2010);
- Relação com o texto: não-familiar e distante (HALLIDAY, 1978; MATTHIESSEN et al., 2010).

Modo:

- Fonte de coleta (HALLIDAY et al., 1964): instituições renomadas do Brasil;
- Modo textual: institucional, pragmático, eletrônico e com objetivo de guiar os leitores para ações específicas (HALLIDAY et al., 1964; ROSE, 2001).

Consoante os procedimentos supracitados, foi possível obter textos pertencentes especificamente ao tipo de texto de interesse deste trabalho, que, de acordo com pesquisas anteriores, será tratado, agora tecnicamente, como ‘manual’ (cf. AZENHA, 1995; LIMA, 2013).

Conforme o estabelecimento desses critérios, a Tabela 1 apresenta os manuais com sua numeração, etiqueta, classificação segundo o processo sociosemiótico e o modo de produção

⁵ Neste artigo, traduz-se ‘tenor’ como ‘sintonia’ e ‘relations’ como ‘relações’.

(HALLIDAY; MATTHIESSEN, 2014), fonte de coleta e número de *tokens* referente à macro-onda de informação⁶ (MARTIN; ROSE, 2007) analisada de cada manual.

Tabela 1 – Perfil do *corpus* de manuais capacitar, escrito/monólogo.

Manual	Etiqueta	Classificação	Instituição	Número de <i>tokens</i>
Manual técnico de Gemas	GEM_01	Capacitar/escrito/monólogo	IBGM- DNPM	1915
Manual de Procedimentos de Enfermagem	ENF_02	Capacitar/escrito/monólogo	Governo do Distrito Federal	793
Manual do Diretor técnico	DIR_03	Capacitar/escrito/monólogo	CREMERJ	1963
Manual técnico de hemovigilância	HEM_04	Capacitar/escrito/monólogo	ANVISA	1105
Manual de Biossegurança em odontologia – UFPB	BIO_05	Capacitar/escrito/monólogo	UFPB	835
Manual de redação da presidência da república	RED_06	Capacitar/escrito/monólogo	Presidência da República	1301
Manual de noções básicas de primeiros socorros	SOS_07	Capacitar/escrito/monólogo	UNICAMP	493
Manual técnico - utilização de produtos fitofarmacêuticos	FIT_08	Capacitar/escrito/monólogo	Direção Geral da Agricultura e Desenvolvimento Ambiental	1302
Manual de normas técnicas - combate à Dengue	DEN_09	Capacitar/escrito/monólogo	FUNASA	789
Manual técnico - criação de pacu	PAC_10	Capacitar/escrito/monólogo	Itaipu Binacional e Unioeste/Campus de Toledo	498
			TOTAL	10994

fonte: os autores.

Após a compilação do *corpus*, foram adotados dois procedimentos de análise: (1) estudo do *corpus*; e (2) anotação sistêmica.

(1) Estudo do *corpus*: primeiramente, para a construção do perfil sistêmico do manual, analisaram-se os textos pelo viés do fluxo de informação. Martin e Rose (2007) apontam que o fluxo de informação é a forma como a metafunção textual organiza seus significados para

⁶ O termo macro-onda de informação apresenta o fluxo de informação no texto em nível macro. A palavra “onda” é utilizada dentro da LSF como termo técnico para se referir à estrutura do fluxo de informação do texto (MARTIN; ROSE, 2007, p. 188).

construir o texto. Nesse sentido, o Quadro 1 apresenta o *corpus* em relação ao fluxo de informação presente em cada texto.

Quadro 1 – Ondas de informação *corpus* de manuais capacitar/escrito/monólogo.

MANUAL	ELEMENTOS PARATEXTUAIS	CONTEÚDO TEXTUAL				REFERÊNCIAS
		Apresentação	Introdução	Desenvolvimento		
				Explicação	Como fazer	
GEM_01	x	X		X	X	
ENF_02	x	X		X	X	x
DIR_03	x	X		X	X	x
HEM_04	x	X	X	X	X	x
BIO_05	x	X		X	X	x
RED_06	x	X		X	X	x
SOS_07				X	X	x
FIT_08	x	X		X	X	x
DEN_09	x	X		X	X	x
PAC_10	x	X	X	X	X	

Fonte: os autores.

O Quadro 1 mostra que o manual possui três ondas de informação: (1) elementos paratextuais, que se referem aos elementos que aparecem antes do texto, como capa, carta ao leitor e índice; (2) conteúdo textual, que é composto por três ondas: apresentação, introdução e desenvolvimento; e (3) referências bibliográficas.

A marcação ‘x’ indica as ondas de informação que estão presentes nos manuais. Ademais, a marcação ‘X’, em negrito e caixa alta, aponta a presença das ondas de informação ‘Explicação’ e ‘Como fazer’ em todos os manuais no ‘Desenvolvimento’. A seção ‘Explicação’ expõe uma explicação sobre o assunto tratado; e a seção ‘Como fazer’ explica os procedimentos de como realizar a atividade determinada no manual.

Como aponta o Quadro 1, a onda de informação ‘Desenvolvimento’ é uma hiperonda de informação (MARTIN; ROSE, 2007), composta por várias ondas menores que possuem a mesma forma padrão: explicação e como fazer. Essas ondas menores de informação são a parte do manual nas quais a língua é usada para facilitar o comportamento do falante em determinada atividade. Sendo assim, essa seção é composta por textos capacitar, os quais compõem o *corpus* de análise deste artigo.

(2) Anotação sistêmica: para essa etapa, o *corpus* foi etiquetado e armazenado segundo os critérios de etiquetamento e armazenamento do Catálogo da Língua Brasileira

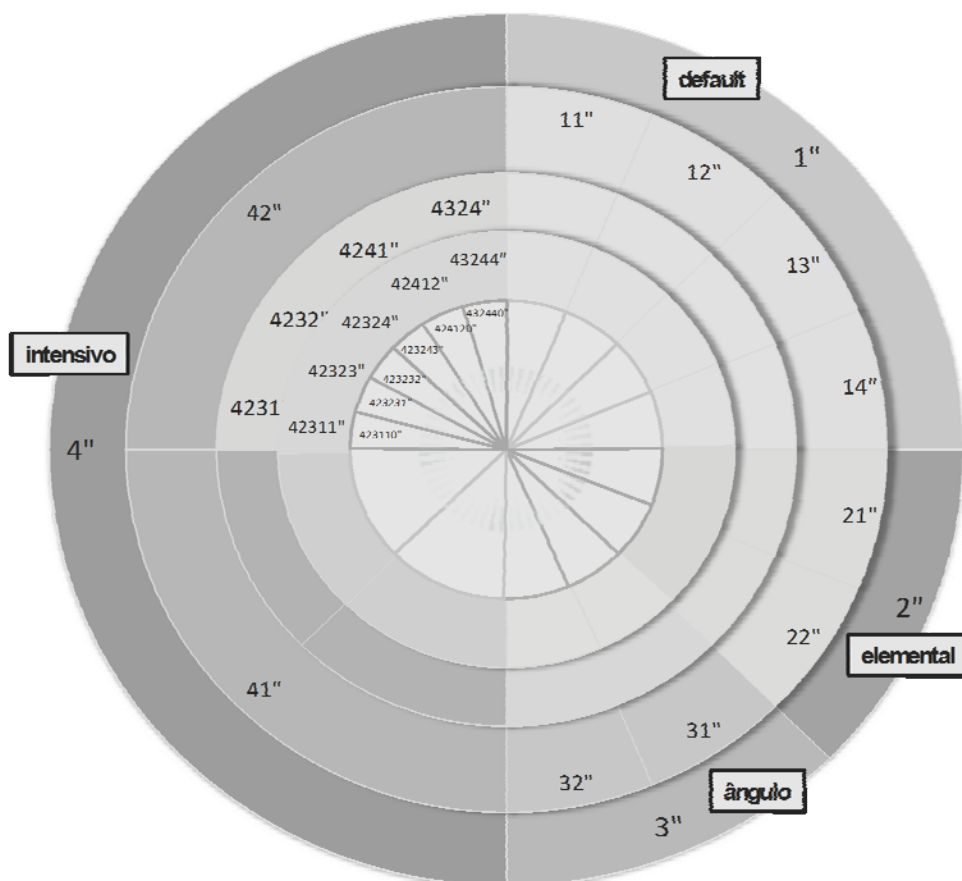
(CALIBRA) e, posteriormente, incorporado ao *software* UAM CorpusTool (O'DONNELL, 2008) para a análise sistêmico-funcional semiautomática⁷.

Após essas duas etapas, foi realizada a anotação das frequências relativas à TRASITIVIDADE, MODO e TEMA, por serem os três sistemas gramaticais da oração relativamente às metafunções ideacional, interpessoal e textual.

Após a análise do *corpus* segundo as categorias da LSF, as orações descritas são colocadas no espaço gramatical segundo o mapa topológico de cada metafunção, estabelecendo-se assim suas distâncias relativas (FIGUEREDO, 2014). Como ilustração, a Figura 2 apresenta o mapa topológico do TEMA em português brasileiro e o Quadro 2 mostra as funções relativas ao mapa.

Figura 2: Mapa topológico do sistema de TEMA em português brasileiro.

⁷ O *software* UAM CorpusTool (O'DONNELL, 2008) permite incorporar os sistemas que serão utilizados na análise do *corpus*. Ao final da análise, *software* apresenta as frequências relativas a cada ocorrência categorizada pelo pesquisador.



Fonte: os autores.

Quadro 2: Funções relativas à topologia do sistema de Tema.

Topologia do sistema de Tema	Funções relativas à topologia (tipologia)
10	Tema: <i>Default</i>
11	Tema: <i>Default</i> : Imperativo
12	Tema: <i>Default</i> : Interrogativo: Polar
13	Tema: <i>Default</i> : Interrogativo: Elemental
14	Tema: <i>Default</i> : Declarativo
20	Tema: Elemental
21	Tema: Elemental: Complemento elemental
22	Tema: Elemental: Adjunto elemental
30	Tema: Ângulo
31	Tema: Ângulo: Fonte
32	Tema: Ângulo: Ponto de vista
40	Tema: Proeminente
41	Tema: Proeminente: Perspectiva
42	Tema: Proeminente: Intensivo
4211	Tema: Proeminente: Intensivo: Relativo: Papel transitivo
4212	Tema: Proeminente: Intensivo: Relativo: Predicação
422	Tema: Proeminente: Intensivo: Absoluto...

Fonte: os autores.

Quadro 3: Perfil sistêmico parcial do Manual de Biossegurança em Odontologia.

ORAÇÃO	REALIZAÇÃO	TRANS (x)	MODO (y)	TEMA (z)
BIO_05_01	Assepsia: é o conjunto de medidas que utilizamos para impedir a penetração de microrganismos num ambiente que logicamente não os tem,	39 ⁸	20	10
BIO_05_02	logo um ambiente asséptico é aquele que está livre de infecção.	39	20	41
BIO_05_03	Antissepsia: é o conjunto de medidas propostas para inibir o crescimento de microrganismos ou removê-los de um determinado ambiente, podendo ou não destruí-los	39	20	10
BIO_05_04	e para tal fim utilizamos antissépticos ou desinfetantes.	14	20	41
BIO_05_05	É a destruição de micro-organismos existentes nas camadas superficiais ou profundas da pele, mediante a aplicação de um agente germicida de baixa causticidade, hipoalergênico e passível de ser aplicado em tecido vivo.	39	20	10
BIO_05_06	Degermação: Significa a diminuição do número de microrganismos patogênicos ou não, após a escovação da pele com água e sabão.	39	20	10
BIO_05_07	É a remoção de detritos e impurezas depositados sobre a pele.	39	20	10
BIO_05_08	Sabões e detergentes sintéticos, graças a sua propriedade de umidificação, penetração, emulsificação e dispersão, removem mecanicamente a maior parte da flora microbiana existente nas camadas superficiais da pele, também chamada flora transitória,	14	20	10
BIO_05_09	mas não conseguem remover aquela que coloniza as camadas mais profundas ou flora residente.	14	20	10
BIO_05_10	Desinfecção: é o processo pelo qual se destroem particularmente os germes patogênicos e/ou se inativa sua toxina ou se inibe o seu desenvolvimento.	39	20	10
BIO_05_11	Os esporos não são necessariamente destruídos.	16	20	10
BIO_05_12	Esterilização: é processo de destruição de todas as formas de vida microbiana (bactérias nas formas vegetativas e esporuladas, fungos e vírus) mediante a aplicação de agentes físicos e ou químicos,	39	20	10
BIO_05_13	Toda esterilização deve ser precedida de lavagem e enxaguadura do artigo para remoção de detritos.	14	20	10

Fonte: os autores.

⁸ Cada código descreve uma função gramatical no texto, a seguir: 39 (Processo: Relacional: Identificativo & Intensivo); 14 (Processo: Material: Transitivo & Transformativo); 16 (Processo: Material: Transitivo & Criativo); 20 (Modo: Declarativo); 10 (Tema: Default) e 41 (Tema: Perspectiva).

O Quadro 3 aponta o número de cada oração composto pelo código da etiqueta do texto. Por exemplo: BIO_05, que significa que este é o quinto texto compilado para o *corpus*, é o manual de biossegurança. Juntamente com essa etiqueta está o número referente a cada oração. Por exemplo: BIO_05_01, significa que esta é a primeira oração do texto BIO_05. Em seguida, a segunda coluna apresenta a realização linguística da oração.

A terceira coluna apresenta as funções do sistema de TRANSITIVIDADE do texto. Por exemplo, para a oração BIO_05_01, “Assepsia: é o conjunto de medidas...”, tem-se um Processo: Relacional Identificativo & Intensivo, que segundo o mapa topológico do sistema de TRANSITIVIDADE em português brasileiro é codificado com o número 39; ‘3’ para Relacional e ‘9’ para Identificativo & Intensivo. Já para a oração BIO_05_04, “e para tal fim utilizamos antissépticos”, tem-se um Processo: Material Transitivo & Transformativo, que, segundo o mapa topológico da TRANSITIVIDADE, recebe o número 14; ‘1’ para Material e ‘4’ para Transitivo & Transformativo.

A quarta coluna apresenta as funções do MODO. Por exemplo, para a oração BIO_05_01, tem-se o Modo Indicativo: Declarativo, classificado como 20; ‘2’ para Indicativo e ‘0’ para Declarativo.

A quinta coluna apresenta as funções do TEMA. Por exemplo, a oração BIO_05_01 possui Tema *Default* e é classificada como 10. No caso da oração BIO_05_04, a função do Tema é Intensivo: Perspectiva e por isso classificada com o número 41; ‘4’ para intensivo e ‘1’ para perspectiva.

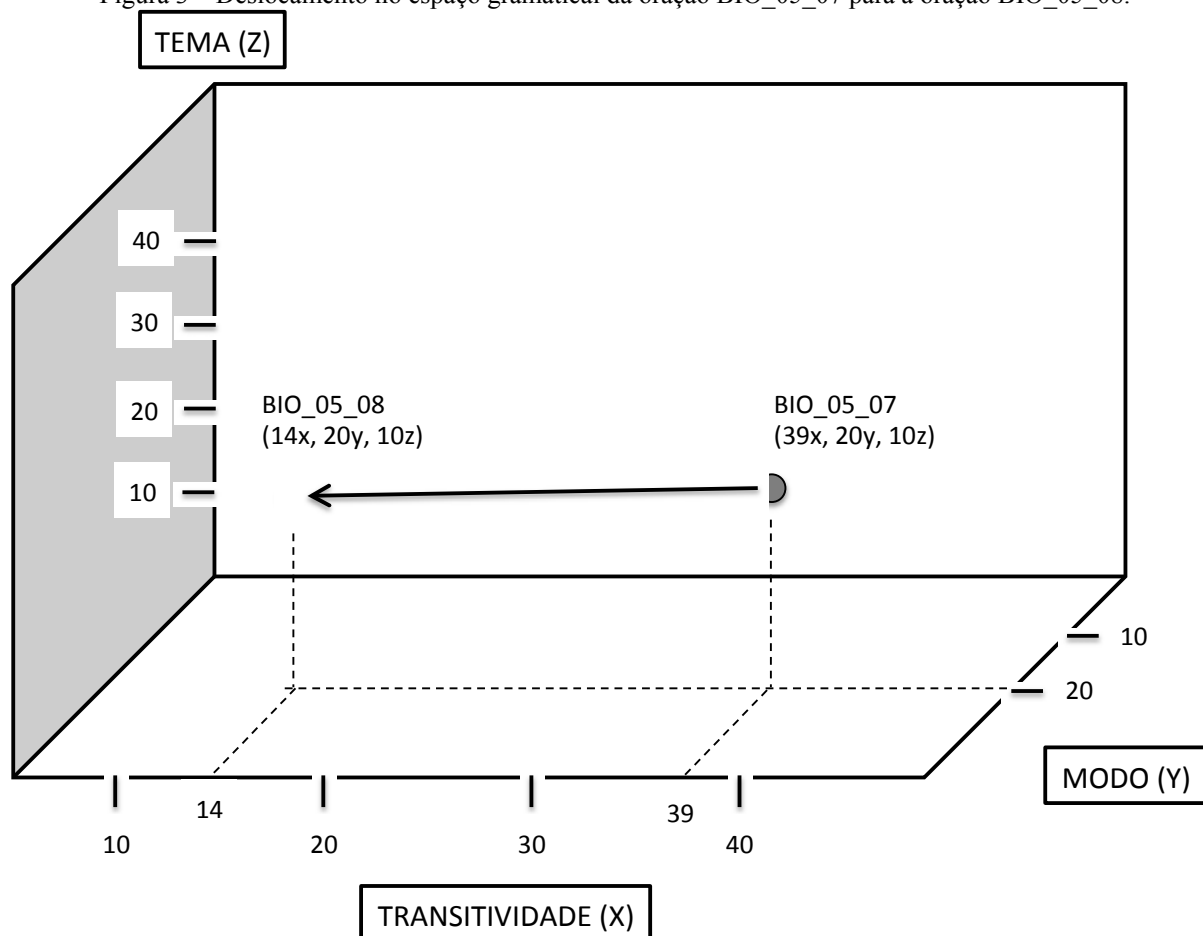
Esse tipo de quadro também foi abordado nos trabalhos de Lima (2013) e Braga (2013). Ele aponta todas as orações do texto com a disposição da TRANSITIVIDADE, do MODO e do TEMA, permitindo assim a visualização da mudança entre orações e, conseqüentemente, as mudanças de significado no manual.

Como complementação, acrescentam-se à perfilação gramatical a identificação e descrição dinâmica da mudança que acontece entre orações e o “movimento” da gramática pelo texto. Cada ponto do espaço representa uma oração composta pelas categorias de TRANSITIVIDADE (eixo x), MODO (eixo y) e TEMA (eixo z). As semirretas que ligam os pontos representam vetores (i.e., magnitude e direção) que apresentam o movimento de uma oração para outra.

Esse movimento pode ser observado, por exemplo, entre as orações BIO_05_07 “É a remoção de detritos e impurezas depositados sobre a pele” e BIO_05_08 “Sabões e

detergentes sintéticos, graças a sua propriedade de umidificação, penetração, emulsificação e dispersão, removem mecanicamente a maior parte da flora microbiana existente nas camadas superficiais da pele, também chamada flora transitória”. O vetor indica uma mudança no espaço gramatical no eixo da TRANSITIVIDADE, de uma oração relacional “[Degermação] é” para uma oração material “sabões... removem”. O Modo para ambas as orações, bem como o Tema se mantêm. Dessa forma, o movimento é zero. Com isso tem-se: BIO_05_07 (39x, 20y, 10z) → BIO_05_08 (14x, 20y, 10z). No espaço, tem-se:

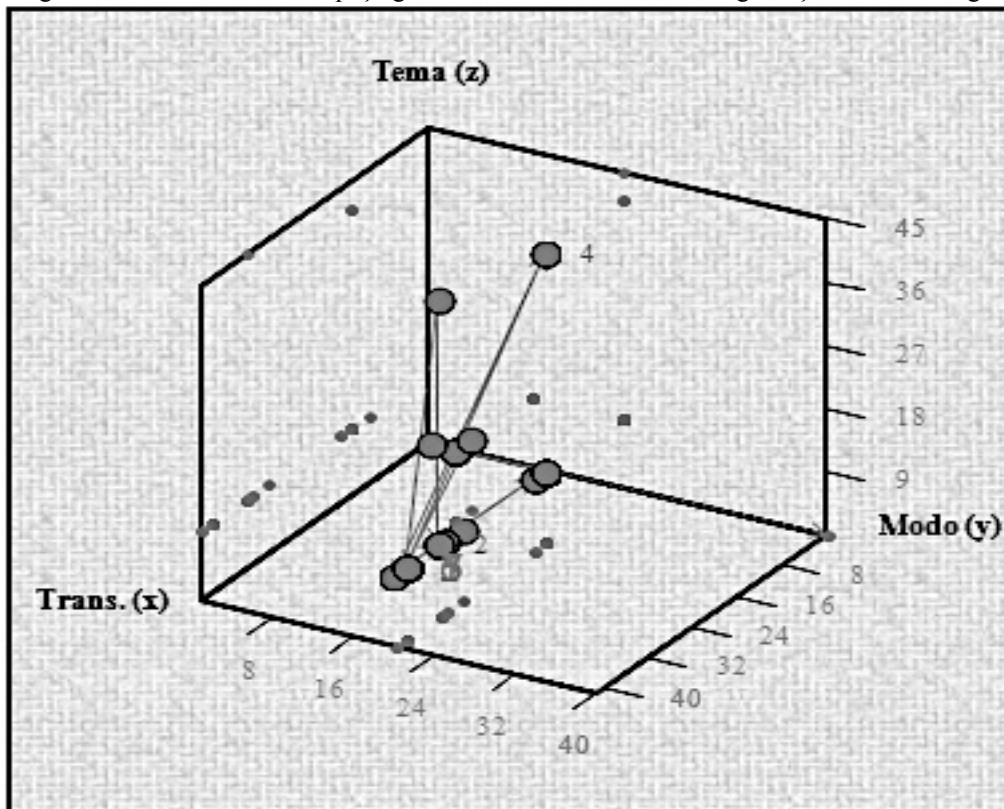
Figura 3 – Deslocamento no espaço gramatical da oração BIO_05_07 para a oração BIO_05_08.



Fonte: os autores.

A Figura 4 apresenta o deslocamento no espaço gramatical do Manual de Biossegurança em Odontologia.

Figura 4: Deslocamento do espaço gramatical do Manual de Biossegurança em Odontologia.



Fonte: os autores.

Esse procedimento foi realizado em todos os textos do *corpus* através de uma macro desenvolvida no Microsoft Office Excel para transferir a perfilação gramatical transcrita em códigos para um gráfico em 3D⁹. Esse tratamento computacional da perfilação permite a visualização e análise das mudanças que a língua faz, do ponto de vista gramatical, para criar o significado do manual (FIGUEREDO, 2014; OLIVEIRA, 2015). Por esse motivo, esse passo metodológico torna-se uma relevante contribuição para a Linguística Computacional e, como consequência, a Linguística de *Corpus*.

4. Resultados

No *corpus* de manuais, após a análise de cada manual no *software* UAM CorpusTool, os resultados de cada categoria foram obtidos em frequências, pela necessidade de se observar a contribuição das diferentes funções e, assim, a probabilidade de escolhas na produção de significado em cada texto. A Tabela 2 exhibe os resultados.

⁹ Obtida em: <http://www.doka.ch/Excel3Dscatterplot.htm>. Acesso em: dezembro de 2016.

Tabela 2: Resultados da análise do *corpus* de manuais em relação à TRANSITIVIDADE, MODO e TEMA retirados do software *UAM CorpusTool*.

TRANSITIVIDADE	
Material	62,17%
Relacional	29,84%
Verbal	5,68%
Existencial	2,31%
Mental	0,00%
TOTAL	100,00%
MODO	
Declarativo	72,47%
Jussivo	27,53%
Sugestivo	0,00%
Polar	0,00%
Elemental	0,00%
TOTAL	100,00%
TEMA	
<i>Default</i>	86,50%
Perspectiva	8,70%
Intensivo	4,26%
Ângulo ponto de vista	0,36%
Ângulo fonte	0,18%
Elemental	0,00%
TOTAL	100,00%

Fonte: os autores.

A Tabela 2 mostra as frequências relativas às delicadezas da TRANSITIVIDADE, do MODO e do TEMA. Na TRANSITIVIDADE, o Processo Material possui 62,17% das ocorrências, o Processo Relacional possui 29,84%, o Verbal 5,68%, o Existencial 2,31% e o Mental 0,00%. No MODO, o Declarativo obteve 72,47%, o Jussivo 27,53% e o Sugestivo, Polar e Elemental 0,00%. No TEMA, o Tema Default obteve 86,50%, o Perspectiva 8,70%, o Intensivo 4,26%, o Ângulo Ponto de Vista 0,36%, o Ângulo Fonte 0,18% e o Elemental 0,00%.

Os dados mostram que o *corpus* de manual é composto principalmente por Processo: Material e Processo: Relacional, Modo: Declarativo e Modo: Jussivo, e Tema: Default, Tema: Perspectiva e Tema: Intensivo. Por conseguinte, esses resultados apresentam o que é o manual: um texto que procura explicar procedimentos, identificar seu objeto de abordagem e apresentar algumas ordens. Halliday e Matthiessen (1999, p. 98) apontam que “nos gêneros

textuais que apresentam procedimentos, as orações materiais são normalmente imperativas e as relacionais indicativas”.¹⁰¹¹ A afirmação dos autores é corroborada pelo que foi encontrado na análise.

Mediante os resultados obtidos através do *corpus*, foi realizada a análise gramatical e semântica das ondas de informação de cada manual. Calcularam-se as frequências relativas a cada categoria supracitada em relação às ondas de informação e, por fim, o modelo, contendo as ocorrências mais frequentes deste gênero textual foi desenvolvido.

O modelo não é, na verdade, nenhum texto específico, criado por usuários do português brasileiro. É, sim, uma generalização das probabilidades maiores de escolhas dos sistemas gramaticais para esse tipo de texto. Assim, a análise que se segue não se refere especificamente a qualquer texto em particular, mas prevê comportamento da gramática oracional do português brasileiro para quaisquer textos caracterizados como manual.

O Quadro 4 a seguir expõe as duas primeiras ondas de informação do modelo. Devido ao fato de o modelo ser uma generalização do *corpus*, ou seja, “formado por partes gramaticais de todos os textos”, ele foi nomeado como Frankenstein:

Quadro 4 — As duas primeiras ondas de informação do Frankenstein.

ONDAS	ORAÇÕES	TRANS. (x)	MODO (y)	TEMA (z)
1	FRANK_01	39 ¹²	20	10
	FRANK_02	33	20	10
	FRANK_03	39	20	10
	FRANK_04	32	20	10
	FRANK_05	39	20	10
	FRANK_06	39	20	10
	FRANK_07	33	20	10
	FRANK_08	14	20	41
	FRANK_09	14	20	10
	FRANK_10	32	20	10
	FRANK_11	32	20	10

¹⁰ ...in procedural registers, material clauses are typically imperative, relational ones indicative.

¹¹ Todas as traduções de citações apresentadas neste artigo, salvo quando for indicado, foram feitas pelo autor.

¹² Cada código descreve uma função gramatical do Frankenstein, a seguir: 39 (Processo: Relacional: Identificativo & Intensivo); 33 (Processo: Relacional: Atributivo & Intensivo); 32 (Processo: Relacional: Atributivo & Intensivo); 14 (Processo: Material: Transitivo & Transformativo); 29 (Processo: Existencial: Permanência); 20 (Modo: Declarativo); 11 (Modo: Imperativo: Jussivo); 10 (Tema: Default); 41 (Tema: Perspectiva); 42 (Tema: Intensivo).

	FRANK_12	14	20	10
	FRANK_13	33	20	10
2	FRANK_14	14	20	42
	FRANK_15	29	20	10
	FRANK_16	14	20	10
	FRANK_17	14	20	10
	FRANK_18	14	20	10
	FRANK_19	14	11	10
	FRANK_20	14	11	10
	FRANK_21	14	11	10
	FRANK_22	14	11	10

Fonte: os autores.

O Quadro 4 apresenta o funcionamento do modelo em relação às duas primeiras ondas de informação. A primeira onda tem como objetivo identificar e atribuir características ao objeto de trabalho do manual. Para isso ela utiliza Processo: Relacional (39 – Identificativo & Intensivo, 33 – Atributivo & Intensivo, 32 – Atributivo & Possessivo) (HALLIDAY; MATTHIESSEN, 1999), o Modo: Declarativo (20) e o Tema: Default (10).

Nessa mesma onda também existem orações com o Processo: Material (14 – Transitivo & Transformativo) que ajudam na explicação do objeto apresentado (HALLIDAY; MATTHIESSEN, 1999) e quando há a troca do tipo de Processo (de uma oração 33 passando para o 14 – 33→14) existe uma mudança no Tema que está presente nessa troca.

A segunda onda é responsável pela explicação do objeto tratado. Para isso, utiliza-se basicamente de Processo: Material (14), do Modo: Declarativo (20), e de alguns comandos para o leitor, utilizando o Modo: Jussivo (11) e Tema: Default (10). No entanto, quando ocorre a troca de onda de informação (1→2) o Tema muda de Default para Intensivo (10/42).

Além do Processo: Material, existe no início da onda um Processo: Existencial (29 - Permanência), que faz a apresentação de um novo participante na explicação, o Existente (HALLIDAY; MATTHIESSEN, 1999). O Quadro 5 mostra as últimas três ondas de informação do Frankenstein:

Quadro 5: As três últimas ondas de informação do Frankenstein.

ONDAS	ORAÇÕES	TRANS. (x)	MODO (y)	TEMA (z)
3	FRANK_23	33 ¹³	20	41
	FRANK_24	14	20	10
	FRANK_25	16	20	10
	FRANK_26	14	20	10
	FRANK_27	14	20	10
	FRANK_28	11	20	10
	FRANK_29	14	20	10
	FRANK_30	14	20	10
	FRANK_31	41	20	41
4	FRANK_32	14	20	10
	FRANK_33	16	11	10
	FRANK_34	14	11	10
	FRANK_35	14	11	10
	FRANK_36	14	11	10
	FRANK_37	14	11	10
	FRANK_38	14	11	10
	FRANK_39	14	11	41
	FRANK_40	14	11	10
	FRANK_41	14	11	10
	FRANK_42	14	11	10
5	FRANK_43	14	20	42
	FRANK_44	32	20	10
	FRANK_45	41	20	10
	FRANK_46	14	20	10

Fonte: os autores.

A terceira onda, como a segunda, é também iniciada com uma mudança de Tema. O que era Default se torna um Tema: Perspectiva (10 → 41) que apresenta no início da oração uma circunstância de intensificação. Na mesma oração, há um Processo: Relacional (31 – Atributivo & Circunstancial) com o Modo: Declarativo (20) que caracteriza o objeto tratado no manual (HALLIDAY; MATTHIESSEN, 1999).

¹³ Cada código descreve uma função gramatical do Frankenstein, a seguir: 32 (Processo: Relacional: Atributivo & Intensivo); 33 (Processo: Relacional: Atributivo & Intensivo); 14 (Processo: Material: Transitivo & Transformativo); 16 (Processo: Material: Transitivo & Criativo); 41 (Processo: Verbal: Semiose & Não-recepção); 20 (Modo: Declarativo); 29 (Processo: Existencial: Permanência); 20 (Modo: Declarativo); 11 (Modo: Imperativo: Jussivo); 10 (Tema: Default); 41 (Tema: Perspectiva); 42 (Tema: Intensivo).

Em seguida, há uma sequência de orações que apresentam o Processo: Material (14 – Transitivo & Transformativo, 16 – Transitivo & Criativo, 11 – Intransitivo & Transformativo), Modo: Declarativo (20) e Tema: Default (10), pois esta onda trata de uma explicação que termina com um Processo: Verbal (41 – Semiose & Não Recepção) e um Tema: Perspectiva (41). Este Processo: Verbal vem para nomear um novo participante, o que ganha suporte na afirmação de Martin e Rose (2007), “Os procedimentos de textos escritos dão ainda um passo além na mediação entre a expertise do autor e o leitor, direcionando cada etapa em relação a objetos e lugares cuidadosamente nomeados”.¹⁴

A quarta onda é responsável pela segunda parte de comandos do manual. Ela começa com explicação (14/20/10) e continua com um fluxo de dez orações de comando realizadas pelo Modo: Jussivo (11) (HALLIDAY; MATTHIESSEN, 1999). Existe uma mudança do tipo de Processo: Material (14 → 16), que acontece devido ao significado do Processo, quando o primeiro transforma um Participante no texto, e o segundo cria ou elimina um outro Participante (HALLIDAY; MATTHIESSEN, 1999).

Além do Processo, o Tema também sofre uma mudança na oração 39, o que era Default (10) passa a ser Tema: Perspectiva (41). Nesse sentido, como ele apresenta uma circunstância de intensificação que é colocada em posição temática como o primeiro elemento da oração. Ela pode ser causal, temporal ou condicional.

E por fim, a última onda trata de finalizar o texto (MARTIN; ROSE, 2007). Para isso ela explica, identifica e nomeia seu objeto do manual usando Processo: Material (14), Processo: Relacional (32) e Processo: Verbal (41) (HALLIDAY; MATTHIESSEN, 1999). Ademais, como ocorreu no início da segunda e terceira ondas, esta também é marcada por um Tema diferente do Default (10). Aqui o Tema: Intensivo (42) aparece na troca de onda de informação. Ele “implica em introdução de, ou contraste com, um novo assunto e subsequente mudança de fase” (FIGUEREDO, 2011, p. 154).

As ondas de informação do modelo, juntamente com a dinâmica existente entre elas e dentro delas, apresentam as probabilidades mais comuns de ocorrência no manual. Sendo assim, este Frankenstein não é o manual e sim um modelo criado para compreender como são realizadas, pelo falante, as escolhas gramaticais (HALLIDAY et al., 1964), a fim de construir o significado do manual. Para exemplificar o modelo, o Quadro 6 mostra a sua primeira onda

¹⁴ “Written procedures go a step further than this to mediate the author’s expertise, directing the learner what to do at each step, in relation to explicitly named objects and locations.” (MARTIN; ROSE, 2007, p. 180).

de informação com exemplos do *corpus*. Os exemplos seguem a descrição de cada oração exibida no Quadro 4:

Quadro 6 — O Frankenstein com exemplos do *corpus*.

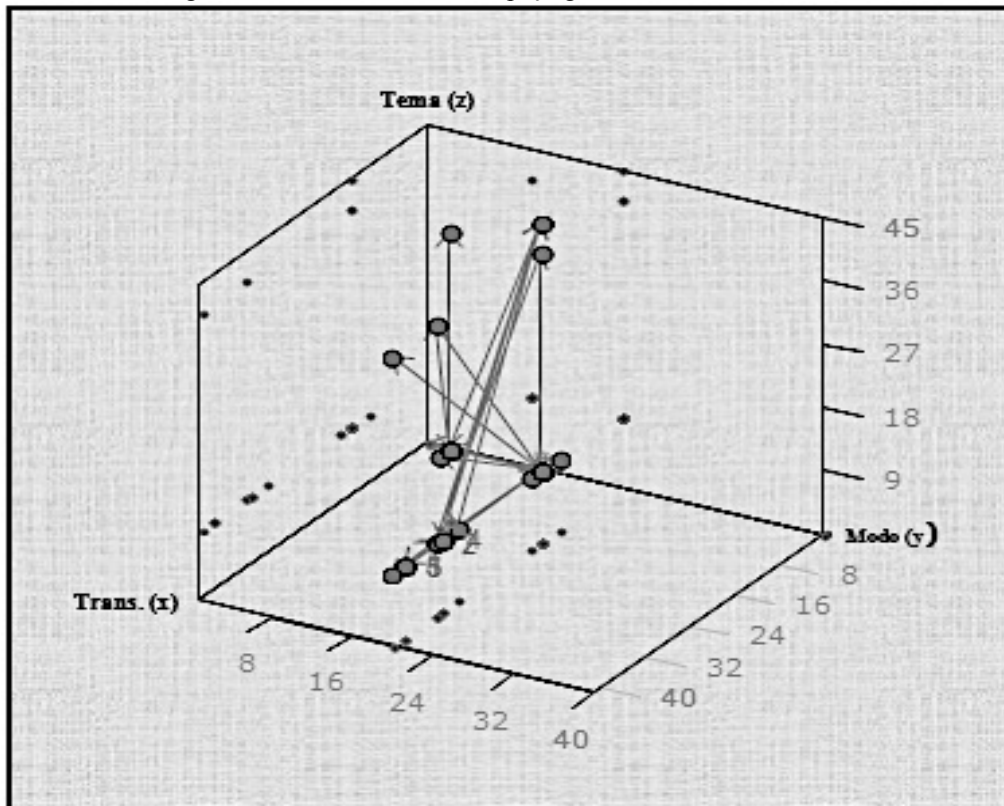
ORAÇÃO	EXEMPLOS
FRANK_01	Assepsia: é o conjunto de medidas que utilizamos para impedir a penetração de microrganismos num ambiente que logicamente não os tem
FRANK_02	Essas complicações na maioria das vezes são inevitáveis.
FRANK_03	Os utensílios utilizados (baldes, bacias, sacos plásticos, entre outros) devem ser limpos sempre que seja observado resíduo (Figura 18).
FRANK_04	Além do itinerário que lhe compete, com os quarteirões numerados, o pessoal da operação deve dispor das tarefas de cada dia e de croquis com o desenho e a disposição de todos os quarteirões (quadras) de cada área.
FRANK_05	Trata-se de uma exigência constitucional (Constituição, art.39,III e 83),
FRANK_06	Esta situação nada mais é do que uma obrigação contratual.
FRANK_07	neste caso a participação dos envolvidos no processo é fundamental.
FRANK_08	Quando necessário, deve-se discriminar a massa (peso) individual das gemas que compõem a joia.
FRANK_09	A marcação se fará com lápis-cera azul ou preto no cateto esquerdo de cada ângulo do quarteirão.
FRANK_10	Ela deve conter o plano de governo, exposição da situação do País e solicitação de providências que julgar necessárias (Constituição, art. 84, XI)
FRANK_11	Os números e os sinais devem ter cinco centímetros de altura.
FRANK_12	Os quarteirões receberão numeração crescente, do número um ao infinito.
FRANK_13	A lavagem de mãos é obrigatória para todos os componentes da equipe de saúde bucal;

Fonte: os autores.

O Quadro 6 mostra o Frankenstein que foi criado com partes distintas dos manuais do *corpus*. Apesar de ele não apresentar os mesmos participantes nas orações e de não haver conectividade entre um termo e outro do ponto de vista lexical, este modelo está gramaticalmente correto no sentido de que ele representa ideacionalmente, encena interpessoalmente e constrói textualmente a primeira onda de informação do *corpus* que trata principalmente da identificação do objeto abordado pelo manual. Ademais, o modelo faz isso a partir das probabilidades maiores de escolha de funções dos sistemas envolvidos para cada oração.

O significado desse tipo de texto, o manual, além de ser compreendido pelas suas funções gramaticais, também é compreendido pela disposição das mudanças que ocorrem de uma oração para outra. A Figura 5 exhibe o deslocamento no espaço gramatical do Frankenstein desta pesquisa:

Figura 5 — Deslocamento no espaço gramatical do Frankenstein.



Fonte: os autores.

Como na Figura 4, a Figura 5 descreve o movimento gramatical realizado pelo Frankenstein. Primeiramente ocorrem as escolhas gramaticais em ‘x’, ‘y’ e ‘z’. Através dessas escolhas são formadas as orações (representadas pelos pontos). Na sequência, observa-se o deslocamento (representado pelas setas) de uma oração para outra. Este conjunto de escolhas e movimentos forma o modelo de descrição dinâmica, ou a dinâmica textual (segundo a dinâmica discursiva apresentada em Lemke no artigo de 1993).

No deslocamento, as ligações entre os pontos podem ser convertidas em vetores, por descreverem movimento. Com isso, apresentam as mudanças gramaticais em relação ao tempo (FIGUEREDO, 2014). Sendo assim, os vetores mostram a quantidade de mudança gramatical de uma oração para outra. O Quadro 7 exhibe as equações vetoriais da primeira onda de informação do Frankenstein:

Quadro 7 — Equações vetoriais do desenvolvimento da primeira onda de informação do Frankenstein.

ORAÇÃO	TRANS./MODO/TEMA	EQUAÇÕES VETORIAIS	QUANTIDADE DE MUDANÇA
FRANK_01	39/20/10	$r = \langle 39, 20, 10 \rangle + t \langle -6, 0, 0 \rangle$	$r = \langle 39 - 6t, 20, 10 \rangle$
FRANK_02	33/20/10	$r = \langle 33, 20, 10 \rangle + t \langle 6, 0, 0 \rangle$	$r = \langle 33 + 6t, 20, 10 \rangle$
FRANK_03	39/20/10	$r = \langle 39, 20, 10 \rangle + t \langle -7, 0, 0 \rangle$	$r = \langle 39 + 7t, 20, 10 \rangle$
FRANK_04	32/20/10	$r = \langle 32, 20, 10 \rangle + t \langle 7, 0, 0 \rangle$	$r = \langle 32 + 7t, 20, 10 \rangle$
FRANK_05	39/20/10	$r = \langle 39, 20, 10 \rangle + t \langle 0, 0, 0 \rangle$	$r = \langle 39, 20, 10 \rangle$
FRANK_06	39/20/10	$r = \langle 39, 20, 10 \rangle + t \langle -6, 0, 0 \rangle$	$r = \langle 39 - 6t, 20, 19 \rangle$
FRANK_07	33/20/10	$r = \langle 33, 20, 10 \rangle + t \langle -19, 0, 31 \rangle$	$r = \langle 33 - 19t, 20, 10 + 31t \rangle$
FRANK_08	14/20/41	$r = \langle 14, 20, 41 \rangle + t \langle 0, 0, -31 \rangle$	$r = \langle 14, 20, 41 - 31t \rangle$
FRANK_09	14/20/10	$r = \langle 14, 20, 10 \rangle + t \langle 18, 0, 0 \rangle$	$r = \langle 14 + 18t, 20, 10 \rangle$
FRANK_10	32/20/10	$r = \langle 32, 20, 10 \rangle + t \langle 0, 0, 0 \rangle$	$r = \langle 32, 20, 10 \rangle$
FRANK_11	32/20/10	$r = \langle 32, 20, 10 \rangle + t \langle -18, 0, 0 \rangle$	$r = \langle 32 - 18t, 20, 10 \rangle$
FRANK_12	14/20/10	$r = \langle 14, 20, 10 \rangle + t \langle 19, 0, 0 \rangle$	$r = \langle 14 + 19t, 20, 10 \rangle$
FRANK_13	33/20/10		

Fonte: os autores.

O mapeamento das funções gramaticais e o estabelecimento do descolamento gramatical apresentado na metodologia por meio da perfilação gramatical constituem, assim, o modelo topológico do manual.

Por fim, cabe ressaltar que as frequências relativas apresentadas para o modelo do manual podem coincidir – e, com efeito, coincidem – com frequências relativas de outros gêneros textuais. A alta incidência de Processos Materiais encontrada no manual é compartilhada, por exemplo, com gêneros que envolvem a representação de mudança material no mundo, tais como a descrição de processos físicos, ou de atividades de trabalho, entre outras. Já a frequência do Modo encontrada no manual pode ser comparada à frequência de outros gêneros tais como os manuais de guia e de como fazer, receitas, procedimentos, entre outros. Esta constatação revela que o manual é, por um lado, um gênero separado de outros;

porém, por outro lado, existe uma contiguidade entre este e outros gêneros, o que se verifica pela frequência comparável entre as funções gramaticais. Assim, a construção de modelos de gêneros e sua relação com outros modelos de gêneros semelhantes abre espaço para a constituição de famílias de gêneros e de processos sociossemióticos: explicar, relatar, recriar, compartilhar, fazer, recomendar, capacitar e explorar (URE, 1989; HALLIDAY; MATTHIESSEN, 2014).

5. Considerações finais

Tendo como foco principal apresentar uma forma de contribuição de aplicações da Linguística de *Corpus* e LSF à análise e descrição da produção de textos, este artigo apresentou a modelagem do gênero textual ‘manual’ de forma que esta possa subsidiar trabalhos futuros em Linguística Aplicada, mais especificamente, trabalhos envolvendo o Processamento de Linguagem Natural.

Através da metodologia de perfilação gramatical, a qual possibilita compilar e analisar corpora gramaticais (FIGUEREDO, 2014), foi possível construir um modelo de produção textual pautado pelas escolhas em sistemas gramaticais. A justificativa para uma metodologia nova para os estudos de produção textual a partir da gramática se deve ao fato de que, em geral, as descrições dos gêneros textuais carecem de descrições relativas à dinâmica textual. Assim, este artigo oferece uma contribuição por conseguir localizar as escolhas nos sistemas e explicar, por meio das probabilidades mais frequentes, o comportamento sistêmico, neste caso específico, do manual.

O modelo, que não representa qualquer texto especificamente, configura-se como uma das formas de generalização das probabilidades maiores de escolhas dos sistemas gramaticais para esse gênero textual. Assim, a análise que se segue tem o potencial para ser aplicada, de uma forma geral, a textos caracterizados como manual por ser capaz de prever, em grande medida, qual é o comportamento futuro do texto. Assim, em qualquer momento do texto, existe o potencial para se saber aonde a gramática “se movimenta”.

Como ilustração, é possível saber que a partir da oração-média 32, ou quarta onda de informação, a explicação continua com um fluxo de dez orações de comandos, com mudança de Modo Declarativo para Imperativo: Jussivo, com uma quantidade de mudança gramatical alta do ponto de vista interpessoal, mas baixa do ponto de vista ideacional e textual, ou $14/20/10 \rightarrow 16/11/10$, ou $r^{\rightarrow} = \langle 2+34t, 9, 0 \rangle$.

Desta maneira, o artigo também contribui, em outra frente, para os estudos em LSF, pois mostra o potencial robusto dessa teoria para lidar com novas formas de representação da produção de textos. Assim, focando-se na função com número maior de ocorrências, o artigo mostrou um conjunto de possibilidades de análise advindas da perfilação gramatical, tais como a frequência da função, a sua distribuição dentro dos textos, a dispersão ao longo dos textos, bem como os agrupamentos que forma, nos padrões oracionais, com outras funções. Para as pesquisas em Linguística Aplicada – de base sistêmico-funcional – a perfilação gramatical mostrou como é possível compilar e extrair dados de corpora, inclusive para a forma como são empregados na dinâmica textual.

Referências bibliográficas

BERBER SARDINHA, T. Linguística de corpus: histórico e problemática. **D.E.L.T.A.**, v. 16, n. 2, p. 323-367, 2000.

BIBER, D. What can a corpus tell us about registers and genres? In: O'KEEFFE, A.; McCARTHY, M. (Ed.). **The Routledge Handbook of corpus Linguistics**. London, New York: Routledge, 2010. p. 241-254. <https://doi.org/10.4324/9780203856949.ch18>

BRAGA, A. **A popularização da popularização da ciência: gramaticalização da interação, avaliação, representação da experiência, construção e organização discursiva**. 2013. 97 f. Monografia (Bacharelado em Tradução) – Departamento de Letras, Universidade Federal de Ouro Preto, Mariana, 2013.

BUNGE, M. **La investigación científica: Su estrategia y su filosofía**. Barcelona: Ediciones Ariel, 1969.

FIGUEREDO, G. **Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues**. 2011. 385 f. Tese (Doutorado em Linguística Aplicada) – Programa de Pós-Graduação em Estudos Linguísticos, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

FIGUEREDO, G. Uma metodologia de perfilação gramatical sistêmica baseada em corpus. **Letras & Letras**, v. 30, n. 2, jul./dec., 2014. p. 17-45. <https://doi.org/10.4324/9780203856949.ch18>

GLEASON, H. **Linguistics and English grammar**. New York: Holt, Rinehart & Winston, 1965.

HALLIDAY, M. A. K. **Language as social semiotic: the social interpretation of language and meaning**. London & Baltimore: Edward Arnold & University Park Press, 1978.

_____. Towards Probabilistic Interpretations. In: VENTOLA, E. (Ed.). **Functional and Systemic Linguistics: Approaches and Uses**. Berlin/ New York: Mouton de Gruyter, 1991. p.39-61. <https://doi.org/10.1515/9783110883527.39>

_____. **On grammar**. London: Continuum, 2002.

_____. **Computational and quantitative studies**. London: Continuum, 2005.

_____. Working with meaning: towards an applicable linguistics. In: WEBSTER, J. (Ed.). **Meaning in Context: implementing intelligent applications of language studies**. London and New York: Continuum, 2008.

_____. Pinpointing the choice: meaning and the search for equivalents in a translated text. In: MAHBOOB, A.; KNIGHT, N. (Ed.). **Applicable Linguistics**. London e New York: Continuum, 2010.

HALLIDAY, M. A. K.; MATTHIESSEN, C. **Construing experience as meaning: a language based approach to cognition**. London: Cassell, 1999.

_____. **Halliday's introduction to functional grammar**. 4ed. Oxford e New York: Routledge, 2014. A

HALLIDAY, M. A. K.; McINTOSH, A.; STREVENSON, P. **The linguist sciences and language teaching**. London: Longmans, 1964.

LEMKE, J. Text Production and Dynamic Text Semantics. In: VENTOLA, E. (Ed.). **Functional and systemic linguistics: approaches and uses**. Berlin: Mouton de Gruyter, 1991. <https://doi.org/10.1515/9783110883527.23>

_____. **Discourse, Dynamics, and Social Change**. Cultural Dynamics. v.6, n. 1, p. 243-275, 1993. <https://doi.org/10.1177/092137409300600107>

LIMA, K.C.S.; **Caracterização de registros orientada para a produção textual no ambiente multilíngue: um estudo baseado em corpora comparáveis**. Tese (Doutorado em Linguística Aplicada) – Belo Horizonte: Faculdade de Letras, UFMG/PosLin, 2013.

MAHBOOB e KNIGHT, N. Applicable linguistics: an introduction. In: _____. (Ed.). **Applicable Linguistics**. London e New York: Continuum, 2010.

MARTIN, J. R. **English text: system and structure**. Philadelphia and Amsterdam: John Benjamins Publishing Company, 1992. <https://doi.org/10.1075/z.59>

_____. **Systemic functional grammar: a next step into the theory – axial relations**. Beijing: Higher Education Press, 2013.

MARTIN, J.; MATTHIESSEN, C. Systemic typology and topology. In: CHRISTIE, F. (Ed.). **Literacy in social processes**. Darwin: Centre for Studies in Language in Education, Northern Territory University, 1991.

MARTIN, J.; ROSE, D. **Working with discourse: meaning beyond the clause**. 2. ed. London: Continuum, 2007.

MATTHIESSEN, C. **Register in the round: registerial cartography**. *Funct Linguist* 2015.

MATTHIESSEN, C.; BATEMAN, J. **Text generation and systemic functional linguistics: experiences from English and Japanese**. London: Pinter Publishers, 1991.

MATTHIESSEN, C.; TERUYA, K.; WU, C. Multilingual studies as a multi-dimensional space of interconnected language studies. In: WEBSTER, J. (Ed.). **Meaning in Context: implementing intelligent applications of language studies**. London and New York: Continuum, 2008.

MATTHIESSEN, C.; TERUYA, K.; LAM, M. **Key terms in systemic functional linguistics**. London and New York: Continuum, 2010.

O'DONNELL, M. The UAM CorpusTool: software for corpus annotation and exploration. In: BRETONES CALLEJAS, C. *et al.* (Ed.). In: **Applied linguistics now: understanding language and mind**. Almería: Universidad de Almería, 2008.

OLIVEIRA, F. S. Comparação linguística e perfilação gramatical sistêmica em um corpus combinado. **Revista Estudos da Linguagem**, v. 23, Edição Especial, p. 727-768, 2015.

PAGANO, A. S. et al. Boas Práticas para a Pesquisa nas Ciências da Saúde. In: TORRES, H. D. C.; REIS, I. A.; PAGANO, A. S. **Empoderamento do Pesquisador nas Ciências da Saúde**. Belo Horizonte: FALE/UFGM, 2015. Cap. 9, p. 139-142.

ROSE, D. **The Western Desert Code: an Australian cryptogrammar**. Canberra, The Australian National University: Pacific Linguistics, 2001.

SAUSSURE, F. **Curso de Linguística Geral**. São Paulo: Cultrix, 2006.

TEICH, E. Contrastive linguistics and translation studies revisited. In: GIL, A. (Org.). **Modelle der Translation: Grundlagen für Methodik, Bewertung, Computer Modellierung**. Frankfurt am Main e Berlin: Lang, 1999.

URE, J. **Text types classified by situational factors**. Manuscrito, 1989.

VIANA, V. Linguística de corpus: conceitos, técnicas e análises. In: VIANA, V.; TAGNIN, S. (Org.). **Corpora no ensino de línguas estrangeiras**. São Paulo: HUB Editorial, 2011, p. 25-95.

WINOGRAD, T. **Understanding natural language**. New York: Academic Press, 1972.

WINOGRAD, T. **Shifting viewpoints: Artificial intelligence and human-computer interaction**. USA: Elsevier B. V., 2006, p. 1256-1258.

Bibliografia

CÂNDIDO-JR, A. **Análise bidirecional da língua na simplificação sintática em textos do português voltada à acessibilidade digital**. 2013. 225f. Tese (Doutorado em Ciências de Computação e Matemática Computacional). Instituto de Ciências Matemáticas e da Computação, Universidade de São Paulo, São Carlos, 2013.

Artigo recebido em: 15.12.2016

Artigo aprovado em: 15.04.2017