

GRUPO DE ESTUDOS LINGUÍSTICOS
DO ESTADO DE SÃO PAULO



ESTUDOS LINGUÍSTICOS (SÃO PAULO. 1978)
v. 54, n. 1

ESTUDOS LINGÜÍSTICOS (SÃO PAULO. 1978)

GRUPO DE ESTUDOS LINGÜÍSTICOS DO ESTADO DE SÃO PAULO (GEL)

Universidade Estadual de Campinas (Unicamp)

Rua Sérgio Buarque de Holanda, 571, sala D.3.15

CEP 13083-859 – Campinas – SP – Brasil

<http://www.gel.org.br/estudoslinguisticos/>

estudoslinguisticos@gel.org.br

Diretoria do GEL (Gestão Unicamp - 2023-2025)

(Presidente) Livia Oushiro

Universidade Estadual de Campinas, Campinas, São Paulo, Brasil

(Vice-Presidente) Dayane Celestino de Almeida

Universidade Estadual de Campinas, Campinas, São Paulo, Brasil

(Secretária) Erica Luciene Alves de Lima

Universidade Estadual de Campinas, Campinas, São Paulo, Brasil

(Tesoureira) Thiago Oliveira da Motta Sampaio

Universidade Estadual de Campinas, Campinas, São Paulo, Brasil

Editor responsável

Profa. Dra. Dayane Celestino de Almeida

Universidade Estadual de Campinas, Campinas, São Paulo, Brasil

Comissão editorial

Profa. Dra. Claudia Zavaglia, Universidade Estadual Paulista

“Júlio de Mesquita Filho” (UNESP), São José do Rio Preto, São Paulo, Brasil

Prof. Dr. Carlos Eduardo Mendes de Moraes, Universidade Estadual Paulista

“Júlio de Mesquita Filho” (UNESP), Assis, São Paulo, Brasil

Prof. Dr. Marcelo Módolo, Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil

Prof. Dr. Oto Araújo Vale, Universidade Federal de São Carlos (UFSCar),

São Carlos, São Paulo, Brasil

Profa. Dra. Luciani Ester Tenani, Universidade Estadual Paulista

“Júlio de Mesquita Filho” (UNESP), São José do Rio Preto, São Paulo, Brasil

Profa. Dra. Maria Irma Hadler Coudry, Universidade Estadual de Campinas (UNICAMP),

Campinas, São Paulo, Brasil

Profa. Dra. Angela Cecília de Souza Rodrigues, Universidade de São Paulo (USP),

São Paulo, São Paulo, Brasil

Profa. Dra. Beth Brait, Pontifícia Universidade Católica de São Paulo (PUC-SP),

São Paulo, São Paulo, Brasil

Conselho editorial

- Prof. Dr. Bertrand Daunay, Universidade de Lille, Lille, França
- Prof. Dr. Eric Laporte, Université Paris-Est Marne-la-Vallée, Champs-sur-Marne, França
- Prof. Dr. Frantome Bezerra Pacheco, Universidade Federal do Amazonas (UFAM),
Manaus, Amazonas, Brasil (*in memoriam*)
- Profa. Dra. Inmaculada Penadés Martínez, Universidad de Alcalá (UAH), Madrid, Espanha
- Profa. Dra. Julia Sevilla Muñoz, Universidad Complutense de Madrid, Madrid, Espanha
- Profa. Dra. Lou-Ann Kleppa, Universidade Federal de Rondônia (UNIR),
Porto Velho, Rondônia, Brasil
- Profa. Dra. Luisa A. Messina Fajardo, Università di Roma Tre, Roma, Itália
- Prof. Dr. Marcos Lopes, Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil
- Profa. Dra. Maria Luisa Ortiz Alvarez, Universidade de Brasília (UnB),
Brasília, Distrito Federal, Brasil
- Profa. Dra. Renira Rampazzo Gambarato, Jönköping University, Jönköping, Sweden
- Prof. Dr. Roberto Francavilla, Università degli Studi di Genova, Genova, Itália
- Prof. Dr. Ronaldo Lima Junior, Universidade Federal do Ceará (UFC), Fortaleza, Ceará, Brasil
- Profa. Dra. Sabela Fernández-Silva, Universidad Católica de Valparaíso (UCV),
Valparaíso, Chile
- Prof. Dr. Salvio Martín Menéndez, Universidad de Buenos Aires (UBA),
Buenos Aires, Argentina
- Prof. Dr. Sirio Possenti, Universidade Estadual de Campinas (UNICAMP),
Campinas, São Paulo, Brasil
- Profa. Dra. Tânia Romero, Universidade Federal de Lavras (UFLa),
Lavras, Minas Gerais, Brasil
- Prof. Dr. Tony Berber Sardinha, Pontifícia Universidade Católica de São Paulo (PUC-SP),
São Paulo, São Paulo, Brasil

Auxiliar editorial

Milton Bortoleto, Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil

Revisão, normatização, projeto gráfico e diagramação

Letraria | www.letraria.net

Catálogo na Publicação elaborada por

Gildenir Carolino Santos (CRB-8ª/5447)

Estudos Linguísticos (São Paulo. 1978). – v.1, n.1 (1978-). – São José do Rio Preto, SP: Grupo de Estudos Linguísticos do Estado de São Paulo, 2002-1 recurso digital : il.

Periodicidade quadrimestral desde volume 36, 2007 (atual).

Periodicidade anual até volume 35, 2006.

Periodicidade semestral até volume 29, 2000.

ISSN 1413-0939 (impresso).

Publicada no formato impresso até volume 29, 2000.

Publicada no formato em CD-ROM dos volumes 30 ao 35 (2001 a 2006).

Disponível online a partir do volume 36, 2007.

Título abreviado: Est. Ling.

Preservada digitalmente no LOCKSS.

Disponível em: <https://revistas.gel.org.br/estudos-linguisticos>

1. Estudos linguísticos – Periódicos. 2. Linguística – Periódicos. I. Grupo de Estudos Linguísticos do Estado de São Paulo.

20-017

CDD: 410.05

CDU: 81 (05)

SUMÁRIO

<i>Apresentação</i>	7
Dayane Celestino de Almeida	
<i>As variáveis linguísticas na realização do objeto direto anafórico de terceira pessoa em legendas audiovisuais</i>	11
Lívia Oliveira Azevedo	
<i>Transcrição automática de entrevistas e anotação Universal Dependencies no Corpus Roda Viva</i>	29
Cláudia Dias de Barros Oto Araújo Vale Gabriela Wick-Pedro	
<i>Rejeição ou inovação: uma discussão a partir de características de manuscritos rejeitados</i>	46
Janayna Carvalho	
<i>Investigar a interação no autismo a partir de uma perspectiva linguístico-interacional</i>	58
Fernanda Miranda da Cruz	
<i>A função social da divulgação científica (e como a Academia pode ajudar ou atrapalhar)</i>	79
Cecília Farias	
<i>De Rowling a Galbraith: as sobrevidas do autor na contemporaneidade</i>	94
Beatriz Masson Francisco	
<i>Negação de atos de fala no português brasileiro por meio de “não”: uma visão da Gramática Discursivo-Funcional</i>	111
Gabriel Henrique Galvão Passetti Erotilde Goreti Pezatti	

<i>Uma nota sobre a hipótese da similaridade construcional</i>	134
Taísa Peres de Oliveira	
<i>Multiletramentos, tecnodiversidade e ensino de Português</i>	152
Ana Elisa Ribeiro	
<i>Morphology in multilingual data resources: A brief survey</i>	169
Magda Ševčíková	
<i>Efeito de impostura: mentira e manipulação no pronunciamento televisionado de Jair Bolsonaro</i>	185
Joarle Magalhães Soares	
<i>Orações subordinadas adjetivas em diferentes gêneros textuais: uma proposta de descrição e análise</i>	205
Ana Carolina Sperança Criscuolo	
Angélica Rodrigues	
<i>Aspectos acústico-prosódicos do alongamento silábico em contextos de intensificação na fala de professores do ensino básico</i>	223
Fernando Luiz Pompeu Varela	
<i>A não uniformidade das bases numerais: uma visão dos contrastes das bases cardinais do português brasileiro</i>	240
Fernando Valls Yoshida	
<i>Corpus-based language comparison: From morphology to dependencies and beyond</i>	259
Daniel Zeman	

Apresentação

É com enorme alegria que apresento o primeiro número do volume 54, da revista *Estudos Linguísticos*, composto por quinze artigos provenientes de trabalhos apresentados no 70º Seminário do Grupo de Estudos Linguísticos de São Paulo (GEL), realizado em julho de 2024. Os trabalhos aqui reunidos refletem a diversidade de áreas e perspectivas que atualmente caracterizam o campo dos estudos da linguagem, abrangendo desde estudos de variação linguística e análise do discurso até questões de ensino, tecnologia, divulgação científica e estudos literários. Este conjunto expressivo de contribuições demonstra a vitalidade da pesquisa em linguística, linguística aplicada, literatura e áreas afins no Brasil e no contexto internacional.

O número inicia-se com o artigo de Azevedo, intitulado “As variáveis linguísticas na realização do objeto direto anafórico de terceira pessoa em legendas audiovisuais”. O trabalho investiga como o objeto direto anafórico de terceira pessoa manifesta-se em legendas da série *Grey’s Anatomy*, comparando legendas profissionais e legendas produzidas por fãs. Integrando a teoria da variação e mudança linguísticas e estudos sobre gêneros textuais-discursivos, Azevedo oferece uma análise estatística detalhada que revela comportamentos específicos do fenômeno dentro do *corpus* audiovisual.

Em seguida, Carvalho nos conduz à reflexão sobre a publicação acadêmica com “Rejeição ou inovação: uma discussão a partir de características de manuscritos rejeitados”. Partindo de sua experiência como editora-chefe da *Revista de Estudos da Linguagem – RELIN*, a autora problematiza aspectos relacionados à rejeição de artigos, retomando discussões apresentadas na mesa-redonda “Por que publicar” do seminário de 2024 do GEL, oferecendo contribuições importantes sobre o processo editorial e os critérios de avaliação científica.

O artigo de Cruz, “Investigar a interação no autismo a partir de uma perspectiva linguístico-interacional”, propõe uma abordagem que enfatiza a abordagem linguístico-interacional para o exame da interação e da linguagem-em-interação no autismo. Foram analisadas interações entre crianças autistas e adultos não-autistas, explorando como as produções verbais não convencionais são negociadas interacionalmente, contribuindo para a compreensão da manutenção da intersubjetividade em contextos de limitação comunicativa.

Dias de Barros *et al.*, com o estudo “Transcrição automática de entrevistas e anotação Universal Dependencies no Corpus Roda Viva”, apresenta um trabalho piloto sobre a transcrição automática de entrevistas do programa Roda Viva. O artigo compara os

dados originais com os resultados obtidos via ferramentas de transcrição e anotação, evidenciando diferenças sintáticas.

No campo da divulgação científica, Farias traz o texto “A função social da divulgação científica (e como a Academia pode ajudar ou atrapalhar)”, discutindo os desafios e as possibilidades de tornar a ciência acessível e relevante para diferentes públicos. Baseando-se em debates da mesa-redonda “Divulgação e Popularização da Linguística” (apresentada no 70º seminário do GEL, 2024), o artigo analisa o papel da academia na promoção da ciência, enfatizando a necessidade de reconhecimento institucional e de estratégias de comunicação eficazes.

Em “De Rowling a Galbraith: as sobrevidas do autor na contemporaneidade”, Francisco oferece uma análise da construção de *personas* autorais a partir da obra de J.K. Rowling e seu pseudônimo Robert Galbraith. São discutidas estratégias performáticas na construção da autoria e os efeitos estéticos dessas escolhas, trazendo à tona reflexões sobre autoria na contemporaneidade literária.

No estudo de Galvão Passetti *et al.*, “Negação de atos de fala no português brasileiro por meio de ‘não’: uma visão da Gramática Discursivo-Funcional”, o foco recai sobre a negação de atos de fala, abordada a partir da Gramática Discursivo-Funcional. A análise detalha como a ação lexical “não” atua na rejeição de atos discursivos, contribuindo para a compreensão das propriedades funcionais e formais desse fenômeno.

Oliveira, em “Uma nota sobre a hipótese da similaridade construcional”, problematiza o Princípio da não-sinonímia dentro da Gramática de Construções, analisando construções que apresentam funções semelhantes apesar de formas distintas. O artigo evidencia como a alternância construcional e a percepção dos falantes sobre similaridades entre construções desafiam interpretações puramente verticais da gramática.

Ribeiro, com “Multiletramentos, tecnodiversidade e ensino de Português”, propõe uma reflexão sobre a integração de tecnologias no ensino da língua materna. Tomando como base a noção de tecnodiversidade de Yuk Hui e documentos fundamentais sobre multiletramentos e diretrizes curriculares, o artigo discute possibilidades para um ensino mais inclusivo e crítico, atento às implicações tecnológicas na educação linguística.

Ševčíková apresenta “Morphology in multilingual data resources: a brief survey”, um panorama de recursos linguísticos multilíngues que contemplam aspectos morfológicos de diferentes línguas. O artigo destaca a importância de *corpora* e bases lexicais para a compreensão de processos de flexão, derivação e estrutura interna das palavras, sugerindo usos relevantes para linguística comparativa e tipologia.

Soares, com “Efeito de impostura: mentira e manipulação no pronunciamento televisado de Jair Bolsonaro”, analisa como o ex-presidente utilizou estratégias de manipulação em pronunciamento oficial, utilizando a Análise do Discurso para demonstrar a construção de uma cena política marcada pela projeção do sujeito como agente da mentira.

O artigo de Sperança-Crisuolo *et al.*, “Orações subordinadas adjetivas em diferentes gêneros textuais: uma proposta de descrição e análise”, explora a correlação entre o uso de orações adjetivas e os propósitos comunicativos de gêneros jornalísticos distintos. A análise funcionalista evidencia diferenças no funcionamento dessas orações em artigos de opinião e notícias, aprofundando o entendimento do uso padrão dessas estruturas.


Em “Aspectos acústico-prosódicos do alongamento silábico em contextos de intensificação na fala de professores do ensino básico”, Varela apresenta uma análise detalhada de parâmetros acústicos, como duração e frequência fundamental, mostrando a relação entre alongamento silábico e efeitos de intensificação, contribuindo para o estudo da fala em contextos educacionais.

Valls Yoshida, com “A não uniformidade das bases numerais: uma visão dos contrastes das bases cardinais do português brasileiro”, problematiza a homogeneidade das bases cardinais do português. O estudo evidencia contrastes sintáticos, morfológicos e categoriais, propondo uma visão mais detalhada da diversidade gramatical dos numerais no português brasileiro.

Por fim, Zeman apresenta “Corpus-based language comparison: From morphology to dependencies and beyond”, uma análise do *corpus* multilíngue Universal Dependencies, destacando suas extensões e utilidade em áreas como linguística comparativa, tipologia e Humanidades Digitais, evidenciando o potencial do corpus para pesquisas avançadas.

É mister ressaltar que a publicação deste volume, com seus quinze artigos, só pôde se concretizar graças à generosa colaboração de várias pessoas, a quem agradeço imensamente: os autores, os pareceristas, os membros da diretoria do GEL (gestões 2023-2025 e 2025-2027). Estendo, ainda, meus agradecimentos ao Milton Bortoleto, auxiliar editorial, pelo seu apoio essencial, e à Editora Letraria.

Antes de finalizar, gostaria de aproveitar este espaço para jogar luzes sobre uma questão que vem sendo bastante debatida e que sempre deve ser realçada: o fato de que, ao menos em nossa área, levar a cabo a publicação de um periódico acadêmico é uma tarefa cada vez mais espinhosa, sem que se tenha, ao menos, o reconhecimento adequado, em contrapartida, para todas as pessoas envolvidas. A crise editorial é real e enraizada na ideologia da busca por uma produtividade desenfreada, própria do neoliberalismo que infelizmente invadiu as universidades, em que números parecem importar mais do



que a qualidade. Isso somado à falta de recursos materiais coloca os periódicos em um desequilíbrio persistente: a conta simplesmente não fecha. Faço votos de que possamos refletir sobre este problema e tomar ações que nos levem a uma real melhora da situação dos periódicos em nossa área.

A guisa de conclusão, saliento que este volume reafirma o compromisso da revista *Estudos Linguísticos* em divulgar pesquisas inovadoras e diversificadas, refletindo a riqueza e a complexidade do campo dos estudos da linguagem em suas múltiplas abordagens. Convido nossos leitores a explorar os trabalhos apresentados e a se engajar nas discussões que eles suscitam, contribuindo para o contínuo avanço de nossa área. Espero que os trabalhos aqui reunidos instiguem reflexões acadêmico-científicas e renovem o interesse e a admiração dos leitores pelas questões de língua e linguagem.

Dayane Celestino de Almeida
Editora-Chefe

As variáveis linguísticas na realização do objeto direto anafórico de terceira pessoa em legendas audiovisuais

DOI: <http://dx.doi.org/10.21165/el.v54i1.3913>

Lívia Oliveira Azevedo¹

Resumo

Este artigo visa a investigar as variáveis linguísticas na realização do objeto direto anafórico de terceira pessoa em legendas da série *Grey's Anatomy*, partindo de uma aproximação entre a Teoria da Variação e Mudança Linguísticas (Weinreich; Labov; Herzog, 2006 [1978]; Labov, 1994, 2001, 2008 [1972]) e as teorias acerca de gêneros textuais-discursivos (Bakhtin, 2016 [1979], Marcuschi, 2008, 2010). Para isso, foram analisadas legendas profissionais, da *Amazon Prime Video*, e legendas feitas por fãs. A análise estatística foi realizada na plataforma *R* (R Core Team, 2023), levando em conta variáveis independentes definidas com base na literatura e nas especificidades do gênero. Os resultados apontaram para um comportamento particular do fenômeno no *corpus*, influenciado pelas características desse gênero textual-discursivo.

Palavras-chave: objeto direto anafórico de terceira pessoa; legenda audiovisual; variação linguística; gêneros-textuais discursivos.

¹ Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; livia.azevedo@estudante.ufscar.br; <https://orcid.org/0009-0001-3137-4991>

Linguistic variables in the realization of the third-person anaphoric direct object in audiovisual subtitles

Abstract

This paper aims to investigate the linguistic variables involved in the use of the third-person anaphoric direct object in subtitles of the series *Grey's Anatomy*, drawing on an approach that combines the Theory of Language Variation and Change (Weinreich; Labov; Herzog, 2006 [1978]; Labov, 1994, 2001, 2008 [1972]) with theories on textual-discursive genres (Bakhtin, 2016 [1979], Marcuschi, 2008, 2010). For this purpose, both professional subtitles from Amazon Prime Video and fan-created subtitles were analyzed. Statistical analysis was conducted on the R platform (R Core Team, 2023), considering independent variables defined based on relevant literature and the specificities of the genre. The results indicated a particular behavior of this phenomenon within the *corpus*, influenced by the characteristics of this textual-discursive genre.

Keywords: third-person anaphoric direct object; audiovisual subtitles; linguistic variation; textual-discursive genres.

Introdução

A noção fundamental para a Sociolinguística Variacionista é a de que todas as línguas são essencialmente heterogêneas e variáveis, sendo a seleção de cada variante condicionada por fatores linguísticos, sociais e estilísticos (Weinreich; Labov; Herzog, 2006 [1968]; Labov, 1994, 2001, 2008 [1972]). Nesse sentido, a análise linguística deve sempre partir de situações concretas de comunicação e interação, que tomam forma por meio de diferentes gêneros textuais-discursivos e se vinculam a estilos, modalidades, e normas linguísticas particulares (Marcuschi, 2008, 2010), de modo que a exploração detalhada das características estruturais e situacionais desses gêneros torna-se fulcral para a compreensão da variação linguística.

Sob essa perspectiva, o pesquisador não pode se limitar a considerar os gêneros textuais-discursivos apenas durante a etapa de elaboração do *corpus* de análise. Longe disso, dado o impacto significativo que eles exercem sobre as escolhas linguísticas dos falantes, é fundamental que a preocupação com os gêneros perpassasse todas as fases da pesquisa sociolinguística, atravessando desde a definição do objeto de estudo, a coleta dos dados e a delimitação das categorias de análise até a interpretação e a discussão dos resultados (Biazolli; Berlinck, 2021).

Com base nisso, o presente artigo apresenta parte dos resultados obtidos por Azevedo (2024) em uma investigação sobre a realização do acusativo anafórico de terceira pessoa

em legendas audiovisuais para a série *Grey's Anatomy* (2004 –), que considerou traduções criadas por profissionais e por fãs a fim de observar a ocorrência das quatro variantes do fenômeno – o clítico acusativo, o objeto nulo, o sintagma nominal (SN) anafórico e o pronome lexical – em conjunto com as especificidades desse gênero. Assim, nessa pesquisa, além de fatores extralinguísticos específicos do gênero legenda audiovisual, foram considerados os condicionadores linguísticos traço semântico do antecedente, forma verbal, estrutura sintática e função sintática do antecedente, com vistas a averiguar se elas se mostrariam tão relevantes para a ocorrência do fenômeno quanto anteriormente reportado em estudos sobre essa variável em outros gêneros textuais-discursivos.

Dessa forma, este artigo está organizado da maneira apresentada a seguir. A seção 2 debate brevemente as relações entre os estudos dos gêneros e da variação, buscando reforçar a importância dessa correlação na observação de fenômenos variáveis. Na seção 3, descreve-se a metodologia aplicada na pesquisa aqui descrita, englobando a elaboração do *corpus*, a seleção do objeto de estudo e das variáveis, e as etapas empregadas nas análises. Em seguida, a seção 4 traz a discussão e a análise dos resultados alcançados por Azevedo (2024), demonstrando como a compreensão da variação do objeto direto (OD) anafórico de terceira pessoa nas legendas foi facilitada pela abordagem utilizada. Por fim, são tecidas as considerações finais e listadas as referências.

Entre gênero e variação

De acordo com a perspectiva bakhtiniana, a língua toma forma por meio de enunciados orais e escritos, associados a diversos campos de atividade humana, e emitidos por falantes em contextos concretos e particulares. Tais enunciados, em estreita relação com as atividades que pertencem a cada um desses campos, expressam condições e propósitos específicos, refletidos em seus conteúdos temáticos, seus estilos e suas construções composicionais, constituindo o tripé que serve de base para a concepção de **gênero do discurso** apresentada por Bakhtin (2016 [1979]).

Para o autor, os elementos que constituem um gênero tornam inseparável a escolha do falante por um tipo de enunciado do contexto interacional em que ele ocorre. Assim, a opção por um gênero não é imparcial, sendo determinada tanto pelo ambiente comunicativo quanto pelas intenções individuais de cada falante. Nas palavras do estudioso,

[...] essa escolha é determinada pela especificidade de um dado campo da comunicação discursiva, por considerações semântico-objetais (temáticas), pela situação concreta da comunicação discursiva, pela composição pessoal dos seus participantes, etc. Em seguida, a intenção discursiva do falante, com toda a sua individualidade e subjetividade, é aplicada e adaptada ao gênero escolhido, constitui-se e desenvolve-se em determinada forma de gênero [...] (Bakhtin, 2016 [1979], p. 37-38).

Baseando-se em aspectos sociointeracionais, Marcuschi (2008, 2010) também estrutura o conceito de **gênero textual** a partir de três elementos centrais – composição funcional, objetivo enunciativo e estilo –, que se concretizam dinamicamente por meio de formas orais, escritas ou híbridas, situadas histórica e socialmente, em articulação com as forças que regem a sociedade. Assim, os gêneros são fenômenos intrinsecamente ligados à vida social e cultural dos falantes, manifestando-se em diferentes domínios discursivos e operando sobretudo como ferramentas para agir e interagir no mundo, buscando alcançar objetivos comunicativos específicos.

A fim de alcançar plenamente esses objetivos, os falantes optam, então, por formas linguísticas distintas, abrindo espaço para diferentes estilos e, por conseguinte, permitindo a ocorrência da variação. Dessa maneira, embora os gêneros apresentem um caráter relativamente estável que pode ditar certas escolhas ao falante (Bakhtin, 2016 [1979]), eles também admitem diferentes níveis de liberdade para criação. Nesse sentido, as escolhas linguísticas “podem, inclusive, subverter certos significados sociais vinculados a usos linguísticos em gêneros mais padronizados ou podem operar como recurso identitário e político em, por exemplo, obras literárias” (Severo, 2014, p. 48).

Portanto, é essencial reconhecer o gênero como o *locus* da variação linguística, sendo impossível dissociar o estudo aprofundado dos processos de variação e mudança linguísticas do estudo dos **gêneros textuais-discursivos**². Nesse contexto, destacam-se trabalhos como o de Biazolli e Berlinck (2021)³, que defendem a integração dessas duas dimensões e discutem os benefícios dessa abordagem, tais como: (i) a possibilidade de analisar dados que representam diferentes níveis de monitoramento e formalidade ao comparar gêneros distintos; (ii) a identificação de ocorrências relevantes para a variável em estudo, considerando as características específicas de um gênero; e (iii) a transformação de características estruturais e situacionais do gênero em variáveis da pesquisa, permitindo quantificar os efeitos do gênero sobre a variação e assegurando maior rigor analítico.

Desse modo, as autoras defendem a relevância de incorporar questões relacionadas aos gêneros em investigações variacionistas, sugerindo estratégias para integrar esses dois âmbitos, assim como foi proposto no estudo sobre a variação do objeto direto anafórico de terceira pessoa em legendas audiovisuais cujos resultados serão reportados neste artigo.

2 Assim como em Biazolli e Berlinck (2021), opta-se, neste artigo, pelo termo *gênero textual-discursivo*, por considerar o gênero a partir de uma visão que engloba a complementaridade entre texto e discurso.

3 Outros trabalhos que também adotam posturas semelhantes são os de Biazolli (2016), Vieira e Lima (2019) e Lima (2022), para citar alguns.

Metodologia

Considerando as características estruturais e situacionais do gênero textual-discursivo legenda audiovisual, foi compilado um *corpus* composto por dois conjuntos de legendas da série televisiva *Grey's Anatomy*, referentes à 6ª, 10ª e 14ª temporadas. O primeiro conjunto englobou legendas feitas por fãs, as *fansubs*, obtidas gratuitamente pelo *site* Legendas TV⁴; já o segundo foi formado por legendas profissionais, extraídas da plataforma *Amazon Prime Video*, por intermédio do *script* de usuário *Amazon Video – subtitle downloader*⁵, versão 1.9.3, disponibilizado pelo usuário Tithen-Firion no *site* *Greasy Fork*⁶.

Em seguida, todos os arquivos de legenda, coletados de 9 episódios, foram convertidos para o formato .srt e passaram por uma limpeza em que foram utilizadas ferramentas do *site* *Subtitle Tools*⁷ para remover a numeração das legendas, as marcações de tempo (*timestamps*) e outras informações que não eram relevantes para a pesquisa. Adicionalmente, abreviações foram incluídas ao início de cada fala, a fim de que, posteriormente, a identificação das personagens pudesse ser facilmente realizada.

Construiu-se, então, uma amostra-piloto por meio da qual decidiu-se trabalhar com a realização do OD anafórico de terceira pessoa (Duarte, 1986; Cyrino, 1994; Freire, 2005; Marques de Sousa, 2021; Lima, 2022). Com base na literatura disponível sobre o fenômeno e nas hipóteses da pesquisa, foram definidas quatro variantes a serem coletadas e analisadas: o clítico acusativo, o objeto nulo, o SN anafórico e o pronome lexical. De maneira geral, esperava-se que a legenda, por ser um gênero que representa produções orais e ocorre de maneira multimodal, refletiria o uso de variantes mais comuns na fala, principalmente devido à necessidade de se representar a língua de maneira razoavelmente natural. Os exemplos (1) a (4) a seguir, retirados do *corpus*, representam cada uma dessas estratégias, respectivamente.

1. M1: Cadê o *Shane*?
Depois que nos forçou
a aceitá-**lo** de volta... (FT10E13)
2. A3: Tem *minha palavra*. Os dois tem **[Ø]**. (FT14E01)

4 Atualmente, o *site* Legendas TV, anteriormente disponibilizado no endereço <http://legendas.tv/>, não está mais em funcionamento. O último acesso ao *site* foi em 08/12/2022.

5 Disponível em: <https://greasyfork.org/pt-BR/scripts/34885-amazon-video-subtitle-downloader>. Acesso em: 18 nov. 2024.

6 O *site* *Greasy Fork* (<https://greasyfork.org/>) é um repositório de *scripts* de usuário, mantido pelo programador Jason Barnabe. Acesso em: 18 nov. 2024.

7 Disponível em: <https://subtitlertools.com/>. Acesso em: 18 nov. 2024.

3. CA1: Quando eu tinha 10 anos,
estava em um restaurante
com meu pai,
comendo *waffles*.
Eu adorava **waffles**. (FT10E24)
4. AM1: Não consigo dormir porque
fico vendo o rosto *dele* o tempo todo.
Vejo **ele** nos pontos de ônibus. (PT06E01)

Além das variáveis previsoras extralinguísticas, que não são descritas aqui por fugirem ao escopo deste artigo⁸, foram definidas, com base no estudo da legenda e nas hipóteses do trabalho, quatro variáveis linguísticas, a fim de observar se elas se correlacionavam de maneira significativa com a ocorrência do fenômeno. Foram elas: (i) traço semântico do antecedente, que contou com os níveis animado e não animado; (ii) estrutura sintática, que considerou os níveis S V OD, S V OD + complementos e S V OD + minioração; (iii) forma verbal, dividida em infinitivo, não finita ou simples flexionada; e (iv) função sintática do antecedente, separada em igual, diferente e não se aplica.

Para o tratamento dos dados, foram realizados tanto análises de estatística descrita como de estatística inferencial, por meio da plataforma *R* (R Core Team, 2024). As do primeiro tipo consistiram no cálculo das frequências e proporções de cada variante de acordo com as variáveis consideradas. Já para as outras análises, foi elaborado um modelo de regressão logística multinomial com todas as variáveis (incluindo as extralinguísticas), a partir do proposto por Levshina (2015). Os resultados dessas análises são apresentados na seção a seguir.

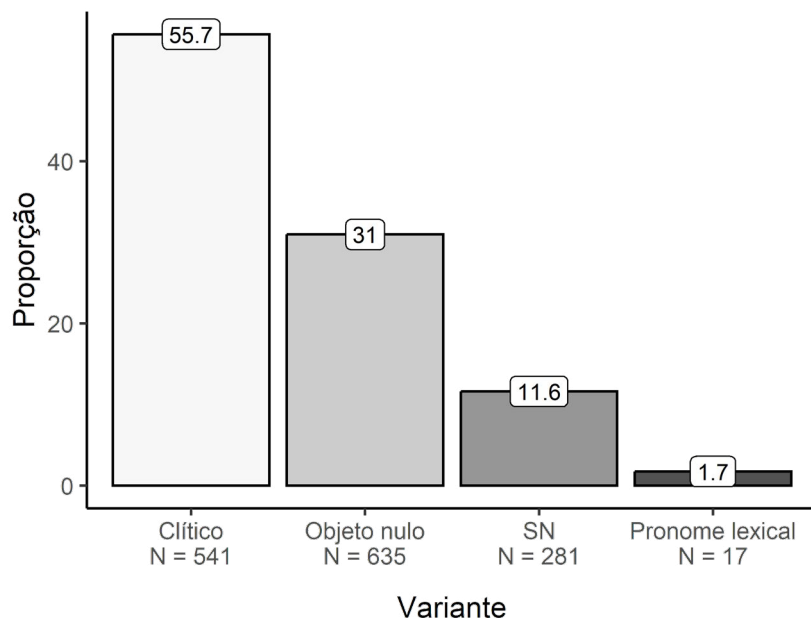
Resultados e discussão

Uma vez excluídos os dados que não foram incluídos na análise⁹, foram identificadas 972 ocorrências do acusativo anafórico de terceira pessoa nas legendas das séries que compuseram do *corpus*. Desse total, 487 ocorrências foram extraídas de legendas profissionais, enquanto 485 foram encontradas em *fansubs*. A distribuição geral desses dados é apresentada na Figura 1.

8 Para mais informações, verificar Azevedo (2024).

9 Não foram incluídas nas análises as ocorrências de acusativos anafóricos que retomavam antecedentes oracionais ou predicativos. Essa decisão foi tomada após a constatação de que apenas objetos nulos e SNs foram empregados nesse contexto, assim como descrito na literatura (Duarte, 1986; Cyrino, 1994; e Freire, 2005, por exemplo). Para conferir a distribuição com a inclusão desses dados, consultar Azevedo (2024).

Figura 1. Frequências e proporções das quatro variantes da realização do acusativo anafórico de terceira pessoa (N = 972)



Fonte: Adaptado de Azevedo (2024, p. 160)

Como ilustrado na Figura 1, o clítico acusativo foi a variante mais frequente no *corpus*, representando 55,7% das ocorrências, seguido pelo objeto nulo (31%), pelo SN (11,6%) e pelo pronome lexical (1,7%). Considerando que os dados são provenientes de um gênero textual-discursivo que busca reproduzir de forma natural a fala das personagens, esperava-se uma menor incidência do clítico, dado que essa variante já não é mais tão comum no português brasileiro (PB) falado (Duarte, 1986; Freire, 2005; Marques de Sousa, 2021, entre outros). Nesse sentido, uma possível justificativa para essas proporções é que, mesmo sendo projetada para refletir a língua oral, a legenda não consegue de fato se afastar do meio escrito em que é produzida.

Em relação às demais variantes, as proporções e frequências observadas estavam dentro do esperado. No caso do objeto nulo, sua alta produtividade já era prevista, considerando que se trata de uma estrutura curta, amplamente difundida no PB e que parece ser um recurso não estigmatizado para realizar a retomada anafórica (Duarte, 1986), o que explica sua presença em cerca de um terço das ocorrências. Da mesma forma, o uso limitado do SN anafórico também era previsto, posto que essa variante, com ou sem modificações, exige um número maior de caracteres em comparação às demais, o que pode levar os legendistas a evitá-la. Por fim, a baixa frequência do pronome lexical era igualmente esperada, já que, embora relativamente comum na fala, ele ainda carrega certo estigma, o que pode desencorajar seu uso em legendas.

Com vistas a compreender plenamente a variação linguística na realização do OD anafórico de terceira pessoa, foi necessário empreender outras análises além das iniciais. A fim de explorar o impacto de cada variável preditora sobre o fenômeno, foi elaborado um modelo de regressão logística multinomial utilizando a função *mlogit* no *R* que incluía todas as variáveis predictoras e três das quatro variantes, após a exclusão dos dados referentes aos pronomes lexicais devido à sua baixa frequência. Os resultados referentes às variáveis linguísticas, apresentados em *logodds*, estão dispostos na Tabela 1¹⁰.

Tabela 1. Resultado da análise de regressão logística multinomial para as variáveis linguísticas na realização do acusativo anafórico de terceira pessoa em legendas audiovisuais (N = 955)

	Estimativa	Erro padrão	Valor-z	p	Aplicação	N
Clítico (ref.)						
Intercept 1: ON	-1,273	0,548	-2,321	0,02		
Intercept 2: SN	-2,663	0,702	-3,791	<0,001		
Traço semântico						
Animado (ref.)					456/530	(86%)
Não animado: ON	3,435	0,226	15,163	<0,001	250/425	(59%)
Não animado: SN	3,086	0,289	10,602	<0,001	90/425	(21%)
Forma verbal						
Infinitivo (ref.)					277/419	(66%)
Não finita: ON	0,777	0,215	3,616	<0,001	14/29	(48%)
Não finita: SN	0,059	0,265	2,108	<0,001	5/29	(17%)
Simples flexionada: ON	1,277	0,659	1,936	0,05	186/507	(37%)
Simples flexionada: SN	0,905	0,771	1,173	0,24	67/507	(13%)
Estrutura sintática						
S V OD (ref.)					294/591	(50%)
S V OD + complementos: ON	-0,166	0,223	-0,748	0,45	86/307	(28%)
S V OD + complementos: SN	-0,924	0,304	-3,030	0,002	20/307	(7%)
S V OD + minioração: ON	-0,948	0,526	-1,801	0,07	7/57	(12%)
S V OD + minioração: SN	-0,703	0,596	-1,178	0,23	4/57	(7%)
Função do antecedente						
Igual (ref.)					234/510	(46%)

10 Os resultados completos da regressão, incluindo as variáveis extralinguísticas, podem ser encontrados em Azevedo (2024).

Diferente: ON	-0,477	0,294	-1,618	0,10	29/193	(15%)
Diferente: SN	-0,670	0,366	-1,826	0,06	13/193	(6%)
NA: ON	0,277	0,242	1,147	0,25	81/252	(32%)
NA: SN	-0,656	0,340	-1,928	0,05	15/223	(6%)

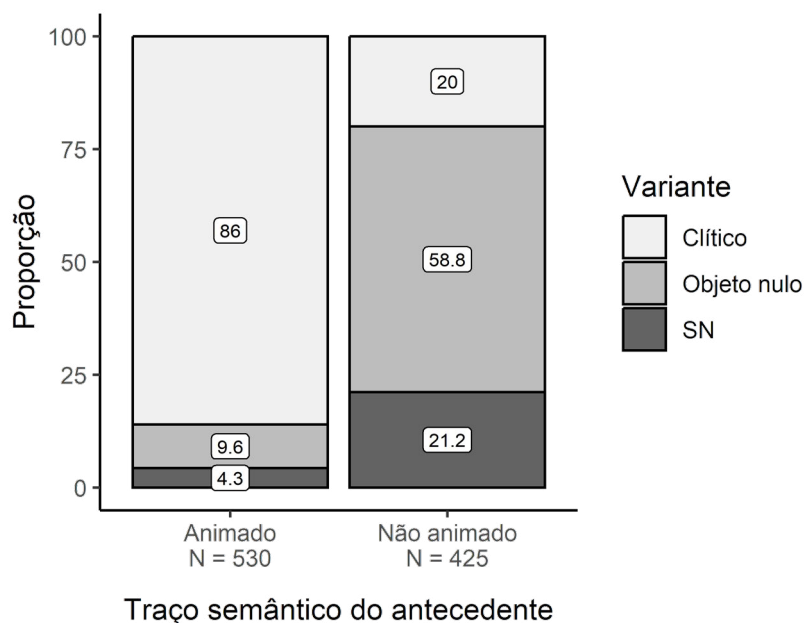
Modelo: $mlogit(VD1 \sim 1 \mid TRACO.ANT + FUNC.ANT + FOR.VERB.A + EST.SINT.A + TOPICO + RELACAO + TIPO.LEGENDA + CPL + CPS, data = dfm2, releve = 1)$.

Fonte: Adaptado de Azevedo (2024, p. 165)

Observando as estimativas reportadas na Tabela 1, tem-se que, de modo geral, a categoria vazia e o SN foram desfavorecidos em -1,273 e em -2,663 *logodds*, respectivamente, quando comparados ao clítico acusativo. Isso indica que, nas legendas audiovisuais analisadas, o uso do clítico é mais provável do que o de um objeto nulo ou um SN anafórico, sendo esta última variante a forma com a menor probabilidade de ocorrer. Além disso, os valores-*p* obtidos para ambas as estratégias confirmam que as proporções inicialmente observadas são estatisticamente significativas, corroborando a distribuição geral dos dados.

Passando à discussão das variáveis linguísticas consideradas, apresenta-se, primeiramente, o efeito do traço semântico do antecedente na expressão do acusativo anafórico de terceira pessoa. Por meio dos coeficientes gerados, é possível notar que o clítico acusativo foi favorecido ao retomar antecedentes animados, ao passo que a categoria vazia e o SN foram favorecidos em 3,435 e 3,068 *logodds*, respectivamente, em contextos nos quais o antecedente era não animado. Percebe-se que o objeto nulo foi a estratégia mais favorecida nesse caso, corroborando a hipótese de que antecedentes de traço [-animado] tenderiam a ser retomados por complementos não lexicalizados. Para que os resultados relacionados à realização do OD anafórico de terceira pessoa em relação ao traço semântico do antecedente sejam mais bem visualizados, pode-se observar as frequências e proporções apresentadas na Figura 2.

Figura 2. Frequências e proporções das variantes do OD anafórico de terceira pessoa de acordo com o traço semântico do antecedente (N = 955)



Fonte: Adaptado de Azevedo (2024, p. 168)

Como ilustrado no gráfico, há diferenças claras na distribuição das variantes conforme o traço semântico do antecedente. Para referentes animados, o clítico foi usado em 86% dos casos, superando amplamente o objeto nulo (9,6%) e o SN (4,3%). Já para antecedentes não animados, o objeto nulo predominou, com 58,8%, enquanto clíticos (20%) e SNs (21,2%) tiveram proporções similares. Esses dados, aliados às estimativas do modelo de regressão, indicam que a realização do OD anafórico não é aleatória, mas se correlacionou a essa variável.

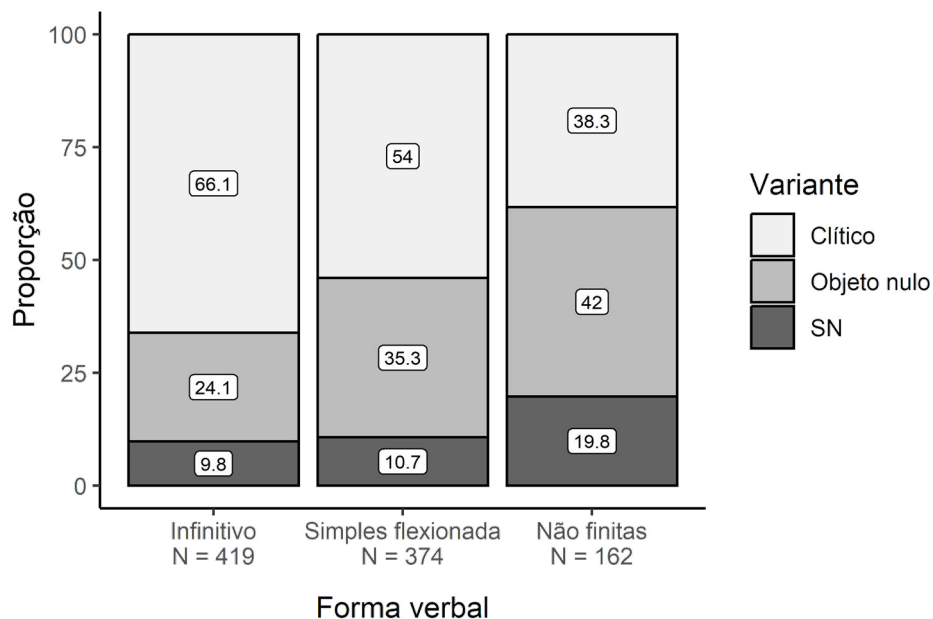
Tal conclusão reforça os achados da literatura, que constata que a predominância do objeto nulo para antecedentes com traço [-animado] se deu por meio do processo de difusão dessa variante no PB. De fato, foi a partir da substituição do clítico neutro pela categoria vazia que essa variante ganhou uma maior projeção, sendo ela a estratégia preferida no PB oral atualmente (cf. Cyrino, 1994; Cyrino; Duarte; Kato, 2000, entre outros). Os exemplos (5) e (6) a seguir trazem casos de antecedentes animados e não animados, respectivamente referenciados pelo clítico e pelo objeto nulo.

5. M3: *Ele* está em Bagdá
com o exército americano.
Não pude adotá-**lo** legalmente,
e ele é um refugiado sírio. (FT14E01)

6. AZ1: Como *um osso*.
 Às vezes é preciso quebrar [Ø]
 para solidificar certo. (PT10E01)

Quanto à forma verbal, as apresentadas na Tabela 5 indicam que as formas verbais no infinitivo (simples ou complexo) favoreceram o uso do clítico acusativo em relação às outras duas variantes. Em contrapartida, para outras formas não finitas, como o particípio e o gerúndio, o SN (0,777 *logodds*) e o objeto nulo (0,059 *logodds*) foram mais favorecidos em comparação com o clítico, sendo o SN a variante preferida. Complementando a interpretação, a Figura 3 traz a distribuição das variantes do acusativo anafórico de terceira pessoa de acordo com a forma verbal.

Figura 3. Frequências e proporções das variantes do OD anafórico de terceira pessoa de acordo com a forma verbal (N = 955)



Fonte: Adaptado de Azevedo (2024, p. 170)

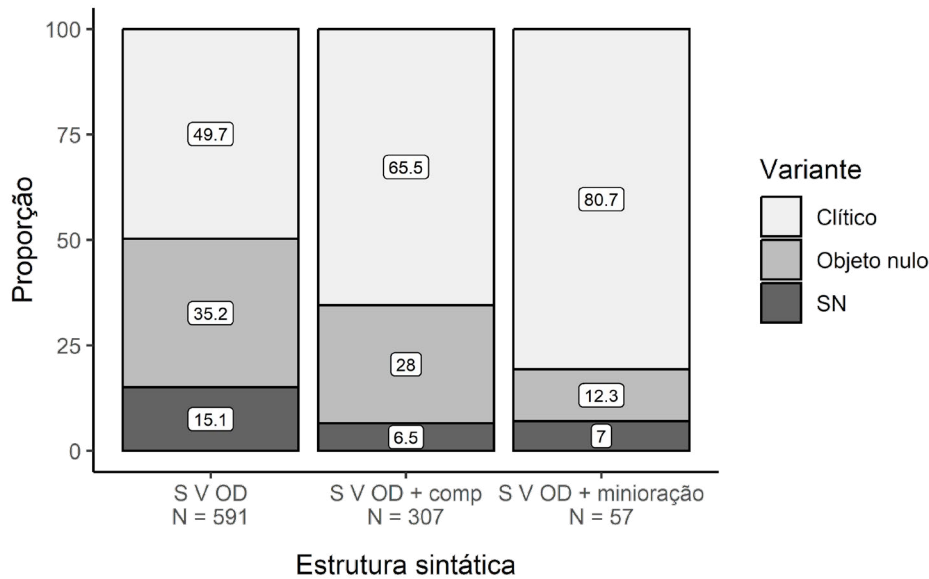
Como demonstrado no gráfico, o clítico acusativo foi a forma mais expressiva quando o OD era argumento de verbos no infinitivo, representando 66,1% das ocorrências, enquanto o objeto nulo apareceu em 24,1% e o SN em 9,8% dos casos. No entanto, diante de outras formas não finitas, o objeto nulo foi a variante mais frequente, com 42%, seguido pelos SNs (19,8%) e o clítico acusativo (38,3%). Assim, à medida que diminui o uso do clítico, parecem aumentar as taxas de objeto nulo e SN, corroborando os resultados da regressão logística multinomial.

Portanto, as expectativas para essa variável foram parcialmente confirmadas. Com efeito, o clítico foi a variante mais favorecida com infinitivos, como reportado anteriormente por Marafoni (2004), Freire (2005) e Lima (2022), por exemplo. Contudo, no contexto de outras formas não finitas, o objeto nulo foi a forma preferida. Além disso, o clítico se manteve em metade das ocorrências com formas simples flexionadas, reforçando sua expressividade em diversos contextos. Os exemplos (7) e (8) ilustram, respectivamente, os contextos favoráveis ao clítico e às demais estratégias.

7. L2: *Ela* perguntou e você não disse nada.
O que, estava esperando o momento
certo para humilhá-**la** publicamente? (PT10E13)
8. Não vou operar *uma paciente*
que não me quer operando **[Ø]**. (FT06E13)

Já no que concerne à estrutura sintática, os coeficientes do modelo de regressão logística multinomial mostram uma correlação significativa desse fator com a variável resposta. Quando o SN foi usado em sentenças de estrutura S V OD + complementos – orações que incluíam, além do OD, um objeto indireto ou complemento oblíquo –, o SN foi desfavorecido em -0,924 *logodds* em relação nível de referência, isto é, o clítico acusativo. A Figura 4 apresenta as frequências e proporções calculadas para cada nível da variável resposta, a fim de auxiliar na compreensão dos resultados.

Figura 4. Frequências e proporções das variantes do OD anafórico de terceira pessoa de acordo com a estrutura sintática (N = 955)



Fonte: Adaptado de Azevedo (2024, p. 171)

As proporções apontadas no gráfico indicam que o clítico foi a estratégia mais empregada em todos os tipos de estrutura sintática considerados, tendo seguido a tendência geral da amostra e superou a soma das outras variantes em estruturas complexas (com complementos ou minioração). Por outro lado, o SN foi a variante menos empregada, alcançando sua menor taxa de aplicação (6,5%) em construções de configuração S V OD + complementos. O excerto (9) traz um exemplo de SN em uma estrutura que o desfavoreceu e o (10) ilustra um caso de clítico usado com uma minioração.

9. C2: Dra. Grey! O *Owen* está morto?
M1: Ele está vivo. Está vivo.
Inconsciente, mas vivo.
C2: Leve **Owen** para a sala em frente e retire
essa bala. (PT06E24)
10. M1: Você manda lá fora,
mas eu mando aqui.
Saia do meu caminho.
RP1: Deixem-**na** passar. (FT10E24)

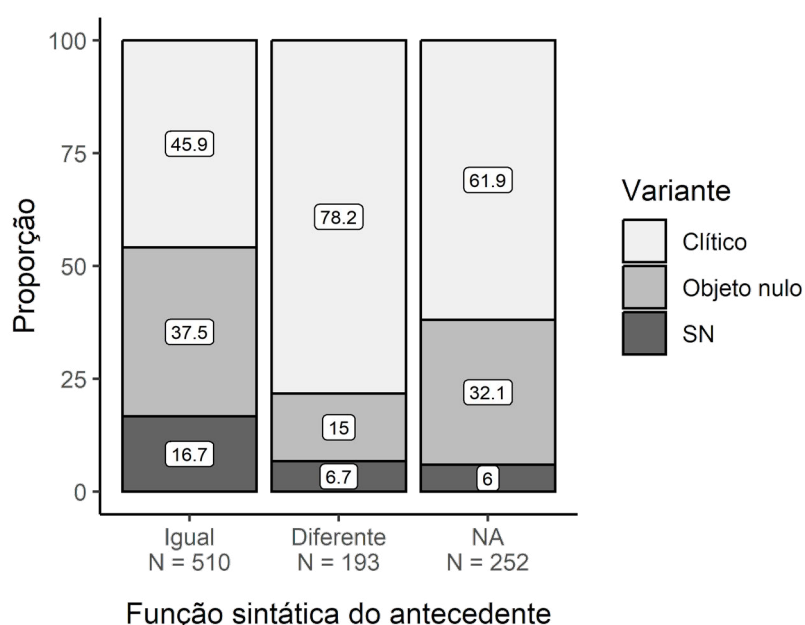
Em relação a esse último tipo de estrutura – o das miniorações –, é interessante mencionar que, embora o pronome lexical não tenha sido incluído na análise multivariada devido à baixa frequência dessa variante no *corpus*, foi precisamente em estruturas com miniorações que ele demonstrou maior produtividade. Das 17 ocorrências encontradas nas legendas, 7 (41%) foram empregadas quando o objeto exercia uma função dupla, confirmando achados anteriores da literatura (cf. Duarte, 1986; Marafoni, 2004). As legendas (11) e (12) a seguir representam casos em que o OD é seguido por um verbo e por um predicativo, nessa ordem.

11. D1: Bipei a *Brooks*.
Mande **ela** fazer uma TC de cabeça,
tórax e abdômen da Srta. Ashford. (PT10E01)
12. LO1: E quando voltarem para casa terei
que contar para eles que *Georgie*... [...]
Não consigo nem ver
ele desse jeito, não dá. (PT06E01)

Dessa forma, foi possível confirmar as hipóteses levantadas para essa variável, já que o clítico se mostrou a variante mais predominante em todos os contextos, enquanto o pronome lexical, apesar de pouco expressivo, apareceu com mais frequência em sentenças com a estrutura S V OD + minioração.

Por fim, faz-se interessante averiguar qual foi a distribuição das variantes de acordo com a função do antecedente, mesmo que essa variável não tenha se correlacionado significativamente com a realização do fenômeno. Conforme apresentado na Figura 5, as frequências e proporções das variantes do OD anafórico de terceira pessoa mostram que o clítico foi a variante mais produtiva nos três contextos analisados. No entanto, é importante notar que o objeto nulo foi mais frequente na retomada de antecedentes que também desempenhavam a função de OD (37,5%), competindo com o clítico, consoante estudos como os de Cyrino (1994), Marafoni (2004) e Santana (2016).

Figura 5. Frequências e proporções das variantes do OD anafórico de terceira pessoa de acordo com a função sintática do antecedente (N = 955)



Fonte: Adaptado de Azevedo (2024, p. 184)

Além disso, é importante observar os casos em que não foi possível identificar o antecedente – seja por ser mostrado visualmente, mencionado em outra cena ou não fazer parte de uma oração recuperável –, posto que esse foi um nível definido a partir da natureza do gênero textual-discursivo analisado. Como mencionado anteriormente, nos casos em que a função sintática do antecedente não pôde ser determinada, o clítico acusativo predominou, representando 61,9% dos dados, seguido pelo objeto nulo (32,1%) e, por último, pelo SN (6%). Os trechos (13) e (14) a seguir exemplificam, respectivamente, a ocorrência com a estratégia menos expressiva nesse contexto e a com o objeto nulo se referindo a um OD apresentado visualmente.

13. H1: *Lidocaína e vários 4x4*.
Pode levar **isso** para o Hunt? (PT10E01)
14. S3: Desculpe.
J2: Por que você não tira **[Ø]** (FT14E01)

Além disso, é importante observar os casos em que não foi possível identificar o antecedente – seja por ser mostrado visualmente, mencionado em outra cena ou não fazer parte de uma oração recuperável –, posto que esse foi um nível definido a partir da natureza do gênero textual-discursivo analisado. Como mencionado anteriormente, nos casos em que a função sintática do antecedente não pôde ser determinada, o clítico acusativo predominou, representando 61,9% dos dados, seguido pelo objeto nulo (32,1%) e, por último, pelo SN (6%). Os trechos (13) e (14) a seguir exemplificam, respectivamente, a ocorrência com a estratégia menos expressiva nesse contexto e a com o objeto nulo se referindo a um OD apresentado visualmente.

Assim, com base na interpretação dos resultados, não foi possível confirmar completamente as hipóteses propostas. Apesar de a categoria vazia realmente ter sido mais utilizada com antecedentes de função sintática similar, não houve correlação estatisticamente significativa, como observado em Lima (2022). Ademais, de maneira oposta às expectativas, o clítico, e não o objeto nulo, foi a variante mais empregada quando a função sintática do antecedente não foi indicada, o que parece estar alinhado à tendência geral de uso do pronome acusativo no *corpus*. Entretanto, a menor expressividade do SN nesse contexto está em consonância com a lógica que motivou a elaboração da hipótese: como os antecedentes podem ser facilmente recuperados, não há necessidade de mencioná-los ou de empregar processos de referência do mesmo modo que em um texto escrito.

Considerações finais

Toda manifestação linguística é necessariamente materializada por meio de um gênero textual-discursivo, o que, naturalmente, influencia o comportamento dos fenômenos observados. Dessa forma, é cada vez mais imprescindível e produtivo voltar um olhar cuidadoso às características dos gêneros utilizados em estudos sociolinguísticos, visando a identificar os fatores específicos que condicionam a ocorrência de uma variável em um determinado gênero.

Com isso em mente, o presente artigo apresentou parte dos resultados de uma pesquisa conduzida por Azevedo (2024) acerca da realização do acusativo anafórico de terceira pessoa em legendas audiovisuais da série *Grey's Anatomy* (2004–), a partir de um *corpus* que englobou tanto legendas profissionais como *fansubs*. Para as análises, foram considerados, além de fatores extralinguísticos específicos do gênero legenda

audiovisual, os condicionadores linguísticos traço semântico do antecedente, forma verbal, estrutura sintática e função sintática do antecedente. O estudo buscou analisar se esses fatores se mostrariam relevantes para a ocorrência do fenômeno, assim como já havia observado em outras pesquisas sobre esse objeto de estudo.

No geral, os resultados da pesquisa aqui descrita tornaram possível a identificação de correlações estatisticamente significativas entre a realização do acusativo anafórico de terceira pessoa e as variáveis linguísticas incluídas na pesquisa. Foi constatado que as legendas audiovisuais demonstram uma clara preferência pelo clítico acusativo, o que pode se relacionar às recomendações normativas e estilísticas dos guias de legendagem que guiam o trabalho dos tradutores, bem como aos contextos de produção e recepção desse gênero. Além disso, os resultados obtidos para as variáveis linguísticas consideradas corroboram achados já reportados na literatura sobre o PB escrito, sugerindo que o comportamento do objeto direto anafórico nas legendas audiovisuais é similar ao observado em outros gêneros da modalidade escrita.

Dessa forma, mesmo concentrando-se apenas nas variáveis linguísticas, foi possível notar que o gênero textual-discursivo exerce uma influência significativa sobre a realização desse fenômeno variável. Tal constatação reforça a necessidade de se considerar as especificidades do gênero ao se estudar a variação linguística, não só no caso da realização do OD anafórico de terceira pessoa em legendas audiovisuais, mas em toda e qualquer investigação sociolinguística que se queira mais precisa e abrangente.

Referências

- AZEVEDO, L. O. *A língua na tela: descrição e análise da variação do objeto direto anafórico de terceira pessoa em legendas profissionais e fansubs*. 2024. 213 f. Dissertação (Mestrado em Linguística) – Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2024.
- BAKHTIN, M. *Os gêneros do discurso*. 1. ed. São Paulo: Editora 34, 2016 [1979].
- BIAZOLLI, C. C.; BERLINCK, R. de A. Por que investigar processos de variação e mudança linguísticas por meio de gêneros textuais-discursivos? In: BIAZOLLI, C. C.; BERLINCK, R. de A. (org.). *Gêneros textuais-discursivos no estudo de processos de variação e mudança*. Campinas: Pontes Editores, 2021. p. 13-38.
- CYRINO, S. M. L. *O objeto nulo no português brasileiro: um estudo sintáticodiacrônico*. 1994. 229 f. Tese (Doutorado em Ciências) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1994.

CYRINO, S. M. L.; DUARTE, M. E. L.; KATO, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: KATO, M. A.; NEGRÃO, E. V. (org.). *Brazilian Portuguese and the null subject parameter*. Madri: Iberoamericana; Frankfurt am Main: Verveurt, 2000. p. 55-73.

DUARTE, M. E. L. *Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no Português do Brasil*. 1986. 73 f. Dissertação (Mestrado em Ciências – Linguística aplicada ao ensino de línguas) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1986.

FREIRE, G. C. *A realização do acusativo e do dativo anafóricos de terceira pessoa na escrita brasileira e lusitana*. 2005. 204 f. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

LABOV, W. *Principles of Linguistic Change*. Vol 1: Int ed. Cambridge: Blackwell, 1994.

LABOV, W. *Principles of Linguistic Change*. Vol 2: Soc ed. Cambridge: Blackwell, 2001.

LABOV, W. *Padrões Sociolinguísticos*. São Paulo: Parábola, 2008 [1972].

LEVSHINA, N. *How to do Linguistics with R*. Amsterdã: John Benjamins, 2015.


LIMA, M. D. A. de O. *Continuum de gêneros textuais jornalísticos para a descrição de norma(s) culta(s): o acusativo anafórico de terceira pessoa e a ordem dos clíticos pronominais*. 2022. 313 f. Tese (Doutorado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2022.

MARAFONI, R. L. *A realização do objeto direto anafórico: um estudo em tempo real de curta duração*. 2004. 112 f. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

MARCUSCHI, L. A. *Da fala para a escrita: atividades de retextualização*. 10. ed. São Paulo: Cortez, 2010.

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

MARQUES DE SOUSA, A. A. *As realizações do acusativo anafórico em variedades do português*. 2021. 201 f. Tese (Doutorado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.



R CORE TEAM. *R: A language and environment for statistical computing*. Vienna: R. Foundation for Statistical Computing, 2023. Disponível em: <http://www.R-project.org/>. Acesso em: 8 fev. 2024.

SANTANA, J. M. C. P. de. *Diagnose e ensino de pronomes: um estudo sobre a retomada anafórica do objeto direto de terceira pessoa no Português brasileiro*. 2016. 205 f. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

SEVERO, C. G. Estilo, variação linguística e discurso. In: GÖRSKI, E. M.; COELHO, I. L.; NUNES DE SOUZA, C. M. (org.). *Variação estilística – reflexões teórico-metodológicas e propostas de análise*. Florianópolis: Insular, 2014. p. 31-49.

WEINREICH, U.; LABOV, W.; HERZOG, M. *Fundamentos empíricos para uma teoria da mudança lingüística*. São Paulo: Parábola, 2006 [1968].

Transcrição automática de entrevistas e anotação Universal Dependencies no Corpus Roda Viva

DOI: <http://dx.doi.org/10.21165/el.v54i1.3851>

Cláudia Dias de Barros¹
Oto Araújo Vale²
Gabriela Wick-Pedro³

Resumo

Neste artigo é apresentada a pesquisa sobre a transcrição automática de quatro entrevistas extraídas do Corpus Roda Viva, que é formado por 713 entrevistas do Programa Roda Viva, da TV Cultura. As entrevistas originais foram transcritas por jornalistas, adquirindo, assim, um *status* de texto escrito, possuindo, ainda, intervenções, como informações enciclopédicas sobre fatos e pessoas citadas. A fim de trabalhar com texto oral, a presente pesquisa realizou um trabalho piloto de transcrição automática de quatro dessas entrevistas, usando a ferramenta Whisper e, posteriormente, as entrevistas foram anotadas automaticamente com a formalização da Universal Dependencies e revisadas manualmente pela ferramenta Arborator Grew ElizIA. Por meio desse trabalho, pôde-se notar as diferenças sintáticas presentes no *corpus* original e nas entrevistas transcritas automaticamente.

Palavras-chave: Universal Dependencies; Sintaxe; Linguística de corpus; reconhecimento automático de fala.

1 Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Sertãozinho, São Paulo, Brasil; claudias84@gmail.com; <https://orcid.org/0009-0003-9388-4297>

2 Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; otovale@ufscar.br; <https://orcid.org/0000-0002-0091-8079>

3 Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; gabiwick@gmail.com; <https://orcid.org/0000-0002-7332-4482>

Automatic transcription of interviews and Universal Dependencies annotation in the Roda Viva Corpus

Abstract

This article presents research on the automatic transcription of four interviews extracted from the Roda Viva Corpus, which consists of 713 interviews from the Roda Viva Program, on TV Cultura. The original interviews were transcribed by journalists, thus acquiring the status of written text, and also presenting interventions, such as encyclopedic information about facts and people mentioned. In order to work with oral text, this research carried out a pilot work of automatic transcription of four of these interviews, using the Whisper tool. Subsequently, the interviews were automatically annotated with the formalization of Universal Dependencies and manually reviewed by the Arborator Grew ElizIA tool. Through this work, it was possible to note the syntactic differences present in the original corpus and in the automatically transcribed interviews.

Keywords: Universal Dependencies; Syntax; Corpus linguistics; automatic speech recognition.

Introdução

As pesquisas com *corpus* são utilizadas há algum tempo na área de Processamento de Línguas Naturais (PLN) e são fundamentais para que fenômenos linguísticos possam ser observados em seu contexto de uso.

Um tipo especial de *corpus* são aqueles formados por língua oral, como entrevistas, por exemplo, nos quais podem ser observadas ocorrências da língua em uso, com maior ou menor nível de formalidade.

O trabalho tem como objetivo estudar a transcrição automática de quatro dentre as 713 entrevistas que formam o Corpus Roda Viva (Miranda Jr. *et al.*, 2024). De fato, as transcrições que se encontram no Corpus Roda Viva são transcrições manuais de entrevistas realizadas no Programa Roda Viva, da TV Cultura, de 1986 a 2009, feitas por jornalistas. Essas entrevistas apresentam informações adicionais, inseridas pelos transcritores, como informações enciclopédicas sobre fatos e pessoas relatadas e correções de desvios gramaticais, por exemplo. Dessa forma, pode-se afirmar que as transcrições não apresentam exatamente as características de um texto oral, mas sim de um texto jornalístico transcrito. A escolha das quatro entrevistas não se deu ao acaso. Foram escolhidas de acordo com um critério de diversidade que fosse de uma fala mais formal, como a entrevista de Mauricio de Souza, passando por uma entrevista política, que também tinha um alto grau de formalidade, que é o caso da entrevista de Benedita

da Silva (então governadora do Rio de Janeiro), chegando a uma entrevista com menos formalidade no caso do jogador de futebol Edmundo, até uma fala bem característica da periferia paulistana na entrevista de Mano Brown.

A fim de se trabalhar com um texto realmente oral, optou-se, neste trabalho, por realizar a transcrição automática das entrevistas por meio de um ASR chamado Whisper (Radford *et al.*, 2023). Essa ferramenta apresentou alguns problemas na transcrição, que serão apresentados com mais detalhe neste artigo.

Após a transcrição automática das quatro entrevistas e a correção manual dos problemas apresentados, realizou-se a anotação automática delas com as etiquetas da Universal Dependencies (De Marneffe *et al.*, 2021) e uma posterior revisão manual. As entrevistas anotadas e revisadas serão inseridas no *corpus* Porttinari (Pardo *et al.*, 2021), a fim de compor a parte oral desse grande *corpus*. Essa escolha se dá em função do formalismo bastante transparente que oferece o quadro das Universal Dependencies.

Na próxima seção será apresentada a fundamentação teórico-metodológica utilizada neste trabalho, com ênfase no gênero oral entrevista e suas características, descrição do Corpus Roda Viva e Corpus Roda Viva TW, ferramentas utilizadas, como o Whisper, parser PortParser, Arborator Grew ElizIA e o formalismo utilizado nas anotações: a Universal Dependencies.

Na seção 3 são apresentados os resultados do trabalho, com ênfase na comparação das transcrições manual e automática das entrevistas.

Por fim, são apresentados os agradecimentos e as referências bibliográficas.

Fundamentação Teórico-Metodológica

Nesta seção serão abordados os fundamentos teórico-metodológicos utilizados neste trabalho.

Gênero oral: entrevista e suas características

Como ponto de partida para se abordar o tema das entrevistas, é necessário definir-se a noção de gênero discursivo. Este trabalho parte da perspectiva de Bakhtin (2016), para quem os gêneros discursivos se constituem a partir de três elementos: o conteúdo temático, o estilo e a construção composicional, além da atividade humana e das situações de interação da vida social dos indivíduos.

Os gêneros discursivos podem ser escritos ou orais, como as entrevistas, sobre as quais Marcuschi (2005, p. 15) destaca alguns pontos importantes, como: a) interação entre pelo menos dois falantes; b) ocorrência de pelo menos uma troca de falantes; c) presença de uma sequência de ações coordenadas; d) execução numa identidade temporal.

Relacionada à oralidade, destaca-se a conversação, através da qual os gêneros orais existem e ocorrem nas diversas situações comunicativas, como as entrevistas, os debates, as notícias. Segundo Marcuschi (1988, p. 319-320), conversação pode ser definida “como uma interação centrada, da qual participam pelo menos dois interlocutores que se revezam, tomando cada qual pelo menos uma vez a palavra, dando-se o evento comunicativo em uma identidade temporal”.

Um conceito importante relacionado à conversação é o **tópico discursivo**, que pode ser definido como o ponto de partida de uma conversa. Segundo Lira (2020), para desenvolver o tópico discursivo, o falante considera seus interlocutores, ou seja, suas representações, reações, intenções, expectativas, grau de intimidade, grau de concordância ou discordância sobre o que está sendo discutido no discurso.

Outro conceito importante é o **turno conversacional**, que tem como função organizar o texto oral, para que haja uma progressão lógica das ideias. Em relação aos turnos, ressalta-se que a conversação pode ser de dois tipos: assimétrica e simétrica. Melo Júnior (2016, p. 15) define que, na conversação assimétrica, “há uma hierarquia linguisticamente marcada entre os interactantes de uma situação discursiva, em que um dos participantes do evento de fala detém o poder da palavra e comanda o turno conversacional”. Já “as relações simétricas acontecem quando dois interactantes têm o mesmo poder de interagir ou o mesmo poder da palavra” (Melo Júnior, 2016, p. 17).

Ainda em relação à conversação, outro conceito essencial são os **pares adjacentes**, definidos por Marcuschi (2007, p. 35) como “uma sequência de dois turnos que co-ocorrem e servem para a organização local da conversação”. Segundo Lira (2020), alguns exemplos dessas sequências são definidos como saudação/saudação, convite/recusa; agradecimento/aceitação; pergunta/resposta.

Finalmente, um último conceito relacionado à conversação que pode ser citado são os **marcadores conversacionais**. Segundo Urbano (1997, p. 81), eles são elementos de “variada natureza, estrutura, dimensão, complexidade semântico-sintática, aparentemente supérfluos ou até complicadores, mas de indiscutível significação e importância para qualquer análise de textos orais e para sua boa e cabal compreensão”.

Marcuschi (1986) apresenta a seguinte classificação em relação aos marcadores conversacionais:

1. **MC simples:** é o marcador que se realiza com um só lexema ou um paralexema, como as interjeições, os advérbios, os verbos, os adjetivos, as conjunções, os pronomes, entre outros.
2. **MC composto:** de caráter sintagmático, com grande tendência à estereotipia e com pouca alteração morfológica no tipo produzido.
3. **MC oracional:** trata-se de pequenas orações, podendo se apresentar em todos os tempos e formas verbais ou modos oracionais (assertivo, indagativo, exclamativo).
4. **MC prosódico:** é o MC formado com recursos prosódicos e normalmente utilizado com algum MC verbal. Encontram-se, nesse contexto, a entonação, a hesitação, o tom de voz, entre outros (Marcuschi, 1986, p. 290-291).

Segundo Lira (2020), o texto oral apresenta alguns mecanismos de textualidade característicos, como:

- a) **Hesitação** – segundo Marcuschi (2006), a hesitação se constitui nos aspectos formais, cognitivos e interacionais. Ela é caracterizada por repetição de palavras, como as pausas e os alongamentos (de vogais, consoantes ou sílabas). O falante se utiliza da hesitação para reformular sua interação, a fim de se comunicar de forma mais eficiente;
- b) **Repetição** – é utilizada para a estruturação do discurso oral e também demonstra características de um planejamento presencial, com características de um texto espontâneo. Segundo Koch (2005, p. 145):

A repetição é particularmente constitutiva do discurso conversacional, no qual os parceiros, conjuntamente e passo a passo, constroem o texto, elaboram as ideias, criam, preservam e negociam as identidades, de tal forma que o texto, de maneira icônica, vai refletir essa atividade de co-produção.

Koch (2017, p. 84) enfatiza que “[...] a repetição de itens lexicais tem por efeito trazer ao enunciado um acréscimo de sentido que ele não teria se o item fosse usado somente uma vez”.

- c) **Paráfrase** – Segundo Lira (2020), a paráfrase pode ser definida como a retomada explícita e consciente de outros textos, ou seja, o texto parafraseando não se diferencia do outro, mas sim compartilha semelhanças. As paráfrases podem ser introduzidas por “ou seja”, ou “isto é”, algumas vezes.

- d) **Correção** – Pode ser definida como uma formulação retrospectiva, que exprime contraste. Fávero *et al.* (2006, p. 258) definem a correção como “um enunciado linguístico que reformula um anterior, considerado ‘errado’ aos olhos de um dos interlocutores”.

As entrevistas fazem parte do gênero oral e são caracterizadas por um monitoramento maior dos falantes. Segundo Costa (2023), esse gênero oral pode apresentar vários subtipos, como entrevista televisiva, radiofônica, de emprego, de seleção, pingue-pongue, entre outras.

Conforme Costa (2023), na construção do gênero entrevista, é necessário levar em consideração razões que vão desde sua situação de produção, como o contexto, o propósito comunicativo, o nível de formalidade e o público-alvo, como também sua recepção.

De acordo com Medina (1990), a entrevista é uma forma de interação social entre duas ou mais pessoas, que tem como alvo obter e difundir informações através dos seus participantes, que são o entrevistador e o entrevistado. Ele considera esse gênero como uma técnica de obtenção de informação e de interação entre os participantes.

De acordo com Minayo (2010, p. 280):

A entrevista é considerada uma modalidade de interação entre duas ou mais pessoas. Essa pode ser definida como a técnica em que o investigador se apresenta frente ao investigado e por meio de perguntas formuladas busca a obtenção dos dados que lhe interessa. É uma conversa a dois, ou entre vários interlocutores, realizada por iniciativa do entrevistador, destinada a construir informações pertinentes para o objeto de pesquisa, e abordagem pelo entrevistador, de temas igualmente pertinentes tendo em vista este objetivo.

Segundo Costa (2023), a entrevista deve ser bem planejada, visando alcançar o seu objetivo principal, que é a obtenção de informações e, por isso, as perguntas devem ser bem formuladas, adequadas à situação e ao tema abordado, levando em consideração o pensamento do entrevistado.

Lira (2020, p. 25) aponta que a entrevista televisiva tem como particularidade o fato de possibilitar a visualização de imagens, o que gera uma maior aproximação entre o entrevistador e o entrevistado e os telespectadores e isso colabora com a interação social entre os sujeitos.

Costa (2023), em seu trabalho, apresenta uma tabela com algumas características das entrevistas, que são apresentadas no Quadro 1:

Quadro 1. Características das entrevistas

CARACTERÍSTICAS DAS ENTREVISTAS	
GÊNERO ENTREVISTA	Número de participantes: deve haver a presença de no mínimo um entrevistador e um entrevistado, porém, em algumas entrevistas, pode conter mais de uma pessoa para cada uma dessas categorias.
	Forma da interação entre entrevistador e entrevistado: a relação entre os interlocutores, ou seja, a condução, a troca de turnos e as interrupções.
	Relação do entrevistado com o tema abordado na entrevista: pode ser sobre sua vida ou um tema de sua área de atuação.
	Enquadramento da fala ao nível de formalidade: entrevistador e entrevistado devem adequar suas falas à situação comunicativa e ao nível de formalidade que ela exige.
	Elementos paralinguísticos e cinésicos: são aspectos como o tom da voz, o ritmo e as pausas na fala. E movimentos corporais, expressões faciais e gestos.

Fonte: Costa (2023, p. 13)

Após serem elencadas as principais características sobre as entrevistas, na subseção que segue, será apresentado o Corpus Roda Viva, composto por entrevistas televisivas.

Corpus Roda Viva

O Corpus Roda Viva⁴ (Miranda *et al.*, 2024) foi criado no âmbito do Projeto Memória Roda Viva, em uma parceria da Fundação Padre Anchieta, Fapesp e Unicamp e apresenta 713 entrevistas do Programa Roda Viva, da TV Cultura, de janeiro de 1986 até julho de 2009. Ele apresenta verbetes, referências, fotos e vídeos. Segundo Miranda *et al.* (2024), esse portal foi alvo de apenas dois trabalhos, Botin (2016) e Pacheco (2020), que tiveram como foco análises linguísticas teóricas. De acordo com Miranda *et al.* (2024), do total de 713 entrevistas do Corpus Roda Viva, apenas 364 delas possuem sua versão em vídeo, sendo que 308 delas (446h 18min 49s) são em português do Brasil e o restante também em outras línguas como inglês, espanhol, português europeu, francês e italiano. As entrevistas são divididas em cinco grandes temas: Ciências, Cultura, Esporte, Economia e Política.

⁴ Acesso por: rodaviva.fapesp.br

Atualmente, o Corpus Roda Viva possui duas versões: a V0.1 apresenta um total de 517.256 sentenças, 9.859.582 *tokens* e 2.633.400 *types*. As entrevistas foram transcritas por jornalistas que fizeram algumas modificações, tornando o *corpus* mais formal e com características de um texto escrito. As inserções textuais no *corpus* são colocadas em colchetes. Alguns exemplos de inserções são:

- A) complementação de palavras omitidas durante a fala: “Eu acho que **[ele]** é o melhor do mundo como chargista»;
- B) direcionamentos de como o fluxo conversacional se desenrola: “Se voce ..., na eleição..., poderia fazer... **[falando junto com o Markun e concordando com ele]**”;
- C) explicação de siglas: PIB **[Produto Interno Bruto]**;
- D) explicação enciclopédica sobre um fato ou alguém: “O Paulinho **[Paulinho da Viola, cantor e compositor]** gravou o quê?”

Na segunda versão do Corpus, a V0.2, foram removidas as intervenções acrescentadas pelos jornalistas no texto original. Essa exclusão representou uma redução de 2,5% de sentenças e 4,6% de *tokens*, segundo apresentam Miranda *et al.* (2024). Ambas as versões estão disponíveis em dois formatos: uma compilação de arquivos CSV (com uma entrevista por arquivo) e um arquivo JSON com todas as entrevistas. O arquivo CSV contém 5 colunas com as informações da data, nome da entrevista, ordem de fala, nome do falante e o texto.⁵

Alguns objetivos da construção desse *corpus* são: ter-se um registro importante da história recente, a preservação das entrevistas, acesso livre a todo conteúdo e retroalimentação das pautas do programa.

Corpus Roda Viva TW

O Corpus Roda Viva TW é um projeto piloto composto por quatro entrevistas extraídas do Corpus Roda Viva (Miranda *et al.*, 2024), totalizando 4.024 sentenças e 69.377 *tokens*. As quatro entrevistas foram escolhidas no intuito de apresentarem uma possível diversidade sintática, pois os entrevistados são pessoas com formações bem diversas, sendo um jogador de futebol, um *rapper*, um escritor de revista em quadrinhos e uma governadora de estado.

Como já foi salientado na subseção anterior, as entrevistas do Corpus Roda Viva foram transcritas manualmente por jornalistas, apresentando, por isso, informações adicionais. Com o objetivo de ser mais fiel à fala, decidiu-se transcrever automaticamente quatro

⁵ Arquivos disponíveis no GitHub: <https://github.com/<ANONYMIZED>/Roda-Viva>

entrevistas (em vistas do curto período de tempo disponível para o trabalho), por meio de um Sistema de Reconhecimento Automático de Fala (ASR) chamado Whisper (Radford *et al.*, 2023), que será apresentado com mais detalhes na próxima subseção.

Após isso, o Corpus Roda Viva TW foi anotado, também automaticamente, com as etiquetas da Universal Dependencies (De Marneffe *et al.*, 2021), por meio do parser PortParser (Lopes *et al.*, 2024) e revisado manualmente com relação à anotação, com o uso da ferramenta Arborator Grew-ElizIA (Guibon *et al.*, 2020).

Nas subseções seguintes, serão apresentadas as ferramentas utilizadas no processamento do Corpus Roda Viva TW, bem como o formalismo da Universal Dependencies.

ASR Whisper

Os Sistemas de Reconhecimento Automático de Fala têm sido muito utilizados nos dias de hoje para realizarem inúmeras tarefas. Com vistas a transcrever de forma automática as quatro entrevistas que formam o Corpus Roda Viva TW, foi utilizado neste trabalho o ASR Whisper (Radford *et al.*, 2023).

De acordo com Radford *et al.* (2023), o Whisper se concentra em ampliar o escopo do pré-treinamento fracamente supervisionado, além do reconhecimento de fala apenas em inglês, sendo tanto multilíngue quanto multitarefa. Das 680.000 horas de áudio do treinamento da ferramenta, 117.000 horas abrangem 96 outras línguas. O conjunto de dados também inclui 125.000 horas de dados de tradução X en.

A utilização de ASR traz algumas considerações éticas, uma vez que esses sistemas costumam gravar e armazenar áudios para treinar modelos e aprimorar-se, o que pode trazer preocupações com relação à invasão de privacidade, retenção excessiva e usos secundários de dado. Para sanar essa questão, é necessário adotar-se consentimento informado, oferecendo clareza sobre como e por que os dados de voz são coletados, por exemplo.

Com relação à variação linguística, observa-se que a cobertura limitada de diferentes sotaques e dialetos no treinamento dos ASR pode frustrar e excluir grupos menos representados. Para que isso não ocorra, os sistemas precisam ser treinados com conjuntos de dados diversificados.

O Whisper sugere que a simples ampliação do pré-treinamento fracamente supervisionado tem sido subestimada até agora para reconhecimento de fala. Os autores alcançaram esses resultados sem a necessidade de técnicas de auto-supervisão ou auto-treinamento, que têm sido uma constante no trabalho recente em reconhecimento

de fala em larga escala, e demonstram como o simples treinamento em um conjunto de dados supervisionado grande e diversificado e o foco na transferência zero-shot podem melhorar significativamente a robustez de um sistema de reconhecimento de fala.

Os vídeos das entrevistas extraídos do YouTube foram submetidos ao Whisper, e a ferramenta gerou um arquivo .txt com o texto transcrito. Esses textos apresentaram alguns problemas, como segmentação errada de frases e alguns erros de transcrição de palavras, principalmente com estrangeirismos e por má dicção dos falantes. Para corrigir isso, todos os textos foram revisados manualmente antes de anotá-los com o parser PortParser (Lopes *et al.*, 2024).

No Quadro 2 são apresentados os números relativos aos erros de identificação de palavras observados em cada uma das entrevistas, bem como um exemplo de erro e seu contexto de ocorrência.

Quadro 2. Números dos erros de identificação de palavras e entrevistas-fonte

Número de erros	Entrevista fonte	Exemplo	Contexto
21	Mano Brown	Afinar da (a finada)	<i>A finada classe média</i>
31	Edmundo	Zazinho (Zezinho)	<i>O seu Zezinho Mansur</i>
14	Maurício de Souza	Passeato (passeata)	<i>a criançada fez passeata em Brasília</i>
14	Benedita de Silva	Disseu (Dirceu)	<i>O Lula, pessoalmente, o Zé Disseu, convenceram a senhora a... a assumir e agora já estaria certa a sua participação no governo?</i>

Fonte: Elaboração própria

Parser PortParser

Após a transcrição automática do Corpus Roda Viva TW e a correção manual dos problemas apresentados pelo ASR Whisper, iniciou-se a etapa de anotação automática do *corpus* com as etiquetas da Universal Dependencies. A ferramenta escolhida para essa tarefa foi o parser PortParser (Lopes *et al.*, 2024).

De acordo com Lopes *et al.* (2024), o Portparser supera os sistemas atuais para textos jornalísticos em português brasileiro. Seguindo o *framework* de Dependências Universais (UD), o modelo foi construído utilizando um *corpus* manualmente anotado recentemente

lançado (Porttinari-base) para treinamento. Os autores testaram diferentes métodos de análise sintática e exploraram configurações de parâmetros com o objetivo de propor um modelo altamente preciso, abrangendo não apenas a anotação de dependências, mas também a marcação de Part-of-Speech, a identificação de lemmas e as características morfológicas relacionadas. O melhor modelo alcançou cerca de 99% de precisão na marcação de Part-of-Speech, lemmas e características morfológicas, com cerca de 95% de precisão na anotação de dependências, superando sistemas conhecidos para o português em até 7% de precisão. Os autores também realizaram uma análise de erros do modelo proposto para mostrar as limitações atuais e os desafios para trabalhos futuros.

O *parser* se mostrou muito capacitado para anotar o *corpus* trabalhado, cometendo apenas alguns erros de anotações com relação a estruturas de fala, como vocativos, marcadores discursivos, objetos diretos antepostos ao verbo, hesitações e truncamentos.

Não foi possível saber quantas correções foram realizadas em cada sentença, pois teria sido necessário não salvar nenhuma alteração nas sentenças em que não foram identificados problemas, o que não foi feito, para que fosse possível se calcular quantas sentenças sofreram alterações e quantas não. Somado a isso, por limitações da ferramenta, também não foi possível se contabilizar quantos *tokens* sofreram uma revisão com relação à anotação.

Arborator Grew-ElizIA

A fim de se revisar a anotação automática do *corpus* realizada pelo *parser* PortParser (Lopes *et al.*, 2024), foi utilizada a ferramenta Arborator-Grew ElizIA (Guibon *et al.*, 2020), que permite visualizar as sentenças anotadas e fazer as correções necessárias.

De acordo com Guibon *et al.* (2020), o Arborator-Grew combina as funcionalidades de duas ferramentas pré-existentes: Arborator e Grew. Arborator é uma ferramenta colaborativa amplamente utilizada para anotação gráfica *online* de árvores de dependência. Grew é uma ferramenta para consulta e reescrita de grafos, especializada em estruturas necessárias em PLN, ou seja, árvores e grafos de dependência sintática e semântica. O Arborator-Grew é um redesenho completo e uma modernização do Arborator, substituindo seu próprio armazenamento interno de banco de dados por uma nova API do Grew, que adiciona uma poderosa ferramenta de consulta às funcionalidades existentes de criação e correção de banco de dados de árvores do Arborator. Isso inclui controle de acesso complexo para anotação paralela por especialistas e crowdsourcing, visualização de comparação de árvores e vários modos de exercício para ensino e treinamento de anotadores. O Arborator-Grew abre novos caminhos para a criação, atualização, manutenção e curadoria coletiva de bancos de dados de árvores sintáticas e bancos de grafos semânticos.

A fim de se realizar a revisão da anotação no Arborator-Grew ElizIA, dois anotadores foram selecionados, sendo um com vasta experiência sobre a anotação Universal Dependencies e outro iniciante, o qual passou por um treinamento prévio antes da tarefa. Os dois anotadores se reuniam regularmente a fim de debaterem sobre casos complexos na anotação. Dessa forma todas as sentenças foram sendo revisadas ao longo do trabalho.

Como forma de se calcular a concordância dos anotadores, foi feito um recorte do *corpus* formado por 200 sentenças aleatórias. O resultado desse cálculo pode ser observado na Tabela 1, em que se nota a porcentagem da concordância entre os anotadores com relação às formas, lemas, etiquetas morfossintáticas, características, núcleos e relações de dependência. Nesses dados, percebe-se que a concordância foi baixa em relação à atribuição dos núcleos e das relações de dependência.

Tabela 1. Porcentagem de concordância entre os anotadores

Tipo de Token	Porcentagem de concordância
FORM (forma)	98,81%
LEMMA (lema)	97,92%
UPOS (etiqueta morfossintática)	95,56%
FEAT (características)	95,24%
HEAD (núcleo)	47,02%
DEPREL (relação de dependência)	55,65%

Fonte: Elaboração própria

Universal Dependencies

Como já foi citado, as entrevistas do Corpus Roda Viva TW foram anotadas com as etiquetas do modelo da Universal Dependencies (UD) (De Marneffe *et al.*, (2021), o qual almeja realizar uma anotação gramatical consistente (etiquetas morfossintáticas, características morfológicas e dependência sintática), entre línguas humanas diferentes. Esse modelo é um esforço colaborativo de cerca de 500 pessoas que produziram quase 200 *treebanks* para aproximadamente 100 línguas.

Neste momento, a UD possui dezessete etiquetas morfossintáticas ou Part-of-Speech (PoS) *tags*, como: NOUN: substantivo, DET: determinante e PRON: pronome.

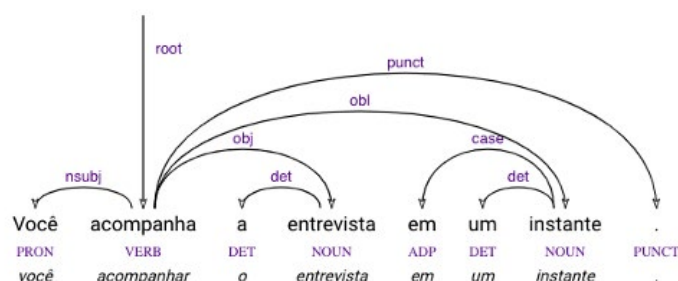
Além disso, a UD também possui 37 etiquetas de relações de dependência – *deprel* (de dependency relation), como PUNCT – pontuação; NSUBJ – sujeito; NMOD – modificador nominal. Uma *deprel* é uma relação que liga dois a dois os elementos (*tokens*) de uma

sentença. Um deles é chamado de *head* (núcleo), que é sempre uma palavra de conteúdo (verbo, substantivo, adjetivo, pronome, numeral e advérbio) e o outro é chamado de dependente.

Normalmente o predicado da oração principal é a raiz de uma sentença, marcada como dependente da *deprel root*. Os arcos das relações não devem se cruzar, pois a atribuição de relações de dependência deve observar o princípio da projetividade.

A Figura 1 apresenta um exemplo de uma sentença anotada com relações de dependência UD.

Figura 1. Exemplo de árvore de dependências anotada com etiquetas da UD



Fonte: Corpus Roda Viva TW

Após a apresentação do arcabouço teórico-metodológico utilizado nesta pesquisa, passa-se, na seção 3 a apresentar-se os resultados do trabalho.

Análise e resultados

Nesta seção serão apresentados alguns resultados de comparação entre trechos de entrevistas do Corpus Roda Viva e do Corpus Roda Viva TW, através dos quais é possível perceber que o primeiro *corpus* não apresenta características de um *corpus* de língua falada, devido à sua transcrição manual. Já o Corpus de Roda Viva TW, transcrito manualmente, mantém essas características.

O Quadro 3 apresenta dois trechos de uma das entrevistas, sendo um deles transcrito manualmente (Corpus Roda Viva) e o outro automaticamente (Corpus Roda Viva TW). Nesses trechos é possível perceber que algumas marcas de oralidade são perdidas na transcrição manual, como a repetição de palavras e a presença de marcadores discursivos, bem como são feitas algumas correções gramaticais, como a inserção de pronomes omitidos na fala.

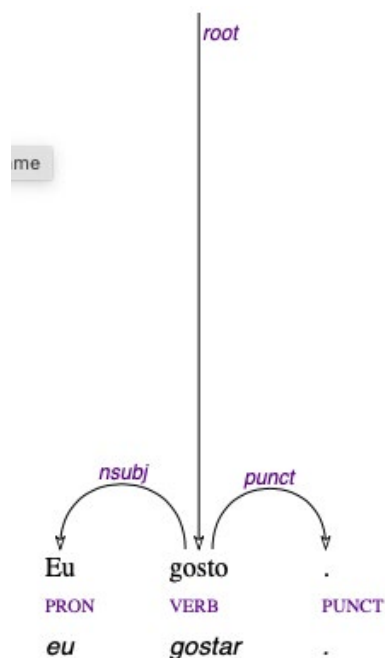
Quadro 3. Diferenças entre a transcrição manual e automática

Transcrição Manual	Transcrição Automática
<p>Edmundo: Eu gosto. Número de camisa, eu nunca tive nenhuma preferência. Coincidentemente, eu marquei pela camisa sete. Fui para o Flamengo, gostaria também de ter mudado quando eu cheguei no Flamengo e lá foi feita uma votação e ganhou por 92% camisa sete. E aqui, o Marcelinho é o dono da camisa sete e gosta de jogar com ela, eu acho que não tinha nem como eu pensar em jogar com a camisa sete. E aí foram escolhidos outros dois números, caiu o número oito. Eu me simpatizo bem com o número oito.</p>	<p>Eu gosto, gosto. Número de camisa, nunca tive nenhuma preferência, né? Coincidentemente, marquei pela camisa 7. Fui para o Flamengo, gostaria também de ter mudado quando cheguei no Flamengo. E lá foi feita uma votação e ganhou por 92% a camisa 7. E aqui o Marcelinho é o dono da camisa 7 e gosta de jogar com a camisa 7. Eu acho que não tinha nem como eu pensar em jogar com a camisa 7. E aí foram escolhidos os outros dois números, caiu o número 8, e eu me simpatizo bem com o número 8.</p>

Fonte: Elaboração própria

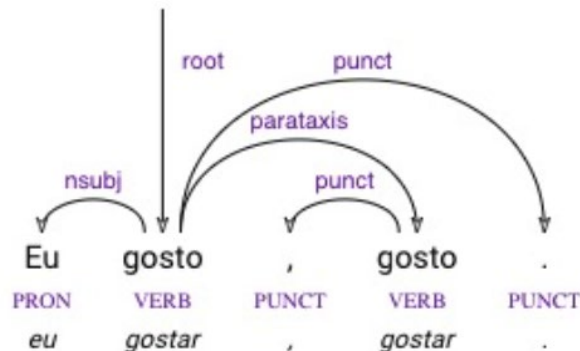
As Figuras 2 e 3 apresentam as diferenças de anotação com as etiquetas da UD de duas sentenças presentes nos trechos apresentados na Tabela 2. Verifica-se que também a anotação de dependências se altera nos dois tipos de transcrição.

Figura 2. Anotação UD de uma sentença transcrita manualmente



Fonte: Corpus Roda Viva

Figura 3. Anotação UD de uma sentença transcrita automaticamente



Fonte: Corpus Roda Viva TW

Por meio desses exemplos, é possível perceber as diferenças que existem entre o *corpus* com as entrevistas transcritas manualmente (Corpus Roda Viva) e o *corpus* com as entrevistas transcritas automaticamente (Corpus Roda Viva TW), pois este último apresenta marcas da oralidade, enquanto o anterior já não as apresenta, motivo pelo qual optou-se pela transcrição automática das entrevistas neste trabalho.

O objetivo final deste trabalho é anotar automaticamente as outras entrevistas do corpus Roda Viva, a fim de se inserir uma porção de *corpus* oral no projeto Porttinari (Pardo *et al.*, 2021), que já possui outros gêneros de textos escritos como *tweets* do mercado financeiro, *reviews* de *e-commerce*, entre outros.

Referências

BAKHTIN, M. *Os gêneros do discurso*. Paulo Bezerra (Organização, Tradução, Posfácio e Notas); Notas da edição russa: Seguei Botcharov. São Paulo: Editora 34, 2016. 164p.

BOTIN, L. M. *Ciência e tecnologia em debate: uma análise das entrevistas do programa Roda Viva, da TV Cultura*. 2016. Tese (Doutorado em Letras), Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, Brasil, 2016.

COSTA, M. B. Uma análise do gênero entrevista: identificação de seus tipos, características e especificidades. *Portal UFERSA*, 2023.

DE MARNEFFE, M. C.; MANNING, C. D.; NIVRE, J.; ZEMAN, D. Universal dependencies. *Computational linguistics*, v. 47, n. 2, p. 255-308, 2021.

FAVERO, L. L. et. al. *Gramática do Português Culto Falado no Brasil: construção do texto falado*. v.1. Campinas: Editora da Unicamp, 2006.

GUIBON, G.; COURTIN, M.; GERDES, K.; GUILLAUME, B. When collaborative treebank curation meets graph grammars: arborator with a grew back-end. *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille, France, European Language Resources Association, p. 5293- 5302, maio 2020.

KOCH, I. V. *Introdução à linguística textual*. 2. ed. São Paulo: Contexto, 2017.

KOCH, I. V. et. al. *Referenciação e discurso*. São Paulo: Contexto, 2005.

LIRA, L. S. *Análise de aspectos orais no gênero entrevista televisiva*. 2020. TCC (Graduação) – Curso de Letras – Língua Portuguesa, Universidade Federal de Alagoas, Arapiraca, 2020.

LOPES, L.; PARDO, T.A.S. Towards Portparser – a highly accurate parsing system for Brazilian Portuguese following the Universal Dependencies framework. *Proceedings of the 16th International Conference on Computational Processing of Portuguese (PROPOR)*, p. 401-410. May. 13-15. Disponível em: <https://aclanthology.org/2024.propor-1.41>. Acesso em: 04 nov. 2024.

MARCUSCHI, L. A. *Análise da Conversação*. 6. ed. São Paulo: Ática, 2007.

MARCUSCHI, L. A. et. al. Fenômenos Intrínsecos da Oralidade: a hesitação. In: MARCUSCHI, L. A. et. al. *Gramática do português falado: construção do texto falado*. Campinas: Editora da Unicamp, 2006.

MARCUSCHI, L. A. et. al. Oralidade e ensino de língua: uma questão pouco “falada”. In: MARCUSCHI, L. A. et. al. *O livro didático de português: múltiplos olhares*. 3. ed. Rio de Janeiro: Lucerna, 2005.

MARCUSCHI, L. A. *Questões atuais na Análise da Conversação*. Recife: ANPOLL, 1988.

MARCUSCHI, L. A. *Análise da conversação*. São Paulo: Ática, 1986.

MEDINA, C. A. *Entrevista: o diálogo possível*. 3. ed. São Paulo: Ática, 1990.

MELO JÚNIOR, J. N. B. *Aspectos textuais e conversacionais na entrevista oral no radiojornalismo alagoano*. 2016. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Programa de pós-graduação profissional em Letras e Linguística, Universidade Federal de Alagoas, Maceió, 2016.

MINAYO, M. C. S. Técnicas de pesquisa: entrevista como técnica privilegiada de comunicação. In: MINAYO, M. C. S. *O desafio do conhecimento: pesquisa qualitativa em saúde*. 12. ed. São Paulo: Hucitec, 2010. p. 261- 297

MIRANDA Jr., I.; PEDRO, G. W.; BARROS, C. D.; VALE, O. A. Roda Viva Boundaries: an overview of an audio-transcription corpus. *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, vol, 2, p. 165-169, mar. 2024.

PACHECO, P. H. *A construção “acontece que” no português brasileiro contemporâneo: uma análise baseada no uso*. 2020. Dissertação (Mestrado em Letras) – Universidade Federal Fluminense, Niterói, Brasil, 2020.

PARDO, T. A. S.; DURAN, M. S.; LOPES, L.; DI FELIPPO, A.; ROMAN, N. T.; NUNES, M. G. V. Porttinari – a large multi-genre treebank for Brazilian Portuguese. In: *Proceedings of the XIII Symposium in Information and Human Language (STIL)*, p. 1-10, 2021.

RADFORD, A; KIM, J. W.; XU, T.; BROCKMAN, G.; MCLEAVEY, C.; SUTSKEVER, I. Robust Speech Recognition via Large-Scale Weak Supervision. *Proceedings of the 40th International Conference on Machine Learning*, PMLR 202, p. 28492-28518, 2023.

URBANO, H. Marcadores conversacionais. In: PRETI, D. (org.). *Análise de textos orais*. 3. ed. São Paulo: Humanitas Publicações – FFLCH/USP, 1997. p. 81-101.

Rejeição ou inovação: uma discussão a partir de características de manuscritos rejeitados

DOI: <http://dx.doi.org/10.21165/el.v54i1.4026>

Janayna Carvalho¹

Resumo

Neste artigo, discuto as características de manuscritos rejeitados a partir da minha experiência como editora-chefe da *Revista de Estudos da Linguagem* (RELIN). Este artigo se desenvolve com base na minha apresentação na mesa redonda “Por que publicar?”, que aconteceu como uma das atividades do 70º seminário do Grupo de Estudos Linguísticos de São Paulo (GEL), realizado na Universidade Estadual de Campinas (Unicamp) entre os dias 16 e 19 de julho de 2024.

Palavras-chave: rejeição de artigos; Revista de Estudos da Linguagem; publicações em revistas.

¹ Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil; janaynacarvalho@gmail.com; <https://orcid.org/0000-0002-2353-1336>

Rejection or innovation: a discussion based on the characteristics of rejected manuscripts

Abstract

In this article, I discuss some characteristics of rejected manuscripts, based on my experience as Editor-in-Chief of *Revista de Estudos da Linguagem* (RELIN). This article is a development of my presentation at the roundtable “Por que publicar?” (“Why publish?”), one of the activities of the 70º seminar of Grupo de Estudos Linguísticos de São Paulo (GEL), which took place at Universidade Estadual de Campinas (UNICAMP) from July 16th to July 19th 2024.

Keywords: paper rejection; *Revista de Estudos da Linguagem*; publications in journals.

Introdução

Neste artigo, discuto as características de manuscritos rejeitados a partir da minha experiência como editora-chefe da *Revista de Estudos da Linguagem* (RELIN). Este artigo se desenvolve com base na minha apresentação na mesa redonda “Por que publicar?”, que aconteceu como uma das atividades do 70º seminário do Grupo de Estudos Linguísticos de São Paulo (GEL), realizado na Universidade Estadual de Campinas (Unicamp) entre os dias 16 e 19 de julho de 2024.

Como minha fala levou em conta a minha experiência como editora da RELIN, devo, primeiramente, apresentar algumas informações da Revista. A RELIN é uma revista da Faculdade de Letras (FALE) da Universidade Federal de Minas Gerais (UFMG), mantida tanto pela Faculdade de Letras quanto pelo Programa de Pós-Graduação em Estudos Linguísticos (POSLIN). Essa Revista está em circulação desde 1992, e é já um periódico consolidado na linguística brasileira, tendo sido avaliada como um periódico A1 na última avaliação do Qualis Periódico divulgado, referente ao quadriênio 2017-2020. O site da Revista é: <https://periodicos.ufmg.br/index.php/relin>.

Os editores da Revista são, na maior parte das vezes, professores credenciados como docentes permanentes no POSLIN e exercem um mandato de editor associado ou chefe de 3 a 4 anos. Sou editora da Revista desde 2021. Primeiramente, fui editora associada entre 2021 e 2022, quando o professor Gustavo Ximenes da Cunha era o editor-chefe. Desde abril de 2023, estou na função de editora-chefe. Na configuração atual, a Revista conta com duas editoras associadas, as professoras Maria Cantoni, também da FALE-UFMG, e a professora Ana Regina Vaz Calindro, professora da Faculdade de Letras da UFRJ.

Levando em conta essa experiência, os pontos que levanto sobre a rejeição em periódicos estão, então, ligados a uma impressão qualitativa que fui formando do processo de revisão cega por pares, um processo em que os pareceristas e os autores não têm sua identidade revelada, o que presumivelmente auxilia uma avaliação idônea e imparcial.

Todavia, os pontos que levanto são, em alguma medida, comuns a todos os periódicos, inclusive aqueles que não adotam a revisão cega, como a Revista da Abralin. Isso é algo que pude atestar levando em conta as conversas com os colegas editores que participaram comigo da mesa “Por que publicar?”.²

Dentre todos os temas envolvidos na editoração de uma revista, o tema da rejeição se mostra particularmente importante por algumas razões. Primeiramente, esse é um dos temas que permeiam nosso contato com os autores e acaba nortendo muito do nosso trabalho na revista. Afinal, um texto que foi rejeitado não passa por todos os estágios de um texto aprovado. Seu processo editorial termina logo após chegar à Revista, no caso de textos rejeitados na etapa de verificação inicial do manuscrito, ou após a avaliação dos pareceristas, não passando para as etapas finais do processo editorial, quais sejam, revisão, diagramação e publicação.

Além disso, todos os autores, quando submetem um texto a uma Revista, esperam que ele seja publicado. Esse desejo vem tanto do fato de que os autores se esforçaram, em algum grau, para entregar um manuscrito coerente, organizado e com uma boa análise para o periódico, quanto pelo fato de que muitos fatores da nossa carreira são regulados por publicações. Quem publica mais pontua mais em concursos, em editais de agências de fomento, pode se credenciar ou se manter como orientador de pós-graduação... Então, obviamente, a rejeição de um texto impacta em todas essas esferas.

Ao falar de rejeições, as tendências que reporto aqui não identificam submissões ou autores, tampouco envolvem porcentagens de qualquer forma. Portanto, exponho os principais fatores que levam um texto à rejeição falando de tendências sem usar qualquer dado factual que as ampare. Essa escolha tem duas razões. A primeira é porque seria antiético de minha parte, já que o processo de avaliação é todo anônimo. A segunda é porque seria bastante custoso fazer qualquer levantamento quantitativo suficientemente representativo de motivos de rejeição em uma revista que recebe textos em fluxo contínuo. Portanto, o que reporto aqui é uma impressão que foi se formando com a leitura que faço uma vez por semana dos pareceres que chegaram à Revista.

Eu poderia, obviamente, comparar essas minhas impressões com textos e manuais sobre as principais razões que levam à rejeição de um texto. Não o faço porque essa

2 A professora Raquel Freitag (UFS/Revista da Abralin) e o professor Marcelo Módolo (USP/Revista do GEL) também participaram da mesa e este artigo incorpora alguns pontos do diálogo frutífero que temos.

comparação não vai nos mostrar alguma verdade oculta. Em todos os casos, são pessoas falando de suas experiências. E na maioria desses casos, o ponto de vista dos pareceristas é o que normalmente aparece nas discussões sobre por que um artigo é recusado. Como esse texto se configura basicamente como um relato de experiência, ele não apresenta referências, já que não me pautei em trabalhos publicados ou em informações de *websites* para fazer as considerações que eu faço aqui. Então, o leitor tem em mãos um artigo que não segue o formato tradicional de artigo ao discutir a rejeição de artigos!

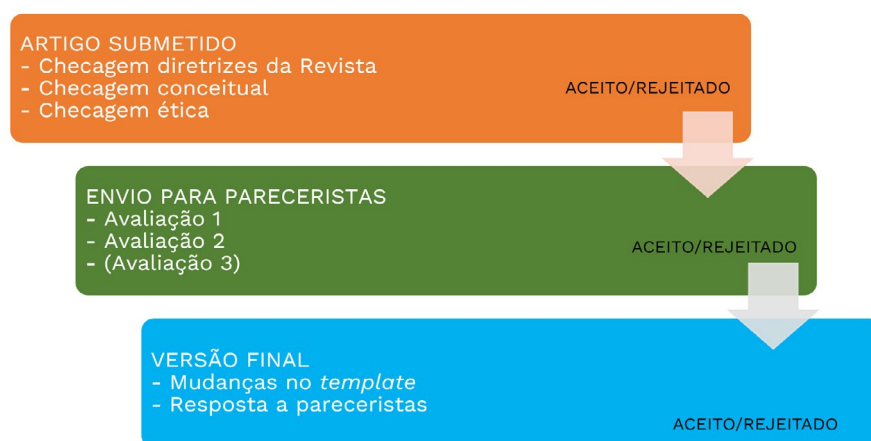
Para discutir esse tema, este artigo está organizado da seguinte forma. Na seção *Como as revistas científicas funcionam*, detalho o processo editorial de um texto, novamente, com base no processo da RELIN. Na seção *Casos de rejeição*, discuto características principais dos textos rejeitados, dividindo-os em dois principais: o chamado *desk rejection*, que é a rejeição pela Comissão Editorial, e a rejeição pelos pareceristas. A seção *Conclusão* sumariza a discussão da rejeição, apresentando algumas reflexões sobre o tema e finaliza o manuscrito.

Como as revistas científicas funcionam

Na maioria das revistas, há possibilidade de aceite ou rejeição de um manuscrito em pelo menos duas etapas: a etapa de submissão inicial e a etapa de apreciação do manuscrito pelos pareceristas.

Especificamente na RELIN, o processo editorial é dividido em três etapas, conforme a figura 1. Em todos eles, há a possibilidade de rejeição do manuscrito.

Figura 1. Fluxograma das submissões



Fonte: Elaboração própria

Detalhemos, aqui, cada uma dessas etapas. A primeira, nomeada como *Artigo submetido*, corresponde à checagem inicial do artigo, que é feita após sua submissão na plataforma. Nessa checagem, observamos se o artigo obedece às diretrizes da Revista, se a discussão feita no artigo está dentro do escopo da Revista e procede-se, finalmente, a uma checagem ética do artigo, com o uso de ferramentas antiplágio.

Estando o artigo adequado nesses três pontos, ele passa para a área de avaliação da Revista e é encaminhado a pareceristas que têm a incumbência de avaliar somente o seu conteúdo e o desenvolvimento da argumentação. Como de praxe também em outras revistas, o artigo é enviado, inicialmente, a dois pareceristas. Caso os dois pareceristas deem parecer positivo, o artigo será aceito; caso os dois deem um parecer negativo ao artigo, ele será rejeitado. Se os pareceristas discordarem na sua avaliação, o texto é enviado a um terceiro parecerista, que, com sua avaliação, subsidiará a Comissão Editorial a decidir pelo aceite ou recusa do manuscrito.

Após o recebimento de todas as avaliações, o artigo passa para a próxima etapa caso seja aceito por dois pareceristas. Nessa etapa final, intitulada *versão final* na figura 1, pede-se ao autor uma versão final do artigo que incorpore ou discuta as sugestões dos pareceristas e na qual haja mudanças para o *template* definitivo do texto. Também nessa etapa, o texto pode ser aceito ou rejeitado.

Assim, as três etapas podem envolver rejeição ou aceite. Todavia, a rejeição está praticamente restrita às duas primeiras etapas do processo editorial: *artigo submetido* e *envio para pareceristas*, e é sobre essas etapas que iremos discorrer nas próximas sessões.

Casos de rejeição

Desk rejection

A chamada *desk rejection* corresponde a uma rejeição na primeira etapa ilustrada na Figura 1, ou seja, uma rejeição na etapa de *artigo submetido*, antes que ele seja enviado para pareceristas. Ao contrário da impressão de muitos autores, que atribuem à avaliação por pares o poder de aceitar ou recusar um artigo, é nessa primeira etapa que a maioria dos artigos ficam retidos. Como essa é uma etapa de verificações dos artigos em três grandes eixos, as razões para rejeição podem ser diversas.

A primeira delas é a de que um texto recusado nessa etapa pode ter sido apresentado com uma organização diferente do *template* da Revista. Nessa etapa, não fazemos uma verificação exaustiva do *template*. Somente se um problema de macroestrutura do *template* fica muito aparente, o texto não é considerado.

Na RELIN, os textos devem ter entre 6000 e 15000 palavras para serem considerados para avaliação. Um texto de 3000 palavras, por exemplo, será recusado nessa etapa por estar muito aquém do limite mínimo de palavras. Mais importante do que o limite de palavras em si é o formato da discussão. Um texto de 3000 a 4000 palavras será muito breve e não trará uma discussão teórica aprofundada. Em termos de espaçamento, os textos devem ter espaçamento simples e margens 3 cm (superior e esquerda) x 2 cm (inferior e direita) em papel A4. Assim, um texto escrito em papel A5 com margens de 2 cm e espaçamento duplo também será desconsiderado nessa etapa.

Note que os textos desconsiderados nessa etapa fogem bastante do esperado em termos de macroestrutura. Curiosamente, a rejeição nessa etapa está majoritariamente ligada à origem do texto. É raro que textos submetidos por pesquisadores brasileiros sejam recusados nessa primeira verificação. É mais comum que textos de pesquisadores estrangeiros sejam desconsiderados nessa etapa. Os textos rejeitados nessa etapa são, possivelmente, textos reaproveitados (ou seja, que não foram aprovados em algum periódico com regras de macroestrutura muito diferentes da RELIN) e que foram submetidos à RELIN do modo como estavam. Isso fica muito claro quando os elementos de macroestrutura são considerados em conjunto. Isto é, não é comum que somente o número mínimo ou máximo de palavras tenha sido desrespeitado. Além da questão do limite de palavras, o artigo terá, normalmente, uma estrutura diferente, como, por exemplo, resumos em três línguas, algo que a RELIN não pede.

Ainda na etapa de *artigo submetido*, também é feita uma checagem conceitual, que se refere a uma rápida checagem temática do texto. Como qualquer outra Revista, a RELIN tem um escopo de interesse, que é sintetizado em sua página da seguinte forma: “[A RELIN] tem como missão fomentar a produção científica na área de Teoria e Análise Linguística, permitindo aos pesquisadores do Brasil e do exterior divulgarem suas pesquisas e contribuírem para o debate e o progresso científico na área”. Portanto, somente trabalhos de linguística teórica e descritiva são considerados. São arquivados trabalhos de linguística aplicada, entendendo-se aqui trabalhos de aplicação de conceitos teóricos ao ensino de língua materna ou estrangeira, sem discussão teórica dos conceitos teóricos que propiciaram a aplicação do estudo ao ensino, ou trabalhos de interface da linguística com outras áreas em que não há proposição de cunho linguístico de qualquer forma. A falta de aderência ao escopo da Revista leva alguns textos a também serem rejeitados nessa etapa.

Por fim, ainda na etapa de verificação inicial do artigo, há uma verificação ética do manuscrito, que pode ser composta por duas etapas, a depender da natureza do manuscrito. Primeiramente, fazemos uma checagem em todos os manuscritos com um *software* antiplágio. O *software* usado captura sequências de palavras na internet que são idênticas ou muito similares às encontradas no artigo. Ele também captura traduções de trechos, possibilitando que traduções daquele texto também sejam examinadas.

É necessário, nessa etapa, um exame qualitativo dos resultados do antiplágio. O *software* usado não conseguirá distinguir uma citação de um trecho efetivamente usado de outro artigo e cabe aos editores examinarem as ocorrências e determinar se o trabalho possui um grau elevado de originalidade ou não.

Enquanto é consensual em nossa área que o plágio significa se apropriar das ideias de outros autores e/ou de seus textos sem dar o devido crédito, não é claro, para uma parcela significativa de autores, que nos apropriar sem dar o crédito de nossas ideias prévias já publicadas e, mais notadamente, dos textos que escrevemos com elas é também um tipo de plágio. Como sabemos, os textos não são escritos a partir do nada, todos nós usamos ideias, argumentos e mesmo estruturas prévias dos nossos trabalhos nos novos, mas saber como usar e referenciar coerentemente essas formas prévias parece ainda uma lacuna de formação em nossa área, que terá de ser suprida.

É comum que autores usem parágrafos e até páginas, sem nenhuma alteração, de publicações anteriores suas. Quando isso é constatado, é pedido que os autores reformulem os manuscritos, no caso de semelhanças localizadas, ou a submissão é arquivada, quando há tantas semelhanças que demandariam uma reestruturação do manuscrito. Quero deixar registrado que eu não acho absolutamente que esses sejam casos de má-fé. Eles são representativos, na verdade, de uma lacuna que há em nossa formação.

Ainda visando a uma checagem ética do texto, também pedimos a aprovação de um comitê de ética ou órgão equivalente caso a pesquisa envolva seres humanos.

Cabem também algumas considerações sobre a verificação ética. De uma forma geral, os autores são bem conscientes da necessidade de aprovação de um comitê de ética de pesquisa quando as pesquisas envolvem seres humanos. Essa não é uma exigência de todas as revistas, mas é possivelmente uma exigência da maioria das instituições, de forma que o pedido de aprovação da pesquisa por algum órgão competente nunca causou estranhamento em nenhum autor.

Há, nesse ponto, casos particulares que devem ser considerados, como o fato de que, em muitas universidades, pesquisas com participantes anônimos feitas na internet não necessitam de uma anuência formal. Há ainda o fato de que, dependendo do país em que a pesquisa é feita, não há necessidade de nenhuma anuência de um órgão de pesquisa. Todavia, de forma geral, há clareza, da parte dos pesquisadores, de quando essa anuência é necessária e de quando não é, o que mostra que nossa formação, como um campo de pesquisa, nessa área tem sido efetiva.

Em suma, as possibilidades diversas de *desk rejection* iluminam muito dos aspectos cruciais que um autor deve levar em conta ao submeter um manuscrito, além de alguns deles também nos mostrarem características da nossa área.

Primeiramente, é imperioso que os autores levem em conta as particularidades do periódico a que submetam. Assim, o material que o autor deseja submeter deve ser examinado de acordo com as exigências da Revista. Nesse sentido, o autor deve se certificar de que o material segue as especificações do que é pedido no *site* da Revista. As especificações de Normas dos Autores (ou variações desse nome) devem ser atendidas tanto quanto possível no manuscrito submetido. Também, o autor tem de se certificar que seu estudo está dentro do escopo de atuação da Revista, para evitar que um texto bem construído seja rejeitado logo após a submissão. Do ponto de vista ético, o autor deve se certificar de que possui todas as aprovações necessárias para o tipo de estudo que visa retratar. Também é necessário que o autor se certifique de que esse trabalho tem autonomia de pensamento e de estrutura em relação aos seus trabalhos prévios.

Tendo esses fatores sido observados, é difícil que haja *desk rejection*. Assim, os envolvidos no processo de publicação podem se concentrar no que é evidentemente importante para todos nós: o mérito da argumentação feita no manuscrito. Na próxima seção, passo a falar sobre as tendências dos pareceres no que diz respeito à rejeição do texto.

Rejeição pelos pareceristas

Uma vez estabelecido, na etapa de verificação inicial do manuscrito, que o texto cumpre minimamente com os requisitos formais, éticos e de conteúdo, ele segue para a apreciação dos pareceristas. Nessa etapa, ocorre a apreciação detalhada do conteúdo do texto. Alguns problemas com esse conteúdo podem levar à rejeição do manuscrito como discutiremos nesta seção. Todavia, problemas conceituais e de análise não são o único caso que levam à rejeição de um manuscrito. Como o próprio título deste texto esclarece, a rejeição de um artigo por parte de um parecerista pode se dever, às vezes, ao fato de ele apresentar alguma inovação em relação ao que é normal na área. Portanto, além de problemas de conteúdo, um texto pode ser rejeitado pela diferença temática ou estrutural que ele apresenta em relação ao que o parecerista considera ser um artigo passível de análise.

Para começar, discuto os casos mais “tradicionais”, por assim dizer, porque eles são mais ilustrativos da revisão por pares. No fim da subseção, discuto rapidamente os casos de rejeições ligadas à inovação.

Focando nos casos tradicionais, a maioria das rejeições tem a ver com um motivo principal: o texto não apresenta uma reflexão (teórica ou empírica) sobre o tema tratado.

Em relação a textos com problemas de reflexão teórica, é muito comum que os pareceristas resumam as considerações que os levaram a rejeitar um texto como “falta de originalidade”. No entanto, uma rápida verificação nas considerações de grande parte dos pareceristas vai nos mostrar que a questão não é exatamente originalidade, mas sim falta de alguma reflexão.

Por exemplo, a maioria dos textos submetidos à RELIN (ou a qualquer outra revista) são aplicações de quadros teóricos a fenômenos. Algumas vezes, essa aplicação, do modo como é feita, já constitui uma novidade: pode ser que os pesquisadores anteriores desse fenômeno nunca pensaram nessa aplicação do quadro teórico para esse tipo de fenômeno. Se um autor, então, aplica um quadro teórico discursivo para análises de morfemas, por exemplo, é difícil que haja a consideração de “falta de originalidade” pelos pareceristas. Isso não significa, naturalmente, que o texto será aceito. Significa tão somente que provavelmente outras questões serão mais salientes para os pareceristas em sua avaliação do que a originalidade.

Essa situação hipotética contrasta com uma situação em que o autor discute, em seu manuscrito, um fenômeno muito conhecido e analisado em determinada corrente teórica usando ferramentas de outra corrente teórica. Essa mudança de paradigma será claramente notada por um parecerista especialista no assunto e deve estar muito bem contextualizada no texto. Do contrário, a análise será lida por alguns pareceristas como simplesmente taxonômica. É em casos como esse que pode aparecer a tão frequente expressão “falta de originalidade”. O problema, nesse caso, não é exatamente a originalidade, mas é uma aplicação sem bastante discussão de sua razão ou pertinência quando já há uma aplicação teórica para o fenômeno bem consolidada no campo.

Também são muito notados pelos pareceristas os casos em que os autores não discutem as limitações muito aparentes da análise e esse é um outro caso crucial em que há uma rejeição alta de manuscritos. Nesses casos, a rejeição parece estar ligada a uma falta de reflexão sobre a análise e não inerentemente aos problemas dela.

Em suma, muitas das rejeições aos textos são justificadas com base na falta de originalidade do manuscrito, mas, na verdade, a rejeição se pauta em uma falta de reflexão do autor sobre a abordagem que fez do fenômeno. Isso fica patente em casos em que o manuscrito lida com um tema muito investigado, mas não discute, com alguma profundidade, a razão pela qual um fenômeno com uma análise estabelecida está sendo tratado em outra abordagem. Há, obviamente, muitas outras razões para a rejeição de um manuscrito, mas as brevemente tratadas aqui aparecem continuamente e merecem nossa atenção porque revelam uma tendência nossa, como área, na avaliação de manuscritos.


Finalmente, abordo o outro caso, que aparece no título deste artigo (e da conferência que deu origem a ele): a rejeição por inovação.

Como os casos clássicos de rejeição nos mostram, há uma conduta implícita em nosso trabalho como pareceristas em artigos que seguem os formatos tradicionais, isto é, em textos que se detêm na análise de um problema linguístico. Todavia, há outras discussões possíveis em artigos, como as revisões de literatura e os artigos que descrevem passos metodológicos na construção de um banco de dados, por exemplo. Embora artigos desse tipo já sejam bem consolidados em outras áreas do conhecimento, eles não o são na linguística e isso fica claro nas avaliações pelos pares.

Pode haver em textos desse tipo uma grande discordância entre os pareceristas. Isto é, como artigos de revisão sistemática da literatura ou artigos sobre a construção de bancos de dados não são a grande parte dos artigos submetidos, alguns pareceres refletem isso. É necessária, também, uma diferenciação desses casos em relação aos casos anteriores, em que a rejeição está ligada a um julgamento de falta de reflexão do(a) autor(a) pelos pareceristas. Naqueles casos, a rejeição por ambos os pareceristas é muito mais frequente do que nos casos sobre os quais discorro agora, em que há inovação de discussão nos artigos. Nos casos que ora comentamos, é mais comum o desacordo entre pareceristas e podemos atribuir a isso à inovação temática em relação à maioria dos artigos submetidos.

Não obstante, a questão da originalidade aparece em muitas avaliações desses artigos com temáticas “inovadoras”. É importante que pensemos como vamos medir a originalidade em uma revisão sistemática da literatura. Se assumirmos, por exemplo, que o texto não é original porque a revisão é acrítica, temos, parece-me, um critério razoável para a objeção de publicação a um texto. Se a queixa residir sobre o próprio mérito da revisão de literatura, esse não parece um critério válido, porque o parecerista espera desse texto que ele tenha os mesmos objetivos de um texto analítico em relação a um fenômeno linguístico, o que não será o caso.

É interessante ressaltar nesse ponto que essa discordância em textos com discussões metodológicas ou com revisões sistemáticas e a crítica à originalidade podem ser usadas como um termômetro para entender o que a disciplina de Linguística está acostumada a valorizar em artigos científicos. No caso das novas temáticas em artigos, nós, como um campo de estudo, devemos fazer critérios que nos nortearão na avaliação de artigos como esses. No caso dos artigos teóricos e empíricos sem uma reflexão aprofundada, já há um parâmetro consolidado, mesmo que ele permaneça implícito e às vezes um tanto quanto opaco por trás de sintagmas como “falta de originalidade”.



De qualquer forma, mudanças no perfil das publicações devem levar a mudanças nas concepções dos pareceristas sobre o que é um estudo original e quais são nossos parâmetros para averiguar isso.

Conclusão

Este artigo analisou brevemente casos de rejeição mais comuns em periódicos, tomando a RELIN como representativa desses casos. Dividi os casos em dois subtipos. Em um estágio de verificação inicial do manuscrito, podem ocorrer as rejeições na etapa inicial de verificação do manuscrito, ligadas a uma série de fatores, desde desrespeito à macroestrutura do artigo até questões éticas. No estágio de avaliação do artigo por pareceristas, foram mais exploradas as rejeições usuais, em que não há uma reflexão no texto sobre a abordagem usada, com destaque para dois casos: (i) já há, na literatura, uma análise muito estabelecida para o fenômeno; (ii) o texto não discute problemas aparentes de análise. Finalmente, falamos dos casos em que há rejeição do manuscrito por haver inovação de formato e temática, em relação à maioria dos manuscritos.

De uma forma geral, os três casos deixam claras algumas das nossas características como área de estudo. Saliento algumas delas. Em relação à rejeição na etapa de verificação inicial do manuscrito, há uma falta de formação explícita sobre os limites entre uma publicação e outra do mesmo autor e sobre os mecanismos de aproveitamento dos nossos trabalhos prévios. Há uma clareza para os autores sobre o problema do plágio de outros autores, mas não há sobre as formas éticas de apropriação da própria produção. Em relação à etapa de pareceres, a reflexão (mais do que a originalidade) por si só é o fator que mais conta para os pareceristas. Aplicar uma teoria nova para um grupo de dados que era usado geralmente em outra teoria não é, para muitos pareceristas, um problema. O problema é a falta de discussão dessa mudança paradigmática e suas questões associadas.

Por fim, há ainda questões para a área de linguística que não estão respondidas e as dúvidas ficam refletidas na avaliação: qual é o papel dos metaestudos? Estudos de experimentos que não deram certo, estudos de descrição de compilação de dados, estudos de revisão sistemática? Como devemos avaliá-los? O que é originalidade nesse caso?

Espero que esse breve relato da minha experiência (que também destoa da discussão que normalmente vemos em artigos científicos!) possa oportunizar discussões pertinentes na nossa área e que os acordos que, implicitamente, fizemos sobre elas apareçam nos pareceres nas próximas décadas.



Agradecimentos

Agradeço à Diretoria atual do Grupos de Estudos Linguísticos de São Paulo, na pessoa de Lívia Oushiro, pelo convite para participação na mesa redonda “Por que publicar?”. Agradeço ainda aos colegas da mesa, Raquel Freitag e Marcelo Módolo, pelo debate proveitoso durante todo o evento.

Investigar a interação no autismo a partir de uma perspectiva linguístico-interacional

DOI: <http://dx.doi.org/10.21165/el.v54i1.4028>

Fernanda Miranda da Cruz¹

Resumo

Este artigo discute os subsídios de uma perspectiva linguístico-interacional êmica para a investigação da interação e da linguagem-em-interação no autismo. Primeiramente são apresentados os pressupostos de uma perspectiva êmica e do método interpretativo dos participantes conforme o campo de estudos da Análise da Conversa Etnometodológica. Em um segundo momento, são analisadas duas situações interativas envolvendo crianças autistas e adultos não-autistas, com ênfase na forma como os participantes lidam interacionalmente com produções verbais de crianças autistas que não correspondem a um repertório linguístico convencionalizado. As análises oferecem um enfoque sobre práticas de reparo. Por fim, discute-se a manutenção intersubjetiva de uma interação quando um dos participantes têm limitações na produção verbal.

Palavras-chave: autismo; intersubjetividade; perspectiva êmica; modelo social da deficiência.

¹ Universidade Federal de São Paulo (Unifesp), Guarulhos, São Paulo, Brasil; fernanda.miranda.cruz@gmail.com; <https://orcid.org/0000-0002-5051-7759>

Investigating interaction in autism from a linguistic-interactional approach

Abstract


This article discusses the contributions of an emic linguistic-interactional perspective to the investigation of interaction and language-in-interaction in autism. First, the assumptions of an emic perspective and the interpretive method of the participants are presented according to the field of study of Ethnomethodological Conversation Analysis. Secondly, two interactive situations involving autistic children and non-autistic adults are analyzed, with an emphasis on how the participants deal with verbal productions interactionally that do not correspond to a conventionalized linguistic repertoire. The analyses focus on repair practices. Finally, the intersubjective maintenance of an interaction is discussed when one of the participants has limitations in verbal production.

Keywords: autism; intersubjectivity; emic perspective; social model of disability.

Introdução: autismo, linguagem e interação social

A antropóloga Elizabeth Fein, em seu artigo “Autismo como modo de engajamento” (Fein, 2019), formula a seguinte pergunta “o que eu estudo quando estudo autismo”? Esta pergunta será um ponto de partida. Neste artigo vamos desenvolver uma reflexão sobre as contribuições de uma perspectiva linguística-interacional para a compreensão das interações envolvendo pessoas autistas. Esta reflexão será construída em torno da análise de duas situações interativas envolvendo crianças autistas e adultos não-autistas. O estatuto dessas interações neste artigo é mais ilustrativo da discussão a ser feita do que um estatuto de uma coleção sistemática de dados empíricos de um determinado fenômeno analítico a ser aprofundado. Dessas duas interações participam crianças autistas que fazem uso limitado da fala em suas trocas interacionais e das quais as produções verbais não correspondem a um repertório linguístico convencionalizado, tais como *tsau pam* (exemplo 01); *nai; i di tê* (exemplo 02). Em termos de caracterizações das alterações e dos déficits linguístico-cognitivos no autismo comparados ao (neuro) desenvolvimento típico da linguagem, produções como essas, associadas ao uso de um repertório limitado de palavras ou frases, podem indicar perfis linguísticos de crianças minimamente verbais ou que não fazem uso da expressão verbal (Tager-Flusberg; Kasari, 2013; Rose *et al.*, 2016; Posar, 2021). Esse tem sido um objeto de investigação na literatura clínico-linguística, incluindo, vale dizer, a própria problematização da categorização minimamente verbal (Tager-Flusberg; Kasari, 2013; para uma revisão da literatura sobre o tema, veja ainda Posar, 2021).

Os pressupostos metodológico-analíticos dos estudos neurolinguísticos de orientação discursiva produzidos por Coudry (1996), Morato (2002), Coudry e Novaes-Pinto (2023)



têm mostrado desde muito que a investigação da linguagem em contextos de alteração linguístico-cognitivo não deve estar apartada da reflexão sobre a produção do que é um dado linguístico. Alinhado a esta reflexão, as análises trazidas neste artigo mostram uma forma de olhar para produções não-convencionalizadas que emergem em interações reais e nos informam sobre experiências com linguagem que tanto os sujeitos autistas quanto seus interlocutores vivenciam quando estão juntos em uma interação. O enfoque descritivo-analítico está sobre o modo pelo qual os participantes se orientam ou não para tais produções, sobre a forma como os participantes se orientam um para o outro durante uma interação e como se orientam para a própria dinâmica da interação em situações em que fala não está disponível para pelo menos um dos participantes.

Um cuidado metodológico-analítico é como olhar para essas produções sem perder de vista os desafios e, principalmente, os processos que experienciam as crianças autistas e seus interlocutores (neste caso, adultos não-autistas) para construírem a intercompreensão mútua e a participação no mundo por meio da fala e tudo afeito a ela. Os estudos linguísticos de distintas abordagens têm contribuído para a compreensão da linguagem, da interação e da cognição no autismo de muitas formas: investigando a natureza das alterações linguístico-cognitivas; descrevendo aspectos fonológicos, morfológicos, sintáticos da linguagem; pensando sobre os processos típicos e atípicos de aquisição e desenvolvimento da linguagem; identificando e descrevendo aspectos pragmáticos e aspectos sociointeracionais; atuando no aprimoramento de instrumentos investigativos e avaliativos. Diante dessa multiplicidade de perspectivas e compromissos investigativos, um movimento inicial que faremos é o de situar o autismo na multiplicidade de possibilidades de investigação da linguagem a ele relacionadas e destacar o que certas escolhas teórico-metodológicas e analíticas deixam ver sobre a linguagem, sobre a interação, mas também sobre o autismo.

O entendimento sobre como conceber e investigar o autismo é um debate que está vigente (Rios; Fein, 2019). O autismo pode ser entendido e investigado como categoria clínica, que organiza e é organizada por práticas situadas como investigação científica clínica, diagnóstico, caracterização, intervenções clínico-terapêuticas, por exemplo. Do ponto de vista clínico, o autismo é referido como Transtorno de Espectro do Autismo (American Psychiatric Association, 2013) e é uma condição biológica do neurodesenvolvimento que se expressa de forma heterogênea nos indivíduos dentro de um espectro. Clinicamente, o autismo é entendido, nos termos do Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM), em sua 5ª edição (APA, 2014), como um comprometimento da interação social que apresenta padrões repetitivos e restritivos de interesses e comportamentos; déficits de processamento sensorial e déficits na linguagem e em outras habilidades de comunicação. Em diálogo e em consonância com os esforços de identificação dos déficits e preditores de autismo, de elaboração de testes de avaliação e rastreamento, da construção de subsídios para as intervenções, as pesquisas em linguística clínica têm se orientado para a construção de uma agenda de investigação que passa pelo

entendimento e pela descrição da natureza dos déficits no autismo (Tager-Flusberg, Paul, Lord, 2005; Brynskov *et al.*, 2016; Naigles, 2017).

No entanto, nos últimos anos, as formas de pensar e conceber o autismo em sua relação com a linguagem e com a interação social também compõem outras agendas de investigação para além da agenda clínica, como, por exemplo, nos estudos antropológicos (Ochs *et al.*, 2004; Rios; Fein, 2019) e nos estudos linguístico-interacionais (Ochs, 2010; Sterponi *et al.*, 2015; Maynard; Turowetz, 2022).

Para irmos um pouco mais longe, a linguagem tem importado de várias formas na sua relação com o autismo. Por exemplo, como se referir ao autismo e às pessoas autistas é um dos aspectos que convoca a investigador a estar atento às suas práticas e a suas posições éticas com relação aos participantes ou sujeitos de suas pesquisas. Um bom exemplo disso está nas políticas de publicação de resultados de pesquisas sobre autismo sensíveis ao debate sobre como se referir a pessoas autistas e as implicações de cada escolha. O periódico científico *Autism, The International Journal of Research and Practices* (Sage Publications) elaborou um guia terminológico destinado aos autores de artigos a serem submetidos no periódico sobre o uso de uma linguagem que respeite as pessoas autistas, suas famílias e cuidadores. Como advertem, esse guia terminológico não é prescritivo, mas ele foi elaborado para incentivar os pesquisadores de várias áreas do conhecimento a se envolverem profundamente com uma literatura mais ampla sobre o autismo a fim de entenderem as opiniões e perspectivas das pessoas autistas, de serem sensíveis e respeitarem suas preferências na linguagem usada para descrevê-las e de, então, tomarem decisões informadas que garantam que seus textos estejam fundamentados em um profundo respeito com pessoas autistas e próximos.

Investigar o autismo e, mais especificamente, a linguagem e a interação social no autismo, nos inscreve também no debate das formas de se conceber as deficiências e as diferenças dentro de modelos ou de perspectivas. No modelo social da deficiência, por exemplo, o autismo tem sido considerado uma diferença e uma deficiência *versus* um modelo biomédico da deficiência. Enquanto o modelo biomédico concebe a deficiência como atributo do indivíduo, o modelo social da deficiência, em uma direção distinta, a concebe como contextualmente situada e como uma condição que se encontra na intersecção do indivíduo, do ambiente e da sociedade. Nessa última perspectiva, a deficiência então não está reduzida a uma questão de déficit ou de alterações de um indivíduo clinicamente diagnosticado a partir de critérios clínico-diagnósticos padronizados, mas é entendida como um fenômeno relacional e contextualizado (Yu; Sterponi, 2022), em sua relação com o ambiente social que, dentre outros, também produz a deficiência via barreiras atitudinais, ambientais, institucionais e sociais que impedem, por exemplo, a participação, a autonomia, a qualidade de vida, os acessos das pessoas. O modelo social da deficiência coloca em saliência os efeitos da relação entre as características e as necessidades (físicas, cognitivas ou emocionais) de um determinado indivíduo e as características

de seu contexto social (den Houting, 2019; Oliver 1996; Hughes, 2016). Esse modelo nos convida a pensar o autismo para além de uma categoria ou condição clínica, o que significa também pensá-lo e investigá-lo para além de uma condição médica (den Houting, 2019), orientando-se, todavia, para a defesa dos direitos à participação plena dos sujeitos autistas na sociedade.

Participação plena envolve também a participação interacional e o que não significa uma participação interacional normativa. Do ponto de vista político e acadêmico-científico, o desenvolvimento da perspectiva dos *disability studies* (Mello; Aydos; Schuch, 2022) e de perspectivas como a da neurodiversidade (Oliver, 1996; Hughes, 2016; Singer, 2017) têm sugerido pensar as deficiências e o autismo como uma categoria social que se intersecciona ainda com outras categorias sociais, como classe, gênero, etnia, por exemplo. Inclusive, é como efeito dessa perspectiva que a forma de se referir a pessoas autistas importa, uma vez que refletem concepções e posições sobre autismo, sobre diferença e sobre estigma (Gernsbacher 2017; Kenny, 2016; Sinclair, 2013; Vivanti, 2020).

Tais abordagens têm fundamentado questões interessantes que passam a ser aspectos (ou variáveis) a serem considerados na investigação de condições como a do autismo. Por exemplo, uma das implicações do movimento político-ativista e acadêmico da neurodiversidade está na proposta de se pensar na natureza situada, socioculturalmente, de uma condição neurológica ou ainda nos fatores contextuais que acomodariam diferenças neurocognitivas com efeitos sobre a forma de conceber as intervenções clínico-terapêuticas; de informar e dialogar com pesquisas científicas (Nicolaidis, 2012; Pellicano; Stears, 2011; Bottema-Beutel *et al.*, 2021) e, finalmente, o ponto que nos interessa, de conceber padrões de comunicação e de interação social (den Houting, 2019; Ochs, 2010; Magnani; Rückert, 2021).

Podemos finalmente retomar a pergunta que abre esta introdução, “o que eu estudo quando estudo autismo?”. Estamos estudando muitas coisas quando estudamos o autismo e uma delas é a própria sociabilidade humana (ou sociabilidades humanas, Ochs, 2010) e as formas possíveis de interação com o outro pela linguagem.

Perspectiva êmica, lógica interna das interações cotidianas e método interpretativo dos participantes

O enfoque sobre a interação da qual participam sujeitos autistas e outros interlocutores têm recebido a atenção de pesquisas sacionteracionais, tais como Ochs e Solomon (2010), Sterponi e Shankey (2014), Korkikangas (2014), Maynard e Turowetz (2022), para citar alguns. Esses estudos dialogam com uma tradição de análises com foco na descrição da organização social das ações por meio da fala-em-interação a partir da perspectiva dos próprios membros da interação, chamada de perspectiva êmica (Sacks, 1992; Garcez; Loder, 2005).

Considerando-se que, na fala-em-interação, sobretudo na conversa cotidiana, os participantes estão continuamente demonstrando uns para os outros seus entendimentos quanto a o que estão fazendo, que por sua vez se tornam disponíveis também para observadores externos (no caso, os analistas), o que interessa para a análise são esses entendimentos revelados, demonstrados, pelos próprios participantes, uns para os outros, na sua interação (Garcez; Loder, 2005, p. 282).

A partir desta perspectiva, a organização da fala-em-interação é co-produzida localmente pelos próprios interagentes para construírem suas ações sociais. Para as interações das quais participam pessoas com condições específicas (como demências, afasias, autismo etc.) a interação em si mesma, como acontecimento, em seu caráter situado e sempre inédito, socioculturalmente inscrito, nos permite ver como tais interações se constroem (acontecem ou não) em termos das ações e de estruturas organizacionais locais daquela interação.

Quando nos atentamos para a forma como as interações envolvendo pessoas autistas se organizam, o fenômeno que nos interessa está menos em um sintoma ou caracterização de uma condição de um indivíduo e mais no que é intersubjetivamente construído com o outro, através do exame/análise constante dos próprios membros daquela interação. Se uma interação ocorreu, o fato mesmo dela ter ocorrido nos interessa, como apontou Schegloff (1993). Sabemos que a manutenção intersubjetiva de uma interação com o outro pode ser desafiante e mesmo não alcançada quando há limitações de comunicação de um dos interlocutores (Eilertsen, 2014). No entanto, não podemos atribuir os desafios ou impossibilidades de uma manutenção intersubjetiva exclusivamente a uma instância de competências de um indivíduo, sob os riscos de ignorarmos formas menos convencionais de interação e de perdermos inteligibilidade sobre as maneiras possíveis de construirmos ou sustentarmos intersubjetivamente uma relação com outro. Uma das principais implicações de uma perspectiva êmica da interação está sobre o trabalho do analista, que tenta discernir a lógica interacional da interação (Geoffrey *et al.*, 2013) que, em outros termos, é a forma como os participantes interpretam as ações uns dos outros e demonstram sua interpretação, localmente, na dinâmica interacional.

Análise: a orientação dos participantes para a dinâmica da interação

As duas interações abaixo referem-se a interações adulto-criança. As crianças são autistas e os adultos não-autistas. Em particular, as duas crianças autistas fazem uso limitado da linguagem verbal expressiva, em termos de palavras e frases, ou em termos de estruturas morfológico-sintáticas. Essas interações foram registradas em vídeo, em ambientes naturalísticos de ocorrência. Esse é um dos procedimentos metodológicos do campo dos estudos sociointeracionais orientados pela Análise da Conversa etnometodológica. Como procedimento analítico, essas interações foram transcritas de

forma a dar conta das ações corporificadas e verbais que a constituem. A convenção de transcrição utilizada foi o modelo de transcrição multimodal proposto por Mondada (2014), disponível *on-line*, que visa dar conta da temporalidade e da sequencialidade de ações corporificadas e verbais no curso de uma ação.

As duas situações foram escolhidas de forma a construir apontamentos de contribuições de uma linguística sociointeracionalmente orientada para o entendimento das interações das quais participam pessoas autistas e não autistas a partir de uma perspectiva êmica com ênfase sobre os métodos interpretativos empregados pelos próprios participantes para exibir e construir a organização local da interação. Nas duas interações, como já dito, temos produções verbais que não correspondem a um repertório linguístico convencionalizado tais como *tsau pam* (exemplo 01); *nai; i di tê* (exemplo 02). Sabemos que os desafios linguísticos e interacionais de crianças que não fazem uso da fala nem de sistemas comunicativos alternativos têm implicações significativas em sua participação no mundo. Mas, quando estão em interação com o outro, como essas interações acontecem?

O excerto (1) refere-se a uma interação entre duas irmãs, Clara (CLA) e Luiza (LUI). Luiza tem nove anos de idade, é autista, expressa uma produção verbal e vocabular restrita se comparadas com padrões típicos de crianças de sua idade e participa das interações cotidianas em família e nos ambientes institucionais que frequenta, como escola, associação de convivência, igreja e outros (Cots, 2018). Neste primeiro excerto, temos uma interação entre Luiza e sua irmã Clara, adulta, de 24 anos, não-autista. O excerto foi extraído de um *corpus* audiovisual de interações entre Luiza e seus familiares com quem mora (pais, irmã, avó, sobrinho). Essa é uma interação entre duas pessoas que tem um grau de familiaridade importante, pois são irmãs, compartilham rotinas, experiências e repertórios. No excerto selecionado, Luiza e Clara estão se preparando para assistir a um filme que irá passar na televisão. Luiza está sentada em um sofá comendo pipoca, enquanto sua irmã Clara se prepara para se sentar ao seu lado. Ao se sentar, Clara derruba, com a ponta de uma coberta, uma pipoca que Luiza havia deixado no canto do sofá. Clara não percebe que seu movimento com o cobertor havia derrubado a pipoca. Luiza então irá contar à Clara o que aconteceu, como vemos no excerto transcrito, formulando seu turno "*tsau (0.2) pam hihhi*" (linha 16).

Excerto (1)

01 * (4.0) *
 lui * olha p/ canto do sofá *

02 * (3.0) *
 lui *Loca canto do sofá *

03 **LUI** ***tsau** *
 lui *gesto de mão lançar *

04 * @ (0.2)
 lui *olha CLA
 cla @olha LUI

05 * pam * hihihi
 lui * gesto de mão bater *

06 **CLA** **você @derrubou no chão?**
 cla @ olha p/chão--->

07 (0.6) *
 ---> *

08 **LUI** h hhh
 09 (0.7)
 10 **CLA** não fui eu, hein?

11 **LUI** hihihi
 12 (.)
 13 **CLA** não fui eu, hein?

14 (4.6)
 15 **CLA** dá aqui, não, eu jogo fora
 16 **LUI** * **PAM** você *
 lui *gesto de bater a mão*

Nesta interação é Luiza é quem inicia uma sequência interacional narrativa de contar o que houve. Os turnos de fala de Luiza são estruturados pela produção verbal *tsau* e *pam* e por ações corporificadas, como direcionamento de olhar e gestos de mão. A produção *tsau*, co-produzida com o gesto de mão de lançar algo no ar, refere-se, temporalmente, ao momento em que Clara senta-se e balança o cobertor, e *pam*, co-produzida com o gesto de mão de bater, refere-se ao momento em que pipoca é derrubada por Clara (pelo movimento de Clara). Do ponto de vista da construção do turno de fala, temos aqui uma construção multimodal de turno (Deppermann, 2013; Mondada, 2016), em que o turno de Luiza constrói-se pela produção verbal *tsau pam* coordenada temporalmente à produção sistemática de gestos de mão (lançar a mão no ar e bater a mão não no ar), como vemos nas linhas 03, 05 e 16 da transcrição multimodal. Assim, embora *tsau pam* não sejam produções lexicais reconhecíveis da língua portuguesa, o valor referencial dessas produções é indexicalizado e os participantes orientam-se para essa indexicalidade ou esta ecologia multimodal da interação (Goodwin, 2010) e para uma órbita microecológica do próprio ambiente interacional (Goffman, 2002 [1964], p. 15) em que tais produções emergem.

Após o turno de Luiza “*tsau pam*”, Clara produz um reparo, que é também um pedido de confirmação “*você derrubou no chão?*”. Para o que nos interessa analisar, fazer uma breve apresentação das práticas de reparo, já bastante descritas na literatura linguística, é pertinente. As práticas de reparo “são recorrentemente exploradas e gerenciadas a cada momento pelos próprios participantes na construção de suas interações” (Garcez; Lorder, 2015, p. 283) e deixam ver como os participantes de uma interação lidam com problemas de produção, audição/percepção, entendimento/compreensão e potencial discordância (Schegloff, 1992).

Em seu formato, o reparo iniciado por Clara demonstra que ele não tem dúvidas sobre o tópico introduzido por Luiza, ou seja, Luiza refere-se a um acontecimento, à pipoca derrubada. Mas Clara formula seu reparo sobre uma informação específica: quem derrubou a pipoca. O reparo iniciado por Clara não incide sobre a produção “*tsau pam*” de Luiza, mas sobre a informação de quem derrubara a pipoca. Das linhas 05 a 16, a interação se desenrola em um enquadre de brincadeira compartilhada sobre a divergência de quem teria derrubado a pipoca. Clara rindo produz então “*não fui eu, hein*” (linha 10). Clara e Luiza estão visivelmente engajadas em uma orientação mútua e compartilhada. Ambas brincam com a situação narrada por Luiza da queda da pipoca. O riso compartilhado é um componente relevante nesta situação, uma vez que exhibe uma afiliação intersubjetiva entre elas. No entanto, Luiza iniciará um reparo em um momento específico da sequencialidade desta interação: após o turno de Clara “*não fui eu, hein*”. Luiza inicia e realiza um reparo no formato “*pam você*” (linha 16). Luiza mobiliza novamente sua produção verbal *pam*. Seu turno (linha 16) então é construído pela produção *pam*, que já havia sido introduzida na linha 05, seguido de *você* que nesta sentença é o agente/sujeito de *pam* (bater a coberta). O uso sistemático de *pam* indica uma coerência interna e circunstanciada desta produção. Ainda que ela não tenha um valor referencial lexical correspondente na língua portuguesa, essa produção tem um valor indexical e sistemático nesta interação. A estruturação sintática desse turno em que o reparo é realizado é analiticamente relevante à medida em que Luiza reformula seu próprio turno anterior inserindo um sujeito “*pam você*”. As práticas de reparo são práticas altamente sofisticadas, uma vez que

Para realizar um reparo, o iniciador do reparo precisa identificar o que rejeitar no turno de origem do problema, além de saber como projetar um reparo relevante (Schegloff *et al.*, 1977). O processo envolve dois estágios de compreensão: “Um falante de alguma conversa anterior deve entender algum *next turn after it* – algum turno reconhecidamente ‘responsivo’ – bastante o suficiente para perceber, em primeiro lugar, que ele se baseia em um mal-entendido daquilo a que se destina a ser responsivo e, em segundo lugar, qual é esse mal-entendido (Schegloff, 1992, p. 1331, tradução própria²).

Esse exemplo (1), nos indica como a) as produções linguísticas não convencionadas ou atípicas de crianças autistas podem ser locais de visibilidade de operações linguísticas (sintáticas) sobre um substrato de fala; b) o estatuto de um enunciado como significativo é atribuído pelos participantes na interação à medida que é com tal enunciado que há o início de uma conversa (iniciativa de fala espontânea); a partir do qual se operam práticas

2 No original: “in order to conduct a repair, the repair initiator needs to identify what to reject in the trouble-source turn, as well as knowing how to design a relevant repair (Schegloff *et al.*, 1977). The process involves two stages of understanding: “A speaker of some prior talk must understand some next turn after it—some recognizably ‘responsive’ turn—well enough to appreciate first, that it is based on a misunderstanding of that to which it is meant to be responsive, and, second, what that misunderstanding is”.

de reparo, operações sintáticas, manutenção de um piso conversacional; c) ocorrências de reparos ao turno de um falante adulto não-autista podem ser iniciados por crianças autistas pouco-verbais de forma espontânea.

Sobre este último item c, embora possa parecer haver uma correlação direta entre essas produções menos convencionais de crianças autistas e práticas de reparos iniciados por adultos para pedir esclarecimentos ou mostrar não-entendimento do que quis dizer a criança, uma perspectiva êmica do que aconteceu nesta interação específica nos indica que as participantes nem sempre consideram problemática (objeto de reparo) uma produção não-convencionalizada ou atípica (*tsau pam*). Em segundo exemplo (2), gostaríamos de tratar mais especificamente das práticas de reparo nestas interações. Vamos ver outra dupla de adulto não-autista e criança autista. Nesse caso os reparos iniciados pela participante adulta não-autista, Amália, emergem como formas de entendimento sobre o que significa “*nai*”, produzido em um turno de fala iniciado por Daniel, uma criança autista.

Amália (AMA), uma educadora não-autista e Daniel (DAN), um educando autista de nove anos, estão sentados lado a lado e conversam. A interação se passa em uma instituição de convivência de crianças e adolescentes autistas. O excerto foi extraído de um *corpus* audiovisual de registro de rotinas interativas dos integrantes desta instituição (educadores não autistas e educandos autistas). De forma semelhante ao excerto 01, há tanto produções verbais atípicas da criança autista quanto a emergência de práticas de reparo que incidem sobre o significado dessas produções.

Da linha 01 a 10, Amália e Daniel conversam e a conversa organiza-se em uma dinâmica de alternância de turnos (Amália-Daniel-Amália) e, neste caso, ambos conversam por meio de produções verbais não-convencionalizadas (*ui, nai, i di tê*).

Excerto (2)

```

DAN      *ui:@
ama      @olha p/ Dan-->
dan      *olha p/ Ama-->
          (.)
AMA      @ui      @ui ((rindo))
ama      -->@olha p/frente@ olha p/dan--> (linha 20)
DAN      hh hOhã:
          (1.0)* (0.4)
dan      ---->*olha p/ baixo -->
AMA      uiui
DAN      na:*i
dan      -->*olha Ama
AMA      hã*?
ama      -->*olha p/ baixo-->
DAN      na:*i
          -->*olha p/Ama
          (0.4)
AMA      nai o quê*?
dan      ---->*olha p/ baixo-->
DAN      i di@ tê
ama      -->@olha p/baixo-->
dan      -->olha p/ frente-->
AMA      ni@ke?
ama      -->@olha p/ dan--> (linha 34)
          (0.9)
AMA      o que? o tênis?
          (0.3)
AMA      que que é* então?
dan      ---->* olha p/ câmara-->
          (1.2)
AMA      hu:m?
          (1.2)* @ (0.7)
dan      ---->*olha p/ lado-->>
ama      -->@olha p/ lado-->>

```

Daniel, na linha 11, produz *nai*, que é objeto de uma iniciação de reparo por Amélia. A iniciação de reparo de Amélia torna inteligível para Daniel que Amélia demonstra um problema (no sentido interacional) com seu turno precedente. Até a linha 11, as produções não-convencionalizadas não foram problemáticas. A partir da linha 11, *nai* será problemático para os participantes. Amélia produz um conjunto de reparos à medida que as produções verbais de Daniel (*nai*, *i di tê*) são produzidas.

Essa troca intersubjetiva entre eles nos convida, como sugere Ochs (2012), a olharmos menos o caráter referencial da linguagem para vermos suas propriedades indexicais e enunciados em si como modos de experimentar o mundo (Ochs, 2012). A prática de reparo neste local específico da interação e não em todos os momentos em que há uma produção não-convencionalizada valida a produção de turno de Daniel como revestida de significado, ainda que não alcançável, naquele momento, por Amélia.

A prática de reparo não diz apenas sobre a fonte do problema (do reparo), mas também sobre o estatuto socio interacional e enunciativo validado pelos participantes do turno anterior e o reconhecimento de seu interlocutor como um falante. Do ponto de vista da sequencialidade desta interação, temos uma estrutura de reparos em três posições:

na primeira posição da sequência de reparo, Amália produz uma iniciação de reparo de classe aberta, sem especificação sobre o que está sendo o objeto de reparo (*hã*), linha 13. Na segunda posição, Daniel se orienta para o reparo iniciado por Amália e realiza uma repetição da produção "*nai*" como resposta ao reparo iniciado por Amália. A responsividade é um aspecto importante da manutenção intersubjetiva de uma interação. Repetir o que já foi dito é uma das trajetórias possíveis das práticas de reparo já descritas na literatura linguística (vide Garcez, 2005), sobretudo quando a fonte do reparo pode ser atribuída a um problema de audição/do que foi ouvido/percebido. Neste caso, temos uma trajetória de reparo iniciada por Amália, mas levada a cabo por Daniel (falante da fonte de problema). Ou seja, Amanda publiciza um problema, para os propósitos locais daquela interação, com o turno anterior de Daniel e é Daniel (o próprio falante da fonte de problema) que se orienta para o problema no turno seguinte ao da iniciação (Garcez; Lorder, 2005) e oferece uma solução (reparo).

A produção repetida de *nai* nesta posição sequencial de reparo realizado tem duas especificidades, a) nos convida a olhar a sistematicidade interna com que essas produções podem acontecer na interação, como já indicado no exemplo (1), b) e nos convida a olhar quais ações e práticas, neste local sequencial, a repetição desta produção projeta como próximo turno ou próxima ação daquele que iniciou o reparo. No que diz respeito a este último item (b), a produção "*nai*" projeta uma segunda iniciação de reparo feita por Amália. No entanto Amália, desta vez, especifica seu objeto de reparo "*nai* o quê?". Na resolução do reparo, Daniel produz um novo enunciado "*i di tê*" (linha 20).

Embora o entendimento mútuo entre os dois sobre o que quer dizer Daniel não tenha sido alcançado, Amália continua tentando compreender o que Daniel quer dizer e ela o faz através das práticas de reparo, em um primeiro momento, mas em um segundo momento, através da oferta de significados potenciais (advinhas) sobre o que seria *nai*. As sequências de ofertas de significados potenciais e as repostas de recusas ou aceites a essas ofertas estruturam-se como sequências de *hint-and-guess* (Laakso; Klipp, 1999), identificadas, por exemplo, em interações entre pessoas afásicas e não-afásicas diante de produções verbais semelhantes. Assim, Amália se reorienta para as produções de Daniel, cujo significado não foi resolvido/esclarecido por meio das trajetórias de reparo, por meio de colaborações de significados possíveis. A trajetória interacional das práticas de reparo construídas por Amália para oferecer um significado para *nai* é visível na própria interação. Amália se orienta para os direcionamentos de olhar de Daniel combinados com a pista sonora de *nai*. Na linha 09, Daniel olha para o chão ou para o tênis. E a produção sonora *nai* tem semelhança com *nike*, a marca de tênis, para o qual está a direção de olhar de Daniel. Conforme Goffman (2002/1964, p. 15), interessado na ecologia material, espacial e corporal das interações, o "indivíduo gesticula usando seu ambiente imediato, não apenas seu corpo. Portanto, devemos apresentar o ambiente de forma sistemática". Neste excerto 02, temos um bom exemplo do ambiente imediato (chão, tênis) apresentado de forma sistemática pela perspectiva dos participantes.

Do ponto de vista da manutenção de um piso comunicacional ou de uma relação intersubjetiva, Amália continua procurando, de forma compartilhada, sentido para as produções verbais de Daniel. Daniel, por sua vez, continua responsivo a essas buscas. O trabalho é recíproco. O trabalho interacional de Amália em indexicalizar as produções de Daniel a pistas do ambiente imediato (como o tênis Nike) e iniciar reparos e o trabalho interacional de Daniel em realizar os reparos, recusar as hipóteses ou interpretações de Amália para o que ele quer dizer mantém uma conexão entre eles sustentada por uma organização sistemática de trocas de turnos (alternância de falantes e sequências de perguntas e respostas). Neste caso, o entendimento mútuo entre Daniel e Amália não é alcançado, mas ambos sustentam um engajamento recíproco um com outro. Os ambientes interacionais que compõem esta arquitetura dos mecanismos de turnos de fala são locais frutíferos em que as expectativas e os métodos interpretativos dos participantes vêm à tona e se tornam mais claramente acessíveis para análise (Geoffrey *et al.*, 2013). As práticas de reparo em suas distintas trajetórias, as ofertas de significados potenciais e a orientação para os movimentos corporais dão visibilidade a uma lógica analítica ou método interpretativo e demonstram uma orientação dos participantes para as produções não-convencionalizadas, para o outro, para a própria dinâmica da interação e para a linguagem.

Sistematização das análises

As análises dos excertos (1) e (2) acima destacaram a dimensão indexical de produções verbais como *tsau pam*; *nai*; *i di tê*. Nas duas interações, exploramos uma relação entre tais produções e práticas de reparo que emergem. Podemos imaginar que as práticas de reparo sejam mais produtivas em contextos que requerem uma verificação de um entendimento mútuo do que o outro diz ou quer dizer. Os dois exemplos analisados confirmam, em certa medida, tal produtividade e confirmam ainda achados da literatura em que as situações de reparo, sobretudo em momentos de quebras ou falhas de comunicação, seriam experienciadas mais frequentemente por crianças autistas do que por crianças não-autistas, com o desenvolvimento típico (Flippin, 2022, p. 519, vide também Alexander *et al.*, 1997; Keen, 2005). Mas vimos também, a partir de uma perspectiva analítica êmica da interação que as produções potencialmente problemáticas por não corresponderem a um vocabulário ou léxico da língua não foram necessariamente tratadas como problemáticas pelos participantes (exemplo 01 e exemplo 02, linhas 01-10). Vimos ainda como tais produções foram incorporadas na construção do turno do interlocutor não-autista (*nai* o quê?, Amália, adulto não-autista, excerto 02) e estruturaram o trabalho de buscas da participante adulta não-autista por pistas no ambiente que pudessem indicar seu significado (*nike*) (exemplo 02). Em 01, a produção de Luiza “*tsau pam*” não foi problemática, do ponto de vista interacional, para os participantes naquela interação. Embora valha destacar que isso não significa dizer que Luiza não encontre limites de comunicabilidade em momentos do seu cotidiano. Mas o que nos interessou ver é que o reparo iniciado por Clara incidiu sobre *quem* derrubou a pipoca, estruturado como um pedido de confirmação “*você derrubou no chão?*” e não

como um reparo de classe aberta, como o que Amália produz em (2) “hã?”, após o turno de Daniel “*na*”. Em 01, o reparo iniciado por Clara é completado por Luiza. Luiza modifica seu turno de forma adequadamente ajustada à fonte do problema apontada por Clara, ou seja, Luiza reestrutura seu turno inserindo o sujeito/agente de *pam*. Ao fazer isso, Luiza opera sintaticamente para oferecer uma solução para a fonte do problema “*pam você*”.

Os dois exemplos (1) e (2) são profícuos para discutirmos a manutenção intersubjetiva de uma interação quando um dos participantes tem limitações na produção verbal e o outro participante tem limites de alcançar o significado dessas produções. Essa não é apenas uma questão de investigação linguística aplicada ao autismo, mas é também um tópico expresso por distintos parceiros comunicacionais (o outro) com os quais crianças autistas, por exemplo, interagem e, mais profundamente, compartilham suas experiências com a linguagem e com a interação. Educadores quando recebem crianças autistas que não fazem uso da fala ou de sistemas aumentativos e alternativos de comunicação estão preocupados. Igualmente pais de crianças autistas e todos aqueles que estão em relação com formas outras de interação.

Estudos clínico-linguísticos também têm se interessado também pelas ocorrências de reparos em interações envolvendo crianças autistas. Orientados para o aprimoramento da avaliação e da intervenção de linguagem, esses estudos articulam as práticas de reparo com as habilidades necessárias e também ausentes reveladas por esses mecanismos em casos de desenvolvimento atípico da linguagem (Alexander *et al.*, 1997; Flippin, 2022). As trajetórias de reparo, destacadas nesta análise, têm implicações variadas para a progressividade da interação, mas também para as quebras que podem acontecer do próprio engajamento da criança autista na interação. A literatura clínico-linguística tem demonstrado os efeitos dessas trajetórias. Em casos em que o que a criança diz não é compreendido, seus interlocutores, por exemplo, os pais, podem responder diferentemente, como por exemplo, fazendo pedidos de esclarecimento (reparos), mudando de tópico ou não ignorando a produção da criança (Flippin, 2022).

Discussão: manutenção intersubjetiva em interações envolvendo crianças autistas

A partir de uma perspectiva êmica, as práticas de reparo, sobretudo as que emergem em contextos de produções verbais atípicas como as que vimos nos dados, fornecem “uma arquitetura operacional, organizada interacionalmente e gerenciada localmente, através da qual os participantes mantêm o entendimento intersubjetivo (Schegloff, 1992)” (Fox, Benjamin, Mazeland, 2012, p.1094³, tradução nossa).


3 No original: “[...] a locally managed, interactionally organized, procedural architecture through which participants preserve intersubjective understanding (Schegloff, 1992)”.

Produções verbais, a princípio, sem um significado correspondente na língua e produzidas por crianças com diagnóstico de autismo e com desenvolvimento linguístico atípico, quando comparado ao desenvolvimento neurotípico, são expressões de alguns perfis linguísticos de crianças autistas (Tager-Flusberg, Kasari, 2013; Rose *et al.*, 2016; Posar, 2021). Os dois exemplos acima instigam nossa reflexão sempre vigente na Linguística sobre produzir sentidos. Essa é uma dimensão com a qual lidam os participantes das duas interações que vimos. Como apontado no início deste artigo, uma das principais implicações de uma perspectiva êmica da interação está sobre o trabalho do analista de reconhecer como os participantes interpretam as ações uns dos outros, ao invés de atribuir sentido ou não sentido a algo e de buscar padrões normativos externos.

A pergunta sobre o que estudamos quando estudamos autismo volta agora após a análise de duas interações que, a uma primeira vista, indicariam mais desconexões do que conexões. Mas o que seria necessário para as conexões humanas em uma dada interação? Interações entre pessoas, mesmo quando a fala não está disponível, acontecem. O *como* acontecem merece ser sempre investigado. A passagem abaixo, um pouco longa, tem ecos com as questões apresentadas sobre o autismo neste artigo, com a escolha dos dados que trouxemos e com a perspectiva interacional êmica adotada para analisá-las:

[...] nunca podemos ter certeza absoluta de que um coparticipante entende a nós e nossas ações. Em vez disso, “até segunda ordem” (Schutz & Luckmann 1979), temos de prever que esse é o caso: “Ao projetar minha pergunta, prevejo que o outro entenderá minha ação (por exemplo, enunciar uma frase interrogativa) como uma pergunta e que esse entendimento o induzirá a agir de tal forma que eu possa entender seu comportamento como uma resposta adequada” (Schutz 1953: 17). Mas mesmo nesses casos bem-sucedidos, considerando nossos históricos individuais, personalidades e perspectivas situacionais, não podemos esperar que os entendimentos dos interlocutores em uma situação sejam exclusivamente compartilhados, mas sim que seus entendimentos sejam compartilhados em um grau que seja “suficiente para todos os propósitos práticos” (Schutz 1953: 8; Garfinkel 1967), ou seja, **o que é prático para a comunicação e a conexão humana nessa situação** (Sorjonen *et al.*, 2021, p. 7, grifo próprio, tradução própria⁴).

4 No original: “[...] we can never be totally sure that a co-participant understands us and our actions. Instead, “until further notice” (Schutz & Luckmann 1979), we have to anticipate that this is the case: “In projecting my question I anticipate that the other will understand my action (for instance my uttering an interrogative sentence) as a question and that this understanding will induce him to act in such a way that I may understand his behavior as an adequate response” (Schutz 1953: 17). But even in such successful cases, given our individual backgrounds, personalities and situational perspectives, we cannot expect that the interlocutors’ understandings in a situation are exclusively shared, but rather that their understandings are shared to a degree that is “sufficient for all practical purposes” (Schutz 1953: 8; Garfinkel 1967), i.e. what is practical for communication and human connection in that situation”.



Uma importante contribuição da Linguística para a investigação do autismo, dentre muitas, está em sua tradição empírica. Uma documentação de práticas interacionais pode diminuir a distância entre uma descrição geral da linguagem e da interação no autismo e a heterogeneidade de formas pelas quais linguagem e interação no autismo se apresentam nos indivíduos. A variabilidade e a diversidade no autismo têm sido um tópico nas várias áreas de pesquisa sobre o autismo. Uma documentação de práticas situadas envolvendo pessoas autistas pode também colocar em questão a ideia de uma forma única de interação e comunicação. Yu e Sterponi (2022) consideram que a ênfase na especificidade e na organização interna da interação é um forte contraste com a maioria dos métodos de avaliação que se baseiam na comparação do desempenho da comunicação social de um indivíduo com algum conjunto de padrões gerais, sejam eles normas estatísticas, uma sequência de desenvolvimento ou classificações de adequação atribuídas ao longo de uma escala. Como apontam as autoras, a partir de uma perspectiva êmica, o valor de um determinado dado está no que podemos descobrir sobre seu próprio funcionamento interno, e não em como isso é comparado a outra coisa (Yu; Sterponi, 2022).

E por fim, uma perspectiva êmica da interação converge com perspectivas como a da neurodiversidade e a dos modelos sociais da deficiência à medida em que oferece uma perspectiva relacional (Yu; Sterponi, 2022) da comunicação autista e da socialidade. Relacional pois não se encerra no indivíduo, como se fosse possível desinscrê-los do mundo social. Sem deixar de reconhecer a importância de se atentar para especificidades linguístico-interacionais no autismo e para a importância de intervenções, ajustes e suportes adequados com vistas a diminuir as barreiras e tornar a participação social mais acessível, uma perspectiva êmica aponta na direção de uma maior compreensão (heurística) das diversas formas de expressão e relações. Este é um ponto de encontro entre as perspectivas não reducionistas do autismo e o campo investigativo da linguística, que tem, desde há muito, em sua agenda empírica e descritiva, um compromisso com uma maior compreensão da forma como participamos no e construímos o mundo e os sentidos do mundo por meio da linguagem e de nossas experiências na linguagem.

Agradecimentos

A pesquisa sobre interação e autismo foi realizada no quadro do projeto *Corporalidade, materialidade e sensorialidade na organização das interações humanas: contribuições para a investigação de interações típicas e atípicas*, financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 23/16810-3, e do projeto *Multimodalidade, corpo e materialidade na organização de interações envolvendo crianças autistas* (2022-2025), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Bolsa de Produtividade em Pesquisa, processo 311299/2022-9.

Referências

- ALEXANDER, D.; WETHERBY, A.; PRIZANT, B. The emergence of repair strategies in infants and toddlers. *Seminars in Speech and Language*, v. 18, n. 3, p. 197-212, 1997.
- AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and statistical manual of mental disorders*. 5. ed. [S. l.]: American Psychiatric Association, 2013. Disponível em: <https://doi.org/10.1176/appi.books.9780890425596>. Acesso em: 07 jul. 2025.
- AMERICAN PSYCHIATRIC ASSOCIATION. *Manual Diagnóstico e Estatístico de Transtornos Mentais: DSM-5*. 5. edição. Porto Alegre: Artmed, 2014.
- BOTTEMA-BEUTEL, K. K.; KAPP, S. K.; LESTER, J. N.; SASSON, N. J.; HAND, B. N. Avoiding ableist language: Suggestions for autism researchers. *Autism in Adulthood*, 2021.
- BRYNSKOV, C.; EIGSTI, I. M.; JØRGENSEN, M.; LEMCKE, S.; BOHN, O. S.; KRØJGAARD, P. Syntax and Morphology in Danish-Speaking Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, v. 47, n. 2, p. 373-383, fev. 2017. DOI: 10.1007/s10803-016-2962-7. PMID: 27844246.
- COTS, C. P. *Recursos interacionais multimodais mobilizados por uma criança com Transtorno do Espectro Autista (TEA) em brincadeiras familiares*. 2018. Dissertação (Mestrado em Letras) – Escola de Filosofia, Letras e Ciências Humanas, Universidade Federal de São Paulo, São Paulo, 2018.
- COUDRY, M. H. O que é dado em Neurolinguística? In: CASTRO, M. F. P. (org.). *O método e o dado no estudo da linguagem*. Campinas: Editora da UNICAMP, 1996. p. 179-194.
- COUDRY, M. I.; NOVAES, R. Neurolinguística e Linguística. *Cadernos de Estudos Linguísticos*, v. 65, e023010, 2023. DOI: 10.20396/cel.v65i00.8673480.
- COUDRY, M. I. Neurolinguística Discursiva: afasia como tradução (Discursive Neurolinguistics: aphasia as translation). *Estudos da Língua(gem)*, v. 6, n. 7, 2008. DOI: 10.22481/el.v6i2.1065.
- DEPPERMAN, A. (ed.). Conversation analytic studies of multimodal interaction. *Journal of Pragmatics*, v. 46, n. 1, p. 1-172, 2013.

DOUGLAS, M.; TUROWETZ, J. *Autistic Intelligence: Interaction, Individuality, and the Challenges of Diagnosis*. Chicago: The University of Chicago Press, 2022.

EILERTSEN, L. J. Maintaining Intersubjectivity when Communication Is Challenging: Hearing Impairment and Complex Needs. *Research on Language and Social Interaction*, v. 47, n. 4, p. 353-379, 2014. DOI: 10.1080/08351813.2014.958278.

FEIN, E. Autismo como um modo de engajamento. In: RIOS, C.; FEIN, E. *Autismo em tradução: uma conversa intercultural sobre condições do espectro autista*. Rio de Janeiro: Papéis Selvagens, 2019.

FLIPPIN, M. Communication Breakdowns and Repairs of Children with Autism Spectrum Disorder with Fathers and Mothers. *International Journal of Disability, Development and Education*, v. 71, n. 4, p. 518-534, 2022. Disponível em: <https://doi.org/10.1080/1034912X.2022.2146069>. Acesso em: 07 jul. 2025.

FOX, B.; BENJAMIN, T.; MAZELAND, H. Conversation Analysis and Repair Organization: Overview, 2012. DOI: 10.1002/9781405198431.wbeal1314.

GARCEZ, P. M.; LODER, L. L. Reparo iniciado e levado a cabo pelo outro na conversa cotidiana em português do Brasil. *DELTA: Documentação de estudos em Linguística Teórica e Aplicada*, v. 21, n. 2, p. 279-312, 2005. Disponível em: <https://doi.org/10.1590/S0102-44502005000200006>. Acesso em: 07 jul. 2025.

GEOFFREY, R.; MAKOTO, H.; SIDNELL, J. *Conversational repair and human understanding*. Cambridge: Cambridge University Press, 2013. Disponível em: https://doi.org/10.26530/OAPEN_630827. Acesso em: 07 jul. 2025.

GERNSBACHER, M. A. Editorial Perspective: The use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry*, v. 58, p. 859-861, 2017. DOI: 10.1111/jcpp.12706.

GOFFMAN, E. A situação negligenciada. In: RIBEIRO, B. T.; GARCEZ, P. M. (org.). *Sociolinguística Interacional*. Tradução de Pedro M. Garcez. São Paulo: Edições Loyola, 2002. p. 13-20. Originalmente publicado em: 1964.

HUGHES, J. *Increasing neurodiversity in disability and social justice advocacy groups* [whitepaper]. Autistic Self Advocacy Network, 2016.

KEEN, D. The use of non-verbal repair strategies by children with autism. *Research In Developmental Disabilities*, v. 26, p. 243-254, 2005.

KENNY, L.; HATTERSLEY, C.; MOLINS, B.; BUCKLEY, C.; POVEY, C.; PELLICANO, E. Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, v. 20, n. 4, p. 442-462, 2016. DOI: 10.1177/1362361315588200.

KORKIAKANGAS, T. *Communication, Gaze and Autism: A Multimodal Interaction Perspective*. London: Routledge, 2014. Disponível em: <https://doi.org/10.4324/9781315621852>. Acesso em: 07 jul. 2025.

LAAKSO, M.; KLIPPI, A. A closer look at the "hint and guess" sequences in aphasic conversation. *Aphasiology*, v. 13, n. 4-5, p. 345-363, 1999. Disponível em: <https://doi.org/10.1080/026870399402136>. Acesso em: 07 jul. 2025.

NAIGLES, L. R. (ed.). *Innovative Investigations of Language in Autism Spectrum Disorder*. Washington, DC: American Psychological Association / Berlin: Walter de Gruyter, 2017.

MAGNANI, L. H.; RÜCKERT, G. H. *Linguagem e autismo: conversas transdisciplinares*. Catu: Editora Bordô-Grená, 2021.

MELLO, A. G. de; AYDOS, V.; SCHUCH, P. Aleijar as antropologias a partir das mediações da deficiência. *Horizontes Antropológicos*, v. 28, n. 64, p. 7-29, 2022. Disponível em: <https://doi.org/10.1590/S0104-71832022000300001>. Acesso em: 07 jul. 2025.

MONDADA, L. Conventions for multimodal transcription. 2014. Disponível em: https://franz.unibas.ch/fileadmin/franz/user_upload/redaktion/Mondada_conv_multimodality. Acesso em: 07 jul. 2025.

MONDADA, L. Challenges of multimodality: language and the body in social interaction. *Journal of Sociolinguistics*, v. 20, n. 30, p. 336-366, 2016.

MORATO, E. M. *et al. Sobre as afasias e os afásicos: subsídios teóricos e práticos elaborados pelo Centro de Convivência de Afásicos (Universidade Estadual de Campinas)*. Campinas: Unicamp, 2002.

NICOLAIDIS, C. What can physicians learn from the neurodiversity movement? *American Medical Association Journal of Ethics*, v. 14, n. 6, p. 503-510, 2012.

OCHS, E.; SOLOMON, O. Autistic Sociality. *Ethos*, v. 38, n. 1, p. 69-92, 2010.

OCHS, E.; KREMER-SADLIK, T.; SIROTA, K. G.; SOLOMON, O. Autism and the Social World: An Anthropological Perspective. *Discourse Studies*, v. 6, n. 2, p. 147-183, 2004. Disponível em: <https://doi.org/10.1177/1461445604041766>. Acesso em: 07 jul. 2025.

OLIVER, M. *Understanding Disability: from Theory to Practice*. Chatham: Mackays of Chatham, 1996.

PELLICANO, E.; STEARS, M. Bridging autism, science and society: moving toward an ethically informed approach to autism research. *Autism Research*, v. 4, n. 4, p. 271-282, 2011.

POSAR, A.; VISCONTI, P. Update about “minimally verbal” children with autism spectrum disorder. *Revista Paulista de Pediatria*, v. 40, e2020158, 2021. DOI: 10.1590/1984-0462/2022/40/2020158. PMID: 34495269; PMCID: PMC8432069.

RIOS, C.; FEIN, E. *Autismo em tradução: uma conversa intercultural sobre condições do espectro autista*. Rio de Janeiro: Papéis Selvagens, 2019.

ROSE, V.; TREMBATH, D.; KEEN, D.; PAYNTER, J. The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, v. 60, p. 464-477, 2016. (Special Issue: Autism Spectrum Disorder). Disponível em: <https://doi.org/10.1111/jir.12284>. Acesso em: 07 jul. 2025.

SACKS, H. *Lectures on Conversation*. Oxford: Basil Blackwell, 1992. v. 1 & 2.

SCHEGLOFF, E. A. Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, v. 97, n. 5, p. 1295-1345, 1992. Disponível em: <http://www.jstor.org/stable/2781417>. Acesso em: 07 jul. 2025.

SCHEGLOFF, E. A. Reflections on Quantification in the Study of Conversation. *Research on Language and Social Interaction*, v. 26, n. 1, p. 99-128, 1993. Disponível em: https://doi.org/10.1207/s15327973rlsi2601_5. Acesso em: 07 jul. 2025.

SINCLAIR, J. Why I dislike “person first” language. *Autonomy, the Critical Journal of Interdisciplinary Autism Studies*, v. 1, 2013.

SINGER, J. *Neurodiversity: the birth of an idea*. 2017.

SORJONEN, M. L.; PERÄKYLÄ, A.; LAURY, R.; LINDSTRÖM, J. Intersubjectivity in action: An introduction. In: LINDSTRÖM, J.; LAURY, R.; PERÄKYLÄ, A.; SORJONEN, M. L. (ed.). *Intersubjectivity in Action: Studies in language and social interaction*. Amsterdam: John Benjamins, 2021.

STERPONI, L.; SHANKEY, J. Rethinking echolalia: Repetition as interactional resource in the communication of a child with autism. *Journal of Child Language*, v. 41, n. 2, p. 275-304, 2014. Disponível em: <https://doi.org/10.1017/S0305000912000682>. Acesso em: 07 jul. 2025.

STERPONI, L.; DE KIRBY, K.; SHANKEY, J. Rethinking language in autism. *Autism*, v. 19, n. 5, p. 517-526, 2015. Disponível em: <https://doi.org/10.1177/1362361314537125>. Acesso em: 07 jul. 2025.

TAGER-FLUSBERG, H.; KASARI, C. Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research: Official Journal of the International Society for Autism Research*, v. 6, n. 6, p. 468-478, 2013. Disponível em: <https://doi.org/10.1002/aur.1329>. Acesso em: 07 jul. 2025.

TAGER-FLUSBERG, H.; PAUL, R.; LORD, C. Language and communication in autism. In: VOLKMAR, F. et al. (ed.). *Handbook of autism and pervasive developmental disorders*. 3. ed. New York: Wiley, 2005. p. 335-364.

THOMAS, R. P.; WITTKE, K.; BLUME, J. et al. Predicting Language in Children with ASD Using Spontaneous Language Samples and Standardized Measures. *Journal of Autism and Developmental Disorders*, v. 53, p. 3916-3931, 2023. Disponível em: <https://doi.org/10.1007/s10803-022-05691-z>. Acesso em: 07 jul. 2025.

VIVANTI, G. Ask the editor: What is the most appropriate way to talk about individuals with a diagnosis of autism? *Journal of Autism and Developmental Disorders*, v. 50, p. 691-693, 2020.

YU, B.; STERPONI, L. Toward Neurodiversity: How Conversation Analysis Can Contribute to a New Approach to Social Communication Assessment. *Language, Speech, and Hearing Services in Schools*, v. 54, p. 27-41, 2022. DOI: 10.1044/2022_LSHSS-22-00041.

A função social da divulgação científica (e como a Academia pode ajudar ou atrapalhar)

DOI: <http://dx.doi.org/10.21165/el.v54i1.4027>

Cecilia Farias¹

Resumo

Este artigo discute a função social da divulgação científica (DC), com foco no campo da linguística, e as possíveis atuações da Academia nessa prática. Partindo de um diagnóstico que evidencia o desconhecimento e/ou a resistência ao discurso científico sobre a linguagem pela sociedade, defendo que a DC é uma via de mão dupla, beneficiando tanto a sociedade civil quanto a Academia. Três funções sociais centrais da DC são propostas: o letramento científico (capacidade de compreender a natureza da ciência e empregar o conhecimento científico) como um direito cidadão; a democratização dos sonhos, isto é, que mais grupos sociais se sintam pertencentes aos ambientes científicos, o que leva à diversidade da comunidade acadêmica; e a dissolução de mitos sobre a língua, cujas consequências podem ser a marginalização e a negação de direitos linguísticos. Ao comparar uma experiência institucional (Museu da Língua Portuguesa) com uma produção independente (Babel Podcast), aponto as vantagens e desvantagens de cada modalidade. Além disso, destaco a importância de a Academia repensar seus critérios de avaliação e reconhecer a importância do trabalho de divulgação, atualmente subvalorizado. A conclusão propõe um novo pacto para a ciência, problematizando a ideia de neutralidade e o que Paolo Demuru chamou de “supremacismo da razão”. O caminho é a ciência cidadã e a comunicação colaborativa, em que o público não é visto apenas como alvo, mas participa ativamente do fazer científico. Por fim, a valorização da divulgação, docência e pesquisa como um ecossistema interdependente é essencial para uma ciência mais inclusiva e próspera.

Palavras-chave: Divulgação científica; popularização da linguística; ciência cidadã; popularização científica.

¹ Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil; cecilia.farias.souza@usp.br; <https://orcid.org/0000-0002-2340-2390>

The Social Role of Science Communication (and How Academia Can Help or Hinder)

Abstract

This article discusses the social function of science communication (SC), focusing on the field of linguistics, and the potential roles of Academia in this practice. Starting from a diagnosis that evidences the lack of knowledge and/or resistance to scientific discourse about language by society, I argue that SC is a two-way street, benefiting both civil society and Academia. Three central social functions of SC are proposed: scientific literacy (the capacity to understand the nature of science and employ scientific knowledge) as a civil right; the democratization of dreams, that is, that more social groups feel they belong in scientific environments, which leads to the diversity of the academic community; and the dissolution of myths about language, the consequences of which can be marginalization and the denial of linguistic rights. By comparing an institutional experience (Museum of the Portuguese Language) with an independent production (Babel Podcast), I point out the advantages and disadvantages of each modality. Furthermore, I highlight the importance of Academia rethinking its evaluation criteria and recognizing the importance of communication work, which is currently undervalued. The conclusion proposes a new pact for science, problematizing the idea of neutrality and what Paolo Demuru called the “supremacism of reason.” The path forward is citizen science and collaborative communication, in which the public is not just seen as a target but actively participates in the scientific process. Finally, valuing communication, teaching, and research as an interdependent ecosystem is essential for a more inclusive and prosperous science.

Keywords: Science communication; popularization of linguistics; citizen science; scientific popularization.

Introdução

Falar sobre divulgação científica pode ter significados distintos para diferentes pessoas. Além disso, existe uma zona cinzenta entre termos como difusão, divulgação e popularização científica, cuja compreensão histórica nos ajuda a analisar os acertos e os desafios enfrentados por linguistas nesse campo. Questões como “De onde vêm os gênios?”, “Quem tem direito a sonhar em ser cientista?” e “Como a ciência é vista hoje?” servem como ponto de partida para refletir sobre as dinâmicas dessa área, abordando também os desafios enfrentados pelas divulgadoras científicas, tanto em contextos institucionalizados quanto na produção independente.

Para discutir a função social da divulgação científica no campo da linguística, este artigo começa com reflexões gerais sobre as dinâmicas da divulgação científica, indo

além dos estudos da linguagem e, em geral, utilizando a linguística como exemplo para fundamentar os argumentos.

Este texto apresenta a discussão realizada na mesa-redonda “Divulgação e Popularização da Linguística”, parte da programação do Seminário do GEL 2024, da qual participei ao lado de Jana Viscardi e Renato Basso. A proposta foi seguir, dentro do possível, a linha argumentativa desse evento, aproveitando a modalidade escrita para reorganizar as ideias e aprofundar as reflexões.

1. Diagnóstico

Em 2001, o deputado Aldo Rebelo propôs um projeto de lei visando proibir estrangeirismos na língua portuguesa, um projeto que se referia a termos de línguas estrangeiras no geral, mas cujo alvo principal, na prática, eram os anglicismos. Para estudantes de linguística, mesmo nos primeiros semestres da graduação, a inviabilidade dessa proposta é evidente, e especialistas da área posicionaram-se publicamente contra o projeto.

Rajagopalan (2004), em um artigo dedicado à relação entre políticas linguísticas e emoções, resgatou comentários de fóruns públicos sobre esse debate. Um dos exemplos diz:

Eu tenho o grande privilégio de não saber quem é quem na linguística brasileira. Por isso, estou em posição de julgar seus textos pelo que valem, sem ser induzido a “pensar” que eles contêm doses de enorme sabedoria... apenas que têm a assinatura de X, Y ou Z. Dos estudiosos que você cita, até agora só li pura bobagem, aqui nesta sala de chat e em outros lugares. Não estou impressionado por seus nomes nem seus PhDs. Esse é o grande privilégio dos cidadãos ordinários que não ligam a mínima pra academia (Rajagopalan, 2004, p. 117, tradução própria).²

Esse exemplo nos leva a questionar: por que tanta resistência a linguistas, o que nós lhe fizemos?

No âmbito da discussão sobre o projeto de lei, as opiniões sobre o PL, as explicações sobre as opiniões sobre o PL e esse ciclo infinito, o linguista Carlos Faraco também comentou o episódio em uma coluna da *Folha de S.Paulo*: “40 anos após sua introdução

² No original: “I have the great privilege of not knowing who is who in Brazilian linguistics. Thus, I am in a position to judge their texts for what they are worth, without being induced to ‘think’ that they contain doses of prodigious wisdom... just that they bear the signature of Tom, Dick or Harry. Of the scholars whose names you cite, up until now I have only read pure nonsense, here in this chat room and elsewhere. I am not impressed by their names nor their PhDs. This is the great privilege of ordinary citizens who do not care a damn about academy.”

como disciplina nas universidades brasileiras, a linguística permanece invisível e inaudível para a sociedade como um todo.” E continua:

Aos linguistas coloca-se o desafio de trabalharem essas questões todas como questões fundamentalmente políticas e de buscarem meios para projetar sua voz, contribuindo, assim, para a instauração de uma necessária guerra cultural entre os discursos que dizem a língua no Brasil (Faraco, 2001, p. 30-31).

Esse episódio ilustra como o público em geral percebe o discurso científico, sobretudo no que diz respeito à linguagem. Em diferentes mídias, circulam ideias sobre a língua – frequentemente apresentadas de forma naturalizada e sem questionamento crítico. Isso demonstra como a língua é um tema amplamente debatido, ainda que nem sempre de forma explícita. Fala-se de língua o tempo todo.

A experiência também revela como o discurso científico, em especial o relacionado à linguística, mobiliza visões de mundo e identidades, frequentemente suscitando paixões intensas. Sem buscar explicar aqui todas as razões para a aversão ao discurso científico ou para a onda anticientificista dos últimos anos, proponho reflexões e encaminhamentos para lidar com esse fenômeno no campo dos estudos da linguagem, com base em experiências de divulgação e popularização da linguística.

1.1 Divulgar para ocupar espaços

Uma figura de destaque na divulgação científica contemporânea é o biólogo Atila Iamarino, que desempenhou um papel crucial durante a pandemia de covid-19. Durante esse período, ele não apenas apresentou descobertas científicas ao público, como também explicou o processo de produção do conhecimento científico. No entanto, seu reconhecimento não foi fruto de uma ação imediata; ele já tinha um histórico consolidado como divulgador.

Em um vídeo de 2019 intitulado “Por que você DEVERIA se expor na internet?”³, o biólogo discute o impacto da internet na privacidade e no anonimato, argumentando que, na era digital, informações sobre qualquer pessoa estão acessíveis. Ele sugere que, ao publicar estrategicamente conteúdos sobre si mesmo, é possível direcionar o que as pessoas encontram, aumentando as chances de controlar a narrativa sobre sua própria imagem.

Em uma palestra reproduzida no vídeo, Atila questiona a plateia: “Quem foi a uma biblioteca hoje?”, recebendo poucas respostas afirmativas. Em seguida, ele pergunta: “E quem deu um Google hoje?”, provocando risadas e um número muito maior de respostas positivas.

³ Disponível em: https://youtu.be/Ja6ZnJUQa_0?si=YC48N8EjuDFkzPqc. Acesso em: 12 nov. 2024.

A partir disso, ele argumenta que a internet se tornou a principal fonte de informações e que, portanto, é fundamental utilizá-la para divulgar o conhecimento científico.

Mais do que ocupar o espaço virtual, o princípio subjacente é a importância de marcar presença nos espaços públicos de discussão. A circulação de informações e opiniões ocorre independentemente da participação de especialistas; no entanto, a ausência deles abre espaço para que desinformações prevaleçam. Assim, ocupar esses espaços aumenta as chances de o público geral ter acesso a conteúdos de qualidade e baseados em critérios científicos.

Em resumo, se nós não falarmos, alguém falará – e nem sempre com o devido rigor ou boas intenções.

2. Função social da Divulgação Científica

Com este trabalho, busco destacar que a função social da divulgação científica (DC) opera como uma *via de mão dupla*. Por um lado, a sociedade civil se beneficia enormemente ao acessar e participar da produção acadêmica. Por outro, a Academia também se enriquece ao incorporar maior diversidade de perspectivas e ao ser reconhecida e validada socialmente. Adicionalmente, contar com o apoio da opinião pública é muitas vezes crucial para a luta pela qualidade do ensino, pelo incentivo à pesquisa e pelo financiamento de projetos científicos.

Nesta seção, proponho três grandes pontos que exemplificam as funções sociais da DC, ciente de que esta não é uma abordagem exaustiva.

I – Letramento científico

Um programa humorístico veiculado no YouTube, “Vrau Cast”, convida especialistas de diferentes áreas para serem entrevistados por dois personagens que são apresentadores de *podcast* despreparados (Mike e Cauê, interpretados por Caíto Mainier e Rafael Saraiva, respectivamente). O que falta de preparo a esses apresentadores sobra em confiança nas próprias convicções. O humor vem justamente das perguntas e comentários que soam absurdos a quem minimamente conhece o tema de cada programa, satirizando a forma como pessoas sem letramento sobre os assuntos e expostas a *fake news* conduzem o debate público sobre questões sociais sérias e importantes.

Em um episódio lançado em outubro de 2024, as convidadas foram Laura Marise e Ana Bonassa, do projeto Nunca Vi 1 Cientista. Comentários como “A ciência garante, mas quem é a ciência?”, ou “Tudo de vocês é ‘normalmente’, ‘segundo estudo tal’; como eu vou saber que o estudo é verdadeiro?” exemplificam a falta de letramento científico

dos personagens (que em certos momentos nem parecem fictícios).⁴ Essas lacunas de compreensão podem levar a erros graves, como a defesa de tratamentos sem comprovação científica, evidenciando os riscos da desinformação.

Uma parte fundamental do trabalho de divulgação científica consiste em ensinar como a ciência é feita, seus métodos, critérios, descobertas e contradições ao longo da história. Cada área do conhecimento busca explicar fenômenos, identificar questões relevantes e estabelecer definições fundamentadas em evidências. No contexto da divulgação, isso é conhecido como letramento científico. Ou seja, parte do trabalho de DC é de letramento científico, definido pelo Inep como:⁵

[...] a capacidade de empregar o conhecimento científico para identificar questões, adquirir novos conhecimentos, explicar fenômenos científicos e tirar conclusões baseadas em evidências sobre questões científicas. Também faz parte do conceito de letramento científico a compreensão das características que diferenciam a ciência como uma forma de conhecimento e investigação; a consciência de como a ciência e a tecnologia moldam nosso meio material, cultural e intelectual; e o interesse em engajar-se em questões científicas, como cidadão crítico capaz de compreender e tomar decisões sobre o mundo natural e as mudanças nele ocorridas. [...] refere-se tanto à compreensão de conceitos científicos como à capacidade de aplicar esses conceitos e pensar sob uma perspectiva científica.

Esse letramento deve ser considerado um direito dos cidadãos. O acesso universal a ele é essencial para a construção de uma sociedade democrática, dado que é uma ferramenta importante para compreender o mundo e intervir ativamente na sociedade. Portanto, se almejamos uma estrutura social mais justa e igualitária, o letramento científico não pode ser tratado como privilégio para poucos.

O letramento científico deve ser tratado como um direito universal, fundamental para a construção de uma sociedade democrática e participativa. Ele não apenas facilita a compreensão do mundo, mas também habilita os cidadãos a tomarem decisões informadas. Além disso, entender a ciência como um campo dinâmico, permeado por erros, acertos e disputas, reduz o descrédito gerado pela não linearidade de suas descobertas.

4 Disponível em: <https://www.youtube.com/watch?v=6DNEXHDMa8k>. Acesso em: 12 nov. 2024.

5 Disponível em: https://download.inep.gov.br/download/internacional/pisa/2010/letramento_cientifico.pdf. Acesso em: 12 nov. 2024.

II – Democratização dos sonhos

“Nunca ganhamos um Nobel. Em compensação, a Seleção Brasileira já ganhou 5 copas do mundo. Aí eu fico me perguntando: como eu falo para uma criança que ela tem que ser cientista e não jogador de futebol?”. Essa frase provocativa, dita pelo personagem Mike no Vraucast, parte de uma intenção provocadora e humorística, no entanto evoca uma pergunta que chega a ser um senso comum: quantas mentes brilhantes deixam de ser estimuladas por não terem sequer o conhecimento do que acontece no universo científico?


Quando se fala de ciência ou de pessoas consideradas geniais, certos nomes imediatamente vêm à mente. Ao analisarmos esses grandes gênios da história e suas trajetórias de vida, observamos que a maioria tem em comum o acesso a uma série de condições: educação formal de qualidade, incentivo familiar, condições materiais adequadas como moradia e alimentação dignas, entre outros fatores. No entanto, há também um aspecto subjetivo crucial: a percepção da ciência como um ambiente acessível, ou seja, esses indivíduos viam o universo acadêmico, da pesquisa e do fazer científico como um domínio ao qual também podiam pertencer. Nesse sentido, a principal função da divulgação científica é garantir acesso às escolhas, ou, em outras palavras, promover a **democratização dos sonhos**, possibilitando que todas as pessoas possam, caso queiram, sonhar com uma trajetória nesse campo.

Os pontos I e II contribuem mais diretamente para a ampliação da diversidade no campo acadêmico, o que contribui diretamente para a inclusão de novas perspectivas e abordagens. O campo científico só tem a ganhar com isso. No caso dos estudos da linguagem, a democratização do acesso também desmistificaria certas crenças, atraindo mais reconhecimento e recursos para a área, frequentemente subvalorizada em comparação às chamadas *hard sciences*.

III – Desfazer mitos sobre língua e suas consequências

Todo campo do conhecimento é permeado por mitos, e os estudos da linguagem não fogem à regra. Como fenômeno universal, a língua suscita opiniões diversas, muitas delas baseadas em pressupostos equivocados. Afinal, a língua, oral ou sinalizada, é um fenômeno compartilhado por todos os seres humanos, e, conscientemente ou não, as pessoas têm concepções sobre esse fenômeno. Embora essas crenças possam parecer inócuas, elas podem ter implicações concretas na vida dos falantes, uma vez que a língua é constitutiva da identidade das pessoas e está inexoravelmente relacionada à cultura e às formas de existência no mundo.

Um mito persistente é a equiparação entre língua e norma-padrão, o que marginaliza e desumaniza falantes de variedades linguísticas não padrão. Outra concepção equivocada



é a personificação da língua como entidade autônoma, negligenciando seu caráter social e dinâmico, construído coletivamente pelos falantes. Essa visão mascara a realidade: nossos julgamentos sobre variedades linguísticas são, na verdade, juízos de valor sobre seus usuários. É mais fácil dizer “Que sotaque feio!” do que “Eu não gosto desse pessoal” (Farias, 2024). Tal perspectiva sustenta ainda a falsa noção de hierarquia entre línguas, como se algumas fossem intrinsecamente superiores ou mais aptas para certos fins. Basta observar como línguas minorizadas raramente são reconhecidas como veículos adequados para atividades intelectuais de prestígio.

Outro mito é a ideologia da escassez, que avalia línguas minorizadas com base no que elas não têm quando comparadas com as línguas hegemônicas, sem considerar suas riquezas únicas. Esse viés afeta não apenas as línguas, mas também as culturas associadas a elas. Respeitar a maneira como as pessoas falam é reconhecer sua humanidade e dignidade (Farias, 2024).

Por fim, a ideologia do monolinguismo também merece destaque. No Brasil, por exemplo, a crença em uma suposta homogeneidade linguística invisibiliza a diversidade real do país, reforçando preconceitos e negando direitos linguísticos. Esses direitos, cada vez mais debatidos no contexto dos direitos humanos, são essenciais para a valorização da diversidade cultural e linguística. Reconhecer a diversidade linguística é um dos passos para promover o respeito à diversidade como um todo, pois se reconhece que a homogeneidade é um ideal abstrato, sem existência concreta.

No limite, ao nos dedicarmos aos fenômenos relacionados à linguagem, podemos questionar: quem tem direito à escrita na nossa sociedade? O fato de, historicamente, serem considerados ‘sérios’ e/ou ‘universais’ escritores de grupos sociais e linguísticos específicos, enquanto outras escritas levam a etiqueta de regionais e/ou identitárias, está baseado numa visão de mundo que tem esses mitos em sua constituição.

Enquanto linguistas, nossa responsabilidade consiste em desconstruir essas concepções equivocadas e evidenciar seus impactos na dignidade dos falantes. A partir desse compromisso ético, podemos efetivamente colaborar para a construção de ambiente social mais equitativo e inclusivo.

3. Institucionalização vs. Produção independente

Dado que o subtítulo deste artigo é “*como a Academia pode ajudar ou atrapalhar*”, trago aqui uma comparação entre duas experiências pessoais no campo da divulgação da linguística: o trabalho como pesquisadora no Centro de Referência do Museu da Língua Portuguesa (MLP) e a produção do Babel Podcast (BP).

O MLP é uma instituição do Governo do Estado de São Paulo, administrada pela Organização Social de Cultura IDBrasil, e conta com certo reconhecimento da sociedade como um espaço de conhecimento oficialmente validado.

De certa forma, museus desempenham o papel de espaços de divulgação científica, aproximando as pesquisas especializadas de seus objetos de estudo ao público não especializado. Entre as vantagens de atuar em uma instituição como o MLP, destacam-se:

- Tempo: Como a divulgação é parte integrante das atividades institucionais, ela pode ser realizada dentro da jornada padrão de trabalho, diferentemente da produção independente, que muitas vezes ocupa horários destinados ao descanso ou lazer.
- Ferramentas e recursos: A instituição fornece os meios necessários para os projetos, enquanto produções independentes dependem de recursos de seus realizadores ou financiamento coletivo, o que pode ser inviável no início de uma iniciativa.
- Rede de divulgação estruturada: A comunicação institucional utiliza canais oficiais e redes de colaboração entre outras entidades culturais e científicas, dispensando a necessidade de construir estratégias de divulgação do zero.
- Credibilidade: A associação a uma instituição reconhecida transfere prestígio aos projetos realizados em seu nome.

Por outro lado, o Babel Podcast representa um exemplo de produção independente. Criado por Bruno Guide e Cecilia Farias em 2019, o programa aborda uma língua por episódio, explorando suas características, formação e a cultura dos povos que a falam, contextualizando-a em seu panorama histórico, social e político. Além do *podcast*, há um *blog* com textos e outras colaborações.⁶

Há duas modalidades de organização do conteúdo: roteiros elaborados pelos apresentadores e entrevistas a convidados que estudam o tema. Talvez se possa dizer que há um terceiro formato, híbrido, em que se convida uma pessoa, não especialista especificamente naquele tema, mas com algum conhecimento para escrever de forma conjunta o roteiro com os apresentadores.

Em contraste com a experiência institucionalizada, as vantagens da produção independente incluem:

⁶ Disponível em: <https://medium.com/babelpodcast>. Acesso em: 12 nov. 2024.

- **Liberdade editorial:** Sem vínculos institucionais, as decisões editoriais seguem as preferências dos criadores. No caso do BP, isso se reflete em priorizar línguas não hegemônicas, evitar a exotização de línguas e culturas mais distantes da nossa e abordar questões sociais e políticas relacionadas às línguas e seus povos. Em especial, não se assume a possibilidade de neutralidade, admitindo que toda abordagem é feita por um ponto de vista.
- **Criação de redes solidárias:** A produção independente conecta criadores que compartilham experiências, dificuldades e estratégias, criando um ambiente de apoio mútuo e divulgação cruzada, isto é, um projeto citando o outro em seus episódios. O BP, por exemplo, recebeu suporte importante do grupo Podosfera Antifascista.
- **Permeabilidade:** Apesar de não receber oficialmente uma validação institucional, essa ausência pode facilitar a conexão com públicos que, de outro modo, se sentiriam alheios ao universo acadêmico. Projetos independentes, sem o peso do nome de uma instituição, muitas vezes chegam a essas pessoas de forma mais acessível.


No caso específico do BP, a estratégia é atuar como um “Cavalo de Troia”: ao oferecer informações sobre história, música, geografia e política, os ouvintes acabam aprendendo conceitos de linguística de maneira incidental, como alinhamento ergativo-absolutivo (Episódio 17 – Karitiana) ou lenição (Episódio 20 – Gaélico escocês).

Podemos considerar que o BP, bem como diversas iniciativas independentes de divulgação/popularização da linguística, fura a bolha da Academia, porém de fora para dentro. De certa forma, muitos projetos de DC que foram ganhando espaço entre o público geral o fizeram sem a intenção inicial de institucionalização – muitos sequer se autodenominavam como tal inicialmente. A institucionalização e a rotulagem vieram posteriormente. Hoje observamos um reconhecimento institucional ainda embrionário, exemplificado pela criação da Comissão de Popularização da Linguística da Abralín e pela crescente inclusão do tema na programação de eventos acadêmico-científicos.

Por enquanto, em comparação com outros espaços frequentados por linguistas, ainda vemos poucos nomes sendo frequentemente chamados para debates sobre divulgação científica, reflexo direto do número ainda limitado de especialistas dedicados ao tema. Contudo, todos os indícios sugerem que esse cenário está em transformação e deverá apresentar mudanças significativas em breve.

3.1 Academia no meio de campo

Em 2021, foi instaurada a CPI da Pandemia, uma comissão parlamentar de inquérito para investigar omissões e irregularidades nas ações do governo de Jair Bolsonaro



durante a pandemia de covid-19. Entre as pessoas chamadas a depor, esteve a bióloga Natalia Pasternak, fundadora do Instituto Questão de Ciência e reconhecida divulgadora científica, que se destacou por comunicar descobertas sobre a doença e, frequentemente, por criticar as ações governamentais durante a crise sanitária. Em seu depoimento, entretanto, sua credibilidade foi questionada com base em seu baixo Índice H, uma métrica que avalia a produtividade e o impacto de publicações científicas. Esse questionamento revela uma contradição: como pesquisadora dedicada principalmente à divulgação científica, era natural que seu índice fosse inferior ao de colegas focadas exclusivamente em pesquisa acadêmica. Esse episódio exemplifica como o trabalho de divulgação, que foge aos parâmetros convencionais de avaliação universitária, continua sendo subvalorizado no meio científico.

Para a Academia contribuir mais efetivamente com a divulgação científica, é essencial:

1. Reconhecer sua importância, destinando tempo e recursos adequados à prática.
2. Repensar os critérios de avaliação acadêmica, que atualmente desconsideram o impacto da divulgação.
3. Teorizar sobre estratégias comunicativas, baseando-se em estudos de gêneros discursivos para propor modelos inovadores.

Esse terceiro ponto tem conexão mais específica com os estudos da linguagem, que têm muito a contribuir para o entendimento de como operam as dinâmicas de veiculação do discurso científico. A teorização, visando à elaboração de ferramentas e instrumentalização de divulgadores, pode se desdobrar em duas frentes: i) levantar e analisar o que já é feito e seus impactos; ii) pensar em novos modelos com base nos estudos dos gêneros discursivos, propondo estratégias comunicativas.

Os contatos iniciais com a divulgação científica ocorrem em momentos precoces de nossas vidas, como o material didático escolar ou revistas como *Ciência Hoje*. Manter o diálogo com escolas e outros meios educacionais, além de parcerias com veículos de comunicação, é crucial para fortalecer essa ponte entre Academia e sociedade.

Outro ponto sensível é o questionamento da dicotomia entre Academia e os genericamente chamados “outros saberes”. A lógica de oposição entre ciência e saberes tradicionais muitas vezes impede diálogos produtivos por criar um abismo na comunicação entre pesquisa científica e sociedade como um todo, minando possibilidades de cooperação. As consequências dessa separação dicotômica são melhor elaboradas em debates sobre descolonização da ciência e da linguística.

Em síntese, embora nem todos os cientistas precisem dedicar-se à divulgação, é fundamental oferecer suporte e reconhecer o valor daqueles que escolhem esse caminho. A pesquisa acadêmica precisa avançar para o desenvolvimento contínuo da linguística, assim como o ensino é fundamental para a formação de novas gerações de linguistas. Compreender que divulgação, docência e pesquisa formam um ecossistema acadêmico interdependente é essencial para construir uma ciência mais inclusiva e próspera.

4. Rumo a um novo pacto

“Nem toda a ciência está na berlinda”, diz a matemática Tatiana Roque em *O dia em que voltamos de Marte*. Trata-se de uma obra muito bem-feita de divulgação científica, em que a autora traça um panorama da história da ciência ocidental desde o Iluminismo, evidenciando sua conexão com as ideologias de cada época. Por meio de uma linguagem acessível e não hermética, Roque utiliza histórias de grandes nomes da ciência (como curiosidades sobre a vida de D’Alembert) e as consequências práticas das descobertas científicas para apresentar suas reflexões.

Ao longo do livro, a autora analisa, entre outras coisas, as razões pelas quais a pesquisa científica é percebida de determinada maneira hoje, destacando a historicidade por trás desse fenômeno. Ela conclui falando sobre qual ciência está sendo questionada e propõe caminhos para reverter esse processo. Em seu diagnóstico, afirma:

[...] o historiador da ciência Steve Shapin sugere que não existe uma crise da verdade, e sim uma “crise do conhecimento social”. Isto é, a sociedade está com dificuldade para discernir o conhecimento científico confiável: reconhecer quem sabe e quem não sabe, quem é e quem não é confiável, que instituições produzem um saber genuíno e sem interesses comerciais (Roque, 2021, p. 272-273).

Roque argumenta que parte do descrédito atual da ciência decorre de seu discurso de neutralidade. Durante o século XX, sob o discurso da neutralidade científica, os financiamentos em pesquisa foram frequentemente direcionados para fins que pouco beneficiavam a humanidade. Grandes somas foram destinadas à criação de armas de destruição em massa e à intensificação da exploração da força de trabalho – elementos que contribuíram para o aumento da desigualdade social –, em vez de investimentos em melhorias de vida das pessoas, como criação de medicamentos, fontes sustentáveis de energia, combate à fome, entre muitos outros.⁷ Toda ciência é permeada pela cultura; ignorar essa interseção é improdutivo. Antes, é mais pertinente mostrar como ciência e cultura se entrelaçam e influenciam mutuamente, em conjunto com outras dimensões da sociedade.

⁷ Um livro que mostra muito bem essas dinâmicas é *O imperador de todos os males: uma biografia do câncer*, de Siddhartha Mukherjee (Companhia das Letras, 2012).

Em vez de supor que um déficit cognitivo ou educacional explique o baixo engajamento do público, falta envolver mais gente nas decisões. A população deve ter mais espaço para expressar e debater suas inquietações, não ser vista como alvo de persuasão por *experts* (Roque, 2021, p. 275).

4.1. Ciência cidadã e comunicação colaborativa

Na discussão sobre a função social de qualquer prática, deve-se rejeitar uma postura salvacionista (o bandeirante, o missionário, o colonizador... temos um vasto vocabulário para isso). No campo da divulgação científica especificamente, assumir o papel de vanguarda iluminada que 'leva a luz do conhecimento às massas' não apenas inviabiliza relações verdadeiramente colaborativas, como pode gerar efeitos contrários aos desejados.

No livro *Políticas do encanto: extrema direita e fantasias da conspiração*, Paolo Demuru explora o conceito de "supremacismo da razão", uma postura que consiste em simplesmente declarar as crenças alheias como equivocadas enquanto se autoproclama portador da verdade. Esse posicionamento, demonstra ele, além de não ser eficaz para combater a desinformação, ainda aprofunda o abismo entre a Academia e a sociedade.

Proferir fatos não basta; precisamos de uma *ciência cidadã*. Essa perspectiva reconhece que

[...] a participação do público não especializado, sem treinamento científico, é essencial ao fazer científico – e esse público não pode ser reduzido à população de interesse e sujeitos de pesquisa, ou seja, a participação vai além da posição de fonte de informação e contribui para a pesquisa de outras formas (De Conto; Sanchez-Mendes; Rigatti, 2022, p. 18).

Essa abordagem se aproxima mais da ideia de popularização científica.⁸ Popularizar não é apenas conquistar aceitação e apreço popular, mas também *tornar-se do povo*. Uma linguística inclusiva e cidadã deve reconhecer que todas as pessoas têm uma língua (oralizada ou sinalizada) e um conhecimento inicial sobre ela, e partir desse reconhecimento para compartilhar saberes e se deixar aprender.

Nesse contexto, ressoa a ideia do professor Lynn Mario Menezes de Souza de "*bring back bodies*" (trazer os corpos de volta). A ciência eurocentrada, ao negar a corporeidade no fazer

8 Para um debate sobre as diferenças entre comunicação, difusão e divulgação científica, ver: Baronas (2010) e Bueno (2010). A Comissão de Popularização da Linguística da Abralín vem realizando uma reflexão sobre o que é popularizar ciência (linguística). Agradeço às conversas sobre o tema em especial com os colegas Lou-Ann Keppla, Vitor Hochsprung e Hélio Oliveira.

científico, adota uma estratégia epistêmica alinhada a um projeto político que racializa e exclui tudo o que não é branco/europeu. A ciência dos colonizadores não se reconhece como racializada, tampouco como produto de corpos situados; por isso, autoproclama-se como neutra – e, portanto, universal. Uma linguística cidadã (ou mesmo descolonial), precisa recuperar a presença do corpo, admitindo que as línguas são faladas por pessoas concretas.

Conclusão

Este artigo partiu do questionamento da função social da divulgação científica, explorando possíveis contribuições da Academia para esse campo. Não se pretendeu esgotar o tema, mas sim levantar pontos que fomentem a discussão pública sobre o assunto.

Reconhecer o valor da divulgação científica é o primeiro passo para que ela receba suporte adequado: tempo, recursos e valorização institucional. Afinal, se a pesquisa sustenta o avanço da ciência, a comunicação eficaz de seus achados à sociedade (e à melhoria da vida) é o que a torna significativa. Assim, o fortalecimento do ecossistema científico como um todo depende do reconhecimento e da valorização de todos os seus elementos.

Referências

BARONAS, R. L. Da necessidade premente de se cometer uma política de divulgação científica qualificada dos trabalhos da linguística do Brasil. *Revista da Anpoll*, v. 1, n. 29, 2010.


BUENO, W. C. Comunicação científica e divulgação científica: aproximações e rupturas conceituais. *Informação & Informação*, v. 15, n. 1 esp., p. 1-12, 2010.

DE CONTO, L.; SANCHEZ-MENDES, L.; RIGATTI, P. C. Quando o falante faz Linguística. *Cadernos de Linguística*, v. 3, n. 2, p. e653-e653, 2022.

DEMURU, P. *Políticas do encanto: extrema direita e fantasias da conspiração*. São Paulo: Elefante, 2024.

FARACO, C. Guerras em torno da língua. *Folha de São Paulo: Caderno 'Mais'*, 30-31, 2001.

FARIAS, C. *A língua é um campo de disputa: ativismos e opressões da linguagem*. Disponível em: <https://escoladeativismo.org.br/lingua-e-ativismos-o-que-quer-e-o-que-pode-essa-lingua/>. Acesso em: 12 nov. 2024.



HOCHSPRUNG, V. Divulgação científica: notas sobre a popularização da linguística na internet e na sala de aula. In: SIBALDO, M. (org.). *Ensino de línguas*. São Paulo: Blucher, 2023. p. 113-130.

MENEZES DE SOUZA, L. M. Decolonial Pedagogies, Multilingualism and Literacies. *Multilingual Margins*, v. 6, n. 1, p. 9-13, 2019.

RAJAGOPALAN, K. Emotion and Language Politics: The Brazilian Case. *Journal of Multilingual and Multicultural Development*, v. 25, n. 2-3, p. 105-123, 2004. DOI: 10.1080/01434630408666524.

ROQUE, T. *O dia em que voltamos de Marte*: uma história da ciência e do poder com pistas para um novo presente. São Paulo: Planeta, 2021.

SENADO. Governo Federal. CIPANDEMIA – CPI da Pandemia. Disponível em: <https://legis.senado.leg.br/atividade/comissoes/comissao/2441/>. Acesso em: 19 nov. 2024.

De Rowling a Galbraith: as sobrevidas do autor na contemporaneidade

DOI: <http://dx.doi.org/10.21165/el.v54i1.3836>

Beatriz Masson Francisco¹

Resumo

Partindo da popularidade e da importância da série de fantasia *Harry Potter* (Rowling, 1997-2007) para o século XXI, este artigo propõe uma análise comparada de sua figura criadora, J.K. Rowling, e Robert Galbraith, pseudônimo masculino criado pela autora britânica para a escrita de livros detetivescos. Por isso, tem-se a intenção de equacionar as perspectivas teóricas que envolvem as questões da construção de um pseudônimo, da biografia do autor, ilusão biográfica e *performance*, a partir dos trabalhos de Foucault (1969), Klinger (2006) e Lejeune (2008), em uma tentativa de entender como Joanne Rowling criou e veiculou diferentes *personas* autorais a depender do tipo de texto que escreve. O trabalho objetiva, portanto, mostrar aproximações e distanciamentos entre essas duas personas, que parecem integrar um movimento estético bastante específico dos escritos contemporâneos: o autor enquanto sujeito performático.

Palavras-chave: autor; Teoria Literária; J.K. Rowling.

¹ Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil; beatrizmasson@hotmail.com; <https://orcid.org/0009-0006-1028-6999>

From Rowling to Galbraith: the author's survivals in contemporaneity

Abstract

Based on the popularity and importance of the fantasy series Harry Potter (Rowling, 1997-2007) in the 21st century, this article proposes a comparative analysis of its creator, J.K. Rowling, and Robert Galbraith, a male pseudonym adopted by the British author to publish her detective novels. The article discusses theoretical perspectives on the construction of a pseudonym, the author's biography, biographical illusion, and performance, drawing on the works of Foucault (1969), Klinger (2006), and Lejeune (2008). The aim is to understand how Joanne Rowling created and conveyed different author personas according to the type of text she writes. The study seeks to show similarities and differences between these two personas, which seem to embody a specific aesthetic movement in contemporary writing: the author as a performing subject.

Keywords: author; Literary Theory, J.K. Rowling.

Introdução

Harry Potter (1997-2007), série literária de fantasia que é objeto de estudo deste trabalho, pauta seu enredo na dualidade entre o protagonista que a nomeia e Voldemort, um bruxo dotado de excepcional inteligência que, desde sua tenra infância, mostrou interesse na magia das trevas. Dentro dos limites da história, o vilão é tão temido que os outros personagens se referem a ele como "Aquele que não deve ser nomeado" ou "Você sabe quem"², por medo de invocarem aquele que é considerado o mal encarnado.

Voldemort também tinha seus próprios temores: ao saber da existência de uma profecia, que apontava o nascimento de um bruxo tão poderoso quanto ele mesmo³, tratou de tentar ceifar a possibilidade de isso acontecer. Procurou pela casa dos Potter, casal que participava da resistência contra suas ações, assassinou James e Lily e tentou matar o filho dos dois, Harry, que sobreviveu à maldição da morte com apenas uma cicatriz no formato de raio sobre a sua testa quando ainda era um bebê.

Os sete romances que compõem a série explicam, ao longo da narrativa, a motivação que levou Voldemort a cometer vários assassinatos, bem como a tentar superar Harry Potter

2 No texto original, Voldemort é designado como "He-Who-Must-Not-Be-Named" ou "You-Know-Who".

3 A profecia que Voldemort teme é apresentada em *Harry Potter and the Order of the Phoenix*: "The one with the power to vanquish the Dark Lord approaches ... born to those who have thrice defied him, born as the seventh month dies ... and the Dark Lord will mark him as his equal, but he will have power the Dark Lord knows not ... and either must die at the hand of the other for neither can live while the other survives ... the one with the power to vanquish the Dark Lord will be born as the seventh month dies ..." (Rowling, p. 742, 2007).

nos níveis mais poderosos da magia. O bruxo desejava se tornar imortal e, para isso, criou objetos mágicos chamados Horcruxes, que guardavam em si pedaços de sua alma, de modo que sua existência física estava conectada a eles. Dentro das leis que regem o Mundo Mágico, para uma Horcrux ser criada e a alma de alguém ser dividida, essa pessoa deveria cometer um assassinato – e foi o que Voldemort fez. Antes do vilão, havia Tom Marvolo Riddle, um rapaz bonito e muito inteligente. Tal sagacidade, em conjunção com a sua frieza, o levou a cometer vários crimes em nome do seu desejo de poder e de imortalidade.

À medida que cindia sua alma para criar Horcruxes, Riddle ia perdendo tudo o que lhe caracterizava como um ser humano: a capacidade de sentir, os atributos físicos e o próprio nome. Voldemort (re)nasce como um ser de aparência ofídica e de olhos vermelhos, depositando sua existência em objetos que serão, eventualmente, destruídos por Harry.

Fora dos limites da narrativa, nota-se uma conjunção entre a ação do vilão e, em vários níveis diferentes, diálogos possíveis entre a história de fantasia e o debate da teoria e crítica literárias a respeito do papel do autor, sobretudo na contemporaneidade. Não foram poucos os teóricos que juntaram o autor de textos literários à ideia de morte e, também, de imortalidade.

Neste texto, intenciona-se equacionar brevemente as perspectivas teóricas que envolvem questões relacionadas a *biografia do autor*, *ilusão biográfica* e *performance*, em uma tentativa de entender como Joanne Rowling veicula diferentes *personas* a depender do tipo de texto que escreve. Para isso, propõe-se uma breve análise de algumas das notas biográficas publicadas nas orelhas dos livros e em *sites* oficiais sobre a voz autoral adotada por Rowling quando da assinatura da série *Potter* e também daquela veiculada por Galbraith acerca da publicação dos romances policiais.

A figura do autor na Teoria Literária

É difícil não começar a revisão da genealogia do autor sem invocar aquele que talvez seja o texto mais conhecido sobre o tema em questão: o ensaio “A morte do autor”, escrito por Roland Barthes em 1968. Aqui, o estudioso valoriza a instância linguística do texto como sendo um produto do “agora da enunciação” em detrimento da figura de um sujeito criador que, para ele, é morto quando essas palavras ganham corpo dentro de um livro.

Em sua visão, no texto, a pessoa-autor não existe, assim como suas intenções para com ele. Nesse sentido, o leitor seria o grande responsável por interpretar as ideias e atribuir significados e sentidos ao trabalho literário: “O leitor é o espaço exato em que se inscrevem todas as citações de que uma escrita é feita – a unidade do texto não está em sua origem, mas em seu destino” (Barthes, 1977, p. 5). Portanto, Barthes é resolutivo

em suas considerações sobre a temática: mata-se o autor para que o texto se configure como materialidade imortal, que ganha novas significações e interpretações a partir do público que o lê.

Transpassando, também, a simbologia da morte, Michel Foucault apresentou suas considerações a respeito do assunto no seminário de 1969 *O Que é um Autor?* que foi convertido em texto e é uma espécie de resposta para as considerações fúnebres de Barthes sobre o tema. Sua fala parte da frase de Beckett: “que importa quem fala, alguém disse que importa quem fala”, porque ela, a seu ver, incorpora a indiferença cotidiana ligada ao debate da figura do autor em seu tempo: não importa quem fala, mas o que se fala – ou, em outras palavras, abandona-se o autor e preserva-se o texto.

O estudioso, ao longo de sua apresentação, se mostra um tanto quanto cético a respeito dessa indiferença e é neste aspecto que ele aloca seu ponto de vista acerca da conexão da escrita com a morte. Nas narrativas do passado, exemplificadas por ele a partir das epopeias gregas e das histórias das *Mil e Uma Noites*, buscava-se imortalizar heróis e adiar a chegada iminente da morte. Refletindo a respeito da postura teórica do tema em seu tempo, Foucault (1969, p. 7) afirma:

[...] Esse tema da narrativa ou da escrita feitos para exorcizar a morte, nossa cultura o metamorfoseou; a escrita está atualmente ligada ao sacrifício, ao próprio sacrifício da vida; apagamento voluntário que não é para ser representado nos livros, pois ele está consumado na própria existência do escritor. A obra que tinha o dever de trazer a imortalidade recebeu agora o direito de matar, de ser assassina do seu autor. [...]

O próprio escritor do texto literário, diz ele, trabalha para o apagamento de sua subjetividade, fundando-se no princípio de que sua marca “[...] não é mais do que a singularidade da sua ausência; é preciso que ele faça o papel de morto no jogo da escrita” (Foucault, 1969, p. 7). Contudo, tal ação denota consequências tanto para a noção de *obra* quanto para a noção de *escrita* que, segundo o estudioso, persevera na existência do autor. É a partir daí, dessa lacuna deixada pela morte do autor pela teoria e crítica literárias, que Foucault fará sua proposição analítica sobre o tema: ao deixar de lado a análise da *personagem do autor*⁴, através de uma perspectiva histórico-sociológica, o estudioso propõe a noção da *função-autor* enquanto elemento que exerce um *papel em relação ao discurso*.

4 Roger Chartier, ao revisitar a conferência de Foucault em *O que é um autor? Revisão de uma genealogia*, explica de forma bastante clara o que é a visão de seu antecessor acerca do personagem-autor: “O primeiro modo de análise (do personagem-autor) se ocupa do estatuto dos autores a partir do sistema de valorização no qual essa postura ou posição é tomada, pôde conduzir, como vocês sabem, a uma acumulação de saberes sobre as trajetórias biográficas, sobre as origens sociais, profissionais, culturais dos autores, não importando seu campo de atividade ou, em uma outra linguagem e com outros conceitos, a uma sociologia do campo literário” (Chartier, 2000, p. 27). Os estudos da teoria literária criaram um paradigma – imputado

Enquanto Barthes crava uma estaca em tudo aquilo que está ligado à existência do autor, Foucault, ainda que descentralize o sujeito-pessoa-física da concepção do texto literário, não propõe seu apagamento. Ele sugere o entendimento de sua existência através da função-autor⁵ que, segundo sua designação diz respeito à “[...] característica do modo de existência, de circulação e do funcionamento de certos discursos no interior de uma sociedade” (Foucault, 1969, p. 14). A nomenclatura que Foucault cria *imortaliza* a instância discursiva proposta por uma figura que *nasce* da cisão entre o autor-real e o locutor fictício.

Destaca-se, também, que a categoria da função-autor se liga à proposta da pluralidade de egos que, segundo Foucault, lhe é característica. Não é fácil definir a “voz” primordial que propulsiona a criação de uma obra, visto que ela é fruto de diversas instâncias discursivas, que passam, como dito, por diversos setores sociais. Na verdade, a tarefa de definir essa voz primordial talvez não seja necessária. Dentro do que é atribuído por Foucault, o que parece valer a pena é entender a função-autor dentro de suas particularidades discursivas⁶.

Entre essas duas visões de autoria, que passeiam entre a morte e a imortalidade, há uma terceira: a de Umberto Eco. No capítulo “Entre autor e texto”, presente em *Interpretação e Superinterpretação*, o escritor italiano parte do mesmo questionamento dos estudiosos franceses para propor sua visão teórica acerca da questão do autor na literatura: “[...] será que ainda podemos nos preocupar com o autor empírico de um texto?” (Eco, 1993, p. 79). Sua argumentação dialoga, em pontos específicos, tanto com questões levantadas por Barthes quanto por Foucault, uma vez que ele aposta tanto nas instâncias linguístico-discursivas presentes no texto literário para fornecer sua própria interpretação quanto no potencial analítico que os leitores têm quando em contato com o poema ou o romance.

pelas considerações da morte do autor de Barthes, que, em escritos posteriores, retoma a relação entre autor e texto ao propor o termo “biografema” – se não se levar em consideração para a análise do texto em si a *biografia do autor*, o que ele faz em sua vida privada. O “abandono” de Foucault e de Chartier disso que eles chamam de “personagem-autor” recai, a meu ver, nesta questão – abandono este com que sigo em concordância. Contudo, em outro momento desse artigo, serão analisadas algumas instâncias discursivas próprias do século XXI acerca da função-autor que se debruçarão na criação consciente de uma *biografia autoral* como parte de uma *performance*. Alerto meu leitor para não confundir as nomenclaturas quando, eventualmente, eu as retomar ao longo do texto.

5 É importante mencionar que Foucault determina quatro traços característicos da função-autor. O primeiro liga a instância discursiva à esfera jurídica e institucional que articula o universo dos discursos; o segundo pressupõe que a função autor nasce e se exerce uniformemente e da mesma maneira sobre todos os discursos; o terceiro propõe que ela é definida por uma operação complexa e, por fim, o quarto estabelece que a função-autor não se refere a um indivíduo real, mas dá lugar a vários egos (Foucault, 1969, p. 20)

6 Como sugerido pelo autor, “Trata-se de retirar do sujeito seu papel de fundamento originário, e de analisá-lo como uma função complexa e variável do discurso” (Foucault, 1969, p. 28).

Tendo isso em vista, ele se apropria da nomenclatura criada por seu aluno Mauro Ferraresi para explicar a sua visão sobre a existência de três tipos diferentes de autores:

Mauro Ferraresi sugeriu que entre o *autor empírico* e o *autor-modelo* (que nada mais é do que uma estratégia textual explícita) existe uma terceira figura, meio fantasmagórica, que batizou de *autor limiar*, ou o *autor no limiar* – o limiar entre a intenção de um determinado ser humano e a intenção linguística revelada por uma estratégia textual (Eco, 1993, p. 82, grifo próprio).

A defesa de Eco sobre a participação do autor para a finalidade de interpretação de uma obra é válida apenas quando ele ainda está vivo e pode ser consultado com o objetivo de a crítica literária mostrar as discrepâncias entre a intenção do autor e a intenção do texto. Dessa forma, Eco não mata nem o autor e nem o leitor, ao passo que dá ao texto protagonismo por este ser revelador de determinadas estratégias linguísticas – sejam elas de um dado tempo histórico, sejam elas frutos da imaginação de um autor empírico.

Os três teóricos aqui apresentados configuram suas inquietações sobre o tema da autoria nos anos de 1967, 1969 e 1993, respectivamente, tempo que os insere nos estudos da literatura produzida na contemporaneidade. A escolha desse recorte teórico específico para abordar a figura do autor na literatura não foi ao acaso: a obra que este trabalho se dedica a analisar, *Harry Potter*, é contemporânea e carrega consigo marcas do nosso tempo em tudo o que envolve seu projeto literário, do texto propriamente dito à figura autoral que o escreveu: J.K. Rowling.

Na contemporaneidade, é impossível falar da obra *Harry Potter* sem debater sobre sua figura criadora, principalmente à luz de alguns episódios ocorridos nos primeiros anos desta segunda década do século XXI. Recentemente, o nome de J.K. Rowling esteve no centro de uma imensa polêmica político-social. Em sua conta no Twitter, a autora de *Harry Potter* escreveu postagens de cunho transfóbico, que geraram uma maciça onda de cancelamento *online* e muita decepção e revolta por parte dos fãs que, até então, tinham encontrado em sua obra um espaço de representatividade e acolhimento. A autora tem apresentado opiniões muito problemáticas acerca da comunidade *trans* ao condenar a existência do outro com base em seu gênero ou orientação sexual, sendo, hoje, uma *persona non grata* em inúmeros círculos sociais.

Neste sentido, as questões teóricas envolvendo a autoria são cabíveis no que concerne à tentativa de entender Joanne Rowling como uma autora que trabalha com questões estéticas próprias da contemporaneidade, como a *biografia do autor*, a *performatividade*, a *ilusão autoral* e a criação de *personas* para veicular diferentes textos, sejam eles literários, relatos pessoais, entrevistas para diferentes mídias, peças de publicidade ou textos de cunho político. Para o tempo histórico em que *Harry Potter* se insere, é uma tarefa muito difícil “matar” Rowling, como sugere Barthes. Embora seu texto seja seminal para o

campo de estudo em que este trabalho se insere, é impossível dissociar a imagem da mulher loira de olhos cansados, que carrega em seu semblante traumas sofridos em outros tempos, da imagem do menino de cabelos negros, de olhos igualmente cansados, que carrega a responsabilidade do mundo bruxo em suas costas com o mesmo peso com que carrega sua cicatriz em forma de raio.

Os pressupostos de Foucault e Eco parecem ser mais certos para equacionar a questão do autor na interpretação que se segue. A análise de declarações providas de fontes diversas, bem como a leitura atenta dos livros de *Harry Potter* mostram que, no caso de Rowling, há a veiculação de *diferentes instâncias discursivas*, oriundas da *função-autor*, produzidas por uma *autora* que está no *limiar* – sobretudo no limiar entre aquilo que permeia a esfera privada de sua vida e a sua esfera pública de ação, com o “tempero” da contemporaneidade, conforme mencionado anteriormente.

J.K. Rowling: a narrativa da *self-made woman*

Durante os anos de publicação de *Harry Potter*, uma outra história era construída em conjunto com as aventuras de Harry, Rony e Hermione: em uma espécie de narrativa de formação, com um certo contorno de conto de fadas, as editoras que publicavam a série veiculavam nas orelhas das primeiras edições dos livros como Rowling havia superado obstáculos para conseguir publicar seus romances, que venceram vários prêmios – um dos quais a ajudou, financeiramente, a terminar seu trabalho. Além disso, há, nessas publicações, um reforço de que o projeto literário de *Harry Potter* parece ter sido, ele mesmo, produto da magia que circunda a série.

Com a finalidade de estabelecer um recorte metodológico e a facilidade de acesso a esse tipo de informação⁷, replicam-se aqui os textos de orelha das primeiras edições de *Harry Potter* que foram publicadas no Brasil. Nesse caso, é importante destacar que Rowling já era um caso de sucesso mundo afora quando os livros chegaram ao nosso país e, portanto, as notas biográficas atreladas aos livros evidenciam essa questão, junto do discurso formativo e fantasioso mencionado anteriormente:

Britânica, 34 anos, J.K. Rowling é autora da série *Harry Potter*, que já foi vendida em 42 países e traduzida para 31 idiomas. *Para se manter em Edimburgo, na Escócia, começa a escrever o primeiro livro da série, no café local. Recebe, então, um prêmio inaudito do Scottish Arts Council, que permite que Rowling conclua seu trabalho. A partir daí, o romance de estreia da autora torna-se um fenômeno internacional, acumulando críticas entusiastas e prêmios importantes como o British Book*

7 Não foi possível encontrar primeiras edições originais dos volumes iniciais de *Harry Potter* circulando por nosso país, por isso optei por trabalhar com as notas biográficas das primeiras edições traduzidas para o português brasileiro.

Awards Children's Book of the Year e o Smarties Prize (Nota biográfica de *Harry Potter e a Pedra Filosofal*, 2000, grifo nosso).

[...]

A ideia de escrever Harry Potter ocorreu a J.K. Rowling em uma viagem de trem de Manchester para Londres, na qual, segundo ela, Harry Potter 'simplesmente entrou na minha cabeça inteiramente formado', e quando ela finalmente chegou à estação de King's Cross, muitos personagens já haviam tomado forma. Nos cinco anos seguintes, ela esboçou as tramas para cada livro e começou a escrever o primeiro da série, *Harry Potter e a Pedra Filosofal*, lançado na Inglaterra em 1997 [...] (Nota biográfica de *Harry Potter e a Ordem da Fênix*, 2003, grifo próprio).

A obra *mater* de Rowling foi publicada em um espaço de dez anos, de 1997 a 2007. Ao longo desse tempo, foi reiterado de diversas formas, tanto pela imprensa quanto pela própria autora em entrevistas, como Potter havia sido um ponto de virada em sua vida⁸, borrando as fronteiras entre aquilo que é entendido como *vida privada* e *vida pública*. A história da mãe-solo desempregada que escrevia aquela que iria se tornar uma das séries mais conhecidas no mundo, em um café em Edimburgo, para manter a filha aquecida, se tornou quase tão famosa quanto *Harry Potter* – história essa ligada exclusivamente à J.K. Rowling, nome pelo qual a autora ficou mundialmente conhecida.

A história por detrás deste nome – J.K. Rowling – é, também, um capítulo integrante da narrativa acerca da vida da autora. Em uma das versões⁹ de seu *site* oficial, na seção biográfica "My Story" ("Minha História"), com uma voz discursiva em primeira pessoa, Rowling (página pessoal, tradução e grifo próprios) explica:

Demorou um ano para que o meu agente Christopher Little encontrasse uma editora. Muitas delas recusaram. Então, em junho de 1997, Bloomsbury publicou *Harry Potter e a Pedra Filosofal* sob o nome de J.K. Rowling. O "K" é de Kathleen, o nome de minha avó paterna. Eu o adicionei e usei as iniciais pelo pedido da editora, uma vez que eles disseram que isso faria os livros serem mais atrativo aos meninos.¹⁰

8 Em entrevista ao Dateline NBC, no ano de 2003, Rowling afirmou: "Eu estava mesmo quebrada e eu sei como é estar quebrada, passar por isso e não saber que está prestes a mudar. Essa é a parte mais importante. Eu não conseguia ver luz alguma no fim do túnel, e todo dia agradeço por nós termos comida na geladeira, por não ter mais que me preocupar com as contas e por saber que posso comprar roupas para a Jessie e... sim, eu agradeço por isso todos os dias". Data de acesso: 09/01/2023.

9 J.K. Rowling possui dois *sites* oficiais sob este nome: um dedicado para o público adulto e outro dedicado aos jovens leitores. As seções que contam sua biografia possuem nomenclaturas e vozes discursivas diferentes a depender da plataforma: para os adultos, a história é contada em terceira pessoa na seção "About"; para as crianças, a história é contada em primeira pessoa na seção "My story."

10 No original: It took a year for my agent Christopher Little to find a publisher. Lots of them turned it down. Then in June 1997 Bloomsbury published *Harry Potter and the Philosopher's Stone* under the name J.K. Rowling.

Seguindo um pedido editorial, ela não pôde se lançar no mercado livreiro com seu verdadeiro nome – Joanne. A inclusão do nome da avó paterna com a posterior supressão dos dois nomes próprios iniciais (J.K.), além de ter designado uma nomenclatura “neutra”, que não afastaria os jovens meninos de lerem a obra, acarretou a criação daquele que viria a ser o mais conhecido pseudônimo da escritora britânica.

Reforça-se o ponto de que a história de vida de J.K. Rowling é contada ao passo em que *Harry Potter* é publicado. São duas narrativas, dois discursos, que se constroem em par e que se interligam de diversas maneiras. Harry encontra seu ponto de virada na vida quando descobre que é um bruxo. Rowling, da mesma forma, ao colocar seu protagonista no mundo, também encontra seu ponto de virada. Sua vida sofrida, marcada por perdas pessoais¹¹ e violência (que faz ecoar as perdas pessoais e a extrema violência pela qual o próprio Harry passa), muda quando Christopher Little encontra na editora inglesa Bloomsbury uma chance de publicação – tal qual Harry encontra uma chance de sair de sua vida medíocre com seus parentes trouxas quando Rúbeo Hagrid o visita e anuncia que sua vaga na Escola de Magia e Bruxaria de Hogwarts está garantida desde o dia de seu nascimento.

É importante frisar, contudo, que não parece haver, em *Harry Potter*, um desejo de se produzir, conscientemente, uma narrativa com pontos autobiográficos ou que dialoguem puramente com o debate da literatura contemporânea das escritas de si. Existem outros gêneros textuais que compõem a obra e que parecem fazer mais sentido para a construção de sua interpretação, como o romance de formação e a fantasia. Porém, ao mesmo tempo, não se pode negar que a imagem autoral de J.K. Rowling, propagada por suas editoras, pela imprensa e, até mesmo, pelo seu *site* oficial, está atrelada, sim, à *Harry Potter*. Nesse sentido – e não perdendo de vista de que estamos falando de um pseudônimo – a *biografia autoral* de Rowling acaba por configurar uma *persona*, uma representação de uma pessoa que, no estilo *self-made woman*, venceu na vida através da arte da escrita. Além disso, essa biografia autoral se mistura com a série criada pela britânica, conectando aspectos de sua vida pessoal à ficção e mostrando que, de certa forma, Rowling é a sua obra.

Para essa *persona* ser criada, foi fundamental difundir entre os leitores aspectos da vida privada da autora. Quando retomamos a leitura da nota biográfica publicada na orelha de *A Pedra Filosofal*, encontramos algo que vai além das informações “britânica, 34 anos”, que já é algo pessoal por si só: “para se manter em Edimburgo, começa a escrever o primeiro livro da série no café local”. Referencia-se, aqui, as dificuldades financeiras como

The “K” stands for Kathleen, my paternal grandmothers name. I added it and used the initials at the publisher’s request, as they said it would make the books more appealing to boys.

11 Segundo informações do *site* voltado para o público adulto, a mãe de Rowling faleceu de esclerose múltipla em 1990, antes de *Harry Potter* ser publicado.

forma de se construir uma base para a narrativa do ponto de virada – este, conquistado, com aquilo que chancela a profissão de um escritor: os prêmios que ele ganhou e as críticas favoráveis que ele recebeu de especialistas.

Robert Galbraith: a narrativa de um ex-militar recluso

Algum tempo após a conclusão da publicação de *Harry Potter*, em abril de 2013, o mercado editorial inglês presenciou o lançamento do livro *The Cuckoo's Calling*, romance policial escrito por Robert Galbraith, autor estreante e desconhecido do público, que foi logo aclamado pela crítica especializada pela maneira confiante de escrever e desenvolver uma narrativa de mistério. Três meses após o lançamento do livro, o *The Sunday Times* revelou que Galbraith era, na verdade, mais um pseudônimo criado por Rowling para a escrita de novas histórias, dessa vez veiculadas para o público adulto. O livro em questão se tornaria o primeiro de uma série protagonizada pelos detetives Cormoran Strike e Robin Ellacott, que investigam crimes envoltos em muita violência dentro de uma Londres contemporânea.

Hoje em dia, os romances policiais contam com sete volumes publicados¹² e uma série adaptada para a televisão intitulada *C.B. Strike*, produzida pela BBC. Seu autor, assim como J.K. Rowling, possui um *site* próprio para a divulgação de seu trabalho, com a presença de uma seção de perguntas e respostas. Gostaria de me valer de algumas das partes desse questionário para dar luz à voz autoral utilizada por Galbraith:

P: Sempre foi ideia sua escrever esses livros sob um pseudônimo?

R: Sim, eu queria voltar para o início de uma carreira na escrita através deste novo gênero, trabalhando sem *hype* ou expectativa e recebendo críticas sinceras. Eu queria que isso fosse sobre a escrita. Foi uma experiência fantástica e eu gostaria que ela tivesse durado um pouquinho mais do que durou. Me sinto eternamente agradecida pelo feedback dos editores e leitores e por muitas resenhas excelentes. Ser Robert Galbraith sempre foi sobre o trabalho, que é minha parte favorita de ser uma escritora. *Desde que minha identidade foi descoberta, eu continuo a escrever como Robert para manter a distinção de outros trabalhos e também porque eu gosto de ter outra persona*¹³ (Galbraith, página pessoal, tradução e grifo próprios).

12 Em ordem cronológica, os volumes publicados são: *The Cuckoo's Calling* (2013), *The Silkworm* (2014), *Career of Evil* (2014), *Lethal White* (2018), *Troubled Blood* (2020), *The Ink Black Heart* (2022) e *The Running Grave* (2023).

13 No original: "P: Was it always your idea to write under a pseudonym for these books? R: Yes, I really wanted to go back to the beginning of a writing career in this new genre, to work without hype or expectation and to receive totally unvarnished feedback. I wanted it to be just about the writing. It was a fantastic experience and I only wish it could have gone on a little longer than it did. I was grateful at the time for all the feedback from publishers and readers, and for some great reviews. Being Robert Galbraith was all about the work, which is my favourite part of being a writer. Since my cover has been blown, I continue to write as Robert to keep the distinction from other writing and because I rather enjoy having another persona".

Vê-se que há uma escolha consciente de Rowling ao lançar no mercado editorial uma outra voz autoral, uma outra *persona* que, através dos temas vinculados nesses novos romances (violência – sobretudo violência contra a mulher – preconceito de classe e raça, discurso de ódio, dentre outros), quer marcar uma diferenciação entre esses trabalhos e os outros publicados sob a *persona* J.K. Contudo, um questionamento vale ser feito aqui: que tipo de “distinção de outros trabalhos” intenciona-se fazer? Faço-me essa pergunta porque, sob a batuta de J.K. Rowling, os lugares narrativos de violência, preconceitos de raça e classe e discurso de ódio aparecem de forma evidente em *Harry Potter* – um tipo de publicação que, através da voz autoral de Galbraith, é distinta dos romances *Strike*.

Na mesma seção de perguntas e respostas, narrativas sobre a vida privada de Galbraith são criadas, assim como as narrativas biográficas de J.K. Questionamentos sobre a escolha de um nome masculino e sobre a carreira militar do autor são feitos, os quais são apresentados abaixo:

P: Por que você escolheu escrever estes livros sob um pseudônimo masculino? Isso influencia a sua escrita de alguma forma?

R: *Eu certamente queria que minha persona escritora se mantivesse o mais distante possível de mim, então um pseudônimo masculino me pareceu uma boa escolha.* Isso não muda conscientemente a forma com que escrevo. *Eu acho que escrevo de maneira diferente porque este é um gênero muito diferente.*

P: Por que você decidiu que o “autor” tivesse um passado militar?

R: *Esse foi o jeito mais fácil e mais plausível para Robert saber como a Divisão de Investigação Especial opera e investiga.* Outro motivo para torná-lo um homem militar trabalhando na indústria de segurança civil era para dar uma desculpa sólida para não o fazer aparecer em público ou providenciar uma fotografia¹⁴ (Galbraith, página pessoal, tradução e grifo próprios).

É interessante notar que parece haver um certo padrão para as narrativas biográficas formuladas para as *personas* de Rowling: todas elas misturam suas sobrevidas autorais às dos protagonistas de seus livros. Nos romances policiais, o personagem Cormoran Strike também possui um passado militar e trabalhava para a Divisão de Investigação Especial. Durante uma operação no Oriente Médio, ele é atingido por uma bomba e perde sua perna, o que o obriga a abandonar a carreira militar, justificando sua atuação como detetive particular no tempo presente da narrativa. Há um eco entre a vida do personagem

¹⁴ No original: “P: Why have you chosen to write these books under a male pseudonym? Does it influence your writing in any way? R: I certainly wanted to take my writing persona as far away as possible from me, so a male pseudonym seemed a good idea. It doesn’t consciously change the way I write. I think I write differently, because it’s a very different genre. P: Why did you decide for the “author” to have a military background? R: It was the easiest and most plausible reason for Robert to know how the Special Investigation Branch operates and investigates. Another reason for making him a military man working in the civilian security industry was to give him a solid excuse not to appear in public or provide a photograph”.

Strike e a vida do autor Galbraith, assim como há um eco entre a vida de J.K. e a de Harry Potter. Além disso, a inclusão de uma seção de perguntas e respostas no *site* de Gailbraith me faz acreditar que há, mais uma vez, uma tentativa de aproximar as esferas privadas e públicas que permeiam a aura do que é ser um autor no tempo presente.

Há, ainda, um outro ponto que é digno de atenção no âmbito da mistura de vida pública e vida privada do autor no que se refere às narrativas dos romances protagonizados por Strike. Os crimes que são investigados pelo detetive são, em sua grande maioria, cometidos contra mulheres¹⁵. Traições, amputações de membros do corpo, estupro e assassinatos – nem Robin, a detetive-sócia de Strike, escapa de vários desses atos de violência, sendo constantemente atormentada pelo trauma de ter sido estuprada quando adolescente e traída por seu marido quando adulta.

As conexões desse *topos* narrativo com a vida privada de Rowling e sua vivência passada em um casamento violento são facilmente mapeáveis, havendo, mais uma vez, a sobreposição daquilo que é privado – mas que, como visto, foi amplamente divulgado – com o que é público. E, mais ainda, o dito popular que diz “a arte imita a vida” parece, nesse caso, ganhar contornos literais. Escrevendo como um *homem*, Rowling faz de sua ficção um meio para a denúncia e exposição da situação de vulnerabilidade das *mulheres*, apostando, mais uma vez, nas marcas cristalizadas da binaridade de gênero.

Retoma-se, portanto, a distinção de trabalho literário que Galbraith procura fazer em relação à J.K.: se em *Harry Potter* há a presença do debate sobre a questão de raça, classe, violência e intolerância através de uma voz autoral feminina, como homem Rowling teria a autorização de falar desses temas de forma mais gráfica? Mais evidente? Um livro escrito por um homem e voltado para o público adulto possui determinadas licenças estilísticas que um livro escrito por uma mulher e dirigido, em um primeiro momento, para as crianças, não tem?¹⁶ Junto disso, não se pode negar a presença de um subtexto nos romances de Galbraith que divulga os mais recentes valores e incômodos de Rowling.

A resposta para essas várias perguntas é, na verdade, bastante clara: Rowling *intenciona*, *afirma* falar de coisas diferentes a depender da voz autoral que comanda seus dois projetos literários, mas, na verdade, ambos estão ligados aos mesmos temas que lhe

15 No artigo “The Twitter Controversy as a Rowling Story: Mirroring Subtext, Narrative Misdirection and Literary Alchemy” (2021), John Granger, um dos maiores estudiosos de Rowling e de seu trabalho ao longo deste nosso século, afirma: “Rowling is a writer whose essential focus is about writing and reading and the transformative power of great writing; it is the focus of this weblog to explain that to serious readers. She does, however, have a *political* issue she wants or needs to put front and center in everything she writes, namely, men abusing women.”

16 É importante lembrar que o pseudônimo J.K. Rowling foi criado para não ser atrelado a nenhum gênero em específico, apostando na neutralidade, com o objetivo de atrair leitores meninos para o mundo de *Harry Potter*.

são caros. Seja como uma *self-made woman*, seja como um militar aposentado e recluso, Rowling sempre usou a sua literatura para dar voz à sua própria vida – que foi amplamente divulgada, como visto, em entrevistas, orelhas de livros e em seus próprios *sítes*.

Joanne Rowling no limiar: uma autora da contemporaneidade

Em *Escritas de si, escritas do outro: autoficção e etnografia na narrativa latino-americana contemporânea*, Diana Klinger faz uma intensa e importante revisão sobre a morte e o retorno do autor, tanto no debate da teoria literária, com as correntes propostas por Barthes, Foucault e Hal Foster, respectivamente, quanto nas implicações disso na construção de romances contemporâneos que se ancoram na perspectiva estilística da escrita de si. Embora a narrativa literária de *Harry Potter* não pareça se basear especificamente nessa tradição, conforme salientado anteriormente, é difícil não pensar que a narrativa biográfica produzida para a *persona* Rowling apresente, sim, características inerentes à contemporaneidade do retorno do autor, principalmente no que se refere à necessidade de misturar a esfera privada com a esfera pública para criar essa *persona* e aproximá-la do público.

Klinger (2006, p. 20) argumenta que muitos romances contemporâneos são marcados pela experiência de vida do autor porque nossa sociedade, desde o final do século XX, caminha para o “avanço da cultura midiática”:

Nela se produz uma crescente visibilidade do privado, uma espetacularização da intimidade e a exploração da lógica da celebridade [...]. Assistimos hoje a uma proliferação de narrativas vivenciais, ao grande sucesso mercadológico das memórias, das biografias, das autobiografias e os testemunhos; aos inúmeros registros biográficos na mídia, retratos, perfis, entrevistas, confissões, *talk shows* e *reality shows* [...] (Klinger, 2006, p. 20).

Para além dos romances que trazem em seu enredo a narrativa vivencial do autor, nota-se que, no tempo presente, há o consumo da história de vida dessas pessoas, sob o viés do que é privado justamente por conta dos recursos midiáticos, que vai da TV às redes sociais. J.K. Rowling se tornou amplamente conhecida na virada do século XX para o XXI, na consolidação da internet como meio rápido de acesso e propagação de informações e como meio de comunicação. Sua *persona* foi criada e moldada pelo tempo histórico em que está inserida; portanto, é “natural” que sua biografia autoral esteja sujeita a tornar visíveis aspectos do que é próprio da esfera particular, assim como a biografia autoral de Robert Galbraith. Isso os aproximam de seu público e das narrativas de *Harry Potter* e de *Cormorant Strike*. Indo além do que Klinger especifica em sua dissertação, não é apenas

a ficção literária que está em consonância com o *zeitgeist*¹⁷: a narrativa que se constrói em torno e sobre a figura do autor caminha em conjunto com o espírito de nossa época.

Phillip Lejeune debate essa questão no ensaio “A imagem do autor na mídia”, que integra o livro *O pacto autobiográfico* (2008). O texto mostra como a construção da imagem do autor é, de uma forma ou de outra, perpassada pelo texto – primeiro, o literário, ao estimular que o leitor mate a sua curiosidade sobre a pessoa que o escreve através de possíveis respostas para perguntas que o texto lança, criando uma “ilusão biográfica” (2008, p. 192); e, posteriormente, o biográfico que se constitui através do literário mas, principalmente, da mídia.

O argumento central de Lejeune para todas essas imagens de autor criada ao longo do tempo é a de que o encontro deste com o leitor, através do texto literário, das correspondências, das entrevistas em jornais e suplementos literários, das falas em rádio e das aparições na televisão, é moldado pela ideia de um *simulacro*: a percepção que temos da pessoa que escreve é mediada, inevitavelmente, pelo texto. A mídia, no tempo presente, inflaciona essa ilusão, porque tendemos a buscar aquilo que o autor diz em suas entrevistas em sua produção literária¹⁸.

Para ilustrar seu ponto, o estudioso francês fala do programa televisivo “Apostrophes”, que recebe romancistas para serem entrevistados. Sobre o que é apresentado nessas entrevistas, Lejeune afirma: “Qualquer autor que compareça ao programa deve, ao mesmo tempo, *apresentar* seu livro (fazer um resumo do conteúdo, declarar sua intenção) e *representá-lo* através de sua pessoa” (2008, p. 198). Se no passado o texto aproximava o leitor do autor, hoje é função do autor aproximar o leitor do texto através da ideia de que sua vida está, de alguma forma, inscrita nele: “É preciso parecer-se com seu livro, imitá-lo, colocá-lo em palavras, *ser ele próprio*” (2008, p. 199, grifo próprio). Com isso, dá-se, mais uma vez, o encontro entre autor – ou melhor, entre a *persona autoral* – e leitor, mediado, sempre, pela ilusão biográfica.

Em consonância com a ideia da criação de uma ilusão biográfica do autor, retorna-se ao texto de Klinger e para o conceito de *performance*, trazido por ela para sua argumentação a respeito da construção da imagem do autor. O conceito implica a ideia de artificialidade, de dramatização. Segundo a estudiosa,

17 “Portanto, uma primeira aproximação à escrita de si na ficção contemporânea deveria, sem dúvida, inscrevê-la no espaço interdiscursivo desses outros textos – não literários – da cultura contemporânea, que evidenciam que esta ficção está em sintonia com o ‘clima da época’” (2006, p. 21).

18 “[...] a mídia incentiva fatalmente a ilusão biográfica que leva a buscar a solução do mistério do próprio autor. O sistema continua sendo o mesmo, simplesmente estamos num período de ilusão e inflação galopantes” (2008, p. 195).

O conceito de *performance* deixaria ver o caráter teatralizado da construção da imagem de autor. Desta perspectiva, não haveria um sujeito pleno, originário, que o texto reflete ou mascara. Pelo contrário, tanto os textos ficcionais quanto a atuação (a vida pública) do autor são faces complementares da mesma produção de uma subjetividade, instâncias de atuação do eu que se tencionam ou se reforçam, mas que, em todo caso, já não podem ser pensadas isoladamente. O autor é considerado enquanto sujeito de uma *performance*, de uma atuação, que “representa um papel” na própria “vida real”, na sua exposição pública, em suas múltiplas falas de si, nas entrevistas, nas crônicas e autorretratos, nas palestras (Klinger, 2006, p. 59).

Quando se olha de volta à J.K. Rowling, pseudônimo da autora de *Harry Potter* e “detentora” das narrativas biográficas apresentadas aqui, construídas no contexto que também foi apresentado anteriormente, parece-nos cabível articular que sua representação pública passa pelas instâncias levantadas pelos dois teóricos trazidos aqui. A biografia da autora, que é conhecida e consumida mundialmente, mostra, de diversas maneiras, que Rowling é *Harry Potter* – menos o personagem, e mais a obra. A *persona* que performa o discurso da mulher que deu a volta por cima está ligada única e exclusivamente ao nome J.K. Rowling.

Em uma suposta contramão, Galbraith quer se manter distante desse entorno através de falas que denotam que sua existência está voltada única e exclusivamente para o trabalho de uma escrita *diferente* da outra, ainda que, na prática, ela não seja tão diferente assim. Tanto *Harry Potter* quando os livros de Strike apresentam *quests* para a resolução de mistérios e enigmas. Ambas as histórias, conforme salientado anteriormente, possuem raízes temáticas parecidas: a vida de Harry é perpassada pela violência do assassinato de seus pais, assim como a vida de Strike e dos vários clientes que ele atende, sem falar em outros temas que norteiam as narrativas.

O que parece haver de diferente entre essas vozes autorais é, justamente, as performatividades criadas em volta delas. J.K. adota uma aura maternal e acolhedora, presente na escrita de outros livros infanto-juvenis. Galbraith, por sua vez, tem um passado militar, trabalha na segurança civil e tem seu nome atrelado a Robert F. Kennedy¹⁹. As

19 Na seção de perguntas e respostas do *site* de Galbraith, há uma explicação para a escolha do nome que compõe o pseudônimo: “I chose Robert because it’s one of my favourite men’s names, because Robert F Kennedy is my hero and because, mercifully, I hadn’t used it for any of the characters in the Potter series or The Casual Vacancy. Galbraith came about for a slightly odd reason. When I was a child, I really wanted to be called ‘Ella Galbraith’, and I’ve no idea why. I don’t even know how I knew that the surname existed, because I can’t remember ever meeting anyone with it. Be that as it may, the name had a fascination for me. I actually considered calling myself L A Galbraith for the Strike series, but for fairly obvious reasons decided that initials were a bad idea. Odder still, there was a well-known economist called J K Galbraith, something I only remembered by the time it was far too late. I was completely paranoid that people might take this as a clue and land at my real identity, but thankfully nobody was looking that deeply at the author’s name.”

personas são criações opostas, que criam simulacros públicos opostos – mas seus projetos literários não parecem ser tão opostos assim.

Considerações finais

Em qualquer uma de suas vozes, Rowling não precisa – e nem deve – ser sacralizada, ainda que tenha sido feito um trabalho coercitivo editorial bastante eficiente para se construir sua imagem de mártir durante os anos de publicação de *Harry Potter*. Reitera-se, mais uma vez, que a construção das *personas* J.K. Rowling e Robert Galbraith, analisadas neste trabalho, são uma manifestação dos escritos contemporâneos em consonância com o retorno do autor ao texto, e por isso há uma mistura de vozes, de biografias e de temas.

No texto *O que é um autor?*, Michel Foucault equaciona a relação da escrita com a morte: na Antiguidade, a escrita relacionava-se com a imortalidade, retratando heróis épicos e jovens donzelas que contavam histórias por mil e uma noites, para não serem assassinadas; no seu próprio tempo, o filósofo mostra-nos que a escrita abraça a mortalidade, uma vez que a literatura apagou do texto escrito qualquer traço de subjetividade. O tempo presente, por sua vez, apresenta contrastes bem menos nítidos a respeito dessa questão: entende-se que a identidade é fragmentária, então por que a do autor não seria diferente? Joanne, à maneira de seu vilão Tom Riddle, dividiu sua identidade autoral em, ao menos, três *personas*, perpetuando-se no tempo e no espaço como J.K. Rowling, @jk_rowling e Robert Galbraith, como um meio de ganhar sobrevidas para além das páginas dos livros que escreve, se tornando, metaforicamente, imortal.

Referências

BARTHES, R. A morte do autor. In: BARTHES, R. *O Rumor da Língua*. São Paulo: Martins Fontes, 2004.

CHARTIER, R. *O que é um autor? revisão de uma genealogia*. São Carlos: EdUFSCar, 2021.

ECO, U. *Interpretação e superinterpretação*. São Paulo: WMF Martins Fontes, 2018.

FOUCAULT, M. *Ditos e Escritos III*. Rio de Janeiro: Forense Universitária, 2009.

GALBRAITH, R. *The Cuckoo's Calling*. Inglaterra: Mulholland Books, 2013.

GALBRAITH, R. *The Silkworm*. Inglaterra: Mulholland Books, 2014.

GALBRAITH, R. *Carrer of Evil*. Inglaterra: Mulholland Books, 2015.

GALBRAITH, R. *Letal White*. Inglaterra: Mulholland Books, 2018.

GALBRAITH, R. *Troubled Blood*. Inglaterra: Mulholland Books, 2020.

GALBRAITH, R. *The Ink Black Heart*. Inglaterra: Mulholland Books, 2022.

GALBRAITH, R. *The Running Grave*. Inglaterra: Mulholland Books, 2023.

GALBRAITH, R. Página Pessoal. Apresenta seção de perguntas e respostas. Disponível em: <https://robert-galbraith.com/>. Acesso em 22 fev. 2024.

GRANGER, J. *The Twitter Controversy as a Rowling Story: Mirroring Subtext, Narrative Misdirection, and Literary Alchemy*. Hogwarts Professor, 2020. Disponível em: <https://www.hogwartsprofessor.com/the-twitter-controversy-as-a-rowling-story-mirroring-subtext-narrative-misdirection-and-literary-alchemy/>. Acesso em: 8 fev. 2024.

KLINGER, D. *Escritas de si, escritas do outro: autoficção e etnografia na narrativa latino-americana contemporânea*. 3. ed. Rio de Janeiro: 7 Letras, 2016.

LEJEUNE, P. *O pacto autobiográfico*. 2. ed. Belo Horizonte: Editora UFMG, 2008.

ROWLING, J.K. *Harry Potter and the Philosopher's Stone*. London: Bloomsbury, 1997.

ROWLING, J.K. *Harry Potter and the Chamber of Secrets*. London: Bloomsbury, 1998.

ROWLING, J.K. *Harry Potter and the Prisoner of Azkaban*. London: Bloomsbury, 1999.

ROWLING, J.K. *Harry Potter and the Goblet of Fire*. London: Bloomsbury, 2000.

ROWLING, J.K. *Harry Potter and the Order of the Phoenix*. London: Bloomsbury, 2003.

ROWLING, J.K. *Harry Potter and the Half-Blood Prince*. London: Bloomsbury, 2005.

ROWLING, J.K. *Harry Potter and the Deathly Hallows*. London: Bloomsbury, 2007.

ROWLING, J.K. *Very Good Lives: The Fringe Benefits of Failure and the Importance of Imagination*. Inglaterra: Little Brown and Company, 2015.

ROWLING, J.K. Página pessoal. Disponível em: <https://www.jkrowling.com/>. Acesso em: 22 fev. 2024.

Negação de atos de fala no português brasileiro por meio de “não”: uma visão da Gramática Discursivo-Funcional

DOI: <http://dx.doi.org/10.21165/el.v54i1.3835>

Gabriel Henrique Galvão Passetti¹
Erotilde Goreti Pezatti²

Resumo

Este estudo analisa a negação de atos de fala expressa por “não” no português brasileiro. Como referencial teórico, adota-se a Gramática Discursivo-Funcional (GDF), de Hengeveld e Mackenzie (2008). Ao tratar da negação, os autores (2018) denominam Rejeição a negação de atos de fala, ou, nos termos da GDF, negação de Ato Discursivo (A-negação). O objetivo é correlacionar as propriedades funcionais da A-negação a seus aspectos formais. O método de análise é dedutivo-indutivo, de caráter qualitativo. Os resultados mostram que A-negação é um Ato Discursivo Interativo, com llocução nucleada pela Ação Lexical “não”, que rejeita qualquer tipo de Ato Discursivo, exceto Expressivo, sendo codificada diretamente no Nível Fonológico como um Enunciado. Ao propor uma definição mais precisa e abrangente da A-negação, este trabalho busca contribuir para o aperfeiçoamento do modelo da GDF.

Palavras-chave: negação; ato de fala; Gramática Discursivo-Funcional.

1 Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), São José do Rio Preto, São Paulo, Brasil; gabriel.galvao@unesp.br; <https://orcid.org/0000-0001-5001-6666>

2 Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), São José do Rio Preto, São Paulo, Brasil; erotilde.pezatti@unesp.br; <https://orcid.org/0000-0001-8822-9587>

Negation of speech acts in Brazilian Portuguese using “*não*”: a Functional Discourse Grammar view

Abstract

This study analyzes the negation of speech acts expressed by “*não*” in Brazilian Portuguese. The Functional Discourse Grammar (FDG) by Hengeveld and Mackenzie (2008) is used as a theoretical framework. When dealing with negation, the authors (2018) call the negation of speech acts Rejection, or, in the terms of the FDG, negation of Discourse Act (A-negation). The objective is to correlate the functional properties of A-negation with its formal aspects. The method of analysis is deductive-inductive and qualitative. The results show that A-negation is an Interactive Discourse Act, with an Illocution centered on the Lexical Deed ‘no’, which rejects any type of Discourse Act, except Expressive. It is directly encoded at the Phonological Level as an Utterance. By proposing a more precise and comprehensive definition of A-negation, this work seeks to contribute to the improvement of the FDG model.

Keywords: negation; speech act; Functional Discourse Grammar.

Introdução³

Segundo Horn (2001), todos os sistemas humanos de comunicação preveem a existência da negação. O caráter universal desse fenômeno tem despertado o interesse de estudiosos da linguagem desde os estudos de Platão e de Aristóteles até os dias atuais. Entretanto, a forma como o tema é abordado distancia o estudo da negação em línguas naturais dos trabalhos desenvolvidos pelos lógicos. Se, na lógica, há uma relação simétrica entre afirmação e negação, sendo a negação entendida como um operador que atua sobre o valor de verdade de uma proposição, isso não se observa nas línguas naturais.

A negação que inverte o valor de verdade de uma proposição é apenas um tipo de negação descrito por Hengeveld e Mackenzie (2018). Os autores elencam uma série de negações, que incidem sobre diversas camadas previstas pela Gramática Discursivo-Funcional (GDF), tanto do Nível Representacional, concernente à semântica, como do Nível Interpessoal, que diz respeito à pragmática.

A negação formulada no Nível Representacional é descritiva e submetida a condições de verdade (*truth-conditional*), desde a negação que inverte o valor de verdade de uma

3 Este artigo divulga parte dos resultados da tese de doutorado de Galvão Passetti (2025).

proposição, denominada Discordância, até a negação de lexema, denominada Antônimo, em que um processo de derivação indica o contrário do sentido de um lexema.

A negação do Nível Interpessoal, por outro lado, não é descritiva, mas sim acional, ou seja, não é submetida a condições de verdade (*non-truth-conditional*), o que significa que ela não expressa significado negativo em sentido estrito, já que não descreve o mundo externo, conforme faz a negação semântica, mas representa uma ação que, em si mesma, contém um valor negativo.

É esse tipo de negação que este trabalho investiga, especificamente a negação acional no português brasileiro (doravante PB) marcada por “não” que ocorre no exemplo (1).

1. aí ele falô(u) – “eu vô(u) pegá(r) minha RO(u)pa e vô(u) tomá(r) banho já e:: passo aqui né? po... pra mim í(r) embora... – “FICA AQUI COM AS MENINA” – ... eu falei – “**não** eu também vô(u) em casa tomá(r) banho” [AI-001; CAS: L. 13]

A negação acional em (1) rejeita o ato de fala “fica aqui com as menina”. O mesmo ocorre em (2), outro exemplo de negação de Ato Discursivo (doravante A-negação).

2. Hengeveld e Mackenzie (2018, p. 36):
A: Go home! (“Vá para casa!”)
B: No! (“Não!”)

Tanto em (1) como em (2), “não” e “no” servem para o Participante B reagir ao Ato Discursivo (A) executado pelo Participante A. Por ser uma ação, a A-negação não deve ser entendida como a não execução ou o desfazimento de um Ato Discursivo. Hengeveld e Mackenzie (2018) apontam para essa característica ao discorrerem sobre negações acionárias, afirmando que “um falante não pode, ao mesmo tempo, realizar uma ação e negar que a está realizando” (p. 35, tradução própria)⁴. Nesse sentido é que os autores definem a A-negação como negação acional, distinguindo-a da proposicional, ambas marcadas por “no” em inglês (cf. Hengeveld; Mackenzie, 2008, p. 148-149; 2018, p. 36), conforme exemplificam (3) e (4), respectivamente.

3. Hengeveld e Mackenzie (2018, p. 36):
A: Did John go home? (“João foi para casa?”)
B: No (he didn't). (“Não (ele não foi).”)
4. Hengeveld e Mackenzie (2018, p. 36):

4 No original: “A speaker cannot at the same time carry out an action and negate that he/she is doing so [...]”

A: Go home! ("Vá para casa!")

B: No!/Forget it!/Get lost! (I don't accept your order.) ("Não!/Sai fora!/Nem vem! (eu não aceito sua ordem.)")

Em (3), "*no*" especifica o valor de verdade de um Conteúdo Proposicional, colocado em questionamento pelo Participante A. Em (4), por outro lado, "*no*" serve para o Participante B rejeitar o Ato Discursivo Imperativo "*Go home!*", já que o Participante A, na concepção do Participante B, não está em posição de lhe dizer o que fazer, ou seja, de lhe dar uma ordem. A partir dessa explicação, oferecida por Hengeveld e Mackenzie (2018), pode-se definir a A-negação como a rejeição de um ato de fala imperativo baseada na relação assimétrica entre os Participantes do discurso.

Para se rejeitar um Ato Discursivo, há várias opções (cf. 4B), que não podem ser usadas para especificar a polaridade negativa de um Conteúdo Proposicional (cf. 5B).

5. Hengeveld e Mackenzie (2018, p. 36):

A: Did Peter go home? ("Peter foi para casa?")

B: No!/*Forget it!/*Get lost! ("Não!/*Sai fora!/*Nem vem!")

Esse, portanto, é um teste profícuo para distinguir a A-negação, e foi o utilizado para selecionar os dados deste trabalho, como a ocorrência em (1), repetida por conveniência em (6), em que "não" pode ser substituído por "nem vem!".

6. aí ele falô(u) – "eu vô(u) pegá(r) minha RO(u)pa e vô(u) tomá(r) banho já e:: passo aqui né? po... pra mim í(r) embora... – "FICA AQUI COM AS MENINA" – ... eu falei – "não[/**nem vem!**]
eu também vô(u) em casa tomá(r) banho" [AI-001; CAS: L. 13]

Como evidenciado em (6), a A-negação nega a realização de um Ato Discursivo, ou seja, desautoriza um Ato Discursivo anteriormente proferido, sendo denominada Rejeição por Hengeveld e Mackenzie (2018) e representada como (7).

7. Hengeveld e Mackenzie (2018, p. 36):

(A₁: no (A₁))

A representação em (7) mostra que a A-negação é formulada como um Ato Discursivo que tem por função rejeitar outro Ato Discursivo anteriormente executado no discurso. Em (7), no entanto, não são oferecidas outras propriedades desse Ato Discursivo, como, por exemplo, que unidade, **núcleo** desse Ato Discursivo, "*no*" compõe.

Além de responder a essa pergunta, o objetivo deste trabalho é investigar a A-negação no PB de modo a correlacionar suas propriedades funcionais às formais sob a ótica da GDF. Os dados são coletados de páginas da *web* sediadas no Brasil, com exceção de (1), única ocorrência de A-negação encontrada no *corpus* Iboruna (Gonçalves, s.d). O método de análise, por sua vez, é dedutivo-indutivo, de caráter qualitativo.

A hipótese é a de que, na A-negação, “não” é formulado, no Nível Interpessoal, como um Ato Discursivo Interativo, que é codificado diretamente no Nível Fonológico por um Enunciado, contornando os níveis Representacional e Morfossintático.

Para cumprir o objetivo e verificar a validade da hipótese, o texto está estruturado da seguinte forma: primeiramente, expõe-se o referencial teórico adotado, a GDF; na sequência, define-se Ato Discursivo; em seguida, elencam-se as llocuções abstratas que Atos Discursivos apresentam no PB; dadas as definições necessárias, passa-se à análise e discussão dos dados; por fim, tecem-se as considerações finais deste estudo.

A Gramática Discursivo-Funcional

Proposta por Hengeveld e Mackenzie (2008), a GDF, que é um desenvolvimento da GF postulada por Dik (1997a, 1997b), mantém, em seu modelo teórico, a natureza situada da comunicação linguística, ou seja, apresenta inter-relação entre linguagem e contexto. Essa inter-relação é expressa pelo Componente Conceptual, pelo Componente Contextual e pelo Componente Articulatório. Esses componentes dão compatibilidade à GDF com uma teoria mais ampla da interação verbal.

O Componente Conceptual é pré-linguístico e responsável pelo desenvolvimento tanto da intenção comunicativa do evento de fala como das conceptualizações associadas relativas a eventos extralinguísticos relevantes, sendo, dessa forma, a força motriz por trás do Componente Gramatical (cf. Hengeveld; Mackenzie, 2012, p. 44), incluindo apenas os aspectos da cognição que afetam a intenção comunicativa imediata.

O Componente Articulatório, por sua vez, é o responsável por gerar as expressões linguísticas (acústicas, escritas ou de sinais) com base na informação fornecida pelo Componente Gramatical. Sua função é traduzir a “informação digital (isto é, categorial, baseada em oposição) na gramática para uma forma analógica (isto é, continuamente variável)” (Hengeveld; Mackenzie, 2012, p. 45).

A Gramática Discursivo-Funcional busca compreender a estrutura das expressões linguísticas em seu contexto discursivo, embora não seja um modelo analítico-discursivo. Assim, a GDF leva em conta o contexto discursivo, já que a intenção do Falante não surge no vácuo, mas num contexto comunicativo multifacetado. Esse contexto é capturado pelo

Componente Contextual, compartilhado e construído conjuntamente pelos interlocutores à medida que a troca comunicativa acontece.

Abrigando apenas informações relevantes para a forma dos enunciados, a GDF adota o que Butler (2008) chama de “postura conservadora” para o Componente Contextual (cf. Connolly, 2007; Alturo; Keizer; Payrato, 2014; Hengeveld; Mackenzie, 2014). Esse acervo de informações, de curto e de longo prazo, alimenta e é alimentado pelas operações de formulação e de codificação do Componente Gramatical, que, por sua vez, constitui a gramática de uma língua natural. As operações de formulação convertem a intenção comunicativa em representações pragmáticas, no Nível Interpessoal (NI), e semânticas, no Nível Representacional (NR), que, em seguida, são convertidas em representações morfossintáticas e fonológicas nos níveis Morfossintático (NM) e Fonológico (NF), respectivamente, por meio das operações de codificação.

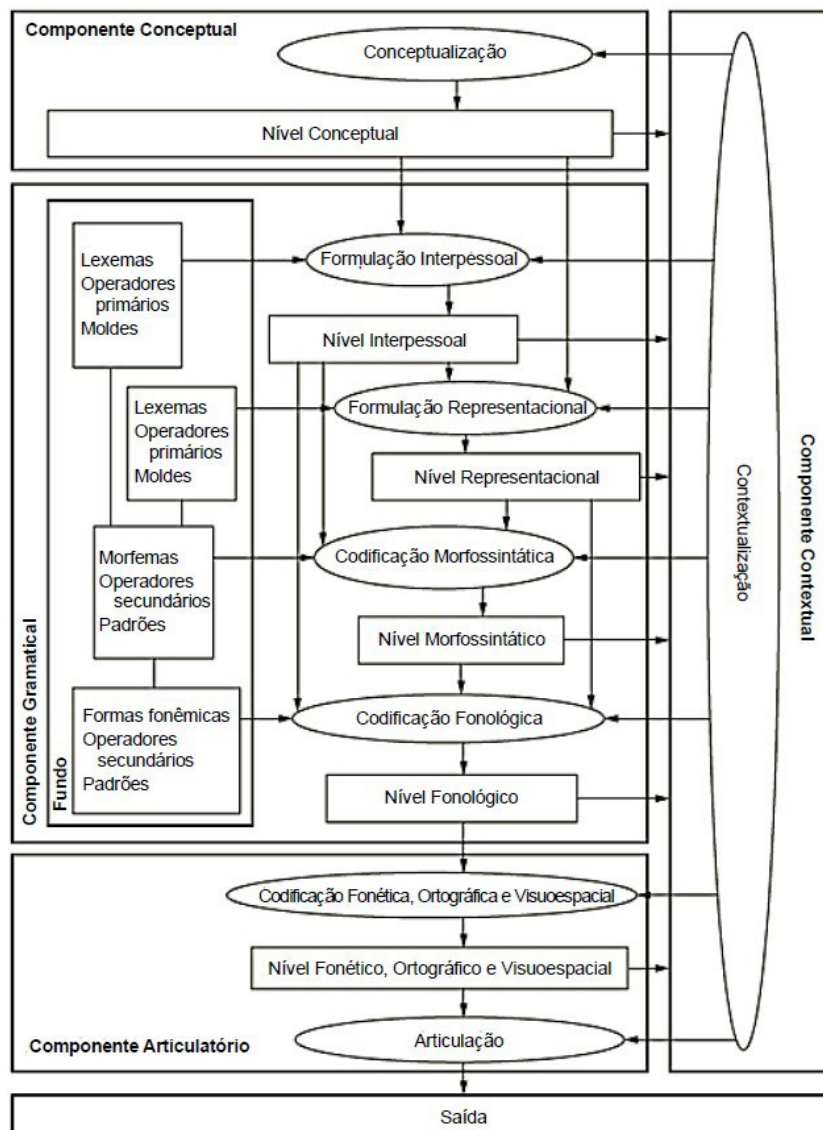
Os níveis que formam o Componente Gramatical são estruturados, cada qual, ao seu modo. O que têm em comum é que são todos dispostos em camadas. Cada camada é composta de um núcleo (h, H), que pode ser restringido por um modificador (σ, Σ) e/ou um operador (π, Π) e ter ainda uma função (φ, Φ). Os núcleos e os modificadores são lexicais, enquanto os operadores e as funções são gramaticais, sendo as funções de carácter relacional, ou seja, elas estabelecem relação entre unidades dispostas numa mesma camada. Assim, a representação geral das camadas dentro dos níveis é como (8), sendo “v” a variável da camada relevante.

$$8. \quad (\pi v_1 : h(v_1) : \sigma(v_1))_{\varphi}$$

Como mostra a Figura 1, o modelo da GDF apresenta uma arquitetura modular com organização descendente, ou seja, da intenção para a forma das expressões linguísticas, direção “motivada pela suposição de que um modelo de gramática será mais eficaz quanto mais sua organização assemelhar-se ao processamento de língua no indivíduo” (Hengeveld; Mackenzie, 2008, p. 1-2, tradução própria)⁵, alcançando, assim, adequação psicológica, como demonstram estudos psicolinguísticos (e.g., Levelt, 1989).

⁵ No original: “[...] motivated by the assumption that a model of grammar will be more effective the more its organization resembles language processing in the individual.”

Figura 1. Arquitetura geral da Gramática Discursivo-Funcional



Fonte: Adaptado de Hengeveld, Keizer e Giomi (em preparação)⁶.

Convém ressaltar que a GDF é “um modelo de intenções e conceptualizações codificadas” (cf. Hengeveld, 2004, p. 366-377; Hengeveld; Mackenzie, 2008, p. 2), e não um modelo de produção da linguagem, como o de Levelt (1989). Assim, o objetivo da GDF é entender como as expressões linguísticas são estruturadas de acordo com as intenções comunicativas e o mundo que descrevem, modelando-as numa “implementação dinâmica” (cf. Bakker; Siewierska, 2004). Nesse sentido, a GDF é um modelo de padronização, e não um modelo

⁶ A adaptação consiste na tradução do original inglês para o português.

de processo da língua, embora inspirado pelo processo, sem, contudo, formalizá-lo (cf. Hengeveld; Mackenzie, 2008, p. 24).

Por ter organização descendente e ser um modelo orientado pelo discurso, a GDF assume o Nível Interpessoal como hierarquicamente acima dos outros níveis e, desse modo, alcança adequação pragmática, como preconiza Dik (1997a, 1997b) em seus princípios de adequação explanatória. Disso, decorre a organização do Componente Gramatical, em que a pragmática governa a semântica, ambas governam a morfossintaxe, e a pragmática, a semântica e a morfossintaxe governam a fonologia, como ilustrado na Figura 1. Assim, “a GDF leva a abordagem funcional de língua ao seu extremo lógico” (Hengeveld; Mackenzie, 2008, p. 13, tradução própria)⁷, já que a postura funcionalista implica a “hipótese de que as categorias formais podem ser criteriosamente explicadas se consideradas em correspondência com as categorias semânticas e pragmáticas originadas na cognição humana e na comunicação inter-humana” (Hengeveld; Mackenzie, 2012, p. 48), correlacionando as funções às estruturas, ambas sedimentadas no repertório da língua ao longo dos tempos como primitivos das operações de formulação e de codificação.

A GDF assume uma abordagem “função-para-forma” orientada para a forma (Hengeveld; Mackenzie, 2008, p. 38-39). É orientada para a forma porque descreve apenas os fenômenos interpessoais e representacionais que são refletidos morfossintática e/ou fonologicamente e é “função-para-forma” porque explica a codificação morfossintática e fonológica com base nas estruturas pragmáticas e semânticas subjacentes, que, por sua vez, modulam as intenções comunicativas do Falante.

O Ato Discursivo

O Ato Discursivo é a unidade básica de análise da GDF e define-se como a menor unidade identificável de comportamento comunicativo, que se constitui de, no máximo, quatro componentes: Ilocução (F), Falante (P)_S, Ouvinte (P)_A e Conteúdo Comunicado (C). Enquanto a Ilocução atribui usos interpessoais convencionalizados a Atos Discursivos na consecução de uma intenção comunicativa do Falante, um Conteúdo Comunicado contém a totalidade do que o Falante deseja evocar do mundo externo na comunicação com o Ouvinte.

Atos Discursivos que contêm apenas Ilocução e Falante são chamados de Expressivos, pois expressam sentimentos do Falante em vez de comunicar algum conteúdo ao Ouvinte, como a interjeição “ai” de dor em (9).

⁷ No original: “FDG takes the functional approach to language to its logical extreme.”

9. Ai!

NI: $(A_i: [(F_i: -ai- (F_i)) (P_i)_S] (A_i))$

Um Ato Discursivo pode, ainda, ser Interativo e conter também o Ouvinte, caso em que o Ato Discursivo é expresso por uma interjeição que representa uma expressão socialmente dirigida, como exemplifica (10).

10. Parabéns!

NI: $(A_i: [(F_i: -parabéns- (F_i)) (P_i)_S (P_j)_A] (A_i))$

Em (10), “parabéns” é um Ato Discursivo Interativo que não tem Conteúdo Comunicado, mas, ao contrário dos Expressivos, os Interativos podem veicular um Conteúdo Comunicado, caso em que são Interativos de Conteúdo, como (11).

11. Parabéns por passar no vestibular!

NI: $(A_i: [(F_i: -parabéns- (F_i)) (P_i)_S (P_j)_A (C_i: -passar no vestibular- (C_i))] (A_i))$

Em (11), a expressão socialmente dirigida “parabéns” é responsável pela intenção comunicativa de congratular o Ouvinte, ao passo que o Conteúdo Comunicado contém o fato pelo qual o Falante expressa congratulações ao Ouvinte.

Já em (12), A_i apresenta Ilocução abstrata, Declarativa (DECL), ou seja, tem o objetivo de informar o Ouvinte sobre o Conteúdo Proposicional evocado pelo Conteúdo Comunicado.

12. (Eu declaro que) Eu não confio neste governador.

NI: $(A_i: [(F_i: DECL (F_i)) (P_i)_S (P_j)_A (C_i: -eu não confio neste governador- (C_i))] (A_i))$

A representação não instanciada do Ato Discursivo é oferecida em (13), em que os constituintes obrigatórios do Ato Discursivo são a Ilocução e o Falante, já que um Ato Discursivo necessariamente satisfaz uma intenção comunicativa do Falante, mas nem sempre prevê a comunicação de algo ao Ouvinte. Já a representação do Ato Discursivo a depender de sua Ilocução e das unidades que o compõem é apresentada no Quadro 1.

13. $(\Pi A_i: [(F_i) (P_i)_S \{(P_j)_A\} \{(C_i)\}] (A_i): \Sigma (A_i))_\Phi$

Quadro 1. Tipo de Ato Discursivo de acordo com as unidades que o compõem

<i>Tipo de Ato Discursivo</i>	<i>Representação não instanciada</i>
Expressivo	(A ₁ : [(F ₁ : (D ₁) (F ₁)) (P ₁) _s] (A ₁))
Interativo	(A ₁ : [(F ₁ : (D ₁) (F ₁)) (P ₁) _s (P ₂) _A] (A ₁))
Interativo de Conteúdo	(A ₁ : [(F ₁ : (D ₁) (F ₁)) (P ₁) _s (P ₂) _A (C ₁)] (A ₁))
de Conteúdo	(A ₁ : [(F ₁ : ILL/(D ₁) (F ₁)) (P ₁) _s (P ₂) _A (C ₁)] (A ₁))

Fonte: Adaptado de Hengeveld e Mackenzie (2008, p. 64).⁸

Conforme o Quadro 1, a Ilocução do Ato Discursivo de Conteúdo pode ser lexical (D₁) ou abstrata (ILL). O último caso é tratado em mais detalhes na seção a seguir.

As Ilocuções abstratas do PB

As Ilocuções abstratas, também chamadas de Ilocuções prontas por Hengeveld e Mackenzie (2008, p. 70), são definidas como “uma coincidência de estrutura gramatical e uso conversacional convencional” (Sadock; Zwicky, 1985, p. 155, tradução própria)⁹. Existem vários tipos de Ilocuções abstratas, que, segundo Hengeveld e Mackenzie (2008), são: Declarativa (DECL); Interrogativa Polar (INTER_p), Interrogativa de Conteúdo (INTER_c); Imperativa (IMP); Proibitiva (PROH); Optativa (OPT); Imprecativa (IMPR); Exortativa (HORT), Dexortativa (DISHORT); Admoestativa (ADMON); Comissiva (COMM); Suplicativa (SUPPL); e Exclamativa (EXCL)¹⁰.

Nem toda língua dispõe de primitivos gramaticais que codificam todas as Ilocuções abstratas. Nesse caso, as Ilocuções abstratas não distinguidas são expressas por meios lexicais ou por outras Ilocuções abstratas. No PB, apenas as Ilocuções Declarativa, Exclamativa, Optativa, Interrogativa (Polar e de Conteúdo), Imperativa e Exortativa têm primitivos gramaticais correspondentes. As cinco primeiras são Ilocuções Proposicionais, ao passo que as duas últimas são Comportamentais.

8 Além da organização dos tipos de Atos Discursivos num quadro, a adaptação consiste na inserção da Ação Lexical (D), proposta por Giomi (2020), no núcleo da Ilocução, quando é expressa por verbos performativos em Atos Discursivos de Conteúdo ou por interjeições em outros tipos de Ato Discursivo.

9 No original: “[...] a coincidence of grammatical structure and conventional conversational use.”

10 Hengeveld e Mackenzie (2008) denominam a Ilocução Exclamativa de Mirativa. Posteriormente, no entanto, Olbertz (2012) e Hengeveld e Olbertz (2012) reconhecem que a miratividade é uma categoria utilizada em Atos Discursivos com diferentes Ilocuções, não sendo, portanto, uma Ilocução *per se*, e admitem a existência da Ilocução Exclamativa.

A Ilocução Declarativa informa o Ouvinte de uma proposição evocada pelo Conteúdo Comunicado. No PB, essa Ilocução é codificada no Nível Fonológico por um contorno entonacional descendente, desencadeado pelo operador fonológico *Falling* na camada da Frase Entonacional. Um exemplo de Ato Discursivo Declarativo é (14).

14. Marçal pagará pelos seus crimes.

(Disponível em: <https://g1.globo.com/sp/sao-paulo/eleicoes/2024/noticia/2024/10/05/boulos-diz-que-vai-pedir-prisao-e-cassacao-de-marcal-apos-adversario-divulgar-suposto-laudo-sobre-drogas.ghtml>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: DECL (F_i)) (P_i)_S (P_j)_A (C_i: –Marçal pagará pelos seus crimes– (C_i))] (A_i))

A Ilocução Exclamativa, por sua vez, manifesta uma “avaliação pessoal” (Olbertz, 2012, p. 80) do Falante em relação ao Conteúdo Proposicional evocado pelo Conteúdo Comunicado. No PB, essa Ilocução é codificada por um operador morfossintático que, no Nível Fonológico, recebe a forma fonêmica de palavras-QU no início do enunciado, como “que” em (15).

15. Que discurso maravilhoso que a senhora fez na sua tribuna.

(Disponível em: <https://www.camaracaxias.rs.gov.br/atividadelegislativa/listarOrdem/3058>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: EXCL (F_i)) (P_i)_S (P_j)_A (C_i: –discurso maravilhoso que [...] tribuna– (C_i))] (A_i))

Já com a Ilocução Optativa, o Falante indica ao Ouvinte o seu desejo de que a situação evocada pelo Conteúdo Comunicado se concretize, como em (16). Essa Ilocução, no PB, é codificada pela Palavra Gramatical “que” no início da Oração somada ao Afixo verbal do modo do subjuntivo. O sentido optativo também é expresso por Atos Discursivos Interativos cujas interjeições são “oxalá” ou “tomara”, podendo ou não ser desenvolvidas por Conteúdos Comunicados.

16. Que Fernanda Torres ganhe.

(Disponível em: <https://www.youtube.com/watch?v=qfa0Y-Ec840>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: OPT (F_i)) (P_i)_S (P_j)_A (C_i: –Fernanda Torres ganhe– (C_i))] (A_i))

As Ilocuções Declarativa, Optativa e Exclamativa são Ilocuções Proposicionais informativas. Com relação às Ilocuções Proposicionais de questionamento, o PB dispõe tanto da Ilocução Interrogativa Polar como da de Conteúdo. A Ilocução Interrogativa Polar, de que a ocorrência em (17) é exemplo, serve para que o Falante solicite ao Ouvinte uma resposta do tipo “sim” ou “não” para o Conteúdo Proposicional evocado pelo

Conteúdo Comunicado. Essa Ilocução, graficamente marcada pelo sinal de interrogação, é codificada no PB por operadores fonológicos específicos.¹¹

17. O trabalho realmente dignifica o homem?

(Disponível em: <https://www.linkedin.com/pulse/o-trabalho-realmente-dignifica-homem-luciele-santana-ymif/>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: INTER_P (F_i)) (P_i)_S (P_j)_A (C_i: -O trabalho realmente dignifica o homem- (C_j))] (A_j))

Na Ilocução Interrogativa de Conteúdo, por sua vez, o Falante também requer do Ouvinte uma resposta para o Conteúdo Proposicional evocado pelo Conteúdo Comunicado, mas essa resposta implica a evocação de uma informação que atualize a informação pragmática disponível ao Ouvinte. No exemplo em (18), essa informação é o tempo em que a floresta amazônica perderá sua capacidade de autorregeneração.

18. *Quando seria*, numa estimativa em anos, *esse ponto de não-retorno?*

(Disponível em: <https://www.ufes.br/conteudo/entrevista-carlos-nobre-afirma-que-mudancas-climaticas-sao-o-maior-desafio-da-humanidade>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: INTER_C (F_i)) (P_i)_S (P_j)_A (C_i: -quando seria esse ponto de não-retorno- (C_j))] (A_j))

A Ilocução Interrogativa de Conteúdo é mapeada, no PB, nos dois níveis de codificação. No Nível Morfossintático, um operador morfossintático é acionado, assumindo a forma de pronomes interrogativos, como “quando” em (18). No Nível Fonológico, por sua vez, a expressão linguística é especificada por operadores fonológicos específicos.¹²

No que se refere às Ilocuções Comportamentais, o PB apresenta apenas as Ilocuções Imperativa e Exortativa. Por meio da Ilocução Imperativa, o Falante orienta o Ouvinte a realizar a ação evocada pelo Conteúdo Comunicado. Essa Ilocução é marcada pela não expressão do sujeito, pela linearização do Sintagma Verbal na posição inicial da Oração e, em alguns casos, pelo Afixo do modo imperativo da terceira pessoa do singular, como o Afixo “-a” no verbo “faça” em (19).

11 Não se define aqui quais operadores fonológicos e que camadas eles especificam porque sentenças interrogativas polares do PB apresentam variação dialetal em seus contornos entonacionais. Silva (2011), por exemplo, descreve esses contornos como ascendente (L+H*__ L+!H*H%) nas regiões norte e nordeste e como circunflexo (L+!H*__ L+H*L%) nas regiões centro-oeste, sudeste e sul do Brasil.

12 Assim como as sentenças interrogativas polares, as de conteúdo também apresentam variação dialetal em seus contornos entonacionais.

19. O lixo é seu. *Faça coleta seletiva*: condição para um mundo melhor.

(Disponível em: <https://brainly.com.br/tarefa/40069654>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: IMP (F_i)) (P_i)_S (P_j)_A (C_i: -faça coleta seletiva- (C_i))] (A_i))

Por fim, ao utilizar a Ilocução Exortativa, o Falante encoraja a si mesmo ou a si e ao Ouvinte ao mesmo tempo a realizar a ação evocada pelo Conteúdo Comunicado, como a ação descrita em (20). A Ilocução Exortativa, assim como a Imperativa, é codificada pela não expressão do sujeito, pela linearização do Sintagma Verbal na posição inicial da Oração e pelo Afixo do modo imperativo, mas da primeira pessoa do plural, como em (20).

20. Sigamos juntos e firmes na greve!

(Disponível em: <https://sindppd-rs.org.br/greve-na-procergs-ninguem-esta-sozinho-porque-nossa-luta-e-coletiva-sigamos-juntos-e-firmes-na-greve/>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: HORT (F_i)) (P_i)_S (P_j)_A (C_i: -sigamos juntos e firmes na greve- (C_i))] (A_i))

Na seção a seguir, descreve-se a A-negação no PB.

A negação de Ato Discursivo no PB: análise e discussão de dados

O objetivo deste trabalho é analisar e descrever as propriedades funcionais e formais da A-negação no PB. Sabe-se, por Hengeveld e Mackenzie (2018), conforme mostrado anteriormente, que a A-negação é um Ato Discursivo. No entanto, os autores não especificam a unidade, **núcleo do Ato Discursivo**, que “não” compõe.

Se “não” compusesse o núcleo de um Conteúdo Comunicado, deveria ser possível reportar sua fonte a outrem, mas, como mostra o teste em (21), isso não é possível, porque não há ações (Subatos) que evoquem uma imagem do mundo, ou seja, não é formulado um Conteúdo Comunicado. Reportar o “não” que expressa Rejeição não é possível.

21. A: Vá para casa!

B: [*Diz que] não!

No núcleo do Ato Discursivo da A-negação, “não” se insere na Ilocução. Uma Ilocução é responsável por atribuir, aos Atos Discursivos, usos interpessoais convencionalizados na consecução de uma intenção comunicativa (cf. Hengeveld; Mackenzie, 2008, p. 69), podendo seu núcleo ser preenchido por uma Ação Lexical (D), “um lexema inserido em uma determinada posição da estrutura pragmática que equivale a uma ação comunicativa

executada pelo Falante" (Giomi, 2020, p. 47, tradução própria)¹³, como "saúde", um lexema que, no Nível Representacional, descreve um estado de bem-estar e equilíbrio físico, mental e psicológico, mas, quando utilizado como uma Ação Lexical, é uma expressão social interativa dirigida a alguém que, por exemplo, espirrou.

A mesma posição na estrutura pragmática ocupa "não" da A-negação. Assim, a A-negação é formulada conforme representado em (22), um Ato Discursivo Interativo.

22. $(A_i: [(F_i: (D_i: /'nawn/ (D_i)) (F_i)) (P_i)_s (P_2)_A] (A_i))$

Os dados coletados demonstram que a A-negação não é utilizada para rejeitar apenas Atos Discursivos de Ilocução Imperativa. Observar a inexistência dessa restrição é importante porque Hengeveld e Mackenzie (2018) exemplificam a Rejeição apenas a Atos Discursivos com Ilocução Imperativa, mas, como se vê na ocorrência (23), a A-negação também rejeita Atos Discursivos Declarativos.

23. [E]le descobriu na hora que era a primeira vez que o músico dirigia. "*Ele está aprendendo ainda, ele vai melhorar*", soltou Sorocaba. "**Não**, não fala isso! Você tá louco?", soltou Thiago, com medo.

(Disponível em: <https://noticiasdatv.uol.com.br/noticia/televisao/thiago-oliveira-leva-susto-em-fazenda-do-cantor-sorocaba-nao-sabem-brincar-101709>. Acesso em: Acesso em: 13 set. 2025.)

NI: $(A_i: [(F_i: DECL (F_i)) (P_i)_s (P_j)_A (C_i: -\text{ele está aprendendo ainda} - (C_i))] (A_i))$

$(A_j: [(F_j: (D_j: /'nawn/ (D_j)) (F_j)) (P_j)_s (P_i)_A] (A_j))$

Em (23), rejeita-se a declaração "ele está aprendendo ainda". Nesse caso, P_j não coloca em questão a verdade dessa declaração, mas sim sua aceitabilidade, já que, para P_j , saber que condutor do veículo ainda não domina totalmente a técnica de dirigir é indesejado e, por isso, P_j rejeita que isso lhe seja declarado, rejeição, inclusive, explicitada pelo Ato Discursivo "não fala isso", subsequente ao Ato Discursivo Interativo "não". Por tratar-se de A-negação, "não", em (23), pode ser substituído por "nem vem", inclusive.

O que os exemplos em (21) e (23) têm em comum é que, em ambos, nega-se a relação que se estabelece entre os componentes do Ato Discursivo rejeitado, um Imperativo e outro Declarativo, respectivamente. Nesse sentido, é negado ao interlocutor a aplicação de uma força ilocucionária a um determinado Conteúdo Comunicado.

¹³ No original: "[...] a lexeme inserted in a given position o pragmatic structure amounts to a communicative action that is performed by the speaker."

De acordo com Hengeveld e Mackenzie (2008) e Hengeveld e Keizer (2025), a Ilocução é entendida como o predicado, núcleo do Ato Discursivo, que tem como argumentos o Conteúdo Comunicado e os dois Participantes do discurso. Assim, na A-negação, o Falante nega a aceitabilidade da relação de predicado-argumento entre os componentes de um Ato Discursivo proferido anteriormente. É precisamente por esse motivo que a A-negação pode negar Atos Discursivos com todos os tipos de Ilocução. As ocorrências (24) e (25) exemplificam, respectivamente, casos de Atos Discursivos com Ilocução Interrogativa e Exclamativa rejeitados.

24. – *Quem é a mulher da relação?*

– **Não.** São dois homens, não tem mulher na relação!

(Disponível em: https://www.bdttd.uerj.br:8443/bitstream/1/11457/1/DISSERTACAO_FINAL_RAYANNI_SAMPAIO_TEIXEIRA%20%281%29.pdf. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: INTER_C (F_i)) (P_i)_S (P_j)_A (C_i: –quem é a mulher da relação– (C_i))] (A_i))

(A_j: [(F_j: (D_i: /'nawn/ (D)) (P_j)_S (P_j)_A] (A_j))

25. – *Bah! Que uísque ruim é esse!*

– **Não,** não diga isso. Não existe uísque RUIM, mas alguns são melhores que outros.

(Disponível em: <https://helenablavatsky.com.br/biografia/william-rudolf-odonovan/>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: EXCL (F_i)) (P_i)_S (P_j)_A (C_i: –uísque ruim é esse– (C_i))] (A_i))

(A_j: [(F_j: (D_i: /'nawn/ (D)) (P_j)_S (P_j)_A] (A_j))

Em tese, qualquer tipo de Ato Discursivo pode ser negado, até mesmo o que não apresenta um Conteúdo Comunicado, como Atos Discursivos Interativos, representados por expressões socialmente dirigidas, como “muito obrigado” em (26), em que P_j, ao proferir “não”, mostra que considera inapropriada a relação de predicado-argumento entre a Ilocução de agradecimento e os Participantes (Falante e Ouvinte).

26. – *Agora eu sei que posso contar com vocês para tudo, absolutamente tudo mesmo. Muito obrigado! [...]*

– **Não...** não precisa agradecer.

(Disponível em: <https://encurtador.com.br/QsTT>. Acesso em: 13 set. 2025.)

NI: (A_i: [(F_i: (D_i: /obri'gad/- D_j: (D_j: /'muito/ (D_j)) (D_j)) (F_i)) (P_i)_S (P_j)_A] (A_i))

(A_j: [(F_j: (D_k: /'nawn/ (D_k)) (F_j)) (P_j)_S (P_j)_A] (A_j))

Os únicos Atos Discursivos que não podem ser rejeitados são os Expressivos, que exprimem algum tipo de emoção do Falante, como “Ai!”, expressando dor em (27).

27. – *Ai!*
 – **Não.*

Os Expressivos não são heterorrelacionados, condição necessária para o Ato Discursivo da A-negação, que ocorre em par dialógico, como a representação do molde interpessoal da A-negação em (28).

28. (M₁: (A₁: [(F₁: (D₁: /'nawn/ (D₁)) (F₁)) (P₁)_S (P₂)_A] (A₁)) (M₁))

Como se nota em (28), o Ato Discursivo de A-negação constitui, por si só, um Movimento (M). Como Movimento, M₁ faz avançar a comunicação, podendo, por isso, ser usado independentemente, ou seja, é uma “unidade mínima livre de discurso” (cf. Kroon, 1995, p. 66), um enunciado completo, que preenche um turno de fala por si só, conforme mostram Hengeveld e Mackenzie (2018, p. 36).

O Movimento composto pela A-negação não apresenta nenhum conteúdo semântico, já que a A-negação não é descritiva, mas acional, não sendo, portanto, submetida a condições de verdade. Assim, a A-negação não é formulada no Nível Representacional e tampouco codificada no Nível Morfossintático, visto que, com “não”, não ocorre processo morfossintático, como concordância ou derivação. Há, portanto, entre os níveis Interpessoal e Representacional/Morfossintático, uma relação de discrepância (*mismatch*) um-para-zero. A A-negação é inserida em sua forma fonêmica no Nível Interpessoal e mandada para o Fonológico para sua codificação prosódica, o que é explicado pelo princípio da Profundidade Máxima, que evita a especificação vazia de níveis de representação que são irrelevantes para a produção de uma expressão linguística.

Devido à ausência de dados orais, para a descrição fonológica da A-negação, recorre-se aos resultados obtidos por Galvão Passetti (2023). Nesse trabalho, um informante produz “não” para diferentes papéis na interação verbal.¹⁴ A partir desse estudo, verifica-se que a A-negação é a única que apresenta dois picos e um vale de frequência fundamental (F0). O movimento LHLHL de F0 desses casos é mostrado no Gráfico 1.¹⁵ A curva de F0 de uma ocorrência, transcrita em (29), que tem o padrão ilustrado pelo Gráfico 1, é apresentada pelo Gráfico 2.¹⁶

14 Certificado de Apresentação para Apreciação Ética (CAAE) nº 73430623.0.0000.5466. Parecer nº 6.570.410.

15 No Gráfico 1, o eixo das abscissas corresponde à passagem do tempo e não é indicado porque os picos e vales de F0 não ocorrem ao mesmo tempo em cada ocorrência. *L* corresponde a *low* (baixo) e *H*, a *high* (alto).

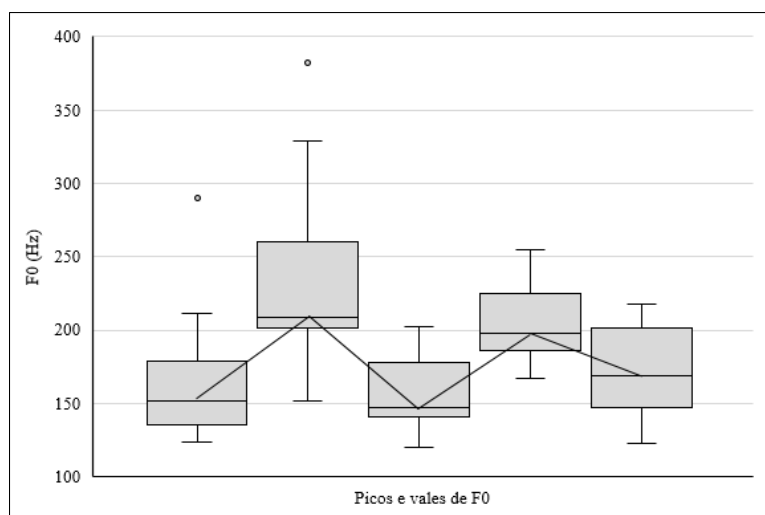
16 Galvão Passetti (2025) identifica padrões entonacionais distintos na rejeição de Atos Discursivos com Ilocuções Comportamentais, como a Imperativa exemplifica em (29), e na de Atos Discursivos com Ilocuções Proposicionais, como a Declarativa.

29. Contexto: Você (B) e um amigo (A) estão em uma festa. Você está desanimado e, por isso, está sentado. Seu amigo lhe ordena:

A: Levanta!

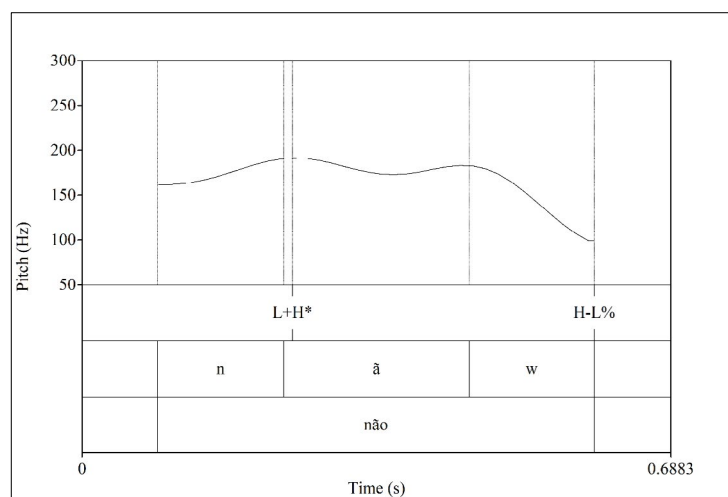
B: Não!

Gráfico 1. Movimento de F0 dos casos de A-negação analisados



Fonte: Galvão Passetti (2023, p. 24).

Gráfico 2. Frequência fundamental em Hz por tempo em segundos de “não” da ocorrência em (29)



Fonte: Elaboração própria.¹⁷

¹⁷ Gráfico gerado por meio do Praat®, versão 6.3.18 (Boersma; Weenink, 2023), utilizando o sistema de notação ToBI (*Tones and Break Indices*) (Beckman; Hirschberg; Shattuck-Hufnagel, 2006).

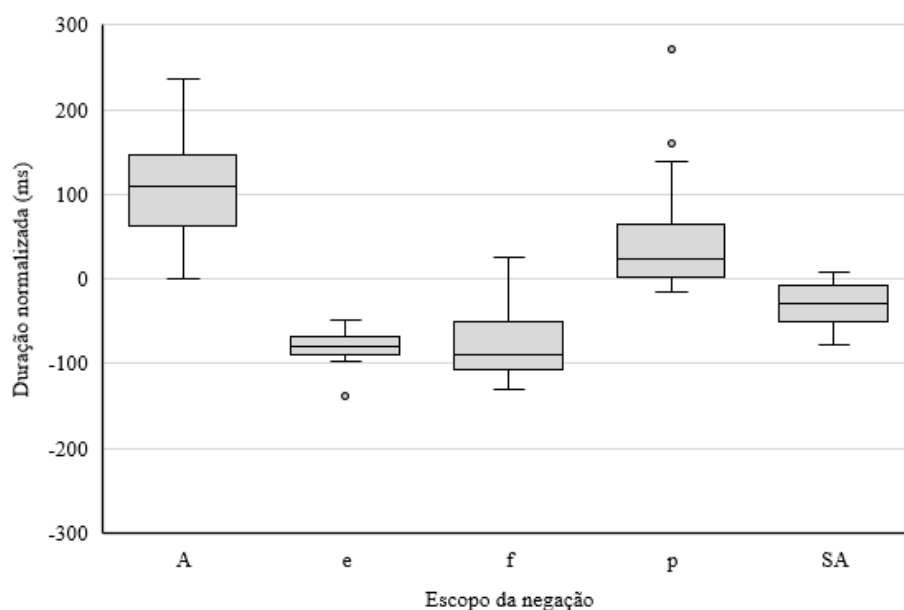
O primeiro pico de F0 corresponde ao acento de tom L+H* e o segundo pico, ao acento frasal H-. Ambos são desencadeados pelo operador *Rising* (r), restringindo a camada da Frase Fonológica (PP). Por fim, o operador *Falling* (f) na camada da Frase Entonacional (IP) é o responsável pelo tom de fronteira L%, que determina a direção final descendente. Esse operador, por especificar a camada da Frase Entonacional, governa globalmente o movimento de F0 e, por isso, também é o responsável pelo vale que se realiza na vogal “ã” de “não”.

Desse modo, a representação da A-negação no Nível Fonológico é como em (30).

30. (U₁: (f IP₁: (r r PP₁: (PW₁: -/’nawn/- (PW₁)) (PP₁)) (IP₁)) (U₁))

Para abrigar todos esses eventos tonais, a duração de “não” que codifica a A-negação é maior frente aos outros tipos de negação. A duração normalizada da realização de “não” é comparada no Quadro 3, de acordo com cada escopo.

Gráfico 3. Duração normalizada de “não” por escopo da negação



Fonte: Galvão Passetti (2023, p. 18)

Sob a perspectiva da Fonologia Prosódica (Nespor; Vogel, 2007), pode-se dizer que a duração maior de “não” que nega Ato Discursivo é explicada pelo acento frasal principal de sentença, já que “não”, nesses casos, constitui Frase Entonacional que compõe, sozinha, um Enunciado, o que bloqueia o processo de redução fonética de “não” para “num”, “nu” ou “n”, de modo que “não” tenha acento primário, representado em (31) pelo operador *Stress* (s), aplicado tanto à camada da Sílabas como à do Pé.

31. (U₁: (f IP₁: (r r PP₁: (PW₁: (s F₁: (s S₁: /nawN/ (S₁)) (F₁)) (PW₁)) (PP₁)) (IP₁)) (U₁))

A representação em (31) é o padrão fonológico instanciado acionado pela A-negação. Enquanto U₁ mapeia o Movimento, IP₁ codifica o Ato Discursivo Interativo “não” formulado no Nível Interpessoal. Essas e outras relações de correspondência entre os níveis, assim como as conclusões da análise da A-negação, são dadas a seguir.

Considerações finais

As relações de correspondência entre os níveis de formulação e os de codificação na A-negação são representadas na Figura 2.

Figura 2. Correspondência entre os níveis de formulação e os de codificação na negação de Ato Discursivo

NI:	(M ₁ : (A ₁ : [(F ₁ : (D ₁ : /'nawN/ (D ₁)) (F ₁)) (P ₁) _s (P ₂) _A] (A ₁)) (M ₁))	↓
NF:	(U ₁ : (f IP ₁ : (r r PP ₁ : -/'nawN/- (PP ₁)) (IP ₁)) (U ₁))	

Fonte: Elaboração própria.

As hipóteses foram confirmadas e os resultados da análise da A-negação são sumarizados a seguir:

- (i) No Nível Interpessoal, a A-negação ocorre no par dialógico Falante-Ouvinte. No PB, ela é formulada como um Ato Discursivo Interativo cuja llocução é preenchida pelo Lexema Acional “não”. Esse Ato Discursivo Interativo compõe um Movimento por si só e rejeita outro Ato Discursivo anteriormente executado no discurso, servindo à estratégia de reprovar sua enunciação, já que o Participante considera não aceitável a relação de predicado-argumento que há entre a llocução e os outros componentes do Ato Discursivo rejeitado, que pode ser de qualquer tipo, exceto Expressivo.
- (ii) No Nível Representacional, a A-negação não é formulada, uma vez que essa negação é estritamente acional e não descritiva. Isso significa que há uma discrepância entre os níveis de formulação, numa relação um-para-zero.
- (iii) No Nível Morfossintático, assim como no Representacional, há uma discrepância estabelecida com o Nível Interpessoal, tendo em vista que a A-negação é diretamente codificada no Nível Fonológico.

(iv) No Nível Fonológico, por fim, o item negativo “não” constitui, em última instância, um Enunciado, que mapeia um Movimento no Nível Interpessoal. Esse Enunciado é composto por uma Frase Entonacional, que apresenta o contorno nuclear L+H* H-L%, desencadeado pelos operadores *Rising* (acento frasal H- e acento de tom L+H*), que especificam duplamente a camada da Frase Fonológica, e pelo operador *Falling* (tom de fronteira L%), aplicado à Frase Entonacional. Além disso, na A-negação, “não” não sofre processo de redução fonética, já que é a única Sílabas de uma Frase Fonológica. O contorno nuclear dessa Frase Fonológica garante a distinção da A-negação de outras formas negativas no PB.

Por fim, esta análise apura a descrição da A-negação oferecida por Hengeveld e Mackenzie (2018), especificando o componente em que é formulada a negação dentro do Ato Discursivo, qual seja, a Ilocução, nucleada pela Ação Lexical “não”. Além disso, fica evidenciado que a A-negação não rejeita apenas Atos Discursivos com Ilocução Imperativa, mas uma série de tipos de Atos Discursivos.

Com isso, propõe-se uma definição mais precisa e abrangente dessa negação acional. Se, antes, ela era conceituada – ao menos implicitamente –, por Hengeveld e Mackenzie (2018), como a rejeição de um Ato Discursivo especificamente Imperativo, rejeição justificada pela relação assimétrica entre os Participantes do discurso (P_2 considera que P_1 não está em posição de lhe dar uma ordem), a A-negação, agora, é definida como a rejeição de um Ato Discursivo anteriormente proferido por P_1 , já que P_2 considera não aceitável a relação de predicado-argumento entre os componentes do Ato Discursivo rejeitado. Em outras palavras, P_2 (Ouvinte – argumento) entende ser inapropriado P_1 (Falante – argumento) dirigir-lhe um ato de fala com determinada força ilocucionária (Ilocução – predicado) aplicada a uma determinada mensagem (Conteúdo Comunicado – argumento).

Espera-se que esses resultados estimulem pesquisas sobre essa negação acional em outras línguas, no contexto de uma abordagem discursivo-funcional da gramática.

Agradecimentos

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento (GD/CNPq Proc. N. 140383/2021-2 e PQ/CNPq Proc. N. 305603/2021-3, bolsas concedidas ao primeiro e ao segundo autor, respectivamente), que permitiram a condução dessa pesquisa e a elaboração deste artigo.

Referências

- ALTURO, N.; KEIZER, E.; PAYRATO, L. The interaction between context and grammar in Functional Discourse Grammar: Introduction. *Pragmatics*, [S. l.], v. 24, n. 2, p. 185-201, 2014.
- BAKKER, D.; SIEWIERSKA, A. Towards a speaker model of Functional Grammar. In: MACKENZIE, J. L.; GÓMEZ-GONZÁLEZ, M. A. (ed.). *A New Architecture for Functional Grammar*. Berlin: Mouton de Gruyter, 2004. p. 325-364.
- BECKMAN, M. E.; HIRSCHBERG, J.; SHATTUCK-HUFNAGEL, S. The original ToBI system and the evolution of the ToBI framework. In: JUN, S-A. (ed.) *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press, 2006. p. 9-55.
- BOERSMA, P.; WEENINK, D. *Praat: Doing Phonetics by Computer*. Version 6.3.18. 2023. Disponível em: <http://www.praat.org/>. Acesso em: 11 out. 2023.
- BUTLER, C. S. Interpersonal meaning in the noun phrase. In: RIJKHOFF, J.; GARCÍA VELASCO, D. (ed.). *The Noun Phrase in Functional Discourse Grammar*. Berlin; New York: De Gruyter Mouton, 2008. p. 221-262.
- CONNOLLY, J. H. Context in Functional Discourse Grammar. *Alfa: Revista de Linguística*, São Paulo, v. 51, n. 2, p. 11-33, 2007.
- DIK, S. C. *The theory of Funcional Grammar*. Part I: The scructure of the clause. Edição de Kess Hengeveld. 2. ed rev. Berlin; New York: De Gruyter Mouton, 1997a.
- DIK, S. C. *The Theory of Funcional Grammar*. Part II: Complex and derived constructions. Edição de Kess Hengeveld. 2. ed rev. Berlin; New York: De Gruyter Mouton, 1997b.
- GALVÃO PASSETTI, G. H. *As funções de “não” na interação verbal e suas propriedades fonético-acústicas: uma análise sob a perspectiva da Gramática Discursivo-Funcional e da Teoria Acústica de Produção da Fala: relatório final de pesquisa apresentado a Comitê de Ética em Pesquisa*. São José do Rio Preto: UNESP, Câmpus de São José do Rio Preto, 2023.
- GALVÃO PASSETTI, G. H. *A negação acional no português brasileiro sob a perspectiva da Gramática Discursivo-Funcional: assalto ao turno de fala e rejeição de ato de fala*. 2025. 205 p. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2025.

GIOMI, R. *Shifting structures, contexts and meanings: a Functional Discourse Grammar account of grammaticalization*. 2020. Tese (Doutor em Linguística) – Universidade de Lisboa, Faculdade de Letras, Lisboa, 2020.

GONÇALVES, S. C. L. G. *Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista*. Disponível em: <http://www.alip.ibilce.unesp.br/iboruna>. Acesso em: 25 jan. 2024.

HENGEVELD, K. Epilogue. In: MACKENZIE, J. L.; GÓMEZ-GONZÁLEZ, M. A. (ed.). *A New Architecture for Functional Grammar*. Berlin: Mouton de Gruyter, 2004. p. 365-378.

HENGEVELD, K.; KEIZER, E. General principles of linearization in Functional Discourse Grammar. In: WOLDE, E.; GIOMI, R.; HENGEVELD, K. (ed.). *Linearization in Functional Discourse Grammar*. Berlin; Boston: Mouton de Gruyter, 2025. p. 43-83.

HENGEVELD, K.; KEIZER, E.; GIOMI, R. *Layering in Functional Discourse Grammar: The hierarchical structure of the language system*. Oxford: Oxford University Press. (em preparação).

HENGEVELD, K.; MACKENZIE, J. L. *Functional Discourse Grammar: a typologically-based theory of language structure*. Oxford: Oxford University Press, 2008.

HENGEVELD, K.; MACKENZIE, J. L. Gramática discursivo-funcional. Tradução Marize Mattos Dall'Aglio-Hattner. In: SOUZA, E. R. (org.). *Funcionalismo linguístico: novas tendências teóricas*. São Paulo: Contexto, 2012. p. 43-85.

HENGEVELD, K.; MACKENZIE, J. L. Grammar and context in Functional Discourse Grammar. *Pragmatics*, [S. l.], v. 24, n. 2, p. 203-227, 2014.

HENGEVELD, K.; MACKENZIE, J. L. Negation in functional discourse grammar. In: KEIZER, E.; OLBERTZ, H. (org.). *Recent Developments in Functional Discourse Grammar*. Amsterdam: John Benjamins Publishing Company, 2018. p. 18-45.

HENGEVELD, K.; OLBERTZ, H. Didn't you know? Mirativity does exist! *Linguistic Typology*, [S. l.], v. 16, n. 3, p. 487-503, 2012.

HORN, L. R. *A natural history of negation*. Stanford, California: CSLI Publications, 2001.

KROON, C. *Discourse Particles in Latin: a study of nam, enim, autem, vero and at.* (Amsterdam Studies in Classical Philology 4). Amsterdam: Gieben, 1995.

LEVELT, W. J. M. *Speaking*. Cambridge, MA: MIT Press, 1989.

NESPOR, M.; VOGEL, I. *Prosodic phonology: with a new foreword*. Berlin; New York: Mouton de Gruyter, 2007.

OLBERTZ, H. The place of exclamatives and miratives in grammar: a functional discourse grammar view. *Revista Linguística*, Rio de Janeiro, v. 8, n. 1, p. 76-98, 2012.

SADOCK, J. M.; ZWICKY, A. M. Speech act distinctions in syntax. In: SHOPEN, T. (ed.). *Language Typology and Syntactic Description* (Vol. I: Clause Structure). Cambridge: Cambridge University Press, 1985. p. 155-196.

SILVA, J. C. B. A Prosódia Regional em Enunciados Interrogativos Espontâneos do Português do Brasil. *Revista Gatilho*, [S. l.], v. 13, n. 1, p. 1-13, 2011.

Uma nota sobre a hipótese da similaridade construcional

DOI: <http://dx.doi.org/10.21165/el.v54i1.4029>

Taísa Peres de Oliveira¹

Resumo

Neste artigo, analiso microconstruções distintas que apresentam alto grau de similaridade semântica e pragmática. A partir da análise de quatro conectores condicionais do português brasileiro, a saber *contanto que*, *desde que*, *a menos que* e *a não ser que*, este estudo demonstra que, embora formalmente diferentes, essas microconstruções exibem padrões de uso, funções discursivas e projeções semânticas amplamente compartilhadas. Assentada na teoria da Gramática de Construções (Goldberg, 1995, 2006; Croft, 2001; Traugott; Trousdale, 2021), proponho a hipótese da similaridade construcional, segundo a qual construções podem ser alternativas funcionais dentro de um mesmo subesquema, entendendo que, para além das generalizações baseadas na construção específica, existem generalizações baseadas em alternâncias que também consistem em informações armazenadas na representação cognitiva dos usuários.

Palavras-chave: gramática de construções; conectores condicionais; sinonímia

¹ Universidade Federal do Mato Grosso do Sul (UFMS), Três Lagoas, Mato Grosso do Sul, Brasil; taisapoliveira@gmail.com; <https://orcid.org/0000-0002-2439-5604>

Abstract

In this paper, I analyze distinct micro-constructions that exhibit a high degree of semantic and pragmatic similarity. Taking into account the description of conditional connectors *contanto que*, *desde que*, *a menos que*, and *a não ser que* in Brazilian Portuguese, this study demonstrates that, although formally different, these micro-constructions display widely shared usage patterns, discourse functions, and semantic projections. Grounded in Construction Grammar theory (Goldberg, 1995, 2006; Croft, 2001; Traugott; Trousdale, 2021), I propose the hypothesis of constructional similarity, according to which constructions may serve as functional alternatives within the same subschema. This approach assumes that, in addition to generalizations based on individual constructions, there are alternation-based generalizations that also constitute information stored in speakers' cognitive representations.

Keywords: construction grammar; conditional connectors; synonym.

1. Introdução

Um dos princípios mais caros à Gramática de Construções (Goldberg, 1995, 2006; Croft, 2001; Traugott; Trousdale, 2021), em qualquer vertente, é o Princípio da não-sinonímia, segundo o qual,

Corolário A: se duas construções são sintaticamente distintas e semanticamente sinônimas, então elas não podem ser pragmaticamente sinônimas.

Corolário B: se duas construções são sintaticamente distintas e pragmaticamente sinônimas, então elas não podem ser semanticamente sinônimas (Goldberg, 1995, p. 67).

De modo mais geral, o princípio sustenta que a sinonímia perfeita (plena), que associaria duas formas diferentes a uma mesma função (significado), não existe. Evidentemente, essa premissa não é nova na linguística, especialmente entre os modelos teóricos ligados ao paradigma funcionalista (Bolinger, 1977, Haiman, 1978, Givón, 1985, Langacker, 2008). De acordo com Bolinger (1977, p. X, tradução própria) “[...] a condição natural da linguagem é preservar uma forma para um significado”². O entendimento geral subjacente é que a linguagem é restrita por um equilíbrio natural entre o poder expressivo (usuários necessitam de formas distintas para expressar com eficiência um vasto conjunto de significados) e a economia linguística (não há necessidade de existirem duas formas diferentes para expressar o mesmo significado), fundamento central em qualquer teoria funcionalista.

2 No original: “[...] the natural condition of language is to preserve one form for one meaning”.

Especificamente no quadro teórico formado pela Gramática de Construções, a premissa está na própria base epistemológica do modelo, que refuta a existência de regras de transformação de estruturas de superfície a partir de uma mesma estrutura profunda (com mesma semântica). Muitas são as construções que, com semântica semelhante e forma diferente, se diferenciam em razão de aspectos pragmáticos, como por exemplo as diferenças entre formas plenas e contraídas, construções passiva e ativa, construções modalizadoras, construções com inversão de ordem dos constituintes, entre outras. No entanto, a questão maior não parece ser identificar as diferenças, mas compreender em que medida essas diferenças se mostram, de fato, relevantes.

Recentemente, um conjunto de pesquisadores ligados ao quadro da Gramática de Construções têm pensado a questão da alternância construcional (Capelle, 2006, Van de Velde, 2014, Uhrig, 2015, Perek, 2015, Diessel, 2019, Laporte *et al.*, 2021). Assumindo posicionamentos semelhantes, os autores compartilham do entendimento de que o Princípio da não sinonímia, tal como formulado em Goldberg (1995, 2006), e amplamente aceito na abordagem construcional, precisa ser, em alguma medida, reelaborado. O princípio se assenta numa perspectiva semasiológica de que um padrão construcional “é mais bem analisado por si só, sem considerar referências implícitas ou explícitas a um possível sintagma alternativo”³ (2006, p. 25, tradução própria). A questão principal aí é que, ao tomar um padrão apenas verticalmente, analisado por si só, desconsidera-se o conhecimento que os falantes têm sobre similaridades entre construções alternativas. Essa questão é discutida em Cappelle (2006, 2009; Perek, 2012, 2015; Devaere, 2020, Oliveira (2025)). Uma teoria gramatical que tenha como objetivo modelar o conhecimento linguístico do falante não deveria ignorar a intuição que os falantes têm sobre similaridades entre construções. Outro ponto que também tem sido discutido é a questão da gradualidade, que parece, de algum modo, perdida quando se assume a relação “uma forma para uma função” tomada numa observação puramente vertical, sem considerar a constelação de construções também na horizontalidade. A gradiência e a gradualidade, tão cara aos modelos funcionalistas, cognitivistas, baseados no uso ficaria restrita, então, apenas na verticalização entre construções mais e menos abstratas e esquemáticas?

É exatamente nesse contexto que se insere o objetivo deste artigo: aqui pretendo explorar as preferências colocacionais para mostrar como, apesar de diferentes em alguma medida, construções podem sim ser usadas com funções muito semelhantes, numa relação de alternância construcional. Especificamente, aqui analiso construções em níveis mais baixo de esquematicidade e abstração para identificar (i) as propriedades que as assemelham e, a partir da análise das preferências distribucionais, (ii) aquelas que as diferenciam.

3 No original: “[...] is best analyzed in its own terms, without relying on explicit or implicit reference to a possible alternative phrase.

A discussão desenvolvida aqui parte da descrição realizada em Oliveira (2008, 2009), em que defendo que conectores condicionais subespecificam três diferentes sentidos associados à zona da condicionalidade. Desses, retomo os subesquemas $[X \text{ que}]_{\text{CondPos}}$ e $[X \text{ se}]_{\text{CondNeg}}$ que, sancionados pelo esquema geral $[X \text{ conj}]_{\text{Cond}}$, compartilham propriedades de formas e de funções. Para cumprir essa tarefa, este artigo está organizado do seguinte modo: primeiro, na seção 2, apresento as bases teóricas da Gramática de Construções, modelo que orienta minha discussão e análise; em seguida, na seção 3, apresento o universo de análise e a descrição dos conectores objeto, na seção 4, na seção final, apresento as implicações dessa análise para os modelos baseados no uso e para a gramática de construções e algumas possibilidades de análises.

2. A Gramática de Construções

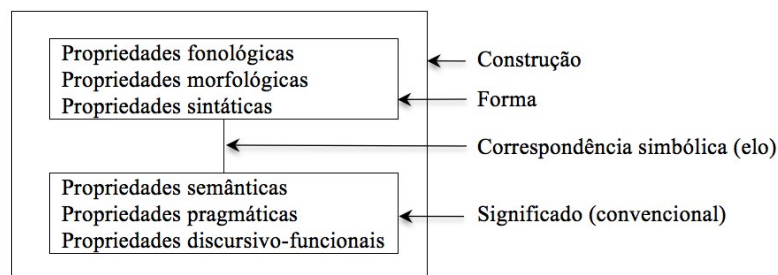
Neste trabalho, assumo como fundamento teórico de base a Gramática de Construções (Croft, 2001; Goldberg, 2006, 2019; Diessel, 2019), que, vinculada aos Modelos Baseados no Uso, entendendo que a estrutura linguística é moldada pela experiência do usuário da língua, refletindo habilidades cognitivas que atuam em outros sistemas do conhecimento humano. A abordagem construcional, em suas diferentes vertentes, se pauta por um conjunto de premissas básicas, dentre os quais destaco:

- 1) O conhecimento sobre a língua envolve reconhecer pares de forma e significado com diferentes graus de complexidade e em vários níveis de abstração, reconhecendo a língua como um sistema de signos convencionais (Croft, 2001; Langacker, 2008; Goldberg, 2006, 2019, Bybee, 2016; Traugott, Trousdale, 2021).
- 2) A relação entre a estrutura linguística e o significado não é fixa. A estrutura linguística tem potencial para o significado, que emerge a partir de rotinas cognitivas de padrões mentalmente acionados pela experiência com a língua.
- 3) A não separação rígida entre léxico e gramática. As categorias são radiais e graduais, sem limites claros (Langacker, 2008; Bybee, 2016; Traugott, Trousdale, 2021).
- 4) O sistema linguístico é organizado e motivado por habilidades cognitivas gerais, porque a estrutura linguística é parte de estruturas conceituais mais amplas (Langacker, 2008; Bybee, 2016).
- 5) A língua faz parte de uma estrutura conceitual mais ampla, organizada numa rede cognitiva em que nenhuma construção (também chamada de nó) está totalmente isolada, mas interconectada por relações associativas que podem ser acionadas, expandidas e (re)elaboradas a partir dos eventos de uso da língua (Langacker, 2008; Hudson, 2010; Goldberg, 2006, 2019; Traugott, 2022).

- 6) A rede como a representação cognitiva do conhecimento linguístico, os diferentes modelos partem do pressuposto de que a rede está assentada sobre habilidades cognitivas gerais, como pensamento analógico, categorização, abstração, rotinização e automação (Langacker, 2008; Hudson, 2010; Goldberg, 2006, 2019; Traugott, 2022).

Nesse contexto, a unidade fundamental de análise é a **construção**, que abriga unidades mínimas, como os morfemas, e unidades complexas, como os idiomatismos e os padrões e as regularidades mais abstratos (padrões de ordenação de constituintes e estrutura argumental, por exemplo). Ao assumir a construção como unidade fundamental da língua, a Gramática de Construções elabora um modelo de análise que dá abrigo tanto às propriedades linguísticas mais centrais como às mais periféricas e idiomáticas, sem recorrer a transformações e derivações, que fatalmente levam à proposição de categorias vazias e posições absolutas, o que é prontamente rejeitado pelo modelo. Por aí também se rejeita a modularidade e compartimentalização da gramática, que é concebida como a organização de estruturas linguísticas interconectadas na rede, o que torna irrelevante a proposição de diferentes níveis de análise. Ainda, ao tomar a construção como unidade de análise afasta-se a rígida distinção entre léxico e gramática, entendendo que as construções podem flutuar entre os dois polos, assumindo a gradualidade em todos os sentidos e direções, seja radialmente numa categoria, seja pela interconexão e mescla de categorias. Assim, a construção, entendida como o pareamento convencional de forma-significado em qualquer nível de complexidade estrutural e linguística, é a unidade fundamental da língua, e pode ser representada, seguindo Croft (2001), pela Figura 1.

Figura 1. A construção



Fonte: Adaptado de Croft (2001, p. 18)

No plano da forma, representam-se as propriedades fonológicas, morfológicas e sintáticas de uma construção; e, no plano do significado, representam-se todos os aspectos convencionalizados de sua função, incluindo suas propriedades semânticas, pragmáticas e, também, as propriedades da situação, do discurso e do contexto da interação verbal em que é usada. O que se vê, portanto, é que qualquer unidade da língua deve ser analisada como construção, do morfema e itens lexicais aos padrões sintáticos mais abstratos e expressões idiomáticas. As construções estão ligadas umas às outras em uma rede cognitiva que, inspirada na neural, busca representar o modo como indivíduos armazenam e processam informações, que é o que passo a discutir em seguida.

3. Os conectores condicionais no português

3.1. Os dados

Os dados foram coletados no banco de dados *Corpus do Português*, disponível em www.corpusdoportugues.org (Davies; Ferreira, 2006). Aqui, utilizei a modalidade Gênero/Histórico, que conta com aproximadamente 45 milhões de palavras distribuídas em textos oral, de ficção, jornalístico e acadêmico, recortados no século XIX e XX, considerando apenas a variedade do português brasileiro. Para a busca, os dados foram pesquisados pelo buscador do próprio cópús a partir dos comandos **contanto que**, **desde que**, **a menos que** e **a não ser que**, selecionando apenas ocorrências dos conectores em contexto de condicionalidade. Foram encontrados um total de 1440 dados. Os dados foram organizados e processados por meio do programa de análise estatística *Studio R* (2024), seguindo os roteiros de análise elaborados em Gries (2014) para análise colostrucional tal como desenvolvida por Gries e Stefanowitsch (2004).

3.2. Descrição dos dados

Aqui retomo as análises desenvolvidas por mim sobre as orações e conectores condicionais em Oliveira (2008, 2009), em que descrevi o comportamento de diversos conectores condicionais em português. Para tanto, considero as microconstruções conectoras [contanto que], [desde que], [a menos que] e [a não ser que], cuja frequência absoluta é apresentada na Tabela 1. Para essa discussão, levanto suas propriedades morfosintáticas, semânticas, pragmáticas e contextuais, para mostrar como os usuários da língua têm conhecimento de sua semelhança e fazem suas escolhas dentro desses contextos. Nesse sentido, e considerando que a gramática consiste numa representação conceitual e cognitiva do conhecimento que falantes têm sobre a língua, essa similaridade deve ser pensada dentro da teoria, que é o que proponho aqui.

Tabela 1. Frequência bruta das microconstruções

Contanto que	305
Desde que	734
A menos que	191
A não ser que	210
Total	1440

Fonte: Elaboração própria.

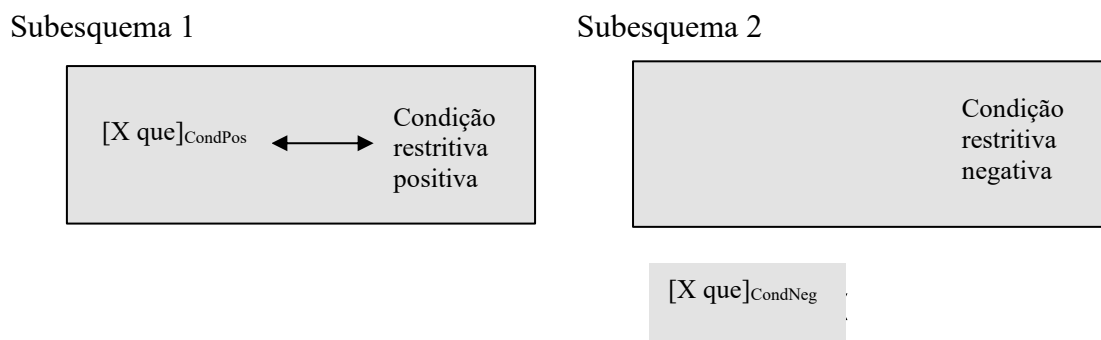
Conforme demonstrei em Oliveira (2008, 2009), conectores condicionais podem subespecificar diferentes valores associados ao significado mais geral de condição. Aqui, trago dois subtipos específicos, os chamados **restritivos positivos** e os **restritivos**

negativos que atuam na indicação de uma condição necessária, daí o rótulo restritivo. Conectores restritivos positivos subespecificam as duas orações numa relação de implicação mútua, isto é, são ambas verdadeiras ou ambas falsas, como se vê no exemplo (1). No exemplo, o conector projeta dois espaços alternativos: um primeiro em que a verdade do condicionante leva necessariamente à verdade do consequente, e outro em que a falsidade do condicionante leva necessariamente à falsidade do consequente, já que exclui qualquer outro cenário para sua validação. Pode-se afirmar, então, que as situações projetadas a partir desse exemplo são “*as testemunhas a compreendam é a única condição para o testamento particular pode ser escrito em língua estrangeira*” e “*as testemunhas não compreenderem é a única condição para o testamento particular não poder ser escrito em língua estrangeira*”. Semelhantemente, conectores restritivos negativos operam o mesmo tipo de projeção, mas com cenários invertidos: isso porque, nesse grupo, além de marcarem uma condição necessária, os também conectores promovem uma inversão de polaridade, introduzindo uma condição que deve ser necessariamente negada para a validação do consequente, como no exemplo em (2). Nesse caso, a condição toma a forma “*atendam nossas exigências justas é a condição necessária para a não validação do consequente rejeitamos uma trégua*”. Há aí, além da bicondicionalidade, uma inversão de polaridade e os cenários possíveis são, portanto, invertidos: se o condicionante é válido, o consequente deve ser negado e se o condicionante é inválido (negado), o consequente se valida.

- (1) O testamento particular pode ser escrito em língua estrangeira, **contanto que** as testemunhas a compreendam. (19Ac:Br:Lac:Misc)
- (2) Rejeitamos uma trégua, **a não ser que** atendam nossas exigências justas (19N:Br:Recf)

O que tenho aqui, portanto, são dois grupos de conectores associados, cada um com uma subespecificação semântica respectiva. O primeiro pode ser generalizado a partir do subesquema [X que]_{CondPos'} que codifica uma condição que deve necessariamente ser cumprida, ou segue-se necessariamente a não validação da situação descrita na oração núcleo. O segundo grupo é generalizado a partir do subesquema [X que]_{CondNeg'} que codifica uma restrição negativa, indicando a situação descrita na oração condicional como única condição em que os sentidos codificados na oração núcleo não se validam, promovendo a inversão de polaridade entre oração condicional e oração núcleo. Pensando em termos do paramento de forma-significado, a partir dessas subespecificações, proponho representar os dois subesquemas na Figura 2:

Figura 2. Pareamento dos conectores condicionais



Fonte: Elaboração própria

A gramática, organizada em rede, representa como o conhecimento linguístico é organizado cognitivamente. Nela, a organização **vertical** em **elos de herança** representa o processo de abstração e categorização e como propriedades comuns/similares de forma e de significado das construções são armazenadas na memória dos falantes. As microconstruções generalizam feixes de propriedades comuns a partir de um conjunto de exemplares similares, os construtos, e, a partir daí, vão se extraindo propriedades de forma e de significado cada vez mais gerais, representadas nos níveis esquematicamente superiores e, por isso mesmo, mais abstratos. As microconstruções representam o primeiro nível de análise propriamente dito, nele os conectores são totalmente especificados. Suas propriedades são generalizadas nos subesquemas [X que]_{CondPos} e [X que]_{CondNeg}, parcialmente esquemáticos e que ainda carregam distinções semânticas que são, por fim, generalizadas pelo esquema [X conj]. Elos verticais de herança ligam essas construções nesses diferentes níveis de abstração e generalização, capturando as propriedades de forma e de significado compartilhadas. A modelagem da gramática, hierarquizada em diferentes níveis de abstração por meio de elos de herança, evidencia exatamente o processo como usuários da língua armazenam informações generalizadas a partir da experiência linguística, aqui representada pelos construtos nos boxes cinzas. A cada experiência, o construto vai deixando rastros na memória do usuário da língua e, à medida que essas experiências vão se repetindo, ou seja, à medida que é exposto, repetidamente, a instâncias similares, o falante começa a identificar padrões e regularidades, que são abstraídos num processo *bottom-up* e estocados como parte da representação cognitiva daquela construção. Em outras palavras, na organização vertical, os esquemas e subesquemas estocam representações cada vez mais esquemáticas, abstraídas sobre propriedades similares de forma e de significado de microconstruções, que, por sua vez, são abstraídas a partir de exemplares reais, instâncias de uso da língua.

Meu real interesse neste trabalho é descrever especificamente os traços das microconstruções, para entender em que medida são similares e se essa similaridade é adequadamente tratada num modelo de gramática que representa o conhecimento linguístico do falante.

Os conectores restritivos positivos constituem unidades complexas formadas a partir de uma base lexical, que pode ser um advérbio ou uma preposição, representado pela posição aberta X no subesquema, e pelo complementizador QUE. Esses conectores conceitualizam um cenário único e específico para a validação da situação descrita na oração núcleo. Partindo das descrições em Oliveira (2008, 2009), as propriedades semânticas e pragmáticas dessas microconstruções conectoras são resumidas na Figura 3.

Figura 3. Conectores restritivos positivos

[DESDE QUE] _{CondPos} CENÁRIO HIPOTÉTICO ÚNICO	[CONTANTO QUE] _{CondPos} CENÁRIO HIPOTÉTICO ÚNICO
<p>Serão boa fonte de recursos desde que realmente se consiga viabilizar o terceiro ponto importantíssimo</p> <p>Sem: Domínios de conteúdo e epistêmico, Projeção preditiva e inferencial, Distância epistêmica e Distância evidencial</p> <p>Prag: Adendo, Tópico</p>	<p>Amâncio trocava, dava-se de graça para as missas cantadas, para os sermões, contanto que o deixassem em paz</p> <p>Sem: Domínios de conteúdo e epistêmico, Projeção preditiva e inferencial, Distância epistêmica e Distância evidencial</p> <p>Prag: Adendo, Tópico</p>

Fonte: Elaboração própria

O mesmo acontece com conectores restritivos negativos. Igualmente, essas microconstruções conectoras constituem unidades complexas, formadas a partir de diferentes bases léxico-gramaticais. Os que analiso aqui, especificamente, se formam a partir da combinação do advérbio de exclusão ‘menos’ ou da estrutura negativa ‘não ser’, combinados com o complementizador QUE. Esses conectores conceitualizam um cenário único impeditivo para a validação da situação descrita na oração núcleo, daí a inversão de polaridade (já que a leitura aí é *se não p, q*). Assumindo a caracterização elaborada em Oliveira (2008, 2009), as propriedades semânticas e pragmáticas dessas microconstruções conectoras são resumidas na Figura 4.

Figura 4. Conectores restritivos positivos

[DESDE QUE] _{CondPos} CENÁRIO HIPOTÉTICO ÚNICO	[CONTANTO QUE] _{CondPos} CENÁRIO HIPOTÉTICO ÚNICO
<p>Não trabalhamos para montar carros velhos, a menos que ele atenda algumas exigências mínimas</p> <p>Sem: Domínios de conteúdo, epistêmico, atos de fala, Projeção preditiva e inferencial, Distância epistêmica e Distância evidencial</p> <p>Prag: Adendo, Tópico</p>	<p>não vão deixá-la em paz muito tempo, a não ser que disso tirem proveito</p> <p>Sem: Domínios de conteúdo, epistêmico, atos de fala Projeção preditiva e inferencial, Distância epistêmica e Distância evidencial</p> <p>Prag: Adendo, Tópico</p>

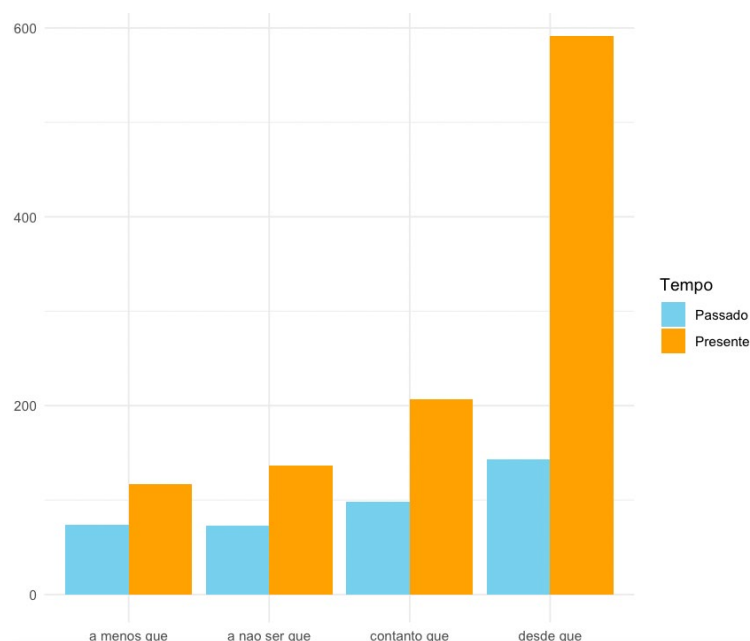
Fonte: Elaboração própria

Vejam que, embora formalmente distintas, essas microconstruções conectoras compartilham uma parte substancial de significado: podem conectar orações nos domínios do conteúdo e epistêmico e, por isso, podem projetar tanto uma condição preditiva (domínio de conteúdo), quanto uma inferencial (domínio epistêmico). No caso das microconstruções restritivas negativas, os conectores também conectam orações no domínio dos atos de fala, projetando a adequação comunicativa para uma asserção. Como consequência de seus usos nesses contextos, as microconstruções sob análise codificam tanto distância epistêmica (quando apenas marca a não factualidade da condicional), como a distância evidencial (quando o falante aciona uma informação disponível no evento de fala por outra fonte, geralmente o ouvinte, mas opta por marcá-la como não factual). Ou seja, estamos diante de pares de microconstruções (considerando cada subesquema) que se diferem estritamente apenas na forma, mas mantém, de algum modo, uma considerável similaridade semântica e pragmática.

Como então dar conta da similaridade semântica e pragmática observada entre as microconstruções conectoras, que, em termos de abstração de propriedades, não pode ser capturada? Para buscar essa resposta, volto aos contextos de uso dessas microconstruções, afinal, a Gramática de Construções é um Modelo Baseado no Uso, e é, então, exatamente no uso, que busco a resposta para esse questionamento. Até aqui, tenho dois subesquemas, instanciados por construções similares semântico e pragmaticamente. A partir daí, busquei diferenças contextuais para analisar os usos dessas microconstruções. Considerei para tanto, fatores clássicos usados na análise de orações adverbiais: a posição, os tempos e modos verbais e a polaridade da oração núcleo. Dessas, dois fatores se mostraram mais relevantes, a posição e a polaridade da oração núcleo.

Para a análise dos tempos verbais, considerei as referências de passado, presente e futuro, associadas aos modos indicativo e subjuntivo. Nos dados analisados, não houve ocorrência das microconstruções conectoras associadas a tempos do indicativo. Esse fato corrobora a análise desenvolvida em Oliveira (2008), em que apresentei dados semelhantes. Isso se deve ao fato de o significado condicional estar apoiado sobre uma base causal hipotética, daí a preferência por tempos e modos verbais não factuais, que concorrem para a construção desse sentido. Nos dados analisados aqui, encontrei referências de presente e passado associadas ao modo subjuntivo, respectivamente o presente e imperfeito do subjuntivo, distribuídos de modo relativamente uniforme entre as diferentes microconstruções conectoras, como demonstra o Gráfico 1. Independentemente do *type*, as microconstruções conectoras ocorrem preferencialmente em condicionais em que figuram o presente do subjuntivo, responsável pela predição, capacidade que a condicional tem de projetar uma causa e, a partir dela, prever uma consequência. A, menos frequente para os quatro *types*, referência de passado é marcada pelo imperfeito do subjuntivo, responsável pela predição com recuo temporal. Como se vê, portanto, no Gráfico 1, os tempos e modos verbais que figuram na oração condicional em que as microconstruções conectoras ocorrem não diferenciam seus contextos de uso.

Gráfico 1. Tempo e modo da oração condicional

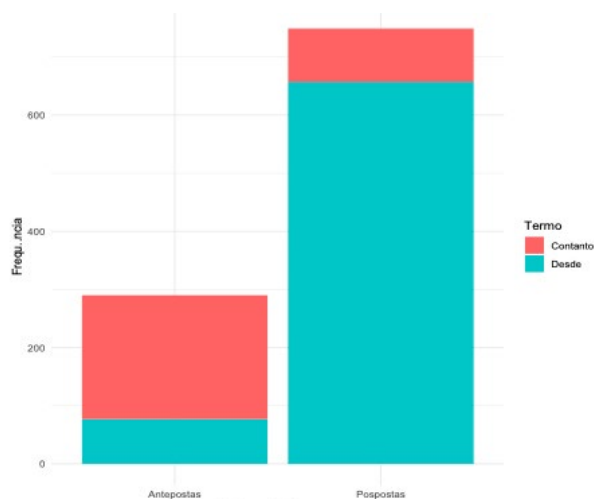


Fonte: Elaboração própria

Um primeiro fator que se mostrou relevante para a análise dos usos das microconstruções [contanto que] e [desde que] é a posição da oração condicional em que esses conectores ocorrem. Os falantes usam a microconstrução [contanto que] preferencialmente em condicionais antepostas e a microconstrução [desde que] em condicionais pospostas.

Essa diferença de contextos de usos fica bem evidenciada no Gráfico 2. A diferença de posição está associada às funções que cumprem as orações condicionais, que atuam como tópico, caso antepostas, ou adendo, casos pospostas (Haiman, 1978; Hirata, 1999). Há portanto, uma diferença no modo como a condicional estrutura a experiência cognitiva codificada na oração complexa. Quando tópicos, atuam como fundo, organizando cognitivamente o cenário alternativo dentro do qual a informação principal, figura, será enunciada. Quando adendos, a oração condicional atua para acrescentar informação suplementar e/ou corrigir alguma informação em relação à unidade informacional veiculada na oração núcleo. Aqui se nota, portanto, uma diferença significativa nos usos dos conectores sob análise e uma distribuição quase complementar entre os dois conectores sancionados pelo subesquema [X que]_{CondPos}.

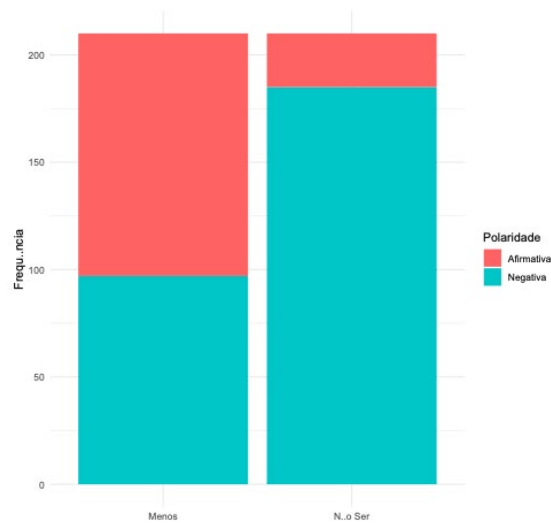
Gráfico 2. Posição da oração condicional



Fonte: Elaboração própria

No caso das microconstruções [a menos que] e [a não ser que], a polaridade da oração núcleo é que se mostrou um fator influente nos contextos de uso em que essas microconstruções aparecem. Conforme demonstra o Gráfico 3, orações condicionais introduzidas pela microconstrução conectora [a menos que] ocorrem preferencialmente combinadas com orações núcleo afirmativas, ao passo que condicionais introduzidas por [a não ser que] ocorrem majoritariamente articuladas a orações núcleo com polaridade negativa. Aqui, novamente, temos contextos em distribuição quase complementar, indicando a especialização dos usos das microconstruções associadas ao subesquema [X que]_{CondNeg}.

Gráfico 3. Polaridade da oração núcleo



Fonte: Elaboração própria

Para resumir, então, as informações recolhidas aqui, há que se ter em mente:

- no recorte analisado, há dois subesquemas associados ao significado condicional, cada um instanciado por duas microconstruções, com diferentes graus de convencionalização, mas não de significado;
- as microconstruções sob análise apresentam similaridades substanciais no que diz respeito à suas propriedades pragmáticas e semânticas, e diferenças contextuais significativas, mas não excludentes.

4. Implicações da similaridade na Gramática de Construções

Ora numa abordagem que seja em alguma medida funcional, as diferenças contextuais não são nenhuma surpresa, como já afirmado na introdução deste artigo. O problema que coloco aqui é, apesar de essas construções preferirem contextos específicos, elas podem, em alguma medida, acontecer em qualquer um dos contextos analisados. Veja que os gráficos não apresentam nenhuma relação *default*, e, no caso, dos tempos verbais essa diferença sequer se mostrou significativa. Como, então, dar conta desse fato, considerando, especialmente, que a abordagem construcional entende que a modelagem da gramática é fundamentalmente acionada por processos cognitivos e pela experiência? O que revela, portanto, a experiência dos falantes do português sobre o uso dos conectores aqui analisados?

Poder-se-ia argumentar que traços do contexto podem estar convencionalizados como propriedade de significado da construção, entendendo, portanto, que esses traços fariam parte do pareamento forma-significado, especificados no nível da microconstrução. No entanto, o que fica demonstrado aqui é que, mesmo com menor frequência, as construções analisadas não se mostraram exclusivas de nenhum contexto, o que significa que esses traços ainda não se convencionalizaram a ponto de subespecificarem as microconstruções. Na descrição das propriedades de forma e significado em níveis altos de abstração, o esquema e o subesquema, o Princípio da Não Sinonímia se mantém. No entanto, em níveis mais baixos de descrição formal, diferenças sintáticas nem sempre correspondem a diferenças funcionais, como por exemplo, os conectores [contanto que] e [desde que], que compartilham a função de restrição positiva e atuam nos mesmos contextos, mas são formalmente distintos. As evidências sugerem que a sinonímia construcional pode ocorrer em níveis mais granulares. Em outras palavras, a descrição realizada aqui identifica padrões de realização formal que não codificam diferenças funcionais.

Aqui entendo que essa indeterminação enfraquece, em alguma medida, o Princípio da não sinonímia, que, portanto, pode não ser aplicável em todos os níveis de descrição linguística. Por isso, diante dessas informações, proponho aqui a **hipótese da similaridade construcional**. Embora as microconstruções analisadas não sejam sinônimas perfeitas, elas compartilham uma parte substancial de significado construcional e podem ser consideradas duas opções construcionais para conceitualizarem um mesmo evento do mundo biossocial. Nesse sentido, o princípio ideal seria, especialmente considerando a gradiência da gramática, uma visão de que a linguagem tende a favorecer uma forma para um significado, mas não de modo tão absoluto. Assim, assumiria que embora apresentem diferenças, construções também exibem semelhanças relevantes que devem ser levadas em consideração em uma teoria gramatical, em especial numa teoria gramatical que se rotula moldada pelo uso.

Ao assumir a hipótese da similaridade construcional e a **alternância construcional**, entende-se que, para além das generalizações baseadas na construção específica, existem generalizações baseadas em alternâncias que também consistem em informações armazenadas na representação cognitiva dos usuários. Rejeito, portanto, o pressuposto de Goldberg (2006, p. 25, tradução própria) de que um padrão construcional “é mais bem analisado por si só, sem considerar referências implícitas ou explícitas a um possível sintagma alternativo”⁴. Posicionamento semelhante pode ser encontrado em Capelle (2006), Perek (2015), Diessel (2019, 2023) e, para o português, em Machado Vieira e Wiedemer (2020).

4 No original: “[...] is best analyzed in its own terms, without relying on explicit or implicit reference to a possible alternative frase”.

Aqui, invoco o processo cognitivo conhecido como **memória enriquecida**, de Bybee (2016, p. 59), segundo o qual os indivíduos não estocam as experiências individualmente, mas todo um conjunto de detalhes dessa experiência, que envolve contexto, inferências, frequência. “Representações cognitivas da gramática incluem informação específica sobre contextos de uso de exemplares e seus significados e implicações nesses contextos”. Ainda, “Uma vez que os exemplares fornecem um registro da experiência do falante com a língua, modelos que assumem exemplares como representações permitem a representação direta tanto da **variação** quanto da **gradiência**” (Bybee, 2016, p. 62).

Considerações finais

Meu ponto de partida neste artigo era analisar uma variedade de padrões de realização das construções conectoras condicionais, que podem ser mais adequadamente descritos como construções alternativas, ou seja, variantes formais que não codificam funções distintas, assumindo assim a hipótese da similaridade construcional. Considerando que a gramática é psicologicamente plausível e que tem como objetivo modular o conhecimento e as experiências do falante, é preciso levar em conta, também, o conhecimento que os falantes têm sobre a semelhança entre construções. E para finalizar, quero destacar algumas vantagens dessa abordagem:

- 1) **Modelagem da mudança diacrônica:** A competição entre construções pode levar a mudanças na frequência de uso, à extensão de padrões e até mesmo ao desaparecimento de construções. A hipótese da similaridade construcional permite entender como essas mudanças ocorrem na rede de construções ao longo do tempo, ilustrando como a competição entre construções similares molda o desenvolvimento das construções. É o que discuti, por exemplo, em trabalho recente sobre mudanças no subesquema [V que] condicional. Neste trabalho, observei que à medida que a microconstrução [supondo que] emerge e se consolida, competindo com outras construções sancionadas pelo mesmo subesquema e tornando-se mais o exemplar mais frequente, a microconstrução [dado que], também sancionada por esse subesquema, diminui significativamente seu uso, até não mais encontrar seu registro no português contemporâneo (Oliveira; Clemente, 2022);
- 2) **Aquisição:** no processo de aquisição, crianças generalizam não apenas padrões e regularidades rumo à abstratização e categorização, mas reconhecem similaridades e, por meio delas, expandem seus repertórios linguísticos;
- 3) **Priming:** o *priming*, processo por meio do qual uma forma linguística pode facilitar a ativação e o processamento cognitivo de formas linguísticas semelhantes, também demonstra a força da hipótese da similaridade construcional, já que a probabilidade de uma construção ativar outra é muito maior quando essas construções são similares, compartilham propriedades e podem ser usadas em contextos semelhantes.

É evidente que aqui não esgotei todos os testes para verificar a similaridade entre as microconstruções conectoras, outras propriedades contextuais como registro de uso da língua, tipos e gêneros textuais, domínio discursivo, testes psicológicos de reconhecimento podem e devem ser incorporados para a verificação da hipótese da similaridade construcional. O que fiz aqui foi tomar passos iniciais para defender a uma visão menos estanque do Princípio da não sinonímia, a favor da hipótese da similaridade construcional. Espero, assim, ter dado alguma contribuição à discussão.

Referências

- BOLINGER, D. *Meaning and form*. Londres: Longman, 1977.
- BYBEE, J. *Language, usage and cognition*. Cambridge: Cambridge University Press, 2010.
- CAPPELLE, B. Particle placement and the case for “allostructions”. *Constructions*. 2006.
- CEZARIO, M. M.; SILVA, T. S.; SANTOS, M. Formação da construção [Xque] conec no Português. *E-scrita-Revista do Curso de Letras da UNIABEU*, v. 6, n. 3, p. 229-243, 2015.
- CLEMENTE, C. G. C. *O subesquema [V_que] condicional no português*. 2020. Dissertação (Mestrado em Letras) – Universidade Federal de Mato Grosso do Sul, Três Lagoas, 2020.
- CROFT, W.; CRUSE, D. A. *Cognitive linguistics*. Cambridge: Cambridge University Press, 2004.
- CROFT, W. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press, USA, 2001.
- DAVIES, M.; FERREIRA, M. *Corpus do Português*. Disponível em: <http://www.corpusdoportugues.org>. 2006.
- DE VAERE, H.; KOLKMANN, J.; BELLIGH, T. Allostructions revisited. *Journal of Pragmatics*, n. 170, p. 96-111, 2020.
- DIESSEL, H. *The constructicon: Taxonomies and networks*. Cambridge University Press, 2023.
- DIESSEL, H. *The grammar network*. Cambridge: Cambridge University Press, 2019.
- GIVÓN, T. *Language, function and typology*. 1985.

GOLDBERG, A. E. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press, 2019.

GOLDBERG, A. E. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press, 2006.

GOLDBERG, A. E. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, 1995.

GRIES, S. *Quantitative corpus linguistics with R: A practical introduction*. Routledge, 2016.

GRIES, S.; STEFANOWITSCH, A. Extending collostructional analysis. A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, v. 9, n. 1, p. 97-129, 2004.

GYSELINCK, E. Modeling shifts and reorganizations in the network hierarchy. Sommerer and Elena Smirnova (ed.). *Nodes and networks in diachronic construction grammar*, v. 27, p. 107, 2020.

HAIMAN, J. Conditionals are topics. *Language*, v. 54, n. 3, p. 564-589, 1978.

HIRATA, F. B. M. *A hipotaxe adverbial condicional no português escrito contemporâneo do Brasil*. 1999. 231 f. Dissertação (Mestrado) – Faculdade de Ciências e Letras, Universidade Estadual Paulista "Júlio de Mesquita Filho, Araraquara, 1999.

HOFFMANN, T. Construction grammar and creativity: Evolution, psychology, and cognitive science. *Cognitive Semiotics*, v. 13, n. 1, 2020.

HUDSON, R. *An introduction to word grammar*. Cambridge: Cambridge University Press, 2010.

LANGACKER, R. W. *Cognitive grammar: A basic introduction*. Oxford University Press, 2008.

LAPORTE, S.; LARSSON, T.; GOULART, L. Testing the Principle of No Synonymy across levels of abstraction: A constructional account of subject extraposition. *Constructions and Frames*, v. 13, n. 2, p. 230-262, 2021.

MACHADO VIEIRA, M. S.; WIEDEMER, M. L. A variação no modelo construcionista da Linguística Funcional-Cognitiva. *Sociolinguística no Brasil: textos selecionados*. Porto Alegre: Editora da PUCRS, 2020. p. 265-304.

OLIVEIRA, T. P. A configuração vertical e horizontal da rede construcional. *Revista de Estudos da Linguagem*, Belo Horizonte, v.33, n.4, p. -27, 2025

OLIVEIRA, T. P. Conjunções adverbiais no português. *Revista de Estudos da Linguagem*, v. 22, p. 45-66, 2014.

OLIVEIRA, T. P. *Conjunções e orações condicionais no português do Brasil*. 2008. 155 f. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara, 2008.

OLIVEIRA, T. P.; DA CRUZ CLEMENTE, C. G. Esquematicidade e produtividade na reconfiguração da rede de conectores condicionais. *Revista do GEL*, v. 19, n. 3, p. 58-84, 2022.

PEREK, F. Alternation-based generalizations are stored in the mental grammar: Evidence from a sorting task experiment. *Cognitive linguistics*, v. 23, n. 3, p. 601-635, 2012.

RSTUDIO TEAM. RStudio: Integrated Development for R. Boston, MA: RStudio, PBC, 2023. Disponível em: <https://posit.co>. Acesso em: 28 set. 2024.

SOMMERER, L.; SMIRNOVA, E. (ed.). *Nodes and networks in diachronic construction grammar*. John Benjamins Publishing Company, 2020.

TRAUGOTT, E. C.; TROUSDALE, G. *Construcionalização e mudanças construcionais*. Tradução de Taísa Peres de Oliveira e Maria Angélica Furtado da Cunha. Campinas: Vozes, 2021.

VAN DE VELDE, F. Degeneracy: The maintenance of constructional networks. *Extending the scope of construction grammar*, v. 1, p. 141-179, 2014.

UHRIG, P. Why the principle of no synonymy is overrated. *Zeitschrift für Anglistik und Amerikanistik*, v. 63, n. 3, p. 323-337, 2015.

ZEHENTNER, E.; TRAUGOTT, E. C. Constructional networks and the development of benefactive ditransitives in English. *Nodes and networks in diachronic construction grammar*, p. 168-211, 2020.

Multiletramentos, tecnodiversidade e ensino de Português

DOI: <http://dx.doi.org/10.21165/el.v54i1.4030>

Ana Elisa Ribeiro¹

Resumo

Este texto se baseia na noção de *tecnodiversidade*, proposta pelo filósofo chinês Yuk Hui, para promover uma reflexão sobre técnicas e tecnologias para o ensino e a aprendizagem, especialmente de língua materna. Partindo da revisão de três documentos importantes – o manifesto da Pedagogia dos Multiletramentos (1996), as Diretrizes Curriculares Nacionais dos cursos de licenciatura em Letras (2001) e a Base Nacional Comum Curricular (BNCC, 2018) –, procuramos apontar o discurso sobre a diversidade em vários âmbitos, mas ainda não diretamente quanto às tecnologias cujos objetivos sejam ensinar e aprender na escola. Hui é convidado a auxiliar na compreensão do que seja uma escola ou uma aula tecnodiversas, assim como a encorajar que sejamos mais diligentes quanto a um futuro menos dependente e vulnerável na relação com empresas e plataformas dominantes.

Palavras-chave: Tecnodiversidade; ensino e aprendizagem de línguas; Multiletramentos.

¹ Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, Minas Gerais, Brasil; anadigital@gmail.com; <https://orcid.org/0000-0002-4422-7480>

Multiliteracies, technodiversity and Portuguese teaching

Abstract

This text is based on the notion of *technodiversity*, proposed by the Chinese philosopher Yuk Hui, to promote a reflection on techniques and technologies for teaching and learning, especially in Portuguese. Based on the review of three important documents – the Pedagogy of Multiliteracies Manifesto (1996), the Brazilian National Curricular Guidelines for undergraduate courses (Languages) (2001) and the National Common Curricular Base for Basic Education (BNCC, 2018) –, we seek to highlight the discourse on *diversity* in various areas, but not yet directly in relation to technologies whose objectives are to teach and learn in basic schools. Yuk Hui and his philosophy help us understand what is a technodiverse school or classroom, as well as to encourage us to be more diligent in relation to a future that is less dependent and less vulnerable in relation to dominant companies and platforms.

Keywords: Technodiversity; Language teaching and learning; Multiliteracies.

Considerações iniciais

Neste trabalho², nossas pesquisas e experiências em sala de aula na Educação Básica inspiram e motivam uma reflexão que parte da análise de aspectos específicos de documentos importantes, tais como a Base Nacional Comum Curricular (BNCC), homologada em 2018, e as Diretrizes Curriculares Nacionais dos cursos de licenciatura em Letras, de 2001, ainda vigentes. Outro documento importante, tanto para esta argumentação quanto para a educação brasileira, é o chamado manifesto da Pedagogia dos Multiletramentos, publicado em 1996, nos Estados Unidos, e traduzido ao português brasileiro apenas em 2021³. Todos eles defendem e reivindicam, de alguma maneira, em

2 Este texto, que bem poderia ser chamado de ensaio, gênero pouco comum entre linguistas, mas que permite especular e pensar com alguma liberdade, deriva de minha participação na mesa intitulada “Multiletramentos, tecnologias e ensino”, ocorrida em 17 de julho de 2024, durante a 70ª reunião do GEL, em Campinas. Na ocasião, me acompanhavam as professoras Márcia Mendonça e Jacqueline Barbosa, às quais dedico estas linhas que expressam poucas certezas. Colegas como elas só aumentam nosso compromisso com a pesquisa em linguística aplicada. Meus agradecimentos ao CNPq (bolsa de produtividade em pesquisa), ao CEFET-MG e ao Grupo de Estudos Linguísticos do Estado de São Paulo (GEL).

3 O manifesto do New London Group, bastante conhecido no Brasil, ganhou duas traduções autorizadas brasileiras, mas apenas mais de 25 anos depois de publicado nos EUA. Tais versões foram feitas simultaneamente, por grupos diferentes. Uma foi publicada como artigo da revista *Linguagem em Foco*, da Universidade Estadual do Ceará (UECE), e foi traduzida por um grupo ligado ao prof. Petrilson Pinheiro (Unicamp); a outra foi publicada como livro pela Led, editora-laboratório do curso de Letras do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), somada a um glossário temático produzido por estudantes da Universidade Federal de Ouro Preto (UFOP). Tanto artigo quanto livro são de acesso aberto.

uma ideia de diversidade que é perseguida até os dias de hoje. Seja quanto às culturas que convivem na escola, seja quanto à multiplicidade de gêneros discursivos/textuais que deveriam ser abordados e ensinados em sala de aula, a noção de *diversidade* atravessa diretrizes, bases, manifestos e, de cima para baixo, provavelmente emerge nas escolas como algo a se realizar e exercer, em termos de tolerância, acolhimento e respeito, mas também de versatilidade. Já em relação às tecnologias presentes (ou ausentes ou escassas) no espaço escolar, raramente se aborda a diversidade como algo importante e até valorizado. É comum que os discursos sobre tecnologias em sala de aula, hoje, digam respeito às tecnologias digitais e carreguem uma nuance competitiva, de evolução, projetando futuros menos analógicos ou sugerindo uma obsolescência geralmente danosa para aqueles e aquelas que ensinam e aprendem.

Aqui, a ideia é retomar os discursos construídos nos documentos, a partir de uma leitura detida e a contrapelo de seus ecos, isto é, desdobramentos de uns documentos em outros, na forma de citações ou de influências declaradas, além de propor uma articulação dessas orientações e normas com a noção de *tecnodiversidade* proposta pelo filósofo chinês Yuk Hui (2020). Além da proposição da *tecnodiversidade* como uma força contracolonialista, antiplataformização e crítica, Hui também aborda noções como as de *cosmotécnicas* e de *sincronização*, além da de *fragmentação*, todas explicadas em sua obra e brevemente retomadas neste texto. Essa costura teórica se justifica pela necessidade de pensarmos criticamente um discurso sobre diversidade e tecnologias, inclusive quanto a documentos que pressionam nossas atividades docentes diárias, levando em consideração aspectos das escolas e de nossas salas de aula – reais – que dizem respeito à infraestrutura e aos recursos (in)disponíveis para a entrada em uma era de multiletramentos (Cazden *et al.*, 2021), na qual as novas mídias estariam contempladas e fariam diferença nos repertórios textuais e discursivos de nossos e nossas estudantes (e nos nossos próprios).

A partir, então, do documento mais antigo aqui mobilizado, isto é, o manifesto da Pedagogia dos Multiletramentos (1996), passando pelas Diretrizes Curriculares Nacionais (2001) e desaguando na BNCC (2018), vamos retomar a ideia de *diversidade* articulada em todos eles quanto a elementos diferentes, e pensar, à luz da proposta filosófica de Yuk Hui, como as salas de aula reais nas escolas brasileiras têm admitido ou reagido ou sobrevivido ou simplesmente atravessado as décadas (precariamente) equipadas para plenos multiletramentos. Um giro no pensamento sobre *tecnodiversidade* pode nos ajudar a entender nossas forças e fraquezas⁴, mesmo em ambientes políticos e educacionais

4 Uso as noções de forças e fraquezas tomadas de empréstimo a uma conhecida matriz da área de Administração, a matriz SWOT ou FOFA (em português), que serve, normalmente, para analisar organizações de vários tipos quanto a suas forças, fraquezas, oportunidades e ameaças. A matriz foi proposta por Kenneth Andrews e Roland Christensen, professores da Harvard Business School, e pode ser encontrada em diversos manuais de introdução à Administração ou de Planejamento estratégico. Neste trabalho, a intenção não é produzir uma análise desse tipo, mas apenas pensar que o que podemos ser levadas a sentir, todo o tempo, como fraqueza, ao contrário, é uma força de nossa atuação e de nossas soluções, dentro das possibilidades que temos.

adversos (ou seja, quase sempre), em nossa relação diária com estudantes em aulas pouco “inovadoras” do ponto de vista infraestrutural ou de recursos tecnológicos, mas que dependem de outros itens ligados à nossa criatividade e interação com jovens e crianças (na Educação Básica regular). Entender as configurações da tecnodiversidade em nossos contextos (escolas, salas, laboratórios...) pode ser útil para que repensemos tanto o esforço de atender aos documentos oficiais quanto certa sensação permanente de insuficiência e mesmo de incompetência, advinda de discursos que, afinal, atuam contra nós, até mesmo, em alguns casos, quando enunciados pela academia.

Multiletramentos e depois: DCN e BNCC

Para uma primeira incursão pela ideia de que a diversidade é positiva no ensino de linguagens, é fundamental retomar a proposta do chamado New London Group (NLG), coletivo de dez professoras e professores de língua materna (inglês, em seus países) que produziu, colaborativamente, um texto publicado em 1996 por um periódico científico de Harvard. O material foi intitulado, pelos próprios autores, de “manifesto”, provavelmente porque, além de explicitar uma análise daquele momento da abordagem do letramento tradicional (*literacy*), propunha uma nova abordagem, suplementar à primeira, tributária dela, mas que agregava novos elementos e possibilidades, considerando todas as mudanças pelas quais passava a sociedade (ou as deles) naqueles anos. Entre tais mudanças estava a das novas mídias, sem grande detalhamento, mas possivelmente já tratando dos computadores, da editoração eletrônica, da Internet e outros ambientes que passaram a existir e a alcançar muita gente, exatamente naquele momento histórico. Contemplada por essa mudança de “mídias” estava a emergência de uma diversidade textual que trazia novidades importantes e, segundo o NLG, deveria ser considerada e abordada nas aulas de língua. Isso significa dizer que as tecnologias digitais estavam subsumidas na noção de multiletramentos, embora não aparecessem sob qualquer destas siglas: TDIC, TIC ou TD. Pode-se depreender, sem muito esforço, que assumir uma visada multiletrante⁵ no ensino de português, por exemplo, pressupõe uma preocupação com as diversidades tecnológica e textual. Dito de outro modo, não se provocam ou propiciam multiletramentos plenamente sem a providência ligada às diversidades de textos e suas tecnologias de produção e circulação. O dito letramento digital, uma das inúmeras derivações ou especializações do debate geral sobre letramentos (Soares, 2002; Coscarelli; Ribeiro, 2005), está contemplado na ideia macro dos multiletramentos, e isso pode ser identificado em trechos do manifesto facilmente encontráveis. Vejamos um deles, no qual as intenções do NLG são declaradas, a partir da observação de elementos como a diversidade linguística e a emergência de tecnologias da comunicação:

5 Tenho pensado a ação de atuar no ensino de Português informada pelos multiletramentos como uma atuação multiletrante, isto é, preferindo o termo com esse sentido que não cessa, que produz ou quer produzir um efeito sobre o outro, com agência. É diferente de pensar numa “visada multiletrada”, no particípio passado ou como um adjetivo.

Os dez autores deste texto são educadores que se encontraram durante uma semana, em setembro de 1994, em Nova Londres, New Hampshire, nos Estados Unidos, para discutir o estado da pedagogia dos letramentos. Os membros do grupo ou tinham trabalhado juntos, ou tinham se alimentado de seus respectivos trabalhos, ao longo de vários anos. As principais áreas de interesse comum ou complementar incluíam a tensão pedagógica entre abordagens de ensino explícitas e imersivas; o *desafio da diversidade linguística e cultural*; a *recente proeminência dos modos e tecnologias da comunicação*; e as *mudanças nos usos dos textos* em espaços de trabalho reestruturados. Quando nos encontramos em 1994, nossa intenção era consolidar e ampliar essas relações a fim de abordar o problema mais amplo dos *propósitos da educação*, e, neste contexto, o problema específico da pedagogia dos letramentos (Cazden et al., 2021, p. 15, tradução e grifo próprios).

Em outros trechos, podemos acompanhar a questão maior que levava ao encontro dos professores e professoras – os “resultados sociais da aprendizagem de língua” –, assim como o exato momento em que o grupo propõe o neologismo que, mais tarde, adotariamos amplamente no Brasil:

Concordamos que precisamos voltar à questão mais ampla dos *resultados sociais da aprendizagem de língua*, e que tínhamos, baseado nisso, que repensar as premissas fundamentais de uma pedagogia dos letramentos, a fim de influenciar *práticas que darão aos estudantes as habilidades e o conhecimento de que precisam para alcançar suas aspirações*. Concordamos que, em cada país anglófono de onde viemos, *o que os estudantes precisavam para aprender estava se transformando*, e que o principal elemento dessa mudança era que não havia mais um inglês singular, canônico, que poderia ou deveria ser ensinado. *Diferenças culturais e a rápida mudança nos meios de comunicação significavam que a própria natureza do assunto – a pedagogia do letramento – estava mudando radicalmente* (Cazden et al., 2021, p. 17, tradução e grifo próprio).

Decidimos que os resultados de nossas discussões poderiam ser encapsulados em uma palavra – multiletramentos – palavra que escolhemos para descrever *dois importantes argumentos* com que podemos abordar as ordens cultural, institucional e global emergentes: a *multiplicidade de canais de comunicação e mídia* e a crescente proeminência da *diversidade cultural e linguística*. A noção de multiletramentos suplementa a pedagogia tradicional do letramento ao abordar esses dois aspectos relacionados à *multiplicidade textual*. (Cazden et al., 2021, p. 18, tradução e grifo próprio).

Note-se que *diversidade* e *multiplicidade* são tomadas quase como sinônimas no texto (a tradução é fiel, embora, como toda tradução, seja uma aproximação), mas o que

importa é observar que a transformação social (tecnológica, cultural) justificava a revisão de uma pedagogia que, segundo o NLG, já não era suficientemente eficaz ou não oferecia os “resultados sociais” compatíveis com as exigências do momento (e do mercado de trabalho). A proposição de um neologismo – multiletramentos – buscava ampliar o escopo do ensino de língua, incluindo então os canais de comunicação e mídia múltiplos (e novos na cena), dos quais decorria uma multiplicidade/diversidade de textos, além da atenção especial à diversidade cultural e linguística gerada pelo contato maior entre pessoas (migrações, ascensão social etc.). Se *diversidade* e *multiplicidade* não foram definidas ou explicadas no manifesto, vemos espaço para tratá-las, aqui, como intercambiáveis, ambas levando à mesma preocupação com novos repertórios, inclusive multimodais e ligados a tecnologias recentes.

Foi esse manifesto, proponente de uma pedagogia de multiletramentos que incluía uma metalinguagem (*design*, *redesign* e outros), que, nos anos seguintes, passou a influenciar a discussão sobre letramentos no Brasil. Não há dúvida de que a abordagem dos multiletramentos fundamenta a BNCC⁶. Isso está explícito lá, embora não detalhado. Mas, antes de chegar à Base Nacional Comum Curricular, passemos pelas Diretrizes Curriculares Nacionais (DCN) para os cursos de Letras, documento publicado em 2001 e vigente no país, com orientações ainda válidas para a formação inicial de professoras e professores de línguas.

O texto das DCN para Letras é breve e se divide em partes que descrevem, por exemplo, o perfil do/da profissional formado/a e as competências e habilidades que ela e ele devem desenvolver ao longo de sua formação superior. Na seção de Perfil, por exemplo, as DCN orientam que a pessoa licenciada em letras “Deve ser capaz de refletir teoricamente sobre a linguagem, de *fazer uso de novas tecnologias* e de compreender sua formação profissional como processo contínuo, autônomo e permanente” (Brasil, 2021, p. 30, grifo próprio). Interessante notar que, pouco depois da virada do milênio, o documento oficial já mencione as TDIC, sem dizer exatamente que usos seriam esses. Podemos perguntar se usos pessoais/individuais ou como ferramenta de trabalho, para o ensino de língua etc.

Já entre as competências e habilidades, o documento orientava para uma “[...] preparação profissional atualizada, de acordo com a dinâmica do mercado de trabalho; [...] *utilização dos recursos da informática*” (Brasil, 2001, p. 30). Nem as tais “dinâmicas do mercado de trabalho”, nem os “recursos da informática” são definidos ou explicitados, assim como suas aplicações. Essa nos parece uma oportunidade de propor algumas perguntas que ajudem na reflexão:

6 Discuti isso mais detidamente em Ribeiro (2020).

- Está claro, nos dias atuais, o “resultado social” que alcançamos com nossas aulas de português ou língua estrangeira?
- Esse resultado nos parece satisfatório? Em que sentidos e em relação a quê?
- Em que medida esse resultado se relaciona ao uso de novas tecnologias ou de recursos da informática?
- Nosso uso pessoal/individual de novas tecnologias se transformou muito nas últimas duas décadas? Como?
- Nossa atividade docente, isto é, em sala de aula, sofreu transformação tecnológica e metodológica de igual impacto, nos recentes vinte anos?

Importa notar que as diretrizes nacionais para a formação docente inicial explicitam, desde 2001, as tecnologias digitais como item fundamental para o desempenho profissional das pessoas de Letras. Aspecto que Magda Soares (2002), mesmo sem referenciar o New London Group, considerava inescapável, pouco depois da publicação das DCN. Para Soares, uma das maiores expoentes dos estudos de letramentos no Brasil, a chegada das tecnologias digitais era uma “oportunidade extremamente favorável” (p. 146) de refinar, precisar e esclarecer o “letramento”. Parecia claro, para ela, que “novas e incipientes modalidades de práticas sociais de leitura e de escrita, propiciadas pelas recentes tecnologias de comunicação eletrônica – o computador, a rede (a web), a Internet” (p. 146), afetavam nosso cotidiano e levavam a outro tipo de letramento, necessariamente do interesse de nossas aulas (e responsabilidade nossa). Era, portanto, fundamental rever o ensino de leitura e escrita (objeto do ensino de língua), incorporando – criticamente – as novidades, a multiplicidade textual que, logo adiante, entenderíamos como uma onda avassaladora e incontornável. Demos conta dela?

Mesmo com indicações em documentos oficiais, as tecnologias digitais parecem continuar pouco presentes na paisagem escolar. Em alguns casos, até com recuos, como foi com laboratórios de informática, hoje desativados em muitas instituições. É evidente que isso tem relação com uma série de dificuldades ligadas à obsolescência dos equipamentos, à falta de investimento financeiro nas escolas, ao sucateamento, à dificuldade (financeira e burocrática) de manutenção adequada, reposição, atualização etc. Por outro lado, também é resultado das próprias mudanças tecnológicas, que ofereceram recursos cada vez menores e mais portáteis/móveis, dispensando equipamentos como computadores de mesa e outros. Os *smartphones*, de uso pessoal, são, hoje, o recurso mais disponível, mas talvez um dos mais controversos, tanto em termos de uso escolar quanto de uso pessoal, em especial por crianças e adolescentes.

Diante da imprecisão dos documentos aqui citados, passemos a pensar em termos de inventários de equipamentos e recursos, repertórios de textos e outras possibilidades. Como visto, as tecnologias são abstrações nas diretrizes. Ficam então as perguntas, de cunho provocador:

- Qual é exatamente o conjunto de tecnologias digitais que poderíamos integrar ao ensino de língua, hoje?
- Esse conjunto diz respeito a equipamentos (*hardware*) e a recursos como programas, plataformas, aplicativos. É possível fazer um inventário sempre renovado deles e realmente relevante para ensinar a ler e a escrever?
- Incurremos em uma ideia equivocada das tecnologias, considerando-as excludentes entre si? (relação competitiva entre analógico e digital, por exemplo)
- Como lidar, diante da pressão pela incorporação de TDIC às aulas, com discursos e evidências na psicologia, na saúde etc. segundo os quais os usos das tecnologias são prejudiciais à infância e à juventude?
- Que conjunto de tecnologias digitais é pertinente ao ensino de língua materna? Há clareza de objetivos e possibilidades nesses usos?
- O que quer dizer uma aula melhor, em termos do emprego de tecnologias? (Aprende-se melhor, mais rápido, mais efetivamente, mais duradouramente...) (Ensina-se melhor, mais rápido, de modo mais claro, com resultados mais perceptíveis etc.?)
- Como medir ou identificar que novos e melhores resultados de ensino e aprendizagem são precisamente uma função do uso de certas tecnologias?
- Como estabelecer relações confiáveis entre o que ensinamos em aulas de línguas e os resultados sociais alcançados pela juventude, em um país tão instável e hostil quanto o Brasil?
- Os usos de TD passam pela preparação de aulas em Word, pelos PDF, pelos sistemas acadêmicos de gestão nas escolas, pelas práticas com apresentações (PowerPoint e assemelhados). Incontáveis professores e professoras são discretamente digitais em seus cotidianos profissionais. No entanto, isso nem sempre é considerado “uso de tecnologia nas aulas”. O que se espera que seja esse uso ostensivo e evidente?

A Base Nacional Comum Curricular (Brasil, 2018) é o mais recente documento em vigor, com força de norma, que pretende influenciar toda a educação brasileira, em frentes como a formação inicial e continuada docente, os materiais didáticos, as avaliações em massa, além, claro, das salas de aula de todo o território do país. Não é, no entanto, um documento completamente transparente. Sua opacidade reside, por exemplo, na falta de referências explícitas, na ausência da explicação de conceitos e noções, na vagueza de algumas indicações. Pode-se afirmar, no entanto, que as seções que dizem respeito ao ensino de língua portuguesa têm forte influência da noção de multiletramentos proposta pelo NLG, já comentado, e apontam explicitamente para o ensino de uma diversidade de gêneros textuais/discursivos relacionada, também, a ambientes digitais de comunicação. Antes dos anos 1990, seria improvável encontrar uma lista de gêneros que incluísse *vlogs* ou *memes*, por exemplo, conforme vemos à página 73 da Base.

A fim de verificar a orientação pela diversidade, basta ver que, por exemplo, na lista de dez competências gerais para a Educação Básica, o documento já trata da “diversidade de saberes e vivências culturais” (Brasil, 2018, p. 9, item 6) e da “diversidade humana” (Brasil, 2018, p. 10, item 8). A formulação “diferentes gêneros” ou “gêneros diversos”, referindo-se aos textos, aparece muitas vezes nos quadros de habilidades a serem desenvolvidas pelos jovens, como à página 72, para um exemplo. Vejamos uma afirmação contundente sobre as orientações em vigor: “As práticas de linguagem contemporâneas não só envolvem *novos gêneros e textos cada vez mais multissemióticos e multimidiáticos*, como também novas formas de produzir, de configurar, de disponibilizar, de replicar e de interagir” (Brasil, 2018, p. 68, grifo próprio). Trata-se de um alinhamento importante à noção de multiletramentos, explícita, por exemplo, na página 70 ou na 72 da BNCC, e completamente alcançada neste parágrafo do documento:

Não se trata de deixar de privilegiar o escrito/impresso nem de deixar de considerar gêneros e práticas consagrados pela escola, tais como notícia, reportagem, entrevista, artigo de opinião, charge, tirinha, crônica, conto, verbete de enciclopédia, artigo de divulgação científica etc., próprios do letramento da letra e do impresso, mas de contemplar também os novos letramentos, essencialmente digitais (Brasil, 2018, p. 69).

A adoção de uma visão *multi* dos letramentos (já no plural) nos ajuda a fazer, portanto, uma articulação produtiva com outras diversidades, agora pensando no que elas produzem de positivo e de competente em termos de “resultados sociais” possíveis. Diante das questões postas e de um esforço para formular algumas respostas, encontramos oportunidade de tecer uma relação entre a noção de *tecnodiversidade*, proposta por Yuk Hui (2020), e nossas práticas docentes em cenários escolares forçosamente tecnodiversos (e quase nunca por escolha nossa e da comunidade acadêmica).

Tecnodiversidade, sincronização e fragmentação

O construto filosófico proposto por Yuk Hui passa por várias noções que podem nos ser úteis para pensar as tecnologias digitais nas escolas da Educação Básica. Esse não é o âmbito ao qual o autor se dedica, mas pode estar contemplado, já que Hui tem contribuído para que repensemos nossas sociedades em termos globais. Entre as noções elencadas em sua obra (Hui, 2020), um compilado de textos antes esparsos, estão as de *cosmotécnicas*, *sincronização*, *fragmentação* e *tecnodiversidade*. De saída, é importante notar que o filósofo chinês pretende provocar criticamente em especial os países não europeus, sobretudo os emergentes, que são vitimados pela plataformação por empresas de tecnologia dominantes (D'Andréa, 2020). A equalização dos processos e acessos levaria ao que Hui chama de *sincronização*, algo a que deveríamos estar atentos (e evitar). Justamente para alcançar um ambiente menos dependente, e até refém, das mesmas empresas, é fundamental preservar nossas *cosmotécnicas* (modos diversos de

fazer as coisas) e garantir a *fragmentação* das técnicas e tecnologias, de maneira que tenhamos sempre uma diversidade saudável e garantidora da sobrevivência (intelectual, material, cultural, entre outras). Seria uma maneira de resistir ao colonialismo de dados, por exemplo, e ao colonialismo geral, que se veste de outras formas hoje (Casati, 2017).

Neste texto, a noção de que nos aproximaremos mais é a de *tecnodiversidade*. Vale retomar um trecho de Hui (2020, p. 130⁷) no qual ele explica o fundamento dessa noção:

O fundamento da ecologia são as diversidades, já que é só a partir das biodiversidades (a variedade e a variabilidade de genes, espécies e ecossistemas) que é possível conceptualizar o sistema ecológico. Para discutir a ecologia das máquinas, precisamos de um conceito análogo ao da biodiversidade: a tecnodiversidade.


A noção de biodiversidade, aliás, tem inspirado outros âmbitos do pensamento, sempre no sentido de que ser diverso é garantir opções, é fortalecer os laços, é respeitar o outro, é prover com repertórios múltiplos, evitando a dependência de algo único. Para o autor, em outro trecho:

É necessário e urgente desenvolver uma tecnodiversidade como orientação ao futuro e como política de descolonização. É ao mesmo tempo uma tarefa de reconstrução das histórias das cosmotécnicas que foram obscurecidas pela busca de uma história universal da tecnologia (e de uma história universal da humanidade), e um chamado à experimentação na arte e na tecnologia do futuro (Hui, 2020, p. 191).

A fim, portanto, de desenvolver tecnodiversidade, o filósofo nos convoca a pelo menos duas ações: reconstruir as “histórias das cosmotécnicas” apagadas por tecnologias que chegam e dominam o cenário; e experimentar novos modos de fazer, produzir, atuar, viver. Isso a fim de que evitemos estar todos reféns das mesmas técnicas e tecnologias todo o tempo, administrados por uma única entidade, sincronizados pelos mesmos *templates* e, afinal, organizados de maneira nada diversa. Parece-nos, então, que assim, nessa mesma chave, também podemos pensar nossas atividades escolares.

Considerando os dois movimentos – de recuperação e de criação – propostos por Hui, podemos inventariar situações reais em nossas salas de aula, em especial quanto às práticas de leitura e escrita ou, melhor, aos nossos modos de ensinar e aprender a ler

7 A obra tem versão em português, lançada em 2021 pela editora Ubu. O título brasileiro é *Tecnodiversidade*. Venho usando a versão argentina porque soube da brasileira apenas depois de iniciar os trabalhos de leitura e investigação das propostas de Yuk Hui e suas possíveis pertinências em nosso contexto.



e escrever, preferencialmente tecnodiversos, em termos de recursos, mas também de métodos e técnicas. Para isso, é importante ter uma visão sistêmica (ecológica talvez) e não concorrente do que acontece nas escolas, nas aulas, nas interações entre humanos e máquinas (analógicas e digitais), a fim de identificar as cosmotécnicas que convivem, talvez até inventariá-las, justapondo-as, em vez de substituir umas por outras sob um discurso insidioso de “antiquado”, “velho”, “desatualizado”, “modernizante” etc., ampliando o leque de possibilidades dos jovens cidadãos e cidadãs que formamos, além do nosso próprio. É claro que devemos pensar na pertinência entre os objetivos de uma aula e os recursos que podem/devem fazer parte dela, mas essa pertinência é sempre avaliada contingentemente, isto é, para cada caso, conforme o que aconteça e seja necessário.

Se a proposta de Yuk Hui para a sociedade, em especial sociedades não dominantes (ou dominadas), parece um tanto abstrata, talvez seja possível dar mais concretude a ela imaginando os repertórios conhecidos nas escolas para ensinar e aprender a ler e escrever, usando tecnologias de variadas naturezas, mescladas ou não, tomando mais consciência da tecnodiversidade possível para cada caso, conforme, claro, as condições infraestruturais reais em nossos territórios.

Talvez seja possível afirmar que grande parte de nossas escolas é tecnodiversa não por opção, mas por necessidade, isto é, combinamos recursos e técnicas que resistiram às mudanças avassaladoras, à ingerência política modernizante e privatizadora, ou foram salvaguardados ao longo do tempo justamente porque, de algum modo, percebemos que se certos itens forem dispensados ou excluídos simplesmente não haverá o que os substitua. Para mencionar um caso extremo, aconteceu de importantes bibliotecas, ao redor do planeta, dispensarem seus acervos impressos sob o argumento de que a microfilmagem ou o escaneamento guardariam tudo. É claro que o arrependimento foi inevitável (Darnton, 2010).

Como é tecnodiversa a minha escola: pensar em forças e fraquezas

Imaginemos o repertório de recursos disponíveis em uma escola pública urbana para que uma professora de Português ensine a ler e a escrever no Ensino Médio. Ela pode mobilizar uma série de itens que servirão de base para a leitura, o debate e a produção textual dos jovens estudantes. Supondo que a docente faça uma curadoria e selecione um texto a ser lido e debatido, nos dias de hoje, é possível imaginar que isso seja feito de muitas maneiras:

- texto impresso fotocopiado entregue aos estudantes;
- pdf ou outro tipo de arquivo do texto enviado por *e-mail*, via sistema acadêmico (se a escola tiver um), grupos de WhatsApp etc.;
- projeção do texto em algum tipo de tela, parede etc.

A professora pode organizar a turma para que os/as estudantes tenham acesso garantido ao texto selecionado, no mesmo dia e horário, a fim de que possam trabalhar/ler juntos em sala; ou ela pode solicitar que as pessoas leiam previamente o material para que a discussão se dê sem delongas ou leituras *in loco*. Caso o texto escolhido seja recebido pelos/as estudantes como arquivo digital, ele pode ser acessado em *smartphones* de muitos modelos, em *tablets*, em *e-readers*, em computadores portáteis ou de mesa, a depender dos acessos que cada pessoa tenha. Já o texto impresso poderá ser um original em livro, em revista, em jornal ou uma cópia em papel A4, por exemplo.

Todo esse repertório de possibilidades não é outra coisa senão um inventário tecnodiverso com que lidamos frequentemente. No entanto, sabemos que ele varia em termos de opções, de acordo com a menor ou maior precariedade de nossas condições de trabalho e estudo⁸. Isto é, é uma *força* quando olhamos seu lado múltiplo e suas possibilidades de articulação, mas é uma *fraqueza* quando nos damos conta de que nos faltam recursos e ferramentas ou quando nos abatemos, nos sentindo desatualizados, desprestigiados e até fora de moda.

Nem toda escola disporá de certas telas, certos computadores e projetores; nem todo docente disporá de máquina de fotocópia na escola; nem todo estudante sabe o que é um *e-reader*, nem todo jovem tem um *smartphone* de último tipo ou conhece uma lousa digital; um número inaceitável de escolas brasileiras não tem biblioteca, portanto não dispõe de um bom e bem-cuidado acervo de livros ou de exemplares para turmas inteiras. O que as escolas precárias e as pessoas que nelas atuam têm feito é pensar sempre um plano B. Nesse sentido, é necessário, em países instáveis como o nosso, que os planos B estejam sempre de prontidão (ON, como diz a juventude), porque é no que nos fiamos quando falta o acervo, falta a luz, falta o equipamento, não existe o laboratório. Nesse sentido, a tecnodiversidade é garantidora da nossa sobrevivência, em condições que quase sempre estão longe do ideal (e longe da BNCC, e não que ela seja a encarnação do ideal).

Continuemos pensando em nossa aula de leitura, da qual decorrerá uma produção textual cujo objetivo pode ser a aprendizagem de certos mecanismos de argumentação, a avaliação de questões gramaticais, de elementos de textualidade, as configurações multimodais de certos textos etc. Além de provocar a leitura de um ou mais textos, que chegam às pessoas de modos diversos e são efetivamente lidos de muitas maneiras, suponhamos que seja preciso apresentar um seminário oral para promover o debate de ideias. Esse seminário pode ser apresentado de algumas maneiras conhecidas:

- com o apoio de cartolina ou outro material celulósico, escrito com canetas, lápis, colagens etc.;

⁸ Estou sempre inquieta com a precariedade de nossas condições. Em pelo menos duas oportunidades, tratei disso com muita indignação. Ver, a propósito, Ribeiro (2021, 2023).

- com o apoio de *banner* ou pster de lona ou similar, geralmente produzido com o apoio de computador e programas específicos;
- com projeção em tela ou parede a partir de um programa tipo PowerPoint ou outro (Prezi, Canva, PiktoChart etc.);
- apresentação oral sem projeção ou demonstração;
- apresentação oral com “cola” (papel, celular ou algum tipo de suporte de anotação e lembrete);
- *hand-out* ou algum tipo de material impresso distribuído às pessoas;
- algum experimento demonstrado no momento.

É claro que a lista pode ser aumentada, mas fiquemos com essas opções, que já parecem bem tecnodiversas. Novamente, elas serão menos e mais acessíveis ou possíveis conforme a escola, mas não apenas em termos de recursos e equipamentos. É importante perceber que a tecnodiversidade está também entre os conhecimentos das pessoas, isto é, deve fazer parte dos repertórios dos docentes e de seus alunos e alunas, são suas cosmotécnicas diversas. É importante que essas opções existam, estejam vivas entre os modos de fazer, embora, é claro, elas suponham processos criativos e mesmo efeitos variados. Produzir uma apresentação em PowerPoint exige conhecimentos um pouco diferentes de produzir a “mesma” apresentação em cartolina. Em termos multimodais, as possibilidades não são exatamente iguais. No entanto, trata-se de uma lista de modos de apresentar um trabalho em sala de aula menos e mais recentes, menos e mais conhecidos no ambiente escolar. Para algumas escolas, a tecnodiversidade será uma questão de opção, de variação; para outras, não haverá muita margem de manobra. E isso, claro, produz exclusão, por um lado, mas também resguarda repertórios com os quais as pessoas sempre puderam aprender. Por que substituí-los completamente ou simplesmente desprestigiá-los?

Se um docente exige que as atividades sejam feitas e apresentadas apenas de uma maneira, ou se ele só admite apresentações de um único jeito, ele provavelmente omite o repertório global de seus estudantes. Assim é com recursos, textos, escritas, gêneros textuais etc. A tecnodiversidade pode ser vista, então, não apenas como uma necessidade em condições precárias, mas como uma maneira de escapar da sincronização e da dependência intelectual ou tecnológica, algo para o que, segundo Yuk Hui, estamos sendo empurrados. Se há muitos modos de fazer – recuperados e criados –, por que não resguardá-los, em vez de substituí-los como se fossem, *a priori*, dispensáveis? Um mapa complexo de possibilidades é trabalhoso, mas pode nos tornar muito menos vulneráveis.

No universo do livro, isto é, no âmbito da produção editorial, a diversidade ou multiversidade (Hawthorne, 2018) vem sendo discutida há décadas. Não apenas há uma consciência da importância do que é chamado de *bibliodiversidade*, como acontece um esforço de

preservação do patrimônio gráfico. Nem tudo precisa ser digital, nem tudo quer ser digital, é importante preservar cosmotécnicas ligadas à produção livreira, num sentido não apenas de arquivamento, mas de ensinar gente jovem a produzir com técnicas passadas, mesmo nos dias de hoje.

Quanto à sala de aula e a outros espaços de ensino e aprendizagem na escola, o que podemos imaginar e produzir? Com base na proposta filosófica de Yuk Hui (2020), parece interessante deixarmos de pensar na substituição (ou quase) de certos repertórios de técnicas e recursos por outros, estes supostamente sempre novos, melhores, mais interessantes, espertos etc. O valor está em preservar o que sempre foi feito e produziu bons efeitos (boas aulas expositivas e as anotações das pessoas atentas) no conjunto pertinente e consciente com outros modos de apresentar ou ouvir (vídeos gravados ou outra forma análoga, mas não a mesma), assim como imaginar, projetar, criar e experimentar maneiras novas de alcançar objetivos educacionais, isso com tecnologias recentíssimas ou não.

Mais uma provocação: que elemento intrínseco nos faz acreditar que o uso de algum recurso digital para certa aula dispensa o uso de outro recurso de natureza diferente? O tecnodiverso se sobrepõe a essa questão, estimulando repertórios cada vez mais diversificados e múltiplos.

Considerações finais

Nesta seção final, pretendemos tecer considerações flexíveis sobre os repertórios e as dinâmicas de ensino e aprendizagem de língua materna em sala de aula, na escola regular brasileira, levando em consideração a noção de tecnodiversidade como algo que, afinal, já promovemos, embora não pelos melhores motivos (a extrema desigualdade na distribuição de recursos nos desafia todo o tempo). No entanto, a consciência do que seja a aula ou a escola tecnodiversas pode fazer nosso trabalho mais diligente, com possibilidades mais interessantes, conjuntos de itens mais múltiplos, com menos julgamento sobre o que é antigo, mas não necessariamente inútil, imprestável ou obsoleto.

A *sincronização* promovida pelas grandes empresas de tecnologia mundiais produz o que conhecemos como “obsolescência programada”, isto é, equipamentos e *softwares* com prazo de validade determinado, geralmente curto, que levarão à frenética substituição de uns recursos por outros que elas mesmas venderão. Se o discurso da obsolescência afeta a escola e suas práticas educacionais, talvez caiamos numa armadilha desnecessária. A proposição filosófica do professor chinês Yuk Hui (2020) aponta para a necessidade de preservação de modos de fazer diversos, a fim de que possamos compor acervos tecnodiversos que, se bem conhecidos e articulados, nos ajudem a ensinar e aprender de várias maneiras, conforme a necessidade, o momento, o objetivo desta ou daquela aula, enfim, também conforme condições de trabalho e de infraestrutura que variam,

especialmente em nosso país, não porque somos tecnodiversos, mas porque somos precarizados, sucateados e pouco expostos ao que há de novo entre as possibilidades tecnológicas (novo, não necessariamente melhor).

Mas e o mercado de trabalho?, dirão os mais preocupados com o futuro produtivo de nossa juventude. Por que manter a cartolina, se o jovem precisa mesmo é conhecer o *software* de último tipo? Pois bem: o *tecnodiverso* é o que ele é. A ideia é a de multiplicidade, repertório, e não a de exclusão. Um jovem que sabe dois provavelmente será mais interessante do que o jovem que sabe apenas um, seja esse um o que for. A solução, muitas vezes, está na caneta, e não no supercursor a *laser*. Outro ponto que importa explicar é que esta reflexão não clama pelo retorno ao extremo analógico (nem dispensa investimentos fundamentais em infraestrutura tecnológica), mas, sim, pretende repensar um discurso sobre a adesão ao digital que por vezes desqualifica outras tecnologias, mantendo-nos num entrelugar, já que se, por um lado, usar o “velho” não é mais interessante, por outro, as melhores novidades não chegam para todos, nem sequer para a maioria.

Neste trabalho, buscamos apresentar as noções de *cosmotécnicas*, *sincronização* e *tecnodiversidade*, a fim de experimentá-las no âmbito educacional, em especial para repensar as questões de técnicas, recursos e ferramentas para a aula de língua materna. Nem sempre nos aprofundamos em diferenças importantes, por exemplo, para a filosofia da tecnologia, como a de técnica e a de tecnologia. Que isso sirva de convite às pessoas que se sentirem tocadas por esta discussão.

A noção de *fragmentação*, menos explicada nesta oportunidade, diz respeito ao que Hui (2020) considera positivo para a construção de um futuro menos dependente de países ricos, menos colonizado, menos vulnerável. *Fragmentar* é, no sentido filosófico construído por ele, dividir em porções funcionais e recombináveis, evitando a dependência de blocos monolíticos e sincronizados de infraestrutura ou tecnologia. Se tudo cair ou se apagar, restarão aqueles e aquelas que mantiveram suas chamas acesas, os/as que sabem tirar energia do vento, da luz ou da água. No caso de pane nas plataformas mundialmente sincronizadas, restará a boa e velha conversa. Em outras palavras: o que é fraqueza pode ser força, se olharmos como pessoas que lidam com instabilidades de todo tipo, diariamente. Não é questão de não aprender o novo; é não esquecer o que já tínhamos, especialmente se o novo nunca vem ou nunca se estabelece. Que analogias podemos procurar quanto aos nossos modos de ensinar e aprender, com tais ou quais recursos? Eis o exercício proposto.

Referências

BRASIL. *Diretrizes Curriculares Nacionais dos cursos de Filosofia, História, Geografia, Serviço Social, Comunicação Social, Ciências Sociais, Letras, Biblioteconomia, Arquivologia e Museologia*. Brasília: Ministério da Educação/Conselho Nacional de Educação, 2001.

Disponível em: <http://portal.mec.gov.br/cne/arquivos/pdf/CES0492.pdf> Acesso em: 26 ago. 2024.

BRASIL. Ministério da Educação. *Base Nacional Comum Curricular*. Brasília: MEC, 2018. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518-versaofinal_site.pdf. Acesso em: 3 set. 2024.

CAZDEN, C. *et al.* A Pedagogy of Multiliteracies: Designing Social Futures. *Harvard Educational Review*, 66, p. 60-92, 1996. [CAZDEN, C. *et al.* *Uma pedagogia dos multiletramentos*. Desenhando futuros sociais. Ana Elisa Ribeiro e Hércules Tolêdo Corrêa (org.); Trad. Adriana Alves Pinto *et al.* Belo Horizonte: LED, 2021. Disponível em: <https://www.led.cefetmg.br/umapedagogia-dos-multiletramentos/> Acesso em: 8 abr. 2023.

CASATI, R. *Elogio del papel*. Contra el colonialismo digital. Tradução Jorge Paredes. Barcelona: Ariel/Planeta, 2017.

COSCARELLI, C. V.; RIBEIRO, A. E. *Letramento digital*. Aspectos sociais e possibilidades pedagógicas. Belo Horizonte: Autêntica, 2005. (Ceale)

D'ANDRÉA, C. *Pesquisando plataformas online: conceitos e métodos*. Salvador: EDUFBA, 2020.


DARNTON, R. *A questão dos livros*. Passado, presente e futuro. Tradução Daniel Pellizzari. São Paulo: Companhia das Letras, 2010.

HAWTHORNE, S. *Bibliodiversidad*. Un manifiesto para la edición independiente. Traducción Sáez Juan Carlos y Alejandro Caviedes. Ciudad Autónoma de Buenos Aires: La Marca Editora, 2018.

HUI, Y. *Fragmentar el futuro*. Ensayos sobre tecnodiversidad. Tradução Tadeo Lima. Ciudad Autónoma de Buenos Aires: Caja Negra, 2020. (Colección Futuros Próximos, 33).


RIBEIRO, A. E. Que futuros redesenhamos? Uma releitura do manifesto da Pedagogia dos Multiletramentos e seus ecos no Brasil para o século XXI. *Diálogo das Letras*, Pau dos Ferros, v. 9, p. 1-19, e02011, 2020. Disponível em: <https://periodicos.apps.uern.br/index.php/DDL/article/view/2196/1985>. Acesso em: 13 abr. 2023.

RIBEIRO, A. E. Educação e tecnologias digitais na pandemia: ciclos da precariedade. *Cadernos de Linguística*, v. 2, n. 1, p. 01-16, 2021. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/270>. Acesso em: 9 set. 2024.



RIBEIRO, A. E. Ciclos da precariedade: revisitando experiências de ensino pandêmicas e além. *Cadernos de Linguística*, v. 4, n. 1, e682, 2023. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/682/835>. Acesso em: 15 set. 2024.

SOARES, M. Novas práticas de leitura e escrita: letramento na cibercultura. *Educação & Sociedade*, v. 23, p. 143-160, 2002. Disponível em: <http://www.scielo.br/pdf/es/v23n81/13935>. Acesso em: 10 set. 2024.



Morphology in multilingual data resources: A brief survey

DOI: <http://dx.doi.org/10.21165/el.v54i1.4031>

Magda Ševčíková¹

Abstract

This paper presents selected multilingual language resources that capture various aspects of morphology. Text corpora, dedicated lexical datasets, and typological databases provide insights into inflection, derivation, and, in some cases, the internal structure of words. With continued improvements in coverage, consistency, and compatibility, these resources have the potential to become a solid basis for a realistic understanding of morphological structures across natural languages.

Keywords: morphology; typology; segmentation; corpora; databases.

¹ Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Prague, Czech Republic; sevcikova@ufal.mff.cuni.cz; <https://orcid.org/0000-0003-4780-7912>

Morfologia em recursos de dados multilíngues: uma breve revisão

Resumo

Este artigo apresenta recursos linguísticos multilíngues selecionados que capturam vários aspectos da morfologia. *Corpora* de texto, conjuntos de dados lexicais dedicados e bancos de dados tipológicos fornecem visão sobre inflexão, derivação e, em alguns casos, a estrutura interna das palavras. Com melhorias contínuas em cobertura, consistência e compatibilidade, esses recursos têm o potencial de se tornar uma base sólida para uma compreensão realista das estruturas morfológicas em todas as línguas naturais.


Palavras-chave: morfologia; tipologia; segmentação; *corpora*; bancos de dados.

1. Introduction

In modern linguistics, the comparison of languages aims not only at revealing similarities and differences between the languages being studied but also to uncover universal features of language as a means of human communication. The number of languages whose structures are compared for the purpose of typological research has grown from dozens to hundreds and thousands over the last six decades; cf. the 30-language sample underlying Greenberg's 1963 universals of grammar and the Universal Dependencies project (de Marneffe *et al.*, 2021) covering more than 160 languages in its current version, or the typological database Grambank (Skirgård *et al.*, 2023) providing selected features for nearly 2,500 languages.

Despite these impressive advances, creating a realistic, data-driven picture of more than seven thousand languages spoken around the world (Eberhard *et al.*, 2025) remains far from achievable. This is not only due to the majority of languages being undocumented but also because even for those that are documented, the information is scattered across various datasets. Data and annotations are often fragmented according to traditional linguistic sub-disciplines such as syntax, morphology, phonology, etc. In the present paper, I will specifically focus on phenomena traditionally classified under morphology in theoretical discussions and provide an overview of the linguistic data resources that capture them.

The paper is structured as follows. Section 2 briefly summons up the broad scope of morphology as a linguistic discipline and its reflection in linguistic data. The following sections are then restricted to datasets that cover a larger number of languages and thus have the potential to be used in linguistic typology. Section 3 introduces morphological annotation in multilingual text corpora, with examples of parallel and non-parallel data. Section 4 gives examples of lexical datasets that contain different types of morphological



information, and in Section 5 typological databases are introduced. A few concluding notes in Section 6 close the paper.

2. Morphology in the theoretical discussion and in language data resources

Morphology “deals with words, their internal structure, and how they are formed” (Aronoff; Fudeman, 2005, p. 2). It is traditionally subdivided into inflectional and derivational morphology. While inflectional morphology involves the modification of words to express grammatical features such as tense, number, or case, without changing the word’s lexical meaning, derivational morphology deals with the creation of new words by adding prefixes, suffixes, or other morphemes to existing words. The scope of morphology can also be broadened to other word-formation processes, esp. conversion, compounding, or blending.

Morphology is positioned between phonology, which deals with how sounds function in languages, and syntax, which examines how words are combined to form sentences. Morphology also has significant ties to semantics and pragmatics. In terms of semantics, morphology plays a crucial role in how meaning is encoded within words. In pragmatics, morphology can influence how meaning is modified in different contexts. The way words are formed and used can depend on social factors, registers, or discourse contexts.

Although linguistic discussions often point out that the disciplines thus defined are merely useful constructs that help grasp the compact linguistic reality, these debates have had only limited influence on the creation of linguistic data. Here, the traditional distinctions are projected into fundamental design decisions: The distinction between derivation and inflection manifests in grouping individual word forms under representative forms (lemmas), the separation of morphology and syntax is reflected in the delineation of layers in corpora with multiple types of linguistic annotation.

3. Morphology in multilingual text corpora

Text corpora are large collections of texts that document words in their natural contexts. Both parallel and non-parallel corpora have been used to compare morphology across languages. While parallel corpora such as the JHU Bible Corpus or the InterCorp project contain equivalent texts in two or more languages that are aligned at the sentence or segment level (Section 3.1), the Universal Dependencies project has developed a unified annotation scheme that is applied to non-parallel texts from individual languages (Section 3.2).

3.1 Parallel corpora

The John Hopkins University Bible Corpus contains more than four thousand translations of the Christian Bible in over 1,600 languages; the texts are verse-aligned. McCarthy *et al.* (2020) summarize Natural Language Processing experiments in which the Bible texts (typically for a specific subset of the languages covered by the corpus) were annotated with part-of-speech tags, dependency relations, or features monitored by the typological database Ethnologue (Eberhard *et al.*, 2025). They also demonstrated the use of unannotated data to calculate the type-token ratio, which was interpreted as an indicator of morphological richness of the languages.

Figure 1. Parallel sentences from the English and Portuguese sections of the InterCorp corpus. The forms of the English verb *smile* are annotated with modified PennTreebank part-of-speech tags (explained in the oval), while the corresponding Portuguese forms are assigned EAGLES tags (in the rectangle)

- A worm ca n't smile /VB .	- Uma minhoca não ri /VMI . Testou o sentido de humor delas ?
Keep on smiling /VBG	Quando você rir /VMN
Why do you smile /VBP ?	Por que está rindo /VMG ?
Why are you smiling /VBG ?	Por que está rindo /VMG ?

VB Verb, base form
VBD Verb, past tense
VBG Verb, gerund or present participle
VBN Verb, past participle
VBP Verb, non-3rd person singular present
VBZ Verb, 3rd person singular present

VERBOS			
Pos.	Atributo	Valor	Código
1	Categoria	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P

Source: Own elaboration

Biblical texts are also part of the InterCorp parallel corpus (Čermák; Rosen, 2012).² In addition, the corpus also includes movie subtitles, legal texts, proceedings of the European Parliament, or fiction. In this corpus, Czech serves as the pivot language, meaning that for every text in a language other than Czech, a corresponding Czech version is available. However, texts can also correspond between other language pairs. The parallel corpora are aligned at the sentence level. The current version of InterCorp (InterCorp Release 16;³ Rosen *et al.*, 2023) includes parallel texts in Czech and 61 other languages. In almost half of the languages, texts are provided with morphological annotation, i.e. the individual word forms are assigned the respective dictionary forms (lemmas) and morphological tags. The morphological tags used in the individual language sections of InterCorp come from various sources, which means that they describe the part of speech and selected morphological categories of individual word forms (such as person, tense, mood) by using

2 <https://intercorp.korpus.cz/?lang=en>

3 https://wiki.korpus.cz/doku.php/en:cnk:intercorp:historie#release_16

labels from different tagsets. Figure 1 provides example sentences from the InterCorp Release 16, where the English section is labeled with modified PennTreebank part-of-speech tags (Marcus *et al.*, 1993), while Portuguese texts use EAGLES tags (EAGLES 1996).

This diversity has been eliminated in the InterCorp Release 16ud,⁴ where the part-of-speech category and morphological features are provided in the unified Universal Dependencies scheme; see Figure 2. On top of morphological annotation, syntactic annotation is provided for each sentence, following the Universal Dependencies format. Some more details on the morphological annotation in Universal Dependencies are given in the following subsection.

Figure 2. Parallel sentences from the English and Portuguese sections of the InterCorp corpus with the unified morphological annotation based on Universal Dependencies

When she smiles /VERB//Sing/3 , you feel like -- I don't know .	Quando ela se ri /VERB//Sing/3 , sentimos ... não sei .
He smiles /VERB//Sing/3 .	Ele ri /VERB//Sing/3 .
When he smiles /VERB//Sing/3 , I can see through his head .	Buracos ! Quando se ri /VERB//Sing/3 , até se lhe vê o cérebro .
- She smiles /VERB//Sing/3 .	- Ela sorri , ela ri /VERB//Sing/3 .

Source: Own elaboration

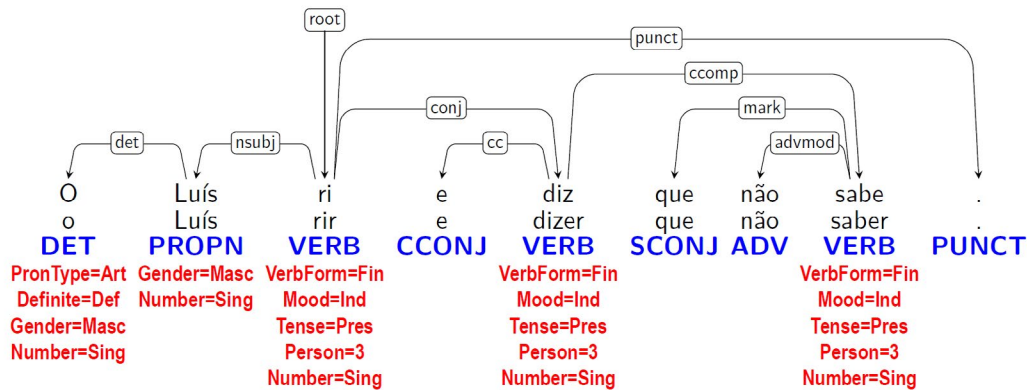
3.2 Universal Dependencies

In its current version 2.15, the Universal Dependencies⁵ collection contains 269 treebanks in 168 languages. The selection of texts and their size vary across the individual treebanks, yet this collection has excellent potential for cross-linguistic comparison due to the fact that the treebanks are annotated with cross-linguistically consistent morphological and syntactic information, which in Figure 3 is displayed below and above the sentence, respectively. The morphological annotation of a word consists of a lemma (displayed in black directly below each word form in the figure), a universal part-of-speech tag (in blue), and a set of universal features (in red). The universal part-of-speech tag is chosen from a fixed list of 17 tags. Universal features further specify the lexical and grammatical properties of the word. If any of the 28 features, each of which has an individual repertoire of values, is not relevant to a given word (e.g. tense with nouns in English), its value is not specified.

4 https://wiki.korpus.cz/doku.php/en:cnk:intercorp:historie#release_16ud

5 <https://universaldependencies.org/>

Figure 3. A Portuguese sentence provided by morphological annotation (below the sentence) and syntactic analysis (above the sentence) according to the Universal Dependencies scheme



Source: Own elaboration

Figure 4. Morphological analyses provided by some of the resource cited

resource	analysis provided				
UniMorph – English (inflections)	collaborationist	collaborationist	N;SG	-	
	collaborationist	collaborationists	N;PL	collaborationist s	
	overwrite	overwriting	V V.PTCP;PRS	overwrite ing	
	overwrite	overwritten	V V.PTCP;PST	overwrite en	
– German	beschreiben	beschreibe	V SBJV;PRS;3;SG	beschreib -e	
	riren	riremos	V IND;PST;PRF;1;PL	riren riremos	
	riso	riso	N;MASC;SG	-	
UniMorph – English (derivations)	collaboration	collaborationist	N:N	-ist	
	central	decentral	ADJ:ADJ	de-	
	centrally	decentrally	ADV:ADV	de-	
	central	subcentral	ADJ:ADJ	sub-	
	subcentral	subcentrally	ADJ:ADV	-ly	
	dezentral	dezentralisieren	ADJ:V	-isieren	
	dezentralisieren	dezentralisieren	V:V	de-	
	riren	riável	V:ADJ	-ável	
– Portug.	riso	risada	N:N	-ada	
CELEX – English	collaborationism	collaboration+ism	(((col)[V].Nx),((labour)[V])[N],(ate)[V xN.])[V],(ion)[N V.])[N],(ism)[N N.])		
	womenfolk	women+folk	((women)[N],(folk)[N])[N]		
	Umgangssprache	Umgang+s+Sprache	(((um)[V].V],(geh)[V])[V])[N],(s)[N N.N],((sprech)[V])[N])[N]		
– German	Grossmachtpolitik	Grossmacht+Politik	(((gross)[A],(Macht)[N])[N],((polit)[R],(ik)[N R.])[N])[N]		
MorphoLex – English	collaborationists	{<co<(labor)>ate>}>ion>>ist>			
UniSegments – English	rewriting	VERB	re+writ+ing (morph: re, span: [0, 1], type: prefix; morph: writ, span: [2, 3, 4, 5], type: root;		
		morph: ing, span: [6, 7, 8], type: suffix)			
	risada	NOUN	ris+ada (span: [0, 1, 2], type: root; span: [3, 4, 5], type: suffix)		
	riável	ADJ	ri+ável (span: [0, 1], type: root; span: [2, 3, 4, 5], type: suffix)		
Corpus Kadiwéu	idinaGataGatinigi				
	i	di	n	aGata	Ga
	1Absolutive	Inverse.voice	Antipassive	hide	Plural
					tinigi
					Applicative

Source: Own elaboration

4. Morphology in lexical datasets

I use “lexical datasets” as an umbrella term for data sources that, in contrast to text corpora just discussed, are repositories of words without contexts. Examples of lexical datasets are given where words are provided with inflectional features (Section 4.1), with information about their derivational history (Section 4.2), and with information about their internal structure (Section 4.3).

4.1 Inflection

The UniMorph⁶ database (Batsuren *et al.*, 2022) currently provides word lists for 169 languages and analyses them in terms of inflectional and/or derivational morphology. The inflectional data consist of quadruples:

- a lemma (cf. the first column in the example entry from the UniMorph inflectional files in Figure 4; e.g. the Portuguese *rir*),
- one of the inflected forms of the lemma (cf. *ríramos* in the second column in the example),
- a part-of-speech category along with morphological features characterizing the particular inflected form (cf. third column). The features are specified by the Universal Morphological Feature Schema (Sylak-Glassman 2016), which is heavily based on the Leipzig Glossing Rules (Comrie *et al.*, 2008). The features are separated from the part-of-speech value if an inflectional affix is identified in the word form, as with *ríramos*, but not with *riso*.
- morphological segmentation of the particular word form (fourth column); in Portuguese and English, a so-called canonical segmentation is provided, where the affix (in this case, inflectional) is separated and the remaining string (root or stem) is replaced by the citation form of the word form analyzed (e.g. *rir|íramos*; cf. Kann *et al.*, 2016). In German, in contrast, the verb form is cut into two substrings without replacing the actual string (*beschreib*) by an infinitive (*beschreiben* ‘to describe’). This column is not used with words without overt inflectional markers.

4.2 Derivation

Derivational analysis is provided for 30 languages in UniMorph 4.0. It differs from the inflectional data produced by this project in several respects. While the inflectional files contain both lemmas and inflected forms, the derivational section is limited to lemmas. As illustrated in Figure 4, the derivational analysis consists of four pieces of information:

⁶ <https://unimorph.github.io/>

- the lemma of the word is listed in the first column that is assumed to enter into derivation to form the word in the second column (cf. the verb *rir* as the motivating word for the derived adjective *riável* in the Portuguese example in Figure 4),
- the part-of-speech category of the input and output words is provided in the third column (cf. V:ADJ related to the above example pair),
- the derivational affix that is assumed to attach to the predecessor to form the derivative is listed in the fourth column.

Some of the analyses provided may be conceived of as inconsistencies; cf. the English examples in Figure 4, where the adverb *decentrally* is formed by a prefix from *centrally*, while *subcentrally* by a suffix from *subcentral*. In the derivational data from German, two immediate predecessors are listed for the German verb *dezentralisieren* 'to decentralize'.

The Universal Derivations⁷ project is another attempt to describe derivational morphology across languages in a unified manner. This resource, which currently covers 21 languages, was created by harmonizing available resources into a unified scheme. According to this scheme, each derivative is linked to its motivating word. By connecting these pairs, a rooted tree structure is formed, with the root node being an unmotivated word. Other words sharing the same root are then organized around it according to increasing morphological complexity; see Figure 5. Nevertheless, the size of the trees and the data for individual languages vary significantly due to the different design decisions made by the original resources, such as limitations to particular part-of-speech categories or to particular types of derivatives (cf. Kyjánek *et al.*, 2020, also for the references to the original datasets).

It should be noted that while the inflectional analysis – often more generally referred to as morphological analysis – in corpora and lexical datasets (cf. Section 3 and 4.1, respectively) commonly involves specification of inflectional meanings, such as number and tense, semantic aspects are not addressed in the derivational data of UniMorph or in Universal Derivations. This is related to the fact that the cross-linguistic discussion of meanings in derivation has only recently been initiated by Bagasheva's (2017) pilot proposal, which was utilized by Körtvélyessy *et al.*, (2020) in their pioneering project on derivational networks.

4.3 Morphological segmentation

Morphological segmentation is understood here as the task of breaking down words into sequences of minimal meaning-bearing units (morphemes), such as roots, prefixes, and suffixes. The delimitation of final inflections and derivational affixes, which has been

⁷ <https://ufal.mff.cuni.cz/universal-derivations>

described as part of the analyses provided by the UniMorph database, can be viewed as a partial segmentation while other resources aim at identifying all morphemes within a word's structure (complete segmentation). Figure 4, which contains examples from the resources that are briefly compared in this section, illustrate some of the differences between the existing approaches.

As described above, UniMorph analyses inflection and derivation of each language separately and according to different principles. In the inflection analysis, an inflectional marker is identified; if not available, no segmentation is carried out. The segmentation complies with the principles of canonical segmentation, where the non-affixal part (root or stem) is replaced by the complete citation form of the word. Derivational analysis attempts to identify solely the morpheme by which a given word may have been formed from a word with a simpler morphological structure. As exemplified above, though, the segmentations are not always consistent within specific languages and across them.

A similar approach to the segmentation of derivatives (and other complex words) was applied by the 1990s project CELEX, which is a forerunner in the field of morphological data (Baayen *et al.*, 1995) and besides English and German (with examples in Figure 4) it covers also Dutch. The resource identifies, in derivatives, the morpheme that is assumed to have been added last (i.e. *-ism* in *collaborationism*) while compounds are decomposed into individual components (*Umgang+s+Sprache*). On top of that, complete segmentation is also provided, i.e. all morphemes in the structure of the word are identified. This second analysis complies with the principles of canonical segmentation (cf. *ate* instead of *at* in *collaborationism*; *geh* instead of *gang* and *sprech* instead of *sprach* in the German compound *Umgangssprache* 'colloquial language'), though some cases deviate from this pattern (cf. *womenfolk*, in which *women* is not replaced by *woman*).

Complete segmentation adhering to the canonical principles is also provided in MorphoLex,⁸ which is available for English and French (Sánchez-Gutiérrez *et al.*, 2018, Mailhot *et al.*, 2020). The MorphoLex example in Figure 4 documents that the final inflection is omitted in the analysis.

In the Universal Segmentations⁹ project (Žabokrtský *et al.*, 2022), a unified annotation scheme for morphological segmentation has been proposed. Words are completely segmented into substrings (morphs), that are not replaced by citation forms. Each morph is classified as prefix, root, or suffix (cf. Figure 4). Segmented data for 32 languages were published in the first version of this resource.

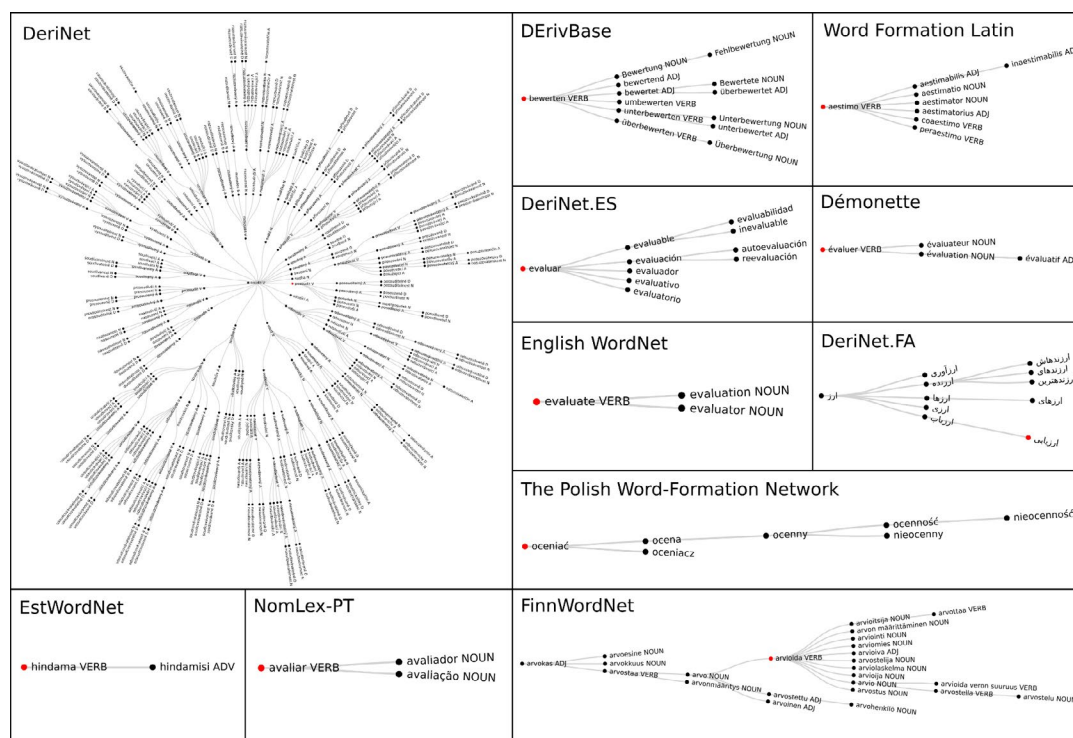
8 <https://github.com/hugomailhot/MorphoLex-en>, <https://github.com/hugomailhot/morpholex-fr>

9 <https://ufal.mff.cuni.cz/universal-segmentations>

While none of the previously mentioned sources providing complete segmentation have identified the meanings of morphemes, there are projects that have undertaken this challenging task for individual languages or, so far, rather modest multilingual collections. It is worth noting that these are projects where annotation at the level of morphemes is part of a multi-level analysis going up to the level of syntactic relations: Corpus Kadiwéu¹⁰ (Galves *et al.*, 2017, Sandalo; Galves 2023), the Universal Dependencies treebank for Beja (Kahane *et al.*, 2021), and Guillaume *et al.*, (2024) proposing annotation guidelines for morpheme-level analysis in Universal Dependencies treebanks and implementing this annotation in three languages. The morphological segmentation of the word form *idinaGataGatinigi* (meaning 'we hid in') from Corpus Kadiwéu is provided in Figure 4 for the sake of comparison.

Moreover, these pioneering projects also pave the way for connecting information about the internal structure of words with analysis at the level of words and sentences, thus escaping the separation that we currently see in linguistic data sources.

Figure 5. Graphs from the Universal Derivations database modelling derivations in Czech (DeriNet), German (DErивBase), Latin (Word Formation Latin), Spanish (DeriNet.ES), French (Démonette), English (English WordNet), Farsí (DeriNet.FA), Polish (Polish Word-Formation Network), Estonia, (EstWordNet), Portuguese (NomLex-PT), and Finnish (FinnWordNet)



Source: Own elaboration

¹⁰ <https://www.tycho.iel.unicamp.br/viewer/C12>

Figure 6. Examples of features and language-specific values from the typological databases
WALS and Grambank

WALS		
Feature 33A: Coding of nominal plurality		
English: plural suffix	Portuguese: plural suffix	Kadiwéu: plural suffix
Feature 112A: Negative morphemes		
English: negative particle	Portuguese: negative particle	Kadiwéu: negative affix
Grambank		
Feature GB030: Is there a gender distinction in independent 3rd person pronouns?		
English: present	Portuguese: present	Hungarian: absent
Feature GB039: Is there nonphonological allomorphy of noun number markers?		
English: present	Portuguese: absent	Hungarian: absent

Source: Own elaboration

5. Typological databases

Typological databases differ from the resources described so far in that they do not contain authentic language material (words and sentences), but they gather structural characteristics extracted from available reference books and other sources for up to thousands of languages. The information in these databases takes the form of metadata on the phonological, morphological, syntactic, and lexical features of individual languages.

Of 192 structural features listed in the World Atlas of Language Structures (WALS; Dryer; Haspelmath, 2013),¹¹ twelve are classified as morphological, another 29 as describing nominal categories, and 17 reporting on verbal categories. All these features are related to inflectional morphology, while derivation is not considered.

The more recent Grambank¹² database (Skirgård *et al.*, 2023) draws on WALS, but differs from it in the inclusion of both inflection and derivation, and also differs in other aspects, including the formulation of features, as illustrated in Figure 6.

As the typological databases do not list all features for every language, the prediction of missing entries has become one of the tasks in the emerging field of computational typology. It has been addressed by using various natural language processing techniques

¹¹ <https://wals.info/>

¹² <https://grambank.clld.org/>

(for summary, cf. Ponti *et al.*, 2019) and became the topic of the 2020 shared task organized by the Association for Computational Linguistics Special Interest Group on Typology (SIGTYP 2020 Shared Task; Bjerva *et al.*, 2020).

Apart from the broad-scope typological resources, dedicated databases have been compiled, such as the Atlas of Pidgin and Creole Language Structures Online (APiCS, Michaelis *et al.*, 2013).¹³

Platforms for sharing typological data have been developed, such as the AUTOTYP (Bickel; Nichols 2002)¹⁴ or URIEL (Litell *et al.*, 2017)¹⁵ meta-databases and, more recently, the CLDF initiative, which additionally calls for data integration through language codes and concepts (Forkel *et al.*, 2018).¹⁶

6. Concluding remarks

The present paper has provided a selective survey of what information about the morphology of natural languages is captured in existing linguistic resources. Since viewed from the perspective of the potential use of these data for language comparison and linguistic typology, I concentrated – apart from a few noteworthy exceptions – on sources covering multiple languages. Existing multilingual resources, into which significant effort has been invested in the recent decades, essentially provide data on all aspects of word structure that are included when defining the scope of morphology as a linguistic sub-discipline. Text corpora document the use of individual inflectional forms in sentence contexts. The available lexical datasets contain information on inflectional morphology, derivational morphology, and the internal structure of words.

A closer look at the selected resources revealed that, regardless of the number of languages covered, maintaining consistent annotation is a significant challenge. When comparing the resources, it became clear that although they aim to address the same phenomenon, the analyses are often based on different principles.

Efforts aimed at improving the internal consistency of the different types of morphological resources and expanding their coverage, as well as ensuring their mutual compatibility and combinability, and, last but not least, linking these resources with those containing other types of linguistic annotations (such as lexical and semantic), could considerably enhance their exploitability for comparative and typological studies.

¹³ <https://apics-online.info/>

¹⁴ <https://www.autotyp.uzh.ch/>

¹⁵ http://www.cs.cmu.edu/~dmortens/projects/7_project/

¹⁶ <https://cldf.clld.org/>

Acknowledgements

I would like to express my sincere gratitude to the organizers of the 70th GEL Seminar, in particular to Filomena Sandalo and Livia Oushiro, for inviting me to participate in the roundtable on Linguistically Annotated Computational Resources for NLP & Language Typology and for providing the opportunity to present my contribution in its published form. The research reported on in the present paper was supported by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. LM2023062 LINDAT/CLARIAH-CZ.

References

- ARONOFF, M.; FUDEMAN, K. *What is Morphology?* Oxford: Blackwell, 2005.
- BAAZEN, R. H.; PIEPENBROCK, R.; GULIKERS, L. *CELEX2*. Philadelphia: Linguistic Data Consortium, 1995.
- BAGASHEVA, A. Comparative Semantic Concepts in Affixation. In: SANTANA-LARIO, J.; VALERA, S. (ed.). *Competing patterns in English affixation*. Bern: Peter Lang, 2017. p. 33-65.
- BATSUREN, K.; GOLDMAN, O.; KHALIFA, S.; HABASH, N.; KIERAŚ, W.; BELLA, G.; LEONARD, B.; NICOLAI, G.; GORMAN, K.; GHANGGO ATE, Y.; RYSKINA, M.; MIELKE, S.; BUDIANSKAYA, E.; EL-KHAISSI, C.; PIMENTEL, T.; GASSER, M.; LANE, W. A.; RAJ, M.; COLER, M.; MONTOYA SAMAME, J. R.; SITICONATZI CAMAITERI, D.; ZUMAETA ROJAS, E.; LÓPEZ FRANCIS, D.; ONCEVAY, A.; LÓPEZ BAUTISTA, J.; SILVA VILLEGAS, G. C.; TORROBA HENNIGEN, L.; EK, A.; GURIEL, D.; DIRIX, P.; BERNARDY, J.-P.; SCHERBAKOV, A.; BAYYR-OOL, A.; ANASTASOPOULOS, A.; ZARIQUIEY, R.; SHEIFER, K.; GANIEVA, S.; CRUZ, H.; KARAHÓĞA, R.; MARKANTONATOU, S.; PAVLIDIS, G.; PLUGARYOV, M.; KLYACHKO, E.; SALEHI, A.; ANGULO, C.; BAXI, J.; KRIZHANOVSKY, A.; KRIZHANOVSKAYA, N.; SALESKY, E.; VANIA, C.; IVANOVA, S.; WHITE, J.; HALL MAUDSLAY, R.; VALVODA, J.; ZMIGROD, R.; CZARNOWSKA, P.; NIKKARINEN, I.; SALCHAK, A.; BHATT, B.; STRAUGHN, C.; LIU, Z.; NORTH WASHINGTON, J.; PINTER, Y.; ATAMAN, D.; WOLINSKI, M.; SUHARDIJANTO, T.; YABLONSKAYA, A.; STOEHR, N.; DOLATIAN, H.; NURIAH, Z.; RATAN, S.; TYERS, F. M.; PONTI, E. M.; AITON, G.; ARORA, A.; HATCHER, R. J.; KUMAR, R.; YOUNG, J.; RODIONOVA, D.; YEMELINA, A.; ANDRUSHKO, T.; MARCHENKO, I.; MASHKOVTSOVA, P.; SEROVA, A.; PRUD'HOMMEAUX, E.; NEPOMNIASHCHAYA, M.; GIUNCHIGLIA, F.; CHODROFF, E.; HULDEN, M.; SILFVERBERG, M.; MCCARTHY, A. D.; YAROWSKY, D.; COTTERELL, R.; TSARFATY, R.; VYLOMOVA, E. UniMorph 4.0: Universal Morphology. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille: ELRA, 2022. p. 840-855.

BICKEL, B.; NICHOLS, J. Autotypologizing databases and their use in fieldwork. In: *Proceedings of the LREC 2002 Workshop on Resources and Tools in Field Linguistics*. Las Palmas: ELRA, 2002.

BJERVA, J.; SALESKY, E.; MIELKE, S. J.; CHAUDHARY, A.; CELANO, G. G. A.; PONTI, E. M.; VYLOMOVA, E.; COTTERELL, R.; AUGENSTEIN, I. SIGTYP 2020 shared task: Prediction of typological features. In: *Proceedings of the Second Workshop on Computational Research in Linguistic Typology*. 2020. p. 1-11.

ČERMÁK, F.; ROSEN, A. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, v. 13, n. 3, 2012. p. 411-427.

COMRIE, B.; HASPELMATH, M.; BICKEL, B. *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Disponível em: <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>. Acesso em: 07 jul. 2025.

DRYER, M. S.; HASPELMATH, M. (ed.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.

EAGLES. *Recommendations for the Morphosyntactic Annotation of Corpora*. Disponível em: <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>. Acesso em: 07 jul. 2025.

EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (ed.). *Ethnologue: Languages of the World*. 28. ed. Dallas, Texas: SIL International, 2025.

FORKEL, R.; LIST, J.-M.; GREENHILL, S. J.; RZYMSKI, C.; BANK, S.; CYSOUW, M.; HAMMARSTRÖM, H.; HASPELMATH, M.; KAIPING, G. A.; GRAY, R. D. Cross-Linguistic Data Formats, advancing data sharing and reuse in comparative linguistics. *Scientific Data*, v. 5, 2018. p. 180205.

GALVES, C.; SANDALO, F.; VERONESI, L. Annotating a polysynthetic language: From Portuguese to Kadiwéu. *Cadernos de Estudos Linguísticos*, v. 59, n. 2, 2017. p. 631-648.

GREENBERG, J. H. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In: GREENBERG, J. H. (ed.). *Universals of Language*. Cambridge, Mass: MIT Press, 1963. p. 73-113.

GUILLAUME, B.; GERDES, K.; GUILLER, K.; KAHANE, S.; LI, Y. Joint Annotation of Morphology and Syntax in Dependency Treebanks. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino: ELRA; ICCL, 2024. p. 9568-9577.

KAHANE, S.; VANHOVE, M.; ZIANE, R.; GUILLAUME, B. A morph-based and a word-based treebank for Beja. In: *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*. Sofia: ACL, 2021. p. 48-60.

KANN, K.; COTTERELL, R.; SCHÜTZE, H. Neural Morphological Analysis: Encoding-Decoding Canonical Segments. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: ACL, 2016. p. 961-967.

KÖRTVÉLYESSY, L.; BAGASHEVA, A.; ŠTEKAUER, P. *Derivational Networks across Languages*. Berlin: De Gruyter, 2020.

KYJÁNEK, L.; ŽABOKRTSKÝ, Z.; ŠEVČÍKOVÁ, M.; VIDRA, J. Universal Derivations 1.0, A Growing Collection of Harmonised Word-Formation Resources. *The Prague Bulletin of Mathematical Linguistics*, v. 115, n. 2, 2020. p. 5-30.

LITTELL, P.; MORTENSEN, D.; LIN, K.; KAIRIS, K.; TURNER, C.; LEVIN, L. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In: *Proceedings of the 15th Conference of EACL*. Valencia: ACL, 2017. p. 8-14.

MAILHOT, H.; WILSON, M. A.; MACOIR, J.; DEACON, S. H.; SÁNCHEZ-GUTIÉRREZ, C. MorphoLex-FR: A derivational morphological database for 38,840 French words. *Behavior Research Methods*, v. 52, n. 4, 2020. p. 1008-1025.

MARCUS, M. P.; SANTORINI, B.; MARCINKIEWICZ, M. A. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, v. 19, n. 2, p. 313-330, 1993.

DE MARNEFFE, M.-C.; MANNING, C. D.; NIVRE, J.; ZEMAN, D. Universal Dependencies. *Computational Linguistics*, v. 47, n. 2, 2021. p. 255-308.

MCCARTHY, A. D.; WICKS, R.; LEWIS, D.; MUELLER, A.; WU, W.; ADAMS, O.; NICOLAI, G.; POST, M.; YAROWSKY, D. The Johns Hopkins University Bible Corpus: 1600+ Tongues for Typological Exploration. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*. Marseille: ELRA, 2020. p. 2884-2892.

MICHAELIS, S. M.; MAURER, P.; HASPELMATH, M.; HUBER, M. (ed.). *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.

PONTI, E. M.; O'HORAN, H.; BERZAK, Y.; VULIĆ, I.; REICHART, R.; POIBEAU, T.; SHUTOVA, E.; KORHONEN, A. Modeling language variation and universals: A survey on typological linguistics for natural language processing. *Computational Linguistics*, v. 45, n. 3, p. 559-601, 2019.

ROSEN, A.; VAVŘÍN, M.; ZASINA, A. J. InterCorp corpus version 16. Prague: Institute of the Czech National Corpus, Faculty of Arts, Charles University, 2023. Disponível em: <http://www.korpus.cz>. Acesso em: 07 jul. 2025.

SÁNCHEZ-GUTIÉRREZ, C. H.; MAILHOT, H.; DEACON, S. H.; WILSON, M. A. MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods*, v. 50, n. 4, p. 1568-1580, 2018.

SANDALO, M. F. S.; GALVES, C. M. C. Anotando sintaticamente uma língua originária do Brasil: o problema de Anchieta. *Cadernos de Estudos Linguísticos*, v. 65, n. 1, p. 1-26, 2023.

SKIRGÅRD, H.; HAYNIE, H. J.; BLASI, D. E.; HAMMARSTRÖM, H.; COLLINS, J.; *et al.* Grambank reveals global patterns in the structural diversity of the world's languages. *Science Advances*, v. 9, 2023. DOI: 10.1126/sciadv.adg6.

SYLAK-GLASSMAN, J. The Composition and Use of the Universal Morphological Feature Schema (UniMorph Schema). Baltimore: John Hopkins University, 2016.

ŽABOKRTSKÝ, Z.; BAFNA, N.; BODNÁR, J.; KYJÁNEK, L.; SVOBODA, E.; ŠEVČÍKOVÁ, M.; VIDRA, J. Towards Universal Segmentations: UniSegments 1.0. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille: ELRA, 2022. p. 1137-1149.

Efeito de impostura: mentira e manipulação no pronunciamento televisionado de Jair Bolsonaro

DOI: <http://dx.doi.org/10.21165/el.v54i1.3883>

Joarle Magalhães Soares¹

Resumo

O propósito deste trabalho é apontar como Jair Bolsonaro usou o aparato de comunicação do Estado para institucionalizar um discurso essencialmente estruturado em estratégias de manipulação. Por meio da análise de um pronunciamento feito em rede nacional de TV e rádio, foi possível constatar que o comportamento discursivo do ex-presidente do Brasil recorreu ao efeito de impostura, ou seja, o sujeito se projetou como um agente da mentira na cena política. O estudo é desenvolvido a partir do suporte teórico-metodológico da Análise do Discurso de linha francesa, mais especificamente a Teoria Semiolinguística e os estudos sobre discurso político de Patrick Charaudeau. Pode-se afirmar que, no pronunciamento oficial, Bolsonaro rompeu com os termos que dão a este contrato de comunicação o *status* de Cena retórica.

Palavras-chave: análise do discurso; Bolsonaro; pronunciamento; manipulação; efeito de impostura.

¹ Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG), Belo Horizonte, Minas Gerais, Brasil; joarle.magalhaes@gmail.com; <https://orcid.org/0000-0002-5316-8072>

The Effect of Imposture: Lies and Manipulation in Jair Bolsonaro's Televised Address

Abstract

The purpose of this study is to demonstrate how Jair Bolsonaro utilized the state's communication apparatus to institutionalize a discourse fundamentally based on manipulation strategies. Through the analysis of a televised and radio broadcast address, it was possible to observe that the discursive behavior of the former president of Brazil relied on the effect of imposture—that is, the projection of the subject as an agent of deceit in the political arena. The study is grounded in the theoretical and methodological framework of French Discourse Analysis, specifically the Semiolinguistic Theory and Patrick Charaudeau's studies on political discourse. It can be asserted that, in the official address, Bolsonaro disrupted the terms that grant this communication contract the status of a rhetorical Scene.

Keywords: discourse analysis; Bolsonaro; address; manipulation; effect of imposture.

Introdução

Os anos do governo de Jair Bolsonaro (2019 a 2022) foram marcados por uma constante tensão discursiva presente em situações cotidianas, que abarcam desde as relações familiares até as notícias dos telejornais. Por meio de postagens em redes sociais, declarações e pronunciamentos, os atores que estavam no poder criaram polêmicas que conduziram a pauta do debate público e os holofotes da imprensa. Em vez de propor discussões que tratassem dos reais problemas brasileiros, o Presidente da República, maior responsável por essa engenharia discursiva, procurava o embate. Bolsonaro coleciona incontáveis manifestações de incivilidade que insuflaram ataques, polarização, preconceito, ódio e polêmicas. O político tinha como característica um modo de governar populista autoritário que desafiava as instituições e evidenciava a falta de compromisso com a verdade. Suas declarações construíam discursos sob um modo peculiar de interpretar a realidade, baseado em crenças e visão de mundo desconectadas de fatos e de conhecimento científico.

Este artigo busca mostrar de que maneira as declarações oficiais do ex-presidente também se pautaram nesse princípio e procuraram institucionalizar a mentira na cena política. Por meio da análise de um pronunciamento transmitido pela TV em 31 de dezembro de 2021, buscamos visualizar uma situação de comunicação na qual Bolsonaro, ocupando a posição de chefe do Executivo, usou o aparato de comunicação do Estado para romper o contrato que comumente é estabelecido nesse dispositivo do discurso político. O estudo é empreendido a partir da perspectiva da Análise do Discurso de filiação francesa, sob a ótica da Teoria Semiolinguística de Patrick Charaudeau e seus trabalhos sobre discurso político e manipulação.

Discurso político e estratégias de manipulação

Em uma democracia, é fundamental que a instância política seja reconhecida pela instância cidadã, uma vez que a adesão dos cidadãos e das instituições de representação da sociedade civil às propostas dos políticos acaba por ressaltar a autoridade que determinado governante conseguiu conquistar. Conforme Charaudeau (2006), na busca pela adesão do público, faz-se necessário que o político seja legitimado em suas ações, sendo reconhecido como um sujeito capacitado a ocupar o seu cargo e a representar os seus eleitores. Em vista disso, os valores defendidos pelos políticos devem estar de acordo com os daqueles que eles representam, para que, desse modo, os feitos realizados sejam justificados pelos princípios que ambos possuem em comum. Tudo isso confere credibilidade ao político frente a seus eleitores, ela relaciona-se à capacidade do sujeito de apresentar suas propostas como verdadeiras.

Para que o político ganhe adesão entre os cidadãos, ele deve apoiar seu discurso em valores transcendentais fundados historicamente e, ao mesmo tempo, se inscrever na volátil regulação das relações entre o povo e seus representantes. Em sua singularidade, o político fala para todos, é voz de todos na sua voz, ao mesmo tempo em que se dirige a todos como se fosse apenas o porta-voz de um terceiro, enunciador de um ideal social. Sendo assim, ele deve saber como inspirar confiança e admiração para que conquiste adesão a essa “[...] imagem ideal do chefe que se encontra no imaginário coletivo dos sentimentos e das emoções” (Charaudeau, 2006, p. 80).

As estratégias discursivas empregadas pelo político para atrair a simpatia do público dependem, então, de vários fatores: sua identidade social; a maneira como ele percebe a opinião pública; o caminho que ele faz para chegar até lá; a posição dos outros atores políticos; e o que ele julgar necessário defender ou atacar (as pessoas, as ideias ou as ações). O político pode perceber a opinião pública como sendo favorável, desfavorável ou incerta e seus discursos deverão ajustar-se a essa percepção e aos diferentes públicos que a constituem. Dessa forma, a atividade de persuasão e de sedução é constitutiva do discurso político, uma vez que, na democracia, é necessário conquistar o poder ou geri-lo a partir da aprovação popular. A fala política tem que ser capaz de “[...] atingir o maior número possível daqueles que compõem esse auditório heterogêneo, cujo denominador comum de compreensão, de análise e de apreciação é, por definição, reduzido” (Charaudeau, 2010).

O discurso político se inscreve, então, dentro da chamada “visada de incitação a fazer”, uma das três intencionalidades do sujeito falante (a expectativa ou o que está em jogo em um ato de linguagem), que Charaudeau (2010) descreve para classificar as finalidades do discurso propagandista. Nesta visada de incitação, o “eu” quer “fazer fazer” alguma coisa a “tu”, mas não está em posição de autoridade. Por isso, o “eu” deve procurar incitar o “tu” a fazer e, para isso, ele deve passar por um “fazer crer”, a fim de persuadir o “tu” de

que será o beneficiário do seu próprio ato, de modo que este aja (ou pense) na direção desejada pelo “eu”. O “tu” se encontra, então, em posição de “dever crer” no que lhe é dito.


No desenvolvimento do enunciado para o convencimento do “tu” pelo “eu”, além das estratégias de persuasão e sedução comuns ao discurso político (e calcadas na relação entre *ethos*, *pathos* e *logos*), são empregadas também estratégias da ordem da “manipulação dos espíritos”. A fronteira entre os dois tipos de estratégia é porosa e não está claramente definida e, por isso, o discurso político se distingue do discurso propagandista e do publicitário, uma vez que nesses dois últimos as instâncias do dispositivo conhecem as condições de produção e a questão da manipulação só pode ser tratada após essas condições serem consideradas. No caso do discurso político, a instância de poder parte do princípio de que “nem tudo pode ser dito” para que se alcance a persuasão, é preciso vestir máscaras.

De acordo com Charaudeau (2010), na busca pelo “fazer crer” (persuasão), a visada de incitação do domínio político (“fazer fazer”) procura suscitar a formação de uma opinião ou alterar a opinião já formada. Para isso, duas características são acrescentadas ao discurso.

Uma é que o manipulador não revela seu projeto de realização e o maquia sob um outro projeto que é apresentado como favorável ao manipulado (quer o benefício seja de ordem individual ou coletiva). A outra é que o manipulador, para melhor impressionar o manipulado, tira partido de certa posição de legitimidade que lhe é dada pela situação e joga com uma credibilidade que ele teria adquirido em outra parte. A consequência disso é que o manipulado, ignorando o verdadeiro teor deste projeto, se deixa persuadir por esta falsa aparência e entra no jogo de persuasão do manipulador sem se dar conta (Charaudeau, 2010).

Sendo assim, as estratégias discursivas empregadas para manipular podem ser resumidas em três tipos: 1) narrativas dramáticas, nas quais são postos em destaque heróis e vítimas com o intuito de produzir às vezes angústia, às vezes exaltação; 2) discursos de promessa (ou até mesmo de profecia), que fazem parte de uma ordem moral ou de um sonho sem que o político dê a impressão de ser autoritário, sob a suspeita de se tornar suspeito, mas suscite a esperança por “dias melhores”; e 3) discursos de provocação do afeto, com o objetivo de tocar a emoção para provocar alegria e simpatia ou temor e medo. Tais estratégias são formuladas por meio de procedimentos formais de simplificação e repetição.

A simplificação se dá com o emprego de fórmulas imagéticas, de *slogans* que têm como efeito “essencializar” os julgamentos, transformá-los em estereótipos e torná-los suporte de identificação ou de apropriação. Segundo Charaudeau (2006), simplificar é tentar reduzir a complexidade do mundo e do universo do pensamento à sua expressão mais



simples. Simplificar pode levar um enunciado a uma falsa verdade, a uma verdade não provada ou mesmo a uma contraverdade, ou seja, a condição de simplicidade acarreta sempre a perda de um pouco de verdade. Já a repetição conferida a esses julgamentos essencializados, por meio de imagens, *slogans*, memes etc., vai da inoculação difusa à repetição excessiva, com a ajuda das mídias sociais e da imprensa, por meio de notícias repetidas à exaustão, até que se tenha o resultado desejado.

A condição de simplificação ocorre ainda a partir de um duplo procedimento: singularização e essencialização. A singularização consiste em evitar a multiplicação de ideias distintas, pois essa multiplicação pode confundir os espíritos não habituados à especulação intelectual. Já a essencialização diz respeito a fazer com que uma ideia seja inteiramente contida, reunida e condensada em uma noção que existiria em si, de maneira natural, como uma essência, independente de outra coisa que não ela mesma. Para tanto, é apresentada sob forma nominalizada. Charaudeau (2006) defende a hipótese psicossociológica segundo a qual quanto mais uma ideia é indeterminada, mais somos atraídos por ela. Esse tipo de fórmula baseada no duplo procedimento de singularização e de essencialização é destinada a produzir um efeito de evidência que acaba causando impacto no público.

Esses procedimentos de persuasão e manipulação concorrem para a produção de um efeito de *slogan* como o encontrado no discurso publicitário, mas com a diferença de que um *slogan* não engana ninguém do ponto de vista de sua força de verdade, já que reconhecemos nele apenas uma força de sonho e de sedução. No discurso político, em contrapartida, uma frase com efeito de *slogan*, como “bandido bom é bandido morto” ou “as minorias têm que se curvar”, é capaz de adquirir uma força de verdade para quem quer crer em sua essencialização. Aqueles que a recebem são expostos a “[...] um efeito de adesão passional mascarada por uma ilusão racional, pois o sentido veiculado está impregnado de uma razão emocional que excede largamente o que é dito explicitamente” (Charaudeau, 2006, p. 100).

Mentira e efeito de impostura

Como estratégia discursiva utilizada para manipular, a simplificação, no contexto político de ascensão da extrema direita, tornou-se um recurso eficiente para a disseminação de *fake news* e distorção de fatos. A “autoverdade” bolsonarista foi formatada a partir dela, dando à mentira um papel fundamental na sustentação das falas presidenciais. Sabemos que o discurso político, de um lado, é um discurso de persuasão, mas muitas vezes, do outro lado, conforme afirma Emediato (2016, p. 21), “[...] há pessoas que esperam ser persuadidas, que se colocam como alvo da persuasão e até se submetem a ela, por interesse, medo ou ignorância”. Como todo discurso de persuasão, ele precisa esconder sua intenção persuasiva por um “mentir verdadeiro”, como postula Charaudeau (2006). Nas palavras de Diderot, citadas por Emediato (2016, p. 21), “[...] engolimos de uma vez a

mentira que nos adula e bebemos gota a gota a verdade que nos amarga”. Para o autor, o enciclopedista francês quis chamar atenção para a nossa relação passional com o mundo, em especial o da política, que nos recruta para uma experiência subjetiva. A mentira que nos adula é mais facilmente aceitável por ser mais desejável do que a verdade, que pode nos frustrar quando não estiver condizente com o que nos traz felicidade. Sendo assim, a verdade possuiria um atributo que a mentira não tem: a indiferença.

Além de incitar as pessoas a fazer, dizer, pensar, o discurso manipulatório, principalmente na cena política, é caracterizado por uma maquiagem intencional e um efeito de impostura que não são próprios de todo ato de persuasão. Entre as muitas estratégias a serviço do “fazer crer”, Charaudeau (2022) aponta quatro figuras que atuam em um ato discursivo de negatividade em relação à verdade: a mentira, a denegação, a má-fé e a impostura. Esta última, no entanto, possui algo específico que a distingue das outras categorias de negação discursiva.

Para explicar essa diferença, o autor recorre ao esquema do ato de linguagem (Charaudeau, 2016b) para lembrar que instância de produção (instância política ou de poder) é constituída pelo sujeito comunicante e pelo enunciador, este investido como um “personagem” que supostamente encena e representa o locutor-pessoa. Sendo assim, em um ato de negação discursiva, o locutor, ao mentir, o faz de modo que o que é dito pelo enunciador mascare, deliberadamente, o que ele pensa como locutor. Já na denegação, o que é dito pelo enunciador mascara o que o sujeito falante pensa, sem que ele tenha consciência disso. Enquanto na má-fé, o que o locutor pensa é colocado entre parênteses. Na impostura, contudo, a relação é mais complexa, pois se trata de um jogo de simulacro entre locutor e enunciador: o eu-locutor-pessoa encena um eu-locutor-personagem que diz o falso dado como verdadeiro; ao fazer isso, o eu-enunciador constrói, positivamente, uma falsa imagem – até mesmo uma identidade – do eu-locutor, mas essa imagem ou identidade é dada como verdadeira.

Em outras palavras, a impostura especula sobre os lugares da instância do sujeito falante, que faz com que a relação entre sujeito locutor e sujeito enunciador não seja em relação aos fatos, mas às identidades. Há, portanto, na impostura, um jogo de usurpação de lugar que não existe na mentira, nem na denegação, nem na má-fé. Se alguém que pratica a denegação ou má-fé costuma ser chamado de impostor, pode ser porque não há palavras para designá-lo e, portanto, um impostor toma o seu lugar (Charaudeau, 2022, p. 77).

Dessa forma, o autor conclui que a impostura resulta de um processo de substituição de instâncias de fala, um jogo de ser e parecer entre essas figuras de sujeitos, uma mistificação que engana os outros. Charaudeau (2022, p. 77) destaca que “[...] o impostor finge ser o que não é, e o receptor (ou seja, a vítima) toma o que vê pelo que é. A impostura supõe, portanto, a duplicidade, na qual o ser, o verdadeiro, é ocultado e o parecer é apresentado como verdadeiro”.

Por não revelar sua intenção, o manipulador a disfarça com um discurso distinto daquele de seu pensamento, enquanto dá indícios, até promessas, de sinceridade. Esse discurso de aparências se apresenta como favorável ou desfavorável ao destinatário, de modo a incitá-lo a agir no sentido desejado pelo manipulador. O efeito de impostura ocorre justamente quando o receptor ignora a intenção do manipulador e se deixa enganar pelo simulacro de sua enunciação. Para atingir seus objetivos, o manipulador recorre a diversas estratégias discursivas, como as que já citamos, e também às diferentes figuras de negação, excluindo-se a denegação, dado que nela o sujeito não tem conhecimento sobre o que está negando.

Dentre os procedimentos discursivos mais utilizados, o manipulador também pode recorrer ao exagero e à generalização, pois são táticas que, se apresentadas com um ar de confiança, reforçam a adesão do público ao lugar para onde o manipulador quer que ele vá. Charaudeau (2022) explica que a generalização permite que as pessoas acreditem em argumentos como “todo mundo sabe que”, que funciona como um artifício para assegurar que o alvo da manipulação adira a uma suposta opinião comum. Além disso, o “amálgama” é outro procedimento que consiste em fazer aproximações entre fatos, causas, consequências, de conectá-los e de dar a ilusão de uma explicação global. “Simples coincidências ou correlações são transformadas em relações de causa e efeito” (Charaudeau, 2022, p. 92). O recurso do amálgama provoca uma amplificação na mente das pessoas que mantêm vivos o medo, o pânico moral, o ódio, o ressentimento. Esses dois procedimentos podem ser considerados como tipos de simplificação (singularização e essencialização, respectivamente), conforme já vimos.

O efeito de impostura abre caminho para a entrada da mentira na política e, conseqüentemente, para a construção de um discurso estruturado em estratégias de manipulação. Charaudeau (2022) considera que o ator político não é *a priori* inimigo da verdade, pelo contrário, já que para exercer sua arte de mentir, é necessário que ele a conheça.

[...] um político, se lhe ocorre dizer a verdade, isto é, dizer o que pensa ou o que vai fazer, encontra múltiplas razões para esconder seus pensamentos ou suas intenções, para o bem do povo, ou mesmo para permanecer no poder. A mentira política pode ser considerada uma garantia de liberdade (Charaudeau, 2022, p. 105).

Análise do pronunciamento de Bolsonaro

O pronunciamento foi adotado pela Secretaria Especial de Comunicação Social como parte de uma estratégia para conferir legitimidade à identidade social de Jair e demonstrar poder por meio de sua identidade discursiva e imagem oficial de chefe do Executivo. Implantada na ditadura militar pelo ex-presidente João Baptista Figueiredo e regulamentada pelo Decreto nº 84.181/1979, a rede nacional de rádio e televisão é

obrigatória a todos os veículos que funcionam a partir de uma concessão pública. A exibição ocorre sempre em horário nobre, interrompendo a programação dos canais. Por possuir essas características, o pronunciamento é parte de uma dimensão institucional do discurso político que confere ao sujeito de fala as condições adequadas para que ele assuma, como enunciador, a identidade discursiva atrelada ao posto de Presidente da República. É um ato de enunciação capaz de criar novos fatos institucionais que poderão interferir nas relações do sistema democrático. Como discurso político, o pronunciamento carrega consigo um sentido de legitimidade que ajuda a conferir autoridade e credibilidade à figura presidencial. É um gênero do discurso que se distingue pela força de veracidade e amplitude institucional que possui.

Levando em conta esses fatores, podemos considerar que o pronunciamento pode ser classificado como uma espécie de “Cena retórica”, conforme o conceito apresentado por Maingueneau (2020). Ele afirma que essa situação de comunicação consiste em um quadro enunciativo no qual os locutores não são simples locutores, mas oradores que pretendem se destacar das situações ordinárias de enunciação. A Cena retórica é monologal e solene, um dispositivo do discurso político para o sujeito “ganhar altura”, ir além da interação imediata com o objetivo de afirmar seus valores e assumir compromissos com o povo. Por isso, no contexto deste estudo, julgamos que essa situação discursiva envolve uma enunciação na qual o presidente espetaculariza a própria fala e, sendo assim, precisa ser feita de modo exemplar, estética e eticamente. “No plano ideal, nessa fala que enuncia o que se tem de enunciar da maneira como se deve, o Belo e o Bem se fundem” (Maingueneau, 2020, p. 111).

Tendo essa concepção como referência, consideramos que, no contexto do discurso político no Brasil, no qual o pronunciamento em rede nacional representa um contrato de comunicação firmado entre as instâncias do dispositivo para ser expressão dos valores do sistema democrático liberal, formatado por meio de elementos visuais e materiais que definem a imagem oficial da Presidência, as regras estabelecidas neste ato de enunciação envolvem, de maneira geral: anunciar projetos, programas e soluções; responder a anseios e angústias da população; informar com dados reais e verificados; celebrar datas; prestar contas das ações realizadas; apresentar-se dentro da liturgia do cargo; respeitar o sistema democrático e as normas constitucionais; e dirigir-se à toda população, sem excluir grupos. Em suma, é preciso que o locutor aja visando a conquistar a confiança da audiência, uma confiança que teoricamente se adquire no discurso quando se busca dizer o verdadeiro.

A fala à nação de 31 de dezembro de 2021 foi a última feita pela rede de rádio e televisão. Apesar de o pronunciamento ter sido um recurso bastante utilizado pelo ex-presidente²,

2 Segundo levantamento do portal Poder 360, Bolsonaro foi o presidente que mais fez pronunciamentos em dois anos de mandato desde a redemocratização, em 1985. Somente em 2019 e 2020, o político impôs um espaço de mais de 42 minutos na rede nacional de TV e rádio (12 pronunciamentos), ficando à frente

o abandono se deu porque o dispositivo também se tornou um canal por meio do qual a população passou a demonstrar insatisfação com a atuação de Bolsonaro perante a crise. O discurso de fim de ano serviu como um compilado em defesa das ações do governo. As declarações do presidente repetiram temas e frases prontas que já tinham sido usadas por ele em outras ocasiões. Gravado em 27 de dezembro, o pronunciamento colocou em segundo plano as fortes chuvas que atingiram cidades da Bahia e de Minas Gerais. Na data da transmissão, Bolsonaro estava de férias em Santa Catarina, visitando o parque temático Beto Carrero World.

Naquela semana, a indiferença do presidente em lidar com a tragédia climática provocou indignação nas redes sociais. Jair estava com a família na cidade de São Francisco do Sul, onde foi filmado e fotografado fazendo passeios de moto aquática, visitando uma pizzeria e dizendo a apoiadores que “esperava não ter que retornar antes” do término das férias³. Na data da declaração, a Bahia registrava 24 mortes provocadas por enchentes, que também destruíram estradas, inutilizaram estoques de medicamentos e vacinas, afetaram cerca de 629 mil pessoas, deixando mais de 90 mil desabrigadas ou desalojadas⁴.

Outro assunto que estava em voga era a vacinação de crianças contra a Covid-19. No pronunciamento, Bolsonaro se posicionou contra a adoção de passaporte vacinal (medida que obrigaria as pessoas a comprovarem que se imunizaram para acessar determinados locais fechados) e defendeu a prescrição médica para a imunização de crianças. Para especialistas, exigir o comprovante de vacinação e facilitar a imunização infantil eram medidas necessárias para o enfrentamento da pandemia, como estava ocorrendo em vários países do mundo. Além disso, uma pesquisa do Instituto Datafolha⁵ mostrou que, para os brasileiros, Bolsonaro era o pior presidente da história. Foi citado por 48% dos entrevistados, o que refletia a queda de sua popularidade em razão da crise econômica e da má gestão da pandemia.

Essas temáticas eram algumas das que estavam em debate e formavam grande parte do contexto no qual se deu o último pronunciamento, reproduzido a seguir.

de FHC (10), Dilma Rousseff (9), Michel Temer (6), Lula (5) e Fernando Collor (3). Disponível em: <https://www.poder360.com.br/governo/bolsonaro-e-o-presidente-que-mais-fez-pronunciamentos-em-2-anos-de-governo>. Acesso em: 25 mar. 2024.

3 Disponível em: <https://www.youtube.com/watch?v=rbxxhWgOPmU>. Acesso em: 14 maio 2024.

4 Disponível em: <https://www.cnnbrasil.com.br/nacional/governo-da-bahia-decreta-situacao-de-emergencia-em-mais-47-cidadespor-enchentes>. Acesso em: 14 maio 2024.

5 Disponível em: <https://www1.folha.uol.com.br/poder/2021/12/datafolha-lula-e-melhor-presidente-da-historia-para-51-e-bolsonaro-o-pior-para-48.shtml>. Acesso em: 14 maio 2024.

Boa noite! Hoje nos preparamos para o início de um novo ano: o bicentenário de nossa Independência. Quis Deus que eu ocupasse a Presidência em 2019 e assumi um Brasil com sérios problemas morais, éticos e econômicos. Formamos um ministério com pessoas capazes para enfrentar a todos os desafios. Ao longo do tempo alguns nos deixaram por livre e espontânea vontade, outros foram substituídos por não se adequarem aos propósitos da maioria que me elegeu. Em 2019 aprovamos a Lei da Liberdade Econômica, simplificamos as normas regulamentadoras, começamos novas obras e concluímos muitas outras inacabadas. Fizemos ressurgir o modal ferroviário, levamos tranquilidade ao campo, flexibilizamos a posse e o porte de arma de fogo para o cidadão e passamos a investir no Brasil, e não mais no exterior com obras bilionárias financiadas pelo BNDES. Completamos 3 anos de governo sem corrupção. Já concluímos, com menor custo, centenas de obras paradas há vários anos. A transposição do Rio São Francisco, finalmente, já é uma realidade e estamos levando mais água para o Nordeste. Somente nos estados de Pernambuco, Paraíba, Ceará e Rio Grande do Norte foram beneficiados 12 milhões de brasileiros em 390 municípios. Já entregamos mais de 1,2 milhão de moradias do Programa “Casa Verde e Amarela” nas três faixas. Em 2020, lamentavelmente, surgiu a pandemia, onde mortes se fizeram presentes no mundo todo. Nessa batalha, o Governo Federal dispensou recursos bilionários para que estados e municípios se preparassem para enfrentar a pandemia. Com a política de muitos governadores e prefeitos de fechar comércios, decretar *lockdown* e toques de recolher, a quebraadeira econômica só não se tornou uma realidade porque nós criamos o PRONAMPE e o BEM, programas para socorrer as pequenas e médias empresas bem como fomentar acordos entre empregadores e trabalhadores para se evitar demissões, com isso, mais de 11 milhões de empregos foram preservados. Para aqueles que perderam sua renda, criamos o Auxílio Emergencial, onde 68 milhões de pessoas se beneficiaram. O total pago em 2020 equivale a mais de 13 anos de gasto com o antigo Bolsa Família, mostramos nossa identidade ao socorrer os mais humildes, que tinham sido abandonados pelos que mandavam fechar tudo. Encerramos o ano de 2021 com 380 milhões de doses de vacinas distribuídas à população. Todas adquiridas pelo nosso governo. Lembro que em 2020 não existia vacina disponível no mercado e a primeira pessoa vacinada foi no Reino Unido em dezembro. Todos os adultos, que assim desejaram, foram vacinados no Brasil. Fomos um exemplo para o mundo! Não apoiamos o passaporte vacinal, nem qualquer restrição àqueles que não desejam se vacinar. Também, como anunciado pelo Ministro da Saúde, defendemos que as vacinas para as crianças entre 5 e 11 anos sejam aplicadas somente com o consentimento dos pais e prescrição médica. A liberdade tem que ser respeitada! Desde o início da pandemia falei que deveríamos combater o vírus, cuidar dos idosos e dos com comorbidades e preservar a renda e o emprego dos trabalhadores. Estamos concluindo 2021 com um saldo de 3 milhões de novos empregos e saldo positivo de 5 milhões de empresas abertas, interrompendo uma série de meia década com saldos negativos. Adentraremos 2022 com esperança de que tudo volte

à normalidade. Já são mais de 800 bilhões de reais contratados pela iniciativa privada, que vão gerar milhões de novos postos de trabalho somente nas áreas de infraestrutura. Isso é uma prova de que reconquistamos a confiança dos investidores, brasileiros e estrangeiros, o que possibilitará, também, a redução da inflação, consequência da equivocada política do “fica em casa, a economia a gente vê depois”. Já começamos a pagar o Auxílio-Brasil, com valor mínimo de 400 reais, programa melhor e mais abrangente do que o antigo Bolsa-Família, onde a média era de apenas 190 reais. O Auxílio-Brasil vai ajudar 17 milhões de famílias mais necessitadas a superar suas dificuldades econômicas e sociais agravadas pela pandemia. Lembro agora dos nossos irmãos da Bahia e do norte de Minas Gerais que nesse momento estão sofrendo os efeitos de fortes chuvas na região. Desde o primeiro momento, determinei que os Ministros João Roma e Rogério Marinho prestassem total apoio aos moradores desses mais de 70 municípios atingidos. Hoje temos um governo que acredita em Deus, respeita seus militares, defende a família e deve lealdade ao seu povo. Um excelente 2022 a todos! Que Deus nos abençoe!⁶

Em um ato de enunciação, as emoções se estruturam a partir dos efeitos de encenação e dramatização que a situação de comunicação propicia. Neste caso, o sujeito funda seu discurso em enunciados que buscam dramatizar os argumentos de modo a conferir força de veracidade aos sentidos. Percebemos que a encenação está associada a uma estratégia que mobiliza imaginários do discurso populista para manipular espíritos a partir da retomada de uma grave situação de declínio social (*assumi um Brasil com sérios problemas éticos, morais e econômicos*). Nesse sentido, o enunciador se coloca como o salvador, adota frases de efeito para essencializar ideias e formar imagens de si capazes de mobilizar na audiência sentimentos de formação de uma identidade originária que é vista como fundadora dos valores que o líder populista se coloca como porta-voz (*A liberdade tem que ser respeitada!; temos um governo que acredita em Deus, respeita seus militares, defende a família e deve lealdade ao seu povo*). A ameaça a essa identidade passa a ser, então, o que precisa ser combatido. O líder deve ser visto de forma incontestável (*outros foram substituídos por não se adequarem aos propósitos da maioria que me elegeu*) para poder se opor aos adversários.

Além disso, os inimigos são associados a atores das instâncias adversária, midiática e cidadã que se colocaram, à época, a favor das medidas sanitárias de enfrentamento à pandemia. Há também um inimigo onipresente, constituído por Lula e o PT, com o qual Bolsonaro precisa se antagonizar para manter a polarização que legitima seu discurso e sua posição de salvador. A menção a esse inimigo aparece de forma velada quando o enunciador compara o Auxílio Emergencial ao Bolsa Família, programa criado no primeiro mandato do presidente Lula, em 2004, e que se tornou um dos principais legados do governo petista. A citação comparativa entre as políticas de transferência de renda surge

6 Disponível em: <https://www.youtube.com/watch?v=YXcqMl2DTY0>. Acesso em: 14 maio 2024.

duas vezes no pronunciamento. Parece haver uma necessidade de enfatizar essa ação específica para associar o que é atribuído ao PT a algo ultrapassado e que agora foi aperfeiçoado. A argumentação é construída omitindo contextos e informações, dentro da mesma estratégia usada para a abordagem de outras temáticas. Intencionalmente, o sujeito suprime dos enunciados a complexidade envolvida no que ele afirma com o objetivo de formar na audiência os sentidos de manipulação que busca produzir.

A condição de simplificação se materializa com a busca dos imaginários mobilizados pela rede de interdiscurso bolsonarista. Essas representações que formam os pilares do discurso da extrema direita brasileira estão todas presentes no pronunciamento, como se uma das intenções tivesse sido resgatar as ideologias que conduziram Bolsonaro à vitória de modo a retomar o apoio e a popularidade que ele vinha perdendo. De acordo com Solano (2019), o ex-presidente construiu sua candidatura a partir da negação das diferenças e da exaltação de um pensamento único e brutalizado (*propósitos da maioria que me elegeu*) moldado por três aspectos enumerados por Maurício Moura (Retrato..., 2020): 1) destruição do sistema político no imaginário da opinião pública em função da Operação Lava Jato e dos escândalos de corrupção (*assumi um Brasil com sérios problemas éticos, morais e econômicos; Completamos 3 anos de governo sem corrupção*); 2) segurança pública como um tema nacional e não mais local ou estadual (*levamos tranquilidade ao campo, flexibilizamos a posse e o porte de arma de fogo para o cidadão*); e 3) necessidade da população votar em alguém novo, um candidato de fora da política tradicional (*Quis Deus que eu ocupasse a Presidência em 2019; temos um governo que... deve lealdade ao seu povo*). Além disso, o pronunciamento contempla enunciados direcionados para agradar os quatro núcleos que Goldstein (2019) aponta como as bases que alicerçaram a vitória bolsonarista: o econômico liberal, o militar, o ruralista e o evangélico. Percebemos no discurso que há um predomínio de medidas tomadas seguindo a agenda econômico-liberal, afinal esse é o quesito que mais afeta a vida da população.

No entanto, como veremos agora, vários dos feitos que o presidente lista foram abordados dentro da condição de simplificação usada como estratégia de manipulação e como construção da mentira e da impostura. Ao fazer sua declaração à nação, o presidente baseou grande parte do que disse no que ele (e sua equipe de governo) pretendia dar veracidade. Vejamos alguns desses momentos.

Completamos três anos de governo sem corrupção.

Essa afirmação é parte essencial do imaginário de intolerância à corrupção e antipetismo formado, principalmente, pela Operação Lava Jato e apropriado por Bolsonaro. Também corresponde a um sentido que integra o imaginário de honestidade criado pelo político para construir seu *ethos* de virtude. Trata-se de uma declaração baseada na afirmação do próprio sujeito e somente nela, já que o enunciado não encontra respaldo nos fatos para se sustentar como um saber de conhecimento (aquele que independe do sujeito). É fruto

da crença formada pelo bolsonarismo de que o ex-capitão do Exército não é corrupto. Na época em que foi dada a declaração, diversos veículos da imprensa a desmentiram e relembrou os casos de corrupção que Jair e seu governo enfrentavam⁷. Também havia suspeitas de corrupção sem relação direta com o governo, mas ligadas à família Bolsonaro e ao partido do presidente.

Formamos um ministério com pessoas capazes para enfrentar a todos os desafios. Ao longo do tempo alguns nos deixaram por livre e espontânea vontade, outros foram substituídos por não se adequarem aos propósitos da maioria que me elegeu.

Esse trecho omite e distorce os frequentes episódios de troca de ministros ocorridos ao longo do mandato. O principal deles (o mais emblemático) diz respeito ao afastamento de Sérgio Moro, em 2020. O ex-juiz da Lava Jato pediu demissão alegando interferência política de Bolsonaro no comando da Polícia Federal e o acusando de crime de responsabilidade. Naquele momento, a saída de Moro representou o desembarque do lavajatismo do governo, um apoio popular e político que foi crucial no desenrolar das eleições de 2018.

Já concluímos, com menor custo, centenas de obras paradas há vários anos. A transposição do Rio São Francisco, finalmente, já é uma realidade e estamos levando mais água para o Nordeste.

Usar informações vagas é parte de um artifício do discurso político, uma estratégia de imprecisão adotada para que seja difícil surpreender o orador em erro ou recriminá-lo por ter mentido. É o que ocorre nessa sentença, o orador estrutura o discurso em cima de uma informação pouco precisa (*centenas de obras paradas há vários anos*). Contudo, ao desenvolver seu argumento, cita a transposição do Rio São Francisco como exemplo de uma das obras inacabadas. Essa citação que torna o enunciado um pouco mais preciso acaba chamando atenção por expor a estratégia de manipulação. Ao classificar a transposição das águas como uma ação de seu governo, Bolsonaro optou por ocultar uma informação importante desse contexto: quando assumiu a Presidência, mais de 90% do projeto estava concluído, vários trechos foram inaugurados em governos anteriores. Mesmo tendo sido um longo projeto, que atravessou mandatos, a transposição estava atrasada, mas não “parada há vários anos”. A obra é fortemente ligada ao imaginário dos governos petistas, pois foi idealizada e iniciada no primeiro mandato do presidente Lula, em 2007⁸. Citar esse projeto, que visava mitigar os severos efeitos da seca no sertão nordestino, acabou gerando um efeito de apropriação de algo não atrelado ao imaginário político de Bolsonaro. Fazer esse gesto de aproximação com a população do Nordeste

7 Disponível em: <https://g1.globo.com/politica/noticia/2021/12/31/na-tv-bolsonaro-critica-passaporte-vacinal-e-defende-prescricao-medica-para-imunizacao-de-criancas.ghtml>. Acesso em: 16 maio 2024.

8 Disponível em: <https://transposicaoosaofrancisco.com.br/index.php/sobre>. Acesso em: 16 maio 2024.

também era uma maneira de o presidente conquistar mais apoio na região onde foi menos votado.

Encerramos o ano de 2021 com 380 milhões de doses de vacinas distribuídas à população. Todas adquiridas pelo nosso governo. Lembro que em 2020 não existia vacina disponível no mercado e a primeira pessoa vacinada foi no Reino Unido em dezembro. Todos os adultos, que assim desejaram, foram vacinados no Brasil. Fomos um exemplo para o mundo!

Conforme apurou a CPI da pandemia, o Governo Federal foi omissivo na busca por vacinas e demorou sete meses para fechar a compra de imunizantes da farmacêutica Pfizer (Senado, 2021). Em seu discurso, Bolsonaro ignorou essas denúncias e as inúmeras declarações negacionistas e antivacinas que deu. Ao simplificar os fatos para encaixá-los em um simulacro onde fosse possível projetar virtude e competência, o presidente procura reforçar sua imagem por meio do efeito de impostura. Ele precisa construir um imaginário favorável à vacinação que seja condizente com o enunciador projetado no discurso. É preciso vestir uma máscara que camufle o político antivacina e a favor da cloroquina. O sujeito sabe que depende da aprovação popular para sustentar seu poder político. Busca, para isso, criar efeitos de veracidade, como dizer que a vacinação no Brasil foi “um exemplo para o mundo”, de modo que o discurso encontre os sentidos de persuasão e manipulação que exaltem sua imagem de líder e, assim, reforcem sua autoridade. Todavia, sustentar essa figura de virtuoso e competente sem levar em conta o contexto da realidade dos fatos, que naquele momento ainda estava vivo na memória coletiva, é deixar de parecer fidedigno e explicitar uma tentativa de enganar e ludibriar o interlocutor. Nessa situação de comunicação, o presidente parece ter sido “pego na mentira” e, por isso, sua declaração enfrentou panelaços e protestos.

Não apoiamos o passaporte vacinal, nem qualquer restrição àqueles que não desejam se vacinar. Também como anunciado pelo Ministro da Saúde, defendemos que as vacinas para as crianças entre 5 e 11 anos sejam aplicadas somente com o consentimento dos pais e prescrição médica. A liberdade tem que ser respeitada!

Nesse trecho, fica materializado novamente o negacionismo científico do ex-presidente. Essa postura foi considerada inadequada por especialistas, que avaliaram a obrigatoriedade da apresentação do cartão de vacinação como uma medida importante para a redução de contaminação pela Covid em ambientes fechados. Sobre a vacinação de crianças, na data do pronunciamento, a Anvisa já tinha fornecido todas as informações técnicas necessárias para justificar a necessidade de imunizar essa faixa etária⁹. É uma fala que vai de encontro àquela que analisamos anteriormente. Na primeira, Bolsonaro busca agradar a audiência que queria se vacinar, já na segunda defende a “liberdade” para

⁹ Disponível em: <https://www1.folha.uol.com.br/poder/2021/12/bolsonaro-distorce-dados-e-omite-fatos-na-tv-e-e-alvo-de-panelaco-antes-do-reveillon.shtml>. Acesso em: 17 maio 2024.

os que não desejam ser imunizados. Essa contradição é um exemplo de como o sujeito busca projetar imagens distintas para persuadir o maior número de indivíduos possível. Como enunciador, Jair precisa se dirigir para a audiência que não o reconhece como um político capaz de exercer o cargo, porém não pode deixar de lado o grupo formado por seus seguidores, que o vê como espelho de seus valores e crenças. É o discurso político buscando figuras distintas para construir identificação e formar imaginários de crenças que sustentem as estratégias de manipulação. O enunciado projeta o presidente como defensor da liberdade a partir de uma frase com efeito de *slogan* (*A liberdade tem que ser respeitada!*), que ilustra a condição de simplificação, segundo a qual uma ideia precisa ser indeterminada para se tornar atrativa e, com isso, conferir dramaticidade ao ato performativo.

Estamos concluindo 2021 com um saldo de 3 milhões de novos empregos e saldo positivo de 5 milhões de empresas abertas, interrompendo uma série de meia década com saldos negativos.

Em relação à economia, Bolsonaro traz para o pronunciamento dados imprecisos que ainda não tinham sido totalizados pelas instituições competentes. Em 2021, de janeiro a novembro, 2,99 milhões de novas vagas foram abertas, mas o resultado de dezembro não havia sido divulgado ainda. Na publicação do número final pelo Ministério do Trabalho e Previdência, um mês depois, o saldo ficou negativo em 265 mil vagas. Com isso, o ano fechou com 2,7 milhões de empregos formais gerados, um quantitativo inferior ao que foi anunciado pelo presidente¹⁰. Da mesma forma, há imprecisão e falta de contexto sobre a quantidade de empresas abertas. Não houve preocupação em esclarecer se o número incluía os microempreendedores individuais (MEIs). Na totalização desse indicador, em fevereiro de 2022, o total de empresas abertas caiu para 4 milhões, sendo 80% delas MEIs¹¹. Mesmo que imprecisos, a inclusão de resultados da economia dá lastro ao discurso com o imaginário do liberalismo econômico e busca projetar, junto à instância de recepção, a figura do economista Paulo Guedes, visto pelo mercado financeiro como fiador do governo Bolsonaro. Há uma busca em recuperar a confiança dos empresários e investidores para forjar o imaginário de que o Brasil tinha superado a pandemia e estava voltando à normalidade (*Isso é uma prova de que reconquistamos a confiança dos investidores, brasileiros e estrangeiros*).

Para aqueles que perderam sua renda, criamos o Auxílio Emergencial, onde 68 milhões de pessoas se beneficiaram. O total pago em 2020 equivale a mais de 13 anos de gasto com o

10 Disponível em: <https://agenciabrasil.ebc.com.br/economia/noticia/2022-01/caged-brasil-gerou-27-milhoes-de-empregos-formais-em-2021>. Acesso em: 17 maio 2024.

11 Disponível em: <https://agenciabrasil.ebc.com.br/radioagencia-nacional/economia/audio/2022-02/80-das-empresas-abertas-no-pais-em-2021-eram-meis>. Acesso em: 17 maio 2024.

antigo Bolsa Família, mostramos nossa identidade ao socorrer os mais humildes, que tinham sido abandonados pelos que mandavam fechar tudo.

Já começamos a pagar o Auxílio-Brasil, com valor mínimo de 400 reais, programa melhor e mais abrangente do que o antigo Bolsa-Família, onde a média era de apenas 190 reais. O Auxílio-Brasil vai ajudar 17 milhões de famílias mais necessitadas a superar suas dificuldades econômicas e sociais agravadas pela pandemia.

Apesar de serem programas lançados pelo governo Bolsonaro, houve resistência na criação tanto do Auxílio Emergencial, quanto do Auxílio-Brasil por parte do Poder Executivo. Em seu pronunciamento, o presidente se apropriou dessas ações sem considerar os embates ocorridos na aprovação das medidas. No início da crise, a equipe econômica pretendia distribuir R\$ 200 aos trabalhadores informais. Bolsonaro não citou as iniciativas do Congresso para ampliar o valor do Auxílio Emergencial, que só foi alterado para R\$ 600 após o governo ter recebido muitas críticas. A criação desse programa foi uma das medidas que seguraram a aprovação de Bolsonaro na pandemia. Transformar o Bolsa Família em Auxílio-Brasil era parte de um artifício de *marketing* para perpetuar o efeito positivo do Auxílio Emergencial e ainda desvincular o programa de assistência social do imaginário que o associava a governos petistas. Na iminência de iniciar mais um ano eleitoral, Jair precisava vincular sua figura ao imaginário de um presidente benevolente que se importava com os pobres (*mostramos nossa identidade ao socorrer os mais humildes; O Auxílio-Brasil vai ajudar 17 milhões de famílias mais necessitadas*). Além de buscar forjar essa imagem, o político insistia na dualidade da polarização, comparando as ações que tomava como sendo melhores que as realizadas nos governos do PT. Também associou o que aconteceu na pandemia aos adversários, mantendo vivo o conflito que construiu entre ele e seus críticos desde o início da crise sanitária (*a quebradeira econômica; abandonados pelos que mandavam fechar tudo*).

Lembro agora dos nossos irmãos da Bahia e do norte de Minas Gerais que nesse momento estão sofrendo os efeitos de fortes chuvas na região. Desde o primeiro momento, determinei que os ministros João Roma e Rogério Marinho prestassem total apoio aos moradores desses mais de 70 municípios atingidos.

Conforme falamos anteriormente, o presidente foi criticado e cobrado por não ter interrompido as férias para liderar as ações de ajuda às cidades que enfrentavam inundações. Bolsonaro gravou sua fala pouco antes de viajar para Santa Catarina. Com isso, seu discurso ficou fora de contexto, não deu a importância devida à tragédia. O presidente cita que mais de 70 municípios foram atingidos pelas chuvas, mas o número estava desatualizado. No dia 31, 136 cidades baianas estavam em situação de emergência, quase o dobro do informado por ele. Segundo o Instituto Nacional de Meteorologia, o

registro do volume de chuvas na Bahia em dezembro foi o maior dos últimos 60 anos¹². Além de menosprezar a catástrofe, Jair não lamentou as mortes ou a situação das centenas de feridos e milhares de desalojados e desabrigados. Ao determinar que os ministros prestassem apoio à população, terceirizou a responsabilidade, isentando-se de participar pessoalmente da ajuda à região. Bolsonaro considerava a Bahia um território “inimigo”, por ter recebido menos votos que Fernando Haddad, em 2018, e porque o governador do Estado era Rui Costa, político filiado ao PT. Ao agir demonstrando certa indiferença, mesmo que sem intenção, o sujeito projetou um enunciador desinteressado em cumprir as obrigações do cargo presidencial. A questão das chuvas foi colocada em segundo plano. Olhar com mais humanidade para a população baiana era um ato que parecia não integrar a cartilha de conduta bolsonarista, pois, ao fazer isso, correria o risco de produzir um sentido cujo significado poderia ser agir em favor daqueles que classificava como “inimigos”.

Apesar da proposta do pronunciamento consistir na apresentação de um balanço das ações realizadas pelo governo (muitas delas citadas de forma distorcida e omitindo o real contexto), o sujeito procura manter vivo em seu discurso o conflito que mobiliza a polarização. Estrutura sua fala dramatizando-a com os elementos populistas que alimentavam os imaginários formadores dos sentidos que intencionava produzir na audiência. Mesmo não estando em período eleitoral, Bolsonaro precisava potencializar sua mensagem com o discurso característico da época de sua candidatura. O capitão é dependente do conflito para existir a partir de um *ethos* que mistura imagens de sua figura política e do cargo presidencial, criando uma realidade paralela na qual consiga se manter relevante no debate público. Mesmo no pronunciamento, que não segue as normas de uma situação de comunicação informal, é preciso alimentar um *ethos* de autenticidade para criar o espelho por onde seus seguidores se veem refletidos. Os imaginários sociodiscursivos são fundamentais na simplificação das ideias e na essencialização dos julgamentos que irão provocar os efeitos de manipulação do processo discursivo.

Os conflitos são, evidentemente, simplificadores do ponto de vista dos julgamentos, das opiniões, dos argumentos que os sustentam e tendem a produzir oposições binárias e maniqueístas. É que qualquer temperança nesses casos, qualquer exame racional das oposições, qualquer troca controlada de argumentos os eliminariam (Charaudeau, 2016a, p. 10).

A polarização política radicalizada fundada antes do mandato bolsonarista e que perdurou ao longo e depois dele dependia das declarações do presidente, da disseminação de mentiras e do discurso de ódio perpetuado principalmente pelas redes sociais para se manter fortalecida. Ao se colocar como missionário divino (*Quis Deus que eu ocupasse a*

12 Disponível em: <https://g1.globo.com/ba/bahia/noticia/2021/12/30/com-maior-volume-de-chuva-nas-ultimas-decadas-no-mes-de-dezembro-bahia-tem-desafio-para-reconstrucao-de-cidades.ghtml>. Acesso em: 19 maio 2024.

Presidência em 2019; Hoje temos um governo que acredita em Deus), Bolsonaro procurava reforçar essa oposição binária e maniqueísta (ligada muito mais à emoção do que à razão) para se blindar das críticas. Como “Messias”, seus atos ou omissões eram parte de uma liturgia governamental que não deveria ser questionada. Por grande parte da instância cidadã, o discurso do ex-presidente era interpretado sob a perspectiva desse imaginário populista, materializado por sentidos vindos de um saber de revelação, sua palavra é a única verdadeira. Além de seduzir seguidores, o discurso buscava produzir significações que colocassem os opositores de Jair como defensores de ideologias, programas políticos ou ações que, na lógica do populismo bolsonarista, estariam contra o que Deus (ou o próprio Bolsonaro) determinou.

Considerações finais

Apesar de buscar tornar efetivo um contrato formatado para o dispositivo político da democracia constitucional, o ex-presidente imprimiu a seu discurso sentidos que fizeram suas falas transgredirem a ordem estabelecida. Essa transgressão característica do discurso populista ocorreu não da forma estética, como apontam Barros e Lago (2022), mas de maneira discursiva. Se, por um lado, Bolsonaro não se expressou de forma vulgar e agressiva; por outro, ele recorreu aos imaginários formadores de seu jeito autêntico. Seu discurso, além de ter sido majoritariamente moldado pela rede de interdiscurso bolsonarista, recorreu a estratégias de persuasão e manipulação para gerar efeitos de sentido que legitimassem fatos e narrativas não compatíveis com o real verdadeiro. Como Cena retórica vinculada a uma situação extraordinária de enunciação, o pronunciamento funciona a partir de normas que impedem que a transgressão verbal de Bolsonaro seja usada no jogo de manipulação populista e, por isso, o sujeito precisa encontrar estratégias discursivas que não sejam esteticamente transgressoras para poder se impor como o líder dominante. Emergiu dos pronunciamentos as representações e características discursivas usadas pela extrema direita.

O discurso no pronunciamento agiu para estabelecer um simulacro no qual a conduta de Bolsonaro e o seu governo foram colocadas como exemplares no enfrentamento à pandemia. O sujeito ainda parece ter agido com impostura ao se solidarizar com as vítimas dos temporais que causaram destruição na Bahia, enquanto passava férias em Santa Catarina, onde afirmou não querer interrompê-las devido à catástrofe.

Ao assumir no pronunciamento a postura de chefe do Executivo, Bolsonaro, como locutor-pessoa, projetou um enunciador que alterna entre duas identidades, a de político e a de presidente, para botar em prática estratégias de manipulação sustentadoras de sua posição como instância detentora de poder. A fala em rede nacional é importante porque representa um contrato no qual não é necessário reagir empiricamente ao interlocutor e, sendo assim, o sujeito pode construir a imagem de si que melhor lhe convier a depender do contexto situacional. O político usou o pronunciamento como um dispositivo para a

legitimação de um simulacro, em que se dirigiu ao público a quem ele deve satisfação e também àquela audiência com a qual ele precisou impor seu *ethos*. Dessa forma, estabeleceu contratos que institucionalizaram o discurso populista autoritário filiado à extrema direita.

A transformação ou desvio dos fatos da realidade, como o negacionismo e as *fake news*, é chamado por Charaudeau (2022) de contraverdade. O autor afirma que, na internet, esse tipo de enunciado é aceito de forma mais natural, porque está inserido no ambiente onde se originou. No entanto, quando a contraverdade é levada para outros ambientes, como o da televisão (regido por um contrato totalmente diferente), pode ser que perca essa aceitação naturalizada comum às mídias sociais. Quando fundamentou seu discurso acionando *ethé* distintos (e até contraditórios) e misturando saberes de conhecimento a saberes de crença para formar a contraverdade, Bolsonaro agiu com a intenção de firmar discursivamente um outro contrato. Parece que houve uma tentativa de levar o enunciador das redes sociais para esse lugar solene onde o comportamento transgressor não é naturalmente aceito. O enunciador do pronunciamento adulava a postura de polemizador do universo digital, nivelando a instância cidadã brasileira pelos parâmetros que usava para se dirigir a seus eleitores. Apesar de mascarado pela projeção da imagem presidencial, esse comportamento discursivo parece ter gerado uma figura dissonante e incongruente que não conseguiu, de fato, falar à nação. O reconhecimento desse sujeito como autoridade digna de crédito acabou ficando comprometido.

Referências

BRASIL. *Decreto nº 84.181, de 12 de novembro de 1979*. Altera a redação do artigo 87, do Regulamento dos Serviços de Radiodifusão, aprovado pelo Decreto nº 52.795, de 31 de outubro de 1963. Brasília, DF: Presidência da República, 1979. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto/antigos/d84181.htm. Acesso em: 25 mar. 2024.

BARROS, T. Z. de; LAGO, M. *Do que falamos quando falamos de populismo*. São Paulo: Cia. das Letras, 2022.

CHARAUDEAU, P. *A manipulação da verdade: do triunfo da negação às sombras da pós-verdade*. Tradução Dóris de Arruda C. da Cunha. São Paulo: Contexto, 2022.

CHARAUDEAU, P. *Discurso político*. São Paulo: Contexto, 2006.

CHARAUDEAU, P. *A conquista da opinião pública: como o discurso manipula as escolhas políticas*. Tradução Angela M. S. Corrêa. São Paulo: Contexto, 2016a.

CHARAUDEAU, P. *Linguagem e discurso: modos de organização*. São Paulo: Editora Contexto, 2016b.

CHARAUDEAU, P. *O discurso propagandista: uma tipologia*. Le site de Patrick Charaudeau, 2010. Disponível em: <https://www.patrick-charaudeau.com/O-discurso-propagandista-uma.html>. Acesso em: 06 mar. 2024.

CHARAUDEAU, P.; MAINGUENEAU, D. *Dicionário de Análise do Discurso*. 3. ed. Tradução Fabiana Komesu. São Paulo: Contexto, 2018.

EMEDIATO, W. Dimensões e faces da mentira no discurso político. In: EMEDIATO, W. (org.). *Análises do discurso político*. Belo Horizonte: Núcleo de Análise do Discurso, Programa de Pós-Graduação em Estudos Linguísticos, Faculdade de Letras da UFMG, 2016, p. 14-49.

GOLDSTEIN, A. *Bolsonaro: la democracia de Brasil en peligro*. Buenos Aires: Marea Editorial, 2019.

MAINGUENEAU, D. *Variações sobre o ethos*. Tradução Marcos Marcionilo. 1. ed. São Paulo: Parábola, 2020.

RETRATO Narrado – Episódio 5: O caminho até o Planalto. Maurício Moura. Reportagem: Carol Pires. [S. l.]: Spotify Studios e revista Piauí, out. 2020. Podcast. Disponível em: <https://open.spotify.com/episode/4hwFT1x4KUcF3cd6lgLwzC?si=c331457b38eb4e43>. Acesso em: 19 dez. 2023.

SENADO FEDERAL. *Comissão Parlamentar de Inquérito da Pandemia: relatório final*. Senado Federal, 2021. Disponível em: <https://legis.senado.leg.br/comissoes/mnas?codcol=2441&tp=4>. Acesso em: 21 fev. 2024.

SOLANO, E. A bolsonarização do Brasil. In: ABRANCHES, S. et al. *Democracia em risco? 22 ensaios sobre o Brasil hoje*. São Paulo: Companhia das Letras, 2019. p. 307-321.

Orações subordinadas adjetivas em diferentes gêneros textuais: uma proposta de descrição e análise

DOI: <http://dx.doi.org/10.21165/el.v54i1.3781>

Ana Carolina Sperança Criscuolo¹
Angélica Rodrigues²

Resumo

Neste artigo, discutimos o uso padrão das orações subordinadas adjetivas em diferentes gêneros textuais, partindo da hipótese de que existe uma correlação entre o funcionamento dessas orações e os propósitos comunicativos de cada gênero. Tradicionalmente, essas orações se caracterizam pela presença do pronome relativo e pela distinção entre orações explicativas e restritivas. Com base em pressupostos funcionalistas, foram estudadas as orações adjetivas em dois gêneros do domínio jornalístico, na modalidade escrita da língua: (1) *artigos de opinião*, de natureza argumentativa, e (2) *notícias*, de natureza narrativa. Os resultados comprovam nossa hipótese inicial, demonstrando haver uma correlação entre o funcionamento das orações adjetivas, o conteúdo expresso, o gênero e o tipo textual em que elas ocorrem, o que permite uma análise mais aprofundada do uso padrão destas orações.

Palavras-chave: oração adjetiva; gêneros textuais; abordagem funcionalista.

1 Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Araraquara, São Paulo, Brasil; ana.speranca@unesp.br; <https://orcid.org/0000-0001-9765-5801>

2 Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Araraquara, São Paulo, Brasil; angelica.rodrigues@unesp.br; <https://orcid.org/0000-0003-1470-4634>

Description and analysis of relative clauses in different textual genres

Abstract

In this article, we discuss the standard use of relative clauses in different textual genres, based on the hypothesis that there is a correlation between the functioning of these clauses and the communicative purposes of each genre. Traditionally, these clauses are characterized by the presence of the relative pronoun and by the distinction between explicative and restrictive clauses. Based on functionalist assumptions, we studied the relative clauses in two genres of the journalistic domain, in the written modality: (1) *opinion articles*, of an argumentative nature, and (2) *news*, of a narrative nature. The results confirm our initial hypothesis, demonstrating that there is a correlation between the functioning of adjective sentences, the expressed content, the genre and the textual type in which they occur, which allows a more in-depth analysis of the standard use of these sentences.

Keywords: relative clause; textual genres; functional approach.

Introdução

O estudo da sintaxe constitui um campo bastante rico para pesquisas acerca do funcionamento da língua, tendo em vista que nos permite observar a maneira como os enunciados são elaborados revelando, muitas vezes, aspectos que não estão presentes explicitamente nos textos e que traduzem, num nível mais abstrato, a maneira como se constroem o pensamento e a interação do homem. Neste artigo, trazemos uma discussão acerca do uso padrão das orações subordinadas adjetivas – ou orações relativas³, com o objetivo principal de mostrar que trazem elementos motivados pelo contexto e pelas intenções comunicativas do enunciador, os quais podem também ser explorados a partir de sua funcionalidade no texto, tal como se observa nos usos não padrão.

A Língua Portuguesa apresenta três estratégias de relativização já consolidadas na literatura: (a) a *relativa padrão*, em que a oração adjetiva possui um referente na oração principal/matriz, recuperado pelo pronome relativo (1); (b) a *relativa cortadora*, em que a preposição regida pelo verbo da oração é apagada (2), e (c) a *relativa copiadora*, em que o referente da oração adjetiva é expresso/repetido por um pronome cópia (3), conforme observamos nos exemplos a seguir (Castilho, 2010, p. 367):

3 Consideraremos os termos “oração (subordinada) adjetiva” e “oração relativa” como sinônimos, tendo em vista que ambos se referem ao mesmo fenômeno sintático. As diferentes nomenclaturas se justificam pelos diferentes critérios utilizados para defini-las.

1. *O livro* **que estou lendo** é de história.
2. Os painéis solares geram *a energia* **que sempre sonhamos**. (Sempre sonhamos com a energia. → Os painéis solares geram *a energia* **com que sempre sonhamos**.)
3. Não há *uma área* em São Paulo **que a polícia não entre nela**.

Segundo Marchesan e Mito (2014, p. 47), há dois grandes tipos de sentenças relativas no português: a relativa com núcleo nominal (que seriam as relativas *padrão*, *cortadora* e *copiadora*) e a relativa livre (4), que não apresenta um nome como antecedente, mas é introduzida por pronomes relativos que trazem embutido o que seria o antecedente expresso na oração principal/matriz:

4. João conhece **quem ganhou na loteria**. (João conhece *a pessoa* **que ganhou na loteria**.)

Embora as estratégias de uso não padrão não constituam o foco deste trabalho, vale destacar que as orações relativas livres, também abordadas por Decat (2011, 2014), possuem uma característica diferente das relativas com núcleo: no período, elas assumem o papel de argumento do verbo da oração principal/matriz, não apenas modificando/caracterizando um referente, mas o constituindo. No exemplo (4), a oração relativa livre assume o papel de complemento – objeto direto – do verbo *conhecer*.

Diversos estudiosos (Mollica, 1977; Tarallo, 1983; Kato, 1993, 1996; Corrêa, 1998; Bastos, 2008, Bispo, 2003, 2009, 2014, entre outros) têm se dedicado às estratégias de relativização não padrão, tendo em vista a preocupação com a descrição dos diversos fenômenos relacionados ao uso da língua. A abordagem da estratégia padrão, contudo, limita-se ao que trazem os manuais de gramática tradicional, a vertentes formais de descrição linguística, que têm o uso padrão como objeto de estudo, e também ao que trazem gramáticas descritivas do português, que contemplam tanto os usos padrão como não padrão.

Em geral, os estudos em torno do uso padrão das orações subordinadas adjetivas têm como enfoque sua estruturação sintática, bem como o aspecto semântico associado ao referente (*explicação* ou *restrição*). Neste artigo, buscamos apresentar outras propriedades desse tipo de oração, reforçando a necessidade de se considerar, na análise sintática, questões relacionadas ao uso. Nesse sentido, nossa hipótese principal é a de que existe uma correlação entre o funcionamento da oração subordinada adjetiva (considerando-se o valor semântico de *explicação* ou *restrição*), o gênero e o tipo textual em que ocorre. Também é relevante o conteúdo expresso na oração adjetiva, uma vez que sua natureza +/- *objetiva* ou +/- *subjativa* é um indício dos propósitos do enunciador ao modificar/avaliar/caracterizar determinado referente (o núcleo nominal ao qual se liga a oração adjetiva).

O *corpus* desta pesquisa é composto por dois gêneros do domínio jornalístico: *artigos de opinião* (do tipo textual “argumentativo”) e *notícias* (do tipo textual “narrativo”), ambos de modalidade escrita, que bem representam o uso padrão da estrutura, analisada a partir de pressupostos funcionalistas (Neves, 2001; Hopper, Traugott, 2003; Castilho, 2010) e cognitivistas (Lakoff; Johnson, 1980; Fauconnier, 1994, 1996; Abreu, 2003, 2012, 2018). Segundo o funcionalismo contemporâneo, o contexto de uso motiva a natureza e o sentido da construção gramatical, ou seja, a maneira como os elementos linguísticos se organizam depende da função que assumem na interação entre os falantes (materializada nos diversos gêneros textuais), o que vale tanto para as formas padrão como não padrão de uso da língua. Num nível mais abstrato, porém integrado a essa visão funcionalista, “a expressão linguística também está intimamente relacionada a motivações cognitivas, no sentido de que codifica a experiência humana” (Bispo, 2014, p. 136). Para Fauconnier (1994, p. 16), a língua é a expressão concreta (não total) de processos cognitivos abstratos que ocorrem na interação discursiva. No caso das orações subordinadas adjetivas, é possível dizer que seu uso (o conteúdo que trazem, o aspecto de *restrição* ou de *explicação*, o gênero em que ocorrem) está atrelado também à maneira como o falante conceptualiza o mundo e fala sobre ele, caracterizando, avaliando, especificando determinado referente frente seu interlocutor.

A fim de apresentar nossas reflexões e os resultados obtidos com nossa pesquisa, seguiremos, neste texto, o seguinte percurso: (i) inicialmente, discutimos como algumas gramáticas tradicionais abordam as orações adjetivas, com o intuito de mostrar lacunas na descrição deste fenômeno sintático; (ii) em seguida, apresentamos a abordagem funcionalista das orações adjetivas que, alinhada a princípios cognitivistas, contribui para um aprofundamento na descrição e na análise destas orações, motivadas pelo uso; (iii) na metodologia, explicamos os critérios utilizados na composição do *corpus* e na análise dos dados; (iv) por fim, trazemos os resultados e uma discussão, que comprovam nossa hipótese inicial de que existe uma correlação entre o funcionamento das orações adjetivas, o conteúdo que expressam, o gênero e o tipo textual em que ocorrem.

A abordagem tradicional das orações adjetivas

As gramáticas tradicionais apresentam a oração adjetiva como uma oração que se subordina a um nome (presente na oração principal do período), o qual é retomado pelo pronome relativo que a introduz (especialmente *que*, *o qual/a qual*, *os quais/as quais*, *cujo(s)/cuja(s)*, *onde*). Segundo Cunha e Cintra (2007, p. 615), “exercem a função de ADJUNTO ADNOMINAL de um substantivo ou pronome antecedente”. De acordo com esses autores, classificam-se em *restritivas* e *explicativas*, “quanto ao sentido”:

As RESTRITIVAS, como o nome indica, restringem, limitam, precisam a significação do substantivo (ou pronome) antecedente. São, por conseguinte, indispensáveis ao sentido da frase; e, como se ligam ao antecedente sem pausa, dele não se separam, na escrita, por vírgula.

[...]

As EXPLICATIVAS acrescentam ao antecedente uma qualidade acessória, isto é, esclarecem melhor sua significação, à semelhança de um aposto. Mas, por isso mesmo, não são indispensáveis ao sentido *essencial* da frase. Na fala, separam-se do antecedente por uma pausa, indicada na escrita por vírgula (Cunha; Cintra, 2007, p. 618, grifo próprio).

Como se pode observar, a classificação da oração adjetiva em *restritiva* ou *explicativa* está associada ao sentido, à natureza da informação nela contida (se uma restrição ou uma explicação em relação ao seu referente). Vale destacar que essa informação é tratada no nível semântico. Nos exemplos (5) e (6), dos autores, temos respectivamente uma oração adjetiva restritiva, que limita o referente de certa forma já restrito pelo adjetivo “raros”, e uma oração adjetiva explicativa, que traz uma “qualidade acessória” ao antecedente “Tio Cosme”:

5. És um dos raros homens / **que têm o mundo nas mãos**. / (A. Abelaira, NC, 121.)
6. Tio Cosme, / **que era advogado**, / confiava-lhe a cópia de papéis de autos. (Machado de Assis, OC, I, 734.)

Bechara (2009, p. 465) discute as orações adjetivas a partir de sua “equivalência semântica e sintática” com adjetivos, considerando o pronome relativo um “elemento transpositor” que faz com que uma oração independente passe a funcionar, num nível inferior (e por isso a oração adjetiva é subordinada), como adjunto adnominal de um substantivo. Esse processo, segundo exemplos do autor (7), ocorre da seguinte maneira:

7. O aluno **estudioso** vence na vida.
O aluno **que estuda** vence na vida.
O aluno estuda. → **que** estuda (oração independente transposta a oração subordinada adjetiva, com equivalência semântica e sintática com o adjetivo “estudioso”)

Em relação aos valores de *explicação* e *restrição*, o autor afirma que a oração adjetiva pode trazer esses dois sentidos. No caso da adjetiva explicativa, tem-se “uma particularidade que não modifica a referência do antecedente e que, por ser mero apêndice, pode ser dispensada sem prejuízo total da mensagem” (Bechara, 2009, p. 467). Na língua falada, é marcada por uma pausa em relação ao antecedente, representada na escrita, em geral, por vírgulas (8):

8. O homem, **que vinha a cavalo**, parou defronte a igreja.

De acordo com o autor, a oração adjetiva em (8) nos permite afirmar que na narração só havia um homem e, nesse caso, a informação “que vinha a cavalo” pode ser dispensada. O contrário ocorre em (9), em que a oração adjetiva restritiva, “proferida sem pausa e não indicada na escrita por sinal de pontuação a separá-la do antecedente”, marca a existência de mais de um homem, mas apenas o “que vinha a cavalo” é que parou defronte a igreja:

9. O homem **que vinha a cavalo** parou defronte a igreja.

É possível perceber que, assim como Cunha e Cintra (2007), Bechara (2009) também define as orações explicativas e restritivas a partir de informações extras acrescentadas sobre um referente ou a partir da sua delimitação dentro de um conjunto, respectivamente, ficando evidente o aspecto semântico desta distinção nas abordagens tradicionais. Esse autor, ainda, aborda as orações relativas que ocupam posição de argumento do verbo da oração principal por um processo que nomeia “substantivação de orações originariamente adjetivas”⁴: tem-se o apagamento do antecedente dos relativos *quem* e *que* e a presença do artigo definido, se o antecedente, pela situação do discurso, é conhecido dos interlocutores ou se lhe quer dar certo ar de generalização: “O homem que cala e ouve não *dissipa o que sabe*, e aprende *o que ignora* (MM)”.

É possível perceber que a abordagem tradicional das orações adjetivas tem seu foco na organização sintática do enunciado, que se dá por meio dos pronomes relativos com função argumental (ao contrário das demais orações subordinadas, que são introduzidas por conjunções sem função argumental) e, especialmente, pela distinção semântica das orações em *explicativas* e *restritivas*. Tal abordagem limita-se ao nível do enunciado, desconsiderando a influência do gênero e dos propósitos pragmáticos do enunciador.

Na próxima seção, discutiremos a abordagem funcionalista das orações adjetivas, com o intuito de verificar em que medida contribui para sanar limitações que encontramos na abordagem tradicional destas orações.

Por uma abordagem funcionalista das orações adjetivas

A oração adjetiva, diferentemente das demais orações subordinadas, é introduzida por pronomes relativos. De acordo com Neves (2000, p. 365), é uma oração de função adnominal, que exprime uma propriedade de uma entidade (o referente), à qual se atribui o predicado expresso na oração principal/matriz (10):

4 Sob a perspectiva linguística, são as chamadas “orações relativas livres” (cf. Marchesan; Mioto, 2014).

10. Mas a mulher **que Aristófanes defende** não tem direito à paixão. (ACM)

Propriedade: “ser defendida por Aristófanes”

Entidade: “mulher”

Predicado: “não tem direito à paixão”

As noções de “entidade”, “propriedade” e “predicação” associam os elementos do enunciado à maneira como o falante percebe o mundo e se manifesta sobre ele, diferentemente da abordagem tradicional.

De acordo com a autora, os pronomes que introduzem as orações adjetivas podem ser fóricos, ou seja, se referem a um antecedente, ou podem não ter um elemento referencial, constituindo eles mesmos a expressão nominal (**quem** *aquela que*). Nestes casos, as orações adjetivas assumem a função de um termo argumental da oração principal.

A oração adjetiva restritiva introduz uma informação que serve para identificar um subconjunto dentro de um conjunto. A oração adjetiva explicativa, por sua vez, introduz “informação suplementar”, não servindo para identificar nenhum subconjunto dentro de um conjunto. A presença de vírgulas assinala uma oração adjetiva explicativa, que introduz uma “informação adicional”. Não se trata, dessa forma, de um “mero apêndice” (Bechara, 2009, p. 467), tal como muitas vezes é apresentado nos manuais de gramática tradicional.

Castilho (2010) também aborda as três estratégias de relativização do português: a *adjetiva padrão*, a *adjetiva copiadora* e a *adjetiva cortadora* (cf. Tarallo, 1983). De acordo com o autor, “as adjetivas ou relativas são sentenças encaixadas num sintagma nominal, em que atuam como Complementadores”, podendo-se encaixar em qualquer expressão núcleo desse sintagma, inclusive a categoria vazia (que constituiria, pois, as construções com relativas livres). O pronome relativo desempenha, simultaneamente, dois papéis: “enquanto pronome, recebe funções argumentais ou de adjunção do verbo da sentença que ele encabeça; enquanto conjunção, liga a adjetiva ao núcleo do sintagma da matriz” (Castilho, 2010, p. 366).

As *adjetivas restritivas* (ou determinativas) “especificam o sentido do sintagma nominal em que estão encaixadas, agregando alguma informação relevante”. O autor retoma Mira Mateus *et al.* (1989, 2003, 2005, p. 655-671), para quem “as restritivas contribuem para a construção do valor referencial da expressão nominal”. As *adjetivas explicativas* (ou apositivas), por sua vez, “operam como um aposto do sintagma nominal, explicitando um *comentário do locutor* [grifo próprio] acerca da entidade denotada por um sintagma nominal, o antecedente da relativa”. Castilho acrescenta, ainda, o que classifica como *adjetivas finais* (11) e *adjetivas causais* (12), uma vez que trazem consigo os valores adverbiais de *finalidade* e *causa*:

11. Mandou retirarem seus sapatos enlameados, **que não sujassem sua sala**.

12. O cão, **que é amigo fiel**, vigiou a casa durante toda a noite.

Embora essas noções possam estar presentes em grande parte das construções com orações adjetivas, não acreditamos ser necessário criar suas respectivas categorias ao lado das explicativas e das restritivas. De qualquer modo, é importante destacar que o propósito do autor vai além da estrutura sintática de organização dessas orações.

Nesse sentido, ainda, o autor questiona qual seria a função adicional das orações adjetivas, uma vez que as línguas já dispõem de adjetivos e expressões adjetivais para especificar o sentido do sintagma nominal. Essa é uma questão de grande valia dentro dos pressupostos deste trabalho, uma vez que se leva em consideração o aspecto discursivo destas construções. Considerem-se os enunciados (13), (14) e (15), de Castilho (2010, p. 371):

13. Vi um vestido **vermelho**.

14. Vi um vestido **de tecido vermelho**.

15. Vi um vestido **que era feito de tecido vermelho**.

É possível ver que a oração adjetiva (15) faz a predicação do sintagma nominal de forma mais complexa que o adjetivo (13) e que a locução adjetiva (14), situando seu referente como participante de um estado de coisas. Além disso, “através das adjetivas, o falante aplica a esse sintagma a categoria de tempo, não encontrada no adjetivo ou nos adjetivais” (Castilho, 2010, p. 371). Como mostraremos neste trabalho, nossos dados confirmam tais pressupostos, indo além no sentido de que muitas informações trazidas pelas orações adjetivas (independentemente de serem explicativas ou restritivas) não poderiam ser expressas por adjetivos ou outras expressões adjetivais.

Abreu (2003, 2012, 2018), partindo de uma perspectiva pragmática (diretamente ligada ao contexto interacional), chama a atenção para o fato de que as orações adjetivas explicativas são, muitas vezes, utilizadas de maneira aparentemente redundante (16):

16. O homem, **que é racional**, ainda é capaz de matar seus semelhantes de forma cruel.

De acordo com o autor, o fato de o homem ser racional “é uma obviedade”, não havendo novidade do ponto de vista informativo. Contudo, é possível perceber uma intenção argumentativa do falante, como se dissesse (16’):

16'. O homem, **apesar de ser racional**, ainda é capaz de matar seus semelhantes de forma cruel.

Nesse caso, "a oração explicativa assume um valor pragmático de *concessão*" (Abreu, 2018, p. 479). No enunciado (17),

17. O homem, **que é racional**, saberá evitar uma catástrofe nuclear.

a mesma oração adjetiva "**que é racional**" assume o valor pragmático de *causa*, semelhante ao que discute Castilho (2010, p. 371), ao propor as *orações adjetivas finais e causais*. Trata-se de sentidos não gramaticais, mas obtidos por inferência pragmática.

Conforme discutido na seção anterior, a abordagem tradicional das orações adjetivas tem seu foco na organização sintática das orações adjetivas e no aspecto semântico de *restrição* ou *explicação*. Essas noções são, essencialmente, tratadas de forma comparativa, em que se tem a delimitação de um elemento dentro de um conjunto (restritivas) ou o acréscimo de informações consideradas "não essenciais" à compreensão do enunciado (explicativas). Tais noções, para Hopper e Traugott (2003, p. 196-199), não refletem apenas o aspecto semântico, mas também denotam o grau de integração sintática (e cognitiva) entre a oração subordinada adjetiva e o seu referente. Logo, a presença ou não de vírgulas/pausa entre as orações não é um recurso suprasegmental, como muitas vezes sugere a abordagem tradicional. Dentro do que propõem esses autores, a articulação de orações se dá num *continuum* que vai da parataxe ou relativa independência (orações coordenadas e justapostas), passando pela hipotaxe ou interdependência (orações adverbiais e relativas apositivas/explicativas) e chegando-se à subordinação (orações substantivas e relativas restritivas). Assim, as *orações explicativas* estabelecem uma ligação "mais frouxa" com o referente da oração principal, à semelhança de um aposto, sendo por isso também chamadas "apositivas"; as *orações restritivas*, por sua vez, assumem o papel de modificadoras/especificadoras de um núcleo nominal, estabelecendo com este um maior grau de integração cognitiva e sintática. Em outras palavras, quanto mais integrado cognitivamente um elemento for de outro, mais o será na forma como é expresso sintaticamente, revelando-se presente nesta relação o princípio da *iconicidade* (Givón, 1984; Croft, 1990), segundo o qual a língua se organiza nos mesmos moldes em que o falante conceptualiza o mundo.

A abordagem funcionalista, alinhada a pressupostos de ordem cognitiva, contribui para o aprofundamento no estudo das orações adjetivas, explicando de que maneira constituem um recurso ao falante na construção dos referentes, na argumentação, na expressão de suas opiniões. Sob essa perspectiva, que embasa a análise e discussões apresentadas, entendem-se as construções com orações adjetivas como um processo de *predicação* (Neves, 2010), contextualizado e materializado nos gêneros textuais. Destaca-se, ainda, que esse processo é mais complexo, uma vez que a oração adjetiva possibilita não apenas

uma caracterização objetiva/descritiva como os adjetivos e expressões adjetivais, mas também uma caracterização subjetiva/argumentativa dos referentes (Castilho, 2010).

Nas próximas seções, apresentamos a Metodologia utilizada no desenvolvimento de nossa pesquisa e os Resultados obtidos, com o intuito de demonstrar que fatores sintáticos, semânticos, pragmáticos e discursivos interagem no funcionamento da língua em diferentes contextos de uso.

Metodologia

Dentro da abordagem em que se insere nosso trabalho, os pressupostos ligados ao conceito de *gênero* e *tipo textual* são fundamentais para a análise do fenômeno estudado, uma vez que pressões características de cada situação comunicativa interferem na maneira como os textos se organizam e, conseqüentemente, como os elementos linguísticos funcionam.

O *corpus* que serviu de base para nossa investigação foi composto por textos de dois gêneros do domínio jornalístico: a) *artigos de opinião* (tipo argumentativo) e b) *notícias* (tipo narrativo). Todos os textos foram extraídos do jornal *O Estado de S. Paulo*, versão *online*. Os *artigos de opinião* foram coletados do caderno *Espaço Aberto*, e as *notícias* foram coletadas dos cadernos *Metrópole* e *Internacional* (janeiro – junho 2018), sempre da primeira edição (dia 01) de cada mês. Desta forma, garantiu-se que os textos selecionados fossem de variados autores. A escolha por esses gêneros deve-se ao objetivo de analisar o uso padrão das orações subordinadas adjetivas, na modalidade escrita da língua.

No total, analisamos 12 artigos de opinião e 24 notícias, tendo em vista que a extensão das notícias é menor. Nos artigos de opinião, a extensão dos textos tem uma média de 947 palavras, variando pouco de um texto para outro, uma vez que o espaço que ocupam na página do jornal é sempre o mesmo. Nas notícias, por sua vez, a média é de 578 palavras, mas a variação de um texto para outro pode ser grande, visto que as notícias ocupam lugares variados nas páginas do jornal (Tabela 1).

Tabela 1. Delimitação do *corpus*

Tipo de texto	Total de palavras	Média de palavras	Caracteres (sem espaço)
<i>Argumentativo</i> (Artigo de Opinião)	11374	947	58530
<i>Narrativo</i> (Notícia)	13889	578	71345

Fonte: Elaboração própria

Após a leitura cuidadosa dos textos, foram coletadas todas as ocorrências de orações subordinadas adjetivas, totalizando 179 orações nos *artigos de opinião* e 190 orações nas *notícias*. As 369 ocorrências estão organizadas na Tabela 2, em número absoluto (N) e sua respectiva porcentagem (%):

Tabela 2. Número de orações adjetivas por tipo de texto

Tipo de texto	Oração restritiva		Oração explicativa		Outros tipos ⁵		Total de orações do corpus	
	N	%	N	%	N	%	N	%
Artigo de opinião (<i>Argumentativo</i>)	116	64,8	44	24,6	19	10,6	179	100
Notícia (<i>Narrativo</i>)	100	52,6	68	35,8	22	11,6	190	100

Fonte: Elaboração própria

Além da análise quantitativa das ocorrências, que demonstra o predomínio das orações restritivas nos dois tipos de textos, procedeu-se a uma análise qualitativa dos dados, pautada nos seguintes critérios:

- o gênero discursivo;
- o tipo textual;
- a natureza do nome ao qual a oração adjetiva está ligada;
- a função sintática do nome ao qual a oração adjetiva está ligada;
- o valor semântico da oração (*restrição* ou *explicação*);
- o valor discursivo da oração no contexto.

A partir desses critérios, foi possível realizar uma descrição mais aprofundada das orações adjetivas, conforme se discutirá na próxima seção.

Resultados e discussão

Na análise dos dados, foram observados tanto aspectos formais quanto funcionais das orações adjetivas. Em relação aos pronomes relativos que introduzem as orações adjetivas, um aspecto formal desta organização sintática, verificou-se que todos eles

⁵ Nesta categoria incluem-se as orações relativas livres, precedidas pelo pronome demonstrativo “o” ou pela expressão “Todos os”.

aparecem nos dados do *corpus*; porém, o pronome “que” predomina em número muito maior que os demais, conforme apresentado nas Tabelas 3 (*Artigo de opinião* – Texto argumentativo) e 4 (*Notícia* – Texto narrativo):

Tabela 3. Pronomes relativos que introduzem as orações adjetivas – *Artigo de Opinião* (Texto argumentativo)

Tipo de pronome relativo	Oração restritiva		Oração explicativa		Outros tipos	
	N	%	N	%	N	%
QUE	103	57,5	36	20,1	-	-
ONDE	5	2,8	2	1,1	-	-
CUJO (e variações)	0	0	4	2,2	-	-
QUAL (e variações, com ou sem preposição)	5	2,8	2	1,1	-	-
QUEM (com ou sem preposição)	3	1,7	0	0	-	-
ORAÇÃO RELATIVA LIVRE (sem antecedente)	-	-	-	-	16	8,9
‘TODOS OS QUE’	-	-	-	-	3	1,7
TOTAL: 179 orações	116	64,8	44	24,5	19	10,6

Fonte: Elaboração própria

Tabela 4. Pronomes relativos que introduzem as orações adjetivas – *Notícia* (Texto narrativo)

Tipo de pronome relativo	Oração Restritiva		Oração Explicativa		Outros tipos	
	N	%	N	%	N	%
QUE	93	49,0	55	28,9	-	-
ONDE	1	0,5	9	4,7	-	-
CUJO (e variações)	1	0,5	1	0,5	-	-
QUAL (e variações, com ou sem preposição)	4	2,1	3	1,5	-	-
QUEM (com ou sem preposição)	1	0,5	0	0	-	-

ORAÇÃO RELATIVA LIVRE (sem antecedente)	-	-	-	-	21	11,0
'TUDO O QUE'	-	-	-	-	1	0,5
TOTAL: 190 orações	100	52,6	68	35,8	22	11,5

Fonte: Elaboração própria

Em relação ao núcleo nominal ao qual se refere a oração adjetiva, verificou-se que todas as funções sintáticas são passíveis do processo de relativização, como se pode observar nos exemplos a seguir (18 a 26):

18. *A coisa **que a Igreja mais necessita agora** é a capacidade de curar feridas e aquecer o coração dos fiéis, aproximar-se.* (ESP, *Espaço Aberto*, 01.01.2018) – SUJEITO
19. *Como milhões de jovens italianos, Elio Vagali enfrenta **opções de carreira que vão de mínimas a inexistentes**.* (ESP, *Internacional*, 01.03.2018) – OBJETO DIRETO
20. *A Corte condenou a legenda e 29 dos membros de sua cúpula a **penas que, somadas, chegam a 351 anos de prisão**.* (ESP, *Internacional*, 01.06.2018) - OBJETO INDIRETO
21. *A aplicação **dessa racionalidade que respeita o Direito como tal** é uma garantia do cidadão.* (ESP, *Espaço Aberto*, 01.05.2018) – COMPLEMENTO NOMINAL
22. *O estudo inédito foi feito com base no Censo Escolar 2016 pelo Interdisciplinaridade e Evidências no Debate Educacional (Iede) e pelo QEDu, **dois Institutos que pesquisam e divulgam dados educacionais**.* (ESP, *Metrópole*, 01.04.2018) – APOSTO
23. *Embora a população tenha aumentado 30% desde o início dos anos 2000, o consumo geral de água permaneceu estável. Muitos dos novos moradores foram assentados **em áreas pobres, onde o consumo de água é menor**, o que acabou baixando o consumo per capita.* (ESP, *Internacional*, 01.02.2018) – ADJUNTO ADVERBIAL (LUGAR)
24. *“Isso é reflexo **de um sistema que não consegue garantir a qualidade da aprendizagem**. A reprovação precisa ser uma exceção e não uma estratégia pedagógica”, diz o diretor do Iede, Ernesto Faria.* (ESP, *Metrópole*, 01.04.2018) – ADJUNTO ADNOMINAL
25. *12 mi de hectares é **a área total que o Brasil prometeu reflorestar no acordo de Paris**, contra o aquecimento global.* (ESP, *Metrópole*, 01.03.2018) – PREDICATIVO
26. *Certamente não é iluminista decisão que decreta prisão temporária de pessoas investigadas – muitas das quais nunca tendo sido antes nem intimadas para depor – com o único e declarado propósito de colher seu depoimento, numa confessada afronta à medida liminar concedida **por ministro do STF que impediu o uso de conduções coercitivas em tais situações**.* (ESP, *Espaço Aberto*, 01.05.2018) – AGENTE DA PASSIVA

É interessante ressaltar que o estatuto sintático do termo ao qual se refere a oração adjetiva raramente é considerado nos manuais de gramática tradicional, muito menos em livros didáticos. Embora haja um predomínio do processo de relativização nas funções de SUJEITO e de OBJETO, definidas pela estrutura argumental do enunciado, termos com funções mais periféricas, como ADJUNTOS e APOSTO, por exemplo, também podem ser especificados por meio desse processo, de acordo com o que o falante julgar importante esclarecer, avaliar/julgar ou informar ao seu interlocutor. No exemplo (18), a oração adjetiva, que tem como referente a palavra “coisa”, é fundamental para designar a entidade sobre a qual se estabelece a predicação, não sendo apenas um processo pelo qual se restringe ou se delimita a significação de um elemento. A oração adjetiva participa da construção do próprio referente, não sendo possível identificá-lo sem ela (Mira Mateus *et al.*, 1989, 2003, 2005 *apud* Castilho, 2010).

Ao observarmos aspectos funcionais dos dados, conforme discutimos nos exemplos a seguir, os resultados demonstram haver uma correlação entre (a) o tipo de oração adjetiva, (b) a informação que ela traz, (c) o gênero em que ela ocorre e a (d) intenção comunicativa do falante, o que confirma nossa hipótese inicial. Nos *artigos de opinião* (textos argumentativos), observamos que as orações adjetivas trazem informações mais subjetivas, ligadas à opinião do enunciador, as quais participam da argumentação que é construída ao longo do texto (ocorrências 27, 28 e 29).

27. Não falo como homem de partido, mas como brasileiro: O Brasil precisa de lideranças **que tenham capacidade de reunificar o país em torno de alguns objetivos comuns**. (ESP, *Espaço Aberto*, 01.04.2018)
28. Faz todo o sentido essa barreira do silêncio porque, situação ou oposição de turno, as chamadas “fontes” do debate nacional são os poderes estabelecidos e essa reforma, uma vez posta para andar, ou vai à questão de fato e muda definitivamente o poder de dono no Brasil, ou continua dando um passo para a frente e dois para trás, como vem acontecendo desde o minuto seguinte à proclamação da República **que nós nunca instituímos de fato**. (ESP, *Espaço Aberto*, 01.06.2018)
29. O tratamento de choque na esbórnia das aposentadorias públicas, **onde o privilégio é lei e o abuso do privilégio é regra**, não é mais uma questão de escolha é um imperativo de sobrevivência. (ESP, *Espaço Aberto*, 01.06.2018)

Nos exemplos acima (27, 28 e 29), fica claro que a informação veiculada pela oração adjetiva está ligada à opinião do enunciador sobre seu referente, em caráter de sugestão ou avaliação. É interessante ainda observar que, nos textos argumentativos, mesmo quando a oração adjetiva traz uma informação aparentemente mais objetiva, esta informação é selecionada estrategicamente na construção do referente. Em (30), a oração adjetiva traz o texto explicativo de uma Lei sendo, portanto, de caráter objetivo. Contudo, essa

informação não é mera caracterização, é também uma explicação para o que vem sendo questionado ao longo do texto:

30. Também vi magistrados defenderem o auxílio-moradia com base na Lei Complementar no. 35, **que no seu artigo 65, II, permite aos magistrados receberem, além dos seus vencimentos, “ajuda de custo, para moradia, nas localidades / em que não houver residência oficial à disposição do Magistrado”**. (ESP, *Espaço Aberto*, 01.03.2018)

Nos textos das *notícias* (textos narrativos), por sua vez, as orações adjetivas trazem, predominantemente, informações mais objetivas, de natureza factual, ligadas aos eventos relatados (ocorrências 31, 32 e 33).

31. Logo depois, uma viatura do Batalhão de Vias Expressas **que trafegava pelas imediações da Cidade de Deus** foi atacada com pedras e paus. (ESP, *Metrópole*, 01.02.2018)
32. Por pouco, a família, **que dormia no imóvel na hora do acidente**, não foi atingida. (ESP, *Metrópole*, 01.05.2018)
33. A partir de fevereiro, os moradores estarão sujeitos a pesadas multas se excederem o novo limite diário de consumo, **que passará dos atuais 87 litros de água por pessoa para menos de 50 litros**. (ESP, *Internacional*, 01.02.2018)

Em (31), (32) e (33), percebemos que a informação apresentada pela oração adjetiva diz respeito a circunstâncias factuais ligadas ao referente, no sentido de melhor contextualizar o evento que está sendo noticiado. E, ainda quando a oração adjetiva traz uma avaliação da situação, muitas vezes essa avaliação é apresentada como um consenso ou tem indicada sua fonte, como se pode perceber na ocorrência (34):

34. A nova onda de protestos parece ter sido deflagrada após um aumento de 40% nos preços dos ovos e das aves, **que um porta-voz do governo atribuiu à restrição de oferta em razão da gripe aviária**. (ESP, *Internacional*, 01.01.2018)

De acordo com os dados do *corpus*, podemos perceber que a informação trazida pela oração adjetiva está, na maior parte das vezes, diretamente relacionada aos propósitos comunicativos dos gêneros em que ocorrem. Embora tanto o *artigo de opinião* quanto a *notícia* sejam gêneros característicos da esfera jornalística, ambos possuem propósitos distintos: o *artigo de opinião*, segundo Rojo (2000, p. 226), é um gênero que busca convencer o outro de determinada ideia, modificando ou formando opiniões. Logo, é um gênero argumentativo por excelência. A *notícia*, por sua vez, tem como objetivo principal o relato de acontecimentos relevantes ao público a que se destina, informando-lhe sobre

fatos da realidade; é essencialmente um texto narrativo, em que informações factuais fazem parte de sua essência.

Retomando Castilho (2010, p. 371) acerca da funcionalidade adicional das orações adjetivas, constata-se de fato que elas “predicam o sintagma nominal de modo mais complexo”, não se limitando à sua mera caracterização, explicação ou restrição dentro de um conjunto. Tampouco são equivalentes a um adjetivo, como preconiza a abordagem tradicional. Pelos dados apresentados, é possível perceber que as informações trazidas pelas orações adjetivas ultrapassam o nível semântico, tendo claramente motivações pragmáticas e discursivas. Na ocorrência (27), por exemplo, vemos claramente a opinião do autor acerca de lideranças na política brasileira. Na maior parte das ocorrências do *corpus*, não teríamos adjetivos ou expressões adjetivais correspondentes que permitiriam outra forma de expressão do conteúdo que trazem as orações adjetivas.

Apartir dos exemplos apresentados, é possível observar que as orações adjetivas assumem funcionalidades que vão além do nível semântico de uma *restrição* ou uma *explicação*, como apresentado na abordagem tradicional. Nos *artigos de opinião*, funcionam como estratégia de argumentação com base em uma caracterização subjetiva – e avaliativa – do falante acerca do referente (28). Nas *notícias*, possibilitam uma melhor descrição e contextualização dos eventos relatados (33), contribuindo para a informatividade do texto.

Considerações finais

Neste artigo, discutimos o uso padrão das orações subordinadas adjetivas, no sentido de explorar a potencialidade comunicativa dessa construção sintática para além da visão canônica trazida pelos manuais de gramática tradicional. Nossos resultados, embasados numa abordagem funcionalista, demonstram que o funcionamento dessas orações é motivado também por elementos do contexto: o gênero e o tipo textual em que ocorrem, o conteúdo que trazem e as intenções comunicativas do falante, o que comprova nossa hipótese de que há uma correlação entre todos estes fatores. A partir dos dados do *corpus*, pudemos mostrar que as orações adjetivas usadas nos artigos de opinião apresentam um caráter mais argumentativo, fortemente associado à expressão da opinião do autor. As adjetivas usadas nas notícias de jornais, por sua vez, veiculam informações mais objetivas, que auxiliam na apresentação dos fatos noticiados e na informatividade dos textos. Esperamos, assim, contribuir para uma abordagem mais reflexiva sobre o uso padrão dessas orações, para além dos critérios semânticos de *restrição* ou *explicação*. Retomando o princípio básico do funcionalismo, as intenções comunicativas do falante motivam suas escolhas linguísticas (tanto lexicais quanto gramaticais), conforme podemos observar nos diferentes gêneros estudados.

Agradecimento

Agradecemos à CAPES, pela concessão da Bolsa de Pós-Doutorado.

Referências

ABREU, A. S. *Gramática mínima para o domínio da Língua Padrão*. 2. ed. Cotia: Ateliê Editorial, 2003.

ABREU, A. S. *Texto e gramática*. Uma visão integrada e funcional para a leitura e a escrita. São Paulo: Melhoramentos, 2012.

ABREU, A. S. *Gramática Integral da Língua Portuguesa*. Cotia: Ateliê Editorial, 2018.

BASTOS, W. A. *A relativização no Português do Brasil: a sentença orientada para o discurso*. 2008. 167 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Estadual Paulista, Araraquara, 2008.

BECHARA, E. *Moderna Gramática Portuguesa*. 37. ed. Rio de Janeiro: Nova Fronteira, 2009.

BISPO, E. B. Estratégias de relativização no PB: motivações discursivo-interacionais e cognitivas. In: BISPO, E. B.; OLIVEIRA, M. R. (org.). *Orações relativas no português brasileiro: diferentes perspectivas*. Niterói: Editora da UFF, 2014.

BISPO, E. B.; OLIVEIRA, M. R. (org.). *Orações relativas no português brasileiro: diferentes perspectivas*. Niterói: Editora da UFF, 2014.

CASTILHO, A. T. *Gramática do português brasileiro*. São Paulo: Contexto, 2010.

CROFT, W. *Typology and universals*. Cambridge, United Kingdom: CUP, 1990.

CROFT, W. Toward a social cognitive linguistics. In: EVANS, V.; POURCEL, S. (ed.). *New directions in Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 2009.

CORRÊA, V. R. *Oração relativa: o que se fala e o que se aprende no português do Brasil*. 1998. Tese (Doutorado) – Instituto de Estudos da linguagem, Universidade Estadual de Campinas, Campinas, 1998.

CUNHA, C.; CINTRA, L. *Nova gramática do Português Contemporâneo*. 4. ed. Rio de Janeiro: Lexikon Editora Digital, 2007.

DECAT, M. B. N. Orações adjetivas explicativas no português brasileiro e no português europeu: aposição rumo ao 'desgarramento'. *SCRIPTA*, Belo Horizonte, v. 5, n. 9, p. 104-118, 2º sem. 2001.

DECAT, M. B. N. *Estruturas desgarradas em Língua Portuguesa*. Campinas: Pontes Editores, 2011.

DECAT, M. B. N. Orações relativas apositivas *desgarradas* no português em uso. In: BISPO, E. B.; OLIVEIRA, M. R. (org.). *Orações relativas no português brasileiro: diferentes perspectivas*. Niterói: Editora da UFF, 2014.

FAUCONNIER, G. *Mental Spaces*. Cambridge: Cambridge University Press, 1994.

FAUCONNIER, G. *Mappings in Thought and Language*. Cambridge: Cambridge University Press, 1997.

GIVÓN, T. *Syntax: a functional-typological introduction*. v. I. New York: Academic Press, 1984.

KATO, M. A. Recontando a história das relativas em uma perspectiva paramétrica. In: KATO, M. A.; ROBERTS, I. (org.). *Português brasileiro: uma viagem diacrônica*. Campinas: UNICAMP, 1996.

MARCHESAN, A. C.; MIOTO, C. Relativas livres. In: BISPO, E. B.; OLIVEIRA, M. R. (org.). *Orações relativas no português brasileiro: diferentes perspectivas*. Niterói: Editora da UFF, 2014.

NEVES, M. H. M. *Gramática de usos do português*. São Paulo: Editora UNESP, 2000.

OLIVEIRA, M. R.; AMORIM, M. B. Orações relativas na sala de aula em perspectiva funcionalista. In: BISPO, E. B.; OLIVEIRA, M. R. (org.). *Orações relativas no português brasileiro: diferentes perspectivas*. Niterói: Editora da UFF, 2014.

ROJO, R. H. *A prática de linguagem em sala de aula: Praticando os PCNs*. Campinas: Mercado de Letras, 2000.

TARALLO, F. L. *Relativization strategies in brasilian portuguese*. 1983. Tese (Doutorado em Linguística) – University of Pennsylvania, Pennsylvania, 1983.

Aspectos acústico-prosódicos do alongamento silábico em contextos de intensificação na fala de professores do ensino básico

DOI: <http://dx.doi.org/10.21165/el.v54i1.3810>

Fernando Luiz Pompeu Varela¹

Resumo

Este trabalho apresenta uma caracterização acústica das sílabas e seus fones entre as posições pretônica, tônica e postônica de palavras sob efeitos de intensificação de sentido. Os parâmetros analisados incluem duração, frequência fundamental (f_0) e ênfase espectral (intensidade relativa). Os dados foram coletados por meio de um experimento de leitura de um texto em três contextos de intensificação (neutro, intensificado e extraintensificado). As medidas acústicas foram extraídas por meio do *script GenAcousticsSegments*, no programa de análise acústica *PRAAT*. Os dados são de seis participantes (3 masculinos e 3 femininos) professores do ensino básico de Campinas/SP e região de São Paulo. Os resultados demonstraram que a duração silábica aumenta conforme são atribuídos os contextos de intensificação, caracterizando uma relação icônica entre alongamento silábico e efeitos de intensificação de sentido.

Palavras-chave: contextos de intensificação; análise acústico-prosódica; alongamento silábico; fala de professores.

¹ Universidade Estadual de Campinas (Unicamp), Campinas, São Paulo, Brasil; fernandopompeuvarela23@gmail.com; <https://orcid.org/0009-0003-2215-4422>

Acoustic-Prosodic Aspects of Syllabic Lengthening in Intensification Contexts in the Speech of Primary School Teachers

Abstract

This study presents an acoustic characterization of syllables and their segments in pretonic, tonic, and postonic positions of words under the effects of meaning intensification. The analyzed parameters include duration, fundamental frequency (f_0), and spectral emphasis (relative intensity). Data were collected through a reading experiment of a text in three intensification contexts: neutral, intensified, and extra-intensified. Acoustic measurements were extracted using the GenAcousticsSegments script in the PRAAT software. The data come from six participants (three male and three female participants), all primary and secondary school teachers from Campinas/SP and the São Paulo region. The results showed that syllable duration increases with increasing levels of intensification, indicating an iconic relationship between syllable lengthening and the intensification of meaning.

Keywords: intensification contexts; acoustic-prosodic analysis; syllabic lengthening; teacher speech.

Introdução

A intensificação de sentido é o reforço, para mais ou para menos, que se atribui às noções graduais de quantidade, tamanho/dimensão, peso/força, localização, estados/sensações biofísico(a)s ou psicoafetivo(a)s, valor/desvalor e pode ser expressa por diferentes planos linguísticos de expressão (lexical, morfológico, fonético, sintático e textual) (Silva, 2014). Do ponto de vista morfológico, por exemplo, o uso da palavra 'muito' com acréscimo de sufixo intensificador – 'muitíssimo' e '**muitíssissíssimo**' – pode expressar níveis progressivos da quantidade de algo. Esse procedimento é regido pelo princípio icônico da quantidade, em que o acréscimo de mais significado é representado pela adição de mais formas linguísticas (Lakoff; Johnson, 2002; Silva, 2014). Porém, do ponto de vista da produção acústica da fala, qual(is) parâmetro(s) acústico(s) manifesta(m) os efeitos de intensificação de sentido e como se caracteriza(m)?

Para responder a esses questionamentos, o presente trabalho apresenta uma caracterização acústica, em termos de duração, frequência fundamental (f_0) e ênfase espectral (intensidade relativa), das sílabas e seus fones das posições pretônica, tônica e postônica de palavras sob efeitos de intensificação de sentido. Os resultados são oriundos de pesquisa anterior baseada na dissertação apresentada ao Mestrado em Linguística da Unicamp. Nessa pesquisa, foram verificadas as seguintes hipóteses:

1) o alongamento da sílaba tônica das palavras em contextos de intensificação de sentido se caracteriza por um pico duracional significativo, que se sobressai aos valores de duração de outras sílabas da vizinhança (Lakoff; Johnson, 2002; Cagliari, 1992; Gomes; Silva, 2014; Silva, Souza; Andrade, 2009, Silva, 2009; Silva, 2015; Carvalho, 2020);

2) a duração da sílaba tônica da palavra sob efeito de intensificação torna-se ainda mais substancial à medida que há um reforço de intensificação ao seu conteúdo semântico/ideacional (Lakoff; Johnson, 2002; Silva, 2014);

3) o alongamento na sílaba tônica é acompanhado de uma redução na sílaba postônica, que serve para maximizar a diferença de duração entre essas sílabas para que a tônica seja percebida como proeminente (Fletcher, 2010), à medida que é atribuída intensificação;

4) a duração da vogal do núcleo da sílaba tônica é mais afetada que os demais segmentos pelas atribuições de intensificação de sentido, pois os efeitos de alongamento prosódico afetam principalmente a vogal tônica (Arantes, 2022; Barbosa, 2012);

5) A frequência fundamental manifesta-se por uma elevação na sílaba pretônica da palavra sob efeito de intensificação, bem como se observa nas palavras intensivas lexicalmente e em algumas derivadas por afixos intensivos (Gonçalves; 2002; Travaglia, 2006);

6) Há reforço de intensidade na pretônica, por serem palavras de semântica intensiva lexicalmente (Gonçalves, 2002) e pela marcação de proeminência inicial ser um recurso para destacar as palavras de seu entorno (Arantes, 2010; Barbosa; Mareüil, 2016), e na tônica (Sluijter; Van Heuven, 1996), por conta da realização de maior esforço vocal esperado nessa posição.

Os dados foram coletados por meio de um experimento de leitura de um texto em três contextos de intensificação (neutro, intensificado e extraintensificado). As medidas acústicas foram extraídas por meio do *script GenAcousticsSegments*, no programa de análise acústica *PRAAT*. Os dados são de seis participantes (3 masculinos e 3 femininos) professores do ensino básico de Campinas/SP e região de São Paulo.

Materiais e métodos

A pesquisa teve sua execução mediante aprovação do projeto pelo Comitê de Ética em Pesquisa de Ciências Humanas da Universidade Estadual de Campinas, cujo parecer foi favorável para coleta de dados (CAAE nº: 59562122.9.0000.8142).

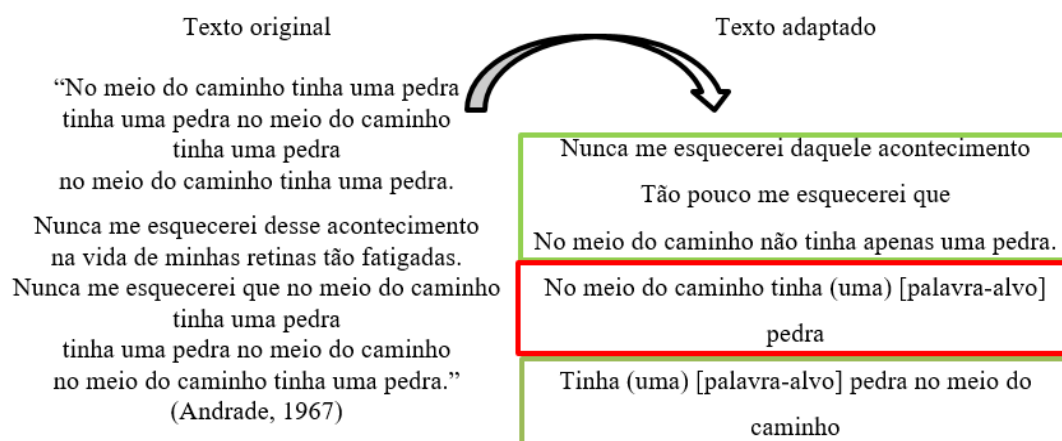
Critérios de seleção dos participantes

Os dados são de três participantes femininos e três participantes masculinos, todos professores da educação básica, de língua portuguesa e da educação infantil (pedagogo(a)), com idades entre 20 e 50 anos, de Campinas e arredores de Campinas, no estado de São Paulo.

Montagem do experimento

É um experimento de leitura em voz alta de pequenos textos. Para a tarefa, adaptou-se o texto de Carlos D. Andrade “No meio do caminho”:

Figura 1. Estratégia de adaptação do texto-veículo para a leitura e realização da intensificação



Fonte: Elaboração própria

Selecionaram-se as palavras-alvo: *muita*, *baita*, *mega*, *profusa* e *vultosa*. Essas palavras possuem semântica intensiva (Gonçalves, 2002; Silva, 2014).

Para a atribuição de intensificação de sentido pelo participante da pesquisa, foram elaborados estímulos imagem-texto com o objetivo de sinalizar os contextos de intensificação (Figura 2):

Figura 2. Apresentação dos estímulos imagem-texto para a leitura dos participantes da pesquisa



Fonte: Elaboração própria

Observe na figura 2 que o texto adaptado é repetido por vezes. Contudo, em contextos distintos. Essa distinção é estabelecida pelo aumento da quantidade². Além das imagens, a cada repetição do texto, a fonte do texto é aumentada e a palavra-alvo é sinalizada em caixa-alta, sugerindo ao participante/leitor a noção de intensificação. Essas estratégias textuais e, especialmente, imagéticas/semióticas constroem uma relação de iconicidade entre as noções quantitativas (e também dimensivas) e o referente da realidade material (*pedra*), que resultam na intensificação dessas noções graduais, à medida que se variam progressivamente as propriedades da imagem. Nesse sentido, são estabelecidos três contextos de intensificação: *neutro*³, *intensificado* e *extraintensificado*.

Para a coleta dos dados, foi criado um *site* para a apresentação *online* dos estímulos ao participante e gravação da leitura em voz alta. O experimento pôde ser realizado individualmente pelo participante em qualquer lugar, desde que tivesse condições mínimas acústicas (atenuação de ruídos externos). A seguir, são apresentadas as etapas de realização do experimento.

Primeiramente, o participante lê as instruções gerais acerca do experimento. Ao prosseguir, o participante precisa preencher um pequeno formulário com informações de identificação do participante (nome, idade, gênero e residência). Em seguida, são apresentadas ao participante instruções específicas a respeito da interpretação graduada

2 Ou, para os casos de *mega*, *baita*, *vultosa* e *profusa*, do tamanho do elemento *pedra* intensificado nas imagens.

3 Do ponto de vista semântico, neutro (faixa roxa) é um contexto de intensificação por conter a palavra de semântica intensiva (palavra-alvo) em nível lexical. Para os conteúdos intensificado e extraintensificado (faixas rosa e dourada) espera-se que a intensificação seja expressa por vias prosódicas, para além do que a palavra remete e por ser o mesmo conteúdo escrito, espera-se que essas situações sejam expressas gradualmente na leitura por vias prosódicas.

das imagens-texto (Cf. Figura 3). Feito isso, o participante pode prosseguir para a leitura em voz alta dos estímulos imagem-texto (Cf. Figura 2) utilizando botões para gravar, parar, tocar e prosseguir para os demais estímulos imagem-texto.

Os dados gerais constam organizados na planilha *Medidas acústico-prosódicas de intensificação de sentido*⁴ elaborada para aplicação das análises descritivas e estatísticas. Para o presente estudo, foram analisadas uma quantidade de 81 palavras-alvo realizadas pelos participantes, considerando as realizações em cada contexto de intensificação. Dessas palavras, foram analisadas 243 sílabas e 486 fones das posições pretônica, tônica e postônica.

Figura 3. Apresentação dos estímulos imagem-texto para a leitura dos participantes da pesquisa

Instruções gerais:

Intensificação de Sentido é um processo linguístico, em que se atribui intensidade ou reforço, para mais ou para menos, à ideia de algum conteúdo, para além do seu sentido básico.

As ilustrações a seguir apresentam diferentes quantidades e/ou tamanhos de pedra. Esses atributos são aumentados em cada ilustração.

Considerando que em cada figura o sentido de quantidade ou de tamanho é intensificado (aumentado), expresse essa noção sobre as palavras em destaque durante sua leitura oral.

Repita o processo nos textos das páginas seguintes.

Prosseguir

Fonte: Elaboração própria

Análise dos dados

As medidas acústicas referentes à duração, frequência fundamental (f_0) e ênfase espectral foram extraídas por meio do *script GenAcousticsSegments*⁵ no programa computacional *Praat* (Boersma; Weenink, 2013) a partir das posições acentuais pretônica, tônica e postônica das palavras-alvo. No caso das palavras dissílabas, considerou-se como posição pretônica a sílaba final da palavra anterior contígua à palavra-alvo. Ex: **uma** baita; **tinha** muita.

As medidas de duração, em milissegundos (ms), das três posições acentuais foram extraídas a partir das(os): a) sílabas fonológicas (DurSyl), considerando que, em situações de ênfase, a consoante de *onset* da sílaba caminha junto à vogal do núcleo (Barbosa, 1996); b) sílabas fonéticas (Unidades VV), pois são unidades do tamanho da sílaba compreendidas do ataque de uma vogal ao ataque da vogal seguinte, integrando fones de duas sílabas distintas, que servem para caracterizar a estrutura duracional em termos

4 Ver em *figshare. Dataset*. <https://doi.org/10.6084/m9.figshare.25874161.v1>

5 Autor do trabalho.

de produção e percepção do enunciado (Barbosa, 2006; Barbosa, 2019); c) segmentos fônicos, os quais podem revelar o impacto/efeito da ocorrência de alongamento no interior da sílaba.

A f_0 é “o correlato acústico da frequência de vibração das pregas vocais” (Barbosa, 2012, p. 17) e corresponde à percepção de altura melódica (*pitch*) da fala (Moraes, 2024; Barbosa, 2012). Portanto, foram analisadas as medidas de f_0 em Hertz (Hz) das três posições acentuais. Este parâmetro mostra-se robusto para gravações com equipamentos não profissionais, sendo menos sensível a variações nos sistemas de gravação, tipos de microfone ou processos de compressão de áudio (Cavalcanti *et al.*, 2023). Foram consideradas duas medidas específicas: (1) a mediana de f_0 (f_{0med}), que por ser uma medida não paramétrica e apresenta vantagem ao ignorar valores espúrios no cálculo da tendência central; e (2) o desvio-padrão de f_0 (f_{0sd}), que quantifica a variabilidade das medidas em relação à média dos segmentos fonéticos analisados (Barbosa, 2022).

Foram analisadas também as medidas de ênfase espectral (intensidade relativa em dB), parâmetro escolhido por sua maior robustez frente a variações na intensidade intrínseca – sensível a fatores como: tipo de equipamento não profissional utilizado, qualidade de captação e distância microfone-boca. Esta medida, conforme demonstrado por Traunmüller e Eriksson (2000) e Barbosa (2019, 2022), permite contornar problemas técnicos como a variação na distância entre falante e microfone. Do ponto de vista perceptivo, a intensidade relaciona-se à sensação de volume (Barbosa, 2012), enquanto fisicamente corresponde à amplitude da frequência de vibração das pregas vocais (Moraes, 2024). A ênfase espectral, por sua vez, é o correlato do esforço vocal do falante, conforme Traunmüller e Eriksson (2000) e Barbosa (2022).

Foi aplicado o teste estatístico não paramétrico de dois fatores SHR (Scheirer-Ray-Hare)⁶ sobre as medidas acústicas por posição acentual (primeiro fator) e por contexto de intensificação (segundo fator). Após a aplicação do teste SHR, as medidas foram submetidas ao teste *post hoc* não paramétrico de Wilcoxon (Cf. Barbosa, 2022). O primeiro teste demonstrou a existência ou não de diferença entre pelo menos um dos níveis de cada fator. O segundo teste demonstrou quais níveis de cada fator apresentam ou não diferença estatística.

Na seção a seguir é apresentado um recorte dos principais resultados da aplicação dos testes estatísticos. A apresentação desses resultados é acompanhada de gráficos

⁶ Este é um procedimento equivalente ao teste não paramétrico ANOVA de dois fatores (Cf. Barbosa, 2022). Uma das condições para uso do teste de ANOVA de dois fatores é que os resíduos do modelo passem no teste de normalidade de Shapiro-Wilks (Razali; Wah, 2011). Considerando o parâmetro duração da sílaba fonológica (DurSyl), o teste de normalidade de Shapiro-Wilks teve sua normalidade rejeitada ($W = 0.94854$, $p = 1.439e-07$).

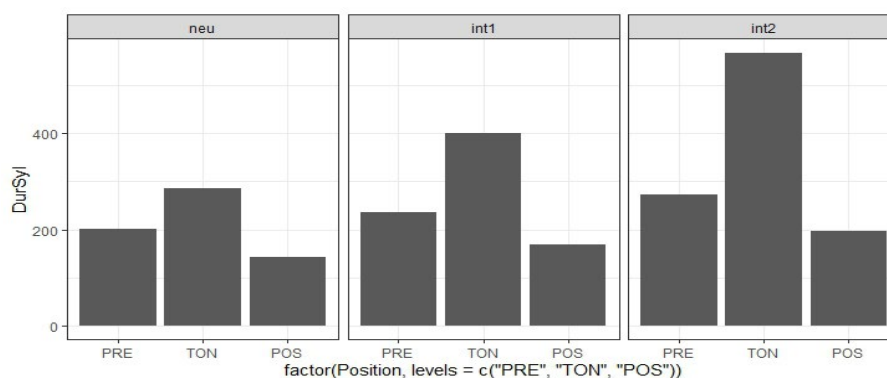
que ilustram a média das medidas acústicas. Os rótulos dos gráficos compreendem: **contextos de intensificação** – neu (Neutro); int1 (Intensificado); int2 (Extraintensificado); e **posições acentuais** – Para as sílabas e vogais, usa-se PRE (Pretônica); TON (Tônica); POS (Postônica). Para as consoantes *onset* da vogal, incluiu-se “C”: PREC, TONC e POSC.

Resultados

Duração

A figura 4 ilustra a duração média em milissegundos das sílabas fonológicas (DurSyl) das posições acentuais pretônica (PRE), tônica (TON) e postônica (POS), por contextos de intensificação neutro (neu), intensificado (int1) e extraintensificado (int2). Por meio dos gráficos, observa-se a relevância da duração da sílaba fonológica tônica. O teste *post hoc* não paramétrico de Wilcoxon demonstrou que a duração da sílaba fonológica tônica é significativamente maior do que das sílabas vizinhas e aumenta significativamente a cada contexto de intensificação.

Figura 4. Duração de sílabas fonológicas pretônica, tônica e postônica por contextos de intensificação

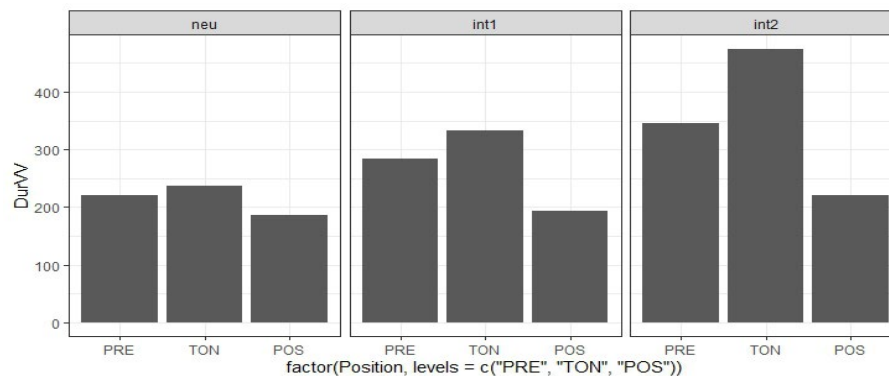


Fonte: Elaboração própria

Quando o nível de análise se dá por sílabas fonéticas (unidades VV), as medidas de duração só se distinguem significativamente entre as sílabas tônica e postônica ($p < 0,05$), especialmente nos casos do contexto intensificado e extraintensificado.

A figura 5 ilustra a duração média em milissegundos das sílabas fonéticas (DurVV) das posições acentuais pretônica (PRE), tônica (TON) e postônica (POS), por contextos de intensificação neutro (neu), intensificado (int1) e extraintensificado (int2). Observa-se que a duração da sílaba fonética tônica é a maior do grupo. Porém, por meio do teste *post hoc* não paramétrico de Wilcoxon, verificou-se que a duração da sílaba tônica é significativamente maior que a postônica, apenas.

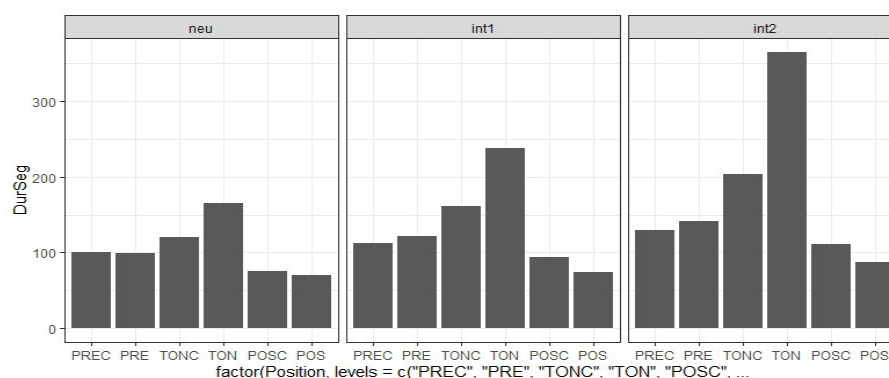
Figura 5. Média de duração de sílabas fonéticas (Unidades VV) pretônica, tônica e postônica por contextos de intensificação



Fonte: Elaboração própria

A figura 6 ilustra a duração média em milissegundos dos segmentos (DurSeg) das vogais e consoantes das posições acentuais pretônica (PRE), tônica (TON) e postônica (POS), por contextos de intensificação neutro (neu), intensificado (int1) e extraintensificado (int2). Lembrando que, para indicar a posição acentual das vogais, usa-se PRE, TON e POS. Para as consoantes *onset* da vogal, incluiu-se "C": PREC, TONC e POSC. O teste *post hoc* não paramétrico de Wilcoxon demonstrou que a duração da vogal tônica é significativamente maior do que as vogais vizinhas e aumenta significativamente a cada contexto de intensificação. A consoante da tônica, por sua vez, possui duração significativamente maior do que a consoante da postônica apenas nos contextos neutro e intensificado. No contexto extraintensificado, a duração da consoante da tônica é significativamente maior que as consoantes vizinhas pretônica e postônica.

Figura 6. Média de duração de consoantes e vogais das sílabas pretônica, tônica e postônica por contextos de intensificação

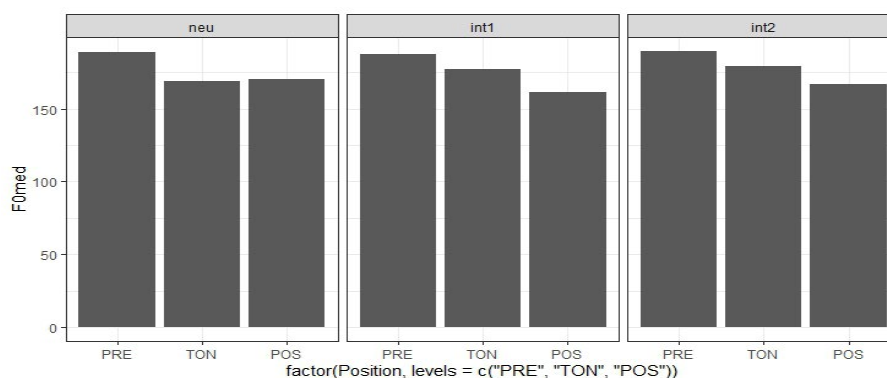


Fonte: Elaboração própria

Frequência fundamental (f_0)

A figura 7 ilustra a média dos valores da f_0 mediana (em Hz) das sílabas fonéticas pretônica (PRE), tônica (TON) e postônica (POS), por contextos de intensificação: neutro (neu), intensificado (int1) e extraintensificado (int2). Observa-se maiores valores na sílaba pretônica. Porém, o teste *post hoc* não paramétrico de Wilcoxon não indicou nenhuma diferença significativa entre as posições acentuais e por contexto de intensificação:

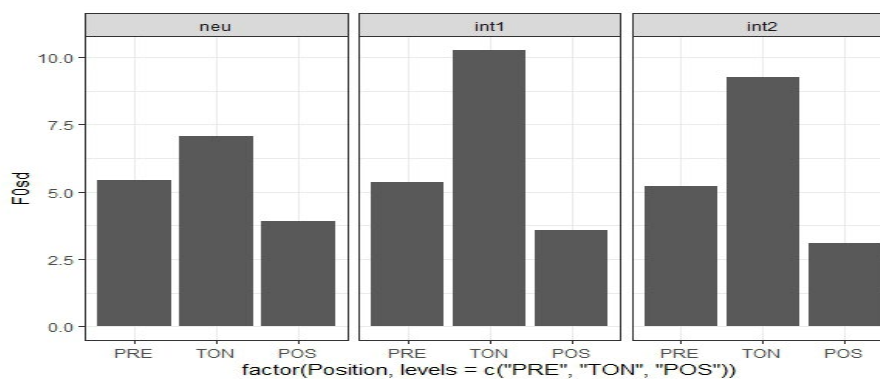
Figura 7. Média de frequência fundamental mediana das sílabas pretônica, tônica e postônica por contextos de intensificação



Fonte: Elaboração própria

Diferente da figura 7, na figura 8, que ilustra os valores médios do desvio-padrão de f_0 (F_0sd) em Hz das sílabas fonéticas pretônica (PRE), tônica (TON) e postônica (POS), por contextos de intensificação neutro (neu), intensificado (int1) e extraintensificado (int2), a sílaba tônica possui maiores valores de desvio-padrão de f_0 . Contudo, teste *post hoc* não paramétrico de Wilcoxon demonstrou que o desvio-padrão é significativo entre as sílabas tônica e postônica nos contextos intensificado e extraintensificado:

Figura 8. Média das medidas de desvio-padrão das sílabas pretônica, tônica e postônica por contextos de intensificação

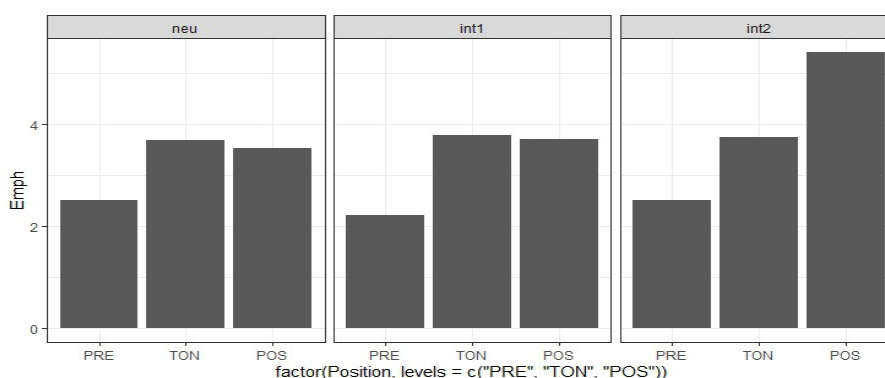


Fonte: Elaboração própria

Ênfase espectral (intensidade relativa)

Na figura 9, que ilustra os valores médios de ênfase espectral (Emph) em dB das sílabas fonéticas pretônica (PRE), tônica (TON) e postônica (POS), por contextos de intensificação neutro (neu), intensificado (int1) e extraintensificado (int2), observa-se medidas elevadas na tônica e postônicas, porém, o teste *post hoc* não paramétrico de Wilcoxon demonstrou que há diferença significativa apenas entre as sílabas pretônica e postônica no contexto extraintensificado:

Figura 9. Média de ênfase espectral das sílabas pretônica, tônica e postônica por contextos de intensificação



Fonte: Elaboração própria

Discussão

Observou-se um comportamento duracional distinto entre a duração silábica fonológica e fonética, que corroborou parcialmente a primeira e a segunda hipótese desse trabalho, a saber: 1) o alongamento da sílaba tônica das palavras em contextos de intensificação de sentido se caracteriza por um pico duracional significativo, que se sobressai aos valores de duração de outras sílabas da vizinhança (Lakoff; Johnson, 2002; Cagliari, 1992; Gomes; Silva, 2014; Silva, Souza; Andrade, 2009, Silva, 2009; Silva, 2015; Carvalho, 2020); 2) a duração da sílaba tônica da palavra sob efeito de intensificação torna-se ainda mais substancial à medida que há um reforço de intensificação ao seu conteúdo semântico/ideacional (Lakoff; Johnson, 2002; Silva, 2014).

A falta de diferença significativa de duração entre as sílabas fonéticas pretônica e tônica pode estar relacionada à constituição da unidade VV (Cf. Barbosa, 2006; Barbosa, 2019). A duração da vogal da sílaba pretônica (PRE) é a segunda maior medida entre as três vogais e é significativamente maior que a duração da vogal postônica (POS). A duração da consoante (TONC) da vogal tônica foi afetada e aumenta significativamente do neutro para o extraintensificado. Portanto, a duração da vogal pretônica integrada à duração da consoante seguinte faz com que a sílaba fonética pretônica obtenha valores elevados

próximos aos da sílaba tônica. Por outro lado, no caso da sílaba fonológica, a duração elevada da consoante da tônica integrada à da vogal do núcleo faz com que a sílaba fonológica tônica seja significativamente maior do que as sílabas da vizinhança em termos duracionais.

Além da própria composição da unidade VV, a falta de diferença significativa de duração entre as sílabas fonéticas pretônica e tônica pode estar relacionada ao alongamento da tônica. Segundo Barbosa e Madureira (2015), a ocorrência de um pico de duração na sílaba tônica de uma palavra acentuada lexicalmente afeta a duração das sílabas anteriores, aumentando-as, ainda que pertençam a outras palavras fonológicas. Conforme prediz o modelo dinâmico do ritmo (MDR) da fala, no português brasileiro, a força que o falante imprime sobre as proeminências prosódicas acarreta um aumento progressivo de duração no grupo acentual até a realização do acento frasal (Barbosa, 2022).

Se, por um lado, a duração da sílaba pretônica é carregada pela tônica, por outro, a duração da sílaba postônica também é afetada pela atribuição de intensificação. A duração tanto da sílaba fonológica quanto da sílaba fonética postônica difere significativamente entre os contextos neutro e extraintensificado ($p < 0,01$; $p < 0,05$). No caso da sílaba fonológica, a consoante da sílaba postônica (POSC) é a principal responsável por essa diferença, uma vez que o *onset* possui diferença significativa de duração entre os contextos neutro e extraintensificado. No caso da sílaba fonética, é provável que a diferença de duração da sílaba postônica [ɐp] entre os contextos neutro e intensificado tenha sido pela influência da duração do *onset* da tônica [p] da palavra *pedra*.

Esses dados indicam que a duração da sílaba postônica também aumenta pela atribuição de intensificação e, desse modo, não corroboram a hipótese 3, a qual afirma que o alongamento na sílaba tônica é acompanhado de uma redução na sílaba postônica, que serve para maximizar a diferença de duração entre essas sílabas para que a tônica seja percebida como proeminente (Fletcher, 2010), à medida que é atribuída intensificação. Não foi verificada a redução da duração dessa sílaba à medida que a tônica se alonga. Pelo contrário, os achados demonstram que o alongamento na sílaba tônica é acompanhado de um aumento duracional da postônica (e também da pretônica), caracterizando uma unidade maior, para além da sílaba tônica, em que incide a intensificação.

Nessa empreitada, a vogal do núcleo da sílaba tônica é mais afetada do que os demais segmentos nas atribuições de intensificação. Portanto, confirma-se que a duração da vogal do núcleo da sílaba tônica é mais afetada que os demais segmentos pelas atribuições de intensificação de sentido, pois os efeitos de alongamento prosódico afetam principalmente a vogal tônica (Arantes, 2022; Barbosa, 2012) (Hipótese 4). O teste *post hoc* não paramétrico de Wilcoxon indicou diferença significativa entre a duração da vogal núcleo da sílaba tônica e das demais vogais e aumento significativo da duração da vogal tônica à medida que são atribuídos os contextos de intensificação.

Esse dado converge com os resultados de Arantes (2022). O autor confirma que os efeitos de alongamento prosódico provocados por acento frasal afetam principalmente a vogal tônica. Por outro lado, segundo Arantes (2022), não somente a duração da vogal, mas a duração dos fones que compõem a sílaba também é afetada por conta dos efeitos do alongamento. Barbosa (2012) também observa que, quando se atribui proeminência em uma palavra, para além do acento lexical, as unidades fônicas de sua sílaba tônica são afetadas, pois atingem valores de duração mais longos.

No sentido do que é colocado acima, verificou-se que a duração da consoante *onset* da sílaba tônica também se mostrou distinta da duração das consoantes das sílabas vizinhas, tendo sua duração afetada significativamente na comparação entre pelo menos dois contextos de intensificação, ainda que com valores inferiores aos da vogal núcleo da sílaba tônica, mas confirmando que os fones da sílaba tônica são afetados pela atribuição de alongamento/proeminência nessa posição.

Ao que se verifica, há uma relação entre os efeitos da atribuição de intensificação e os correlatos de duração, no que diz respeito ao alongamento silábico. Porém, no caso das medidas de frequência fundamental, nota-se duas formas de realização. A primeira forma se caracteriza por medidas de *f0* mediana similares entre as sílabas e entre os contextos de intensificação. A segunda forma se caracteriza pela diferença significativa de desvio-padrão de *f0* nos contextos intensificado e extraintensificado: o desvio-padrão da sílaba tônica é significativamente mais elevado que o da postônica. Desse modo, não foi corroborada a hipótese (5), a qual afirma que a frequência fundamental se manifesta por uma elevação na sílaba pretônica da palavra sob efeito de intensificação, bem como se observa nas palavras intensivas lexicalmente e em algumas derivadas por afixos intensivos (Gonçalves; 2002; Travaglia, 2006). A mediana e o desvio-padrão de *f0* não demonstraram relevância da sílaba pretônica em relação às sílabas tônica e postônica, como nas palavras intensivas ou intensificadas sufixalmente, às quais Gonçalves (2002) e Travaglia (2006) atribuíram elevação de *f0* na sílaba pretônica ou em início de palavra. Os resultados acerca das duas medidas demonstram um padrão diferente do que se afirmou na quinta hipótese (5).

Os resultados da *f0* mediana também podem ser relacionados ao que Travaglia (2006) verificou acerca da realização de tonicidade em algumas palavras intensivas. Segundo Travaglia (2006), há a existência de uma gradação progressiva da atribuição do relevo (ou proeminência) pelo que chama de “altura de voz”, que ocorre na maioria das vezes na sílaba tônica, mas pode ocorrer em mais de uma sílaba da palavra sob focalização. Segundo o autor, a maior tonicidade pode abranger todas as sílabas de uma palavra de semântica intensiva: “SENSACIONAL”, “EXÓTICO”, “PRECISA” e “MERAMENTE”⁷ (Travaglia, 2006).

7 A letra em caixa alta representa a tonicidade das sílabas.

A respeito do correlato acústico de intensidade, Gonçalves (2002) verifica que há atuação conjunta entre f_0 e um aumento de intensidade geralmente na sílaba pretônica de palavras de semântica intensiva. Isso foi verificado por ele tanto nas palavras lexicalmente intensivas quanto nas derivadas por sufixo intensivo. Para Barbosa (2012), o aumento de intensidade é um dos correlatos quando a proeminência recai em uma palavra, para além do acento lexical. Contudo, não foi verificado o aumento de intensidade nas sílabas pretônica e tônica nos dados sobre a intensificação. A medida de ênfase espectral (intensidade relativa) das posições acentuais não se distingue nos contextos neutro e intensificado. O que ocorre é um aumento significativo de ênfase espectral na postônica em relação à pretônica no contexto extraintensificado. Por essa razão, não se confirma a hipótese (6), que há reforço de intensidade na pretônica, por serem palavras de semântica intensiva lexicalmente (Gonçalves, 2002) e pela marcação de proeminência inicial ser um recurso para destacar as palavras de seu entorno (Arantes, 2010; Barbosa; Mareüil, 2016), e na tônica (Sluijter; Van Heuven, 1996), por conta da realização de maior esforço vocal esperado nessa posição.

Algo que não está conforme aos achados de Barbosa (2022), que uma queda em ênfase espectral na vogal postônica contribui para que a vogal tônica seja percebida como acentuada lexicalmente. Esperava-se também um reforço de intensidade na sílaba pretônica, por serem palavras lexicalmente intensivas (Gonçalves, 2002) e porque a marcação de proeminência inicial é um recurso frequentemente utilizado para destacar palavras em relação ao seu entorno (Arantes, 2010; Barbosa; Mareüil, 2016). Além disso, esperava-se também um aumento de intensidade na sílaba tônica, devido ao maior esforço vocal tipicamente associado a essa posição (Sluijter; Van Heuven, 1996).

Há outros fatores que podemos associar aos resultados de ênfase espectral apresentados na presente pesquisa. Um deles é a familiaridade que alguns falantes têm com determinadas palavras na fala. Apesar de “muito”, “baita” e “mega” serem tão produtivas (Lopes, 2007; Carvalho, 2019; Foltran; Nóbrega, 2016) quanto as palavras “profusa” e “vultosa”, a frequência com que as palavras ocorrem na língua (Broadbent, 1967) e/ou a densidade de vizinhança, conforme a quantidade de palavras com sons semelhantes existentes no arcabouço lexical (Luce, 1986), são fatores que podem implicar modificações no modo de produzimos tais palavras. Um maior monitoramento da produção da fala pode resultar em um maior esforço vocal e, conseqüentemente, maiores valores de ênfase espectral (Miranda *et al.*, 2017).

Conclusão

O presente artigo teve como objetivo apresentar uma caracterização acústico-prosódica da intensificação de sentido realizada na fala de professores do ensino básico. Os resultados demonstraram que, entre as medidas, a duração caracteriza a relação icônica de quantidade entre alongamento silábico e efeitos de intensificação de sentido. A

frequência fundamental e a ênfase espectral apresentaram um comportamento distinto do esperado, sugerindo que, em contextos de intensificação, a frequência fundamental pode não diferir significativamente entre as posições acentuadas ou que a diferença relevante ocorra especificamente entre as sílabas tônica e postônica. Além disso, os dados indicam que a ênfase espectral se caracteriza por um aumento na postônica da palavra intensificada, reforçando a distinção entre os padrões observados nessas medidas acústicas.

Referências

ANDRADE, C. D. *Uma pedra no meio do caminho*: biografia de um poema. Rio de Janeiro: Editora do Autor, 1967.

ARANTES, P. Acento. In: OLIVEIRA-JR, M. *Prosódia, prosódias*: uma introdução. 1. ed. São Paulo: Contexto, 2022. p. 9-25.

BARBOSA, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes/Fapesp, 2006.

BARBOSA, P. A. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. *Cadernos de Estudos Linguísticos*, v. 31, p. 33-53, 1996.

BARBOSA, P. A. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 20, n. 1, p. 11-27, jan./jun. 2012. Disponível em: <http://manualdefoneticaacusticaexperimental.com/assets/barbosa2012-3.pdf>. Acesso em: 10 nov. 2025.

BARBOSA, P. A. *Prosódia*. 1 ed. São Paulo: Parábola, 2019.

BARBOSA, P. A. *Manual de prosódia experimental*. 1. ed. Campinas: Editora da Abralin, 2022.

BARBOSA, P. A.; MADUREIRA, S. *Manual de Fonética Acústica Experimental: Aplicações a dados do português*. São Paulo: Cortez, 2015.

BOERSMA, P.; WEENINK, D. *Praat*: doing phonetics by computer [Computer program], 2013. Versão 5.3.51, baixado em 2 Junho de 2013 do site [www.praat.org].

BROADBENT, D. E. Word-frequency effect and response bias. *Psychological review*, v. 74, n. 1, p. 1, 1967.

CAGLIARI, L. C. Prosódia: algumas funções dos supra-segmentos. *Cadernos de estudos linguísticos*, v. 23, Campinas, p. 137-151, 1992. Disponível em: <http://revistas.iel.unicamp.br/index.php/cel/article/view/2901/4188>. Acesso em: 10 nov. 2025.

CARVALHO, H. A. *Expressão da gradação aumentativa na fala manauara*. 2020. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, Programa de Pós-Graduação em Linguística, Florianópolis, 2020.

CARVALHO, S. D. *Os advérbios graduadores na fala rural e na fala urbana de minas gerais*. 2019. Tese (Doutorado) – Programa de Pós-Graduação em Estudos Linguísticos, Faculdade de Letras da UFMG, Belo Horizonte, 2019.

CAVALCANTI, J. C. *et al.* Microphone and audio compression effects on acoustic voice analysis: A pilot study. *Journal of Voice*, v. 37, n. 2, p. 162-172, 2023.

CINTRA, G. Distribuição de padrões acentuais no vocábulo em português. *Confluência: Boletim do Departamento de Linguística*, v. 5, n. 3, p. 82-93, 1997.

FLETCHER J. The prosody of speech: Timing and rhythm. *In*: HARDCASTLE, W. J.; LAVER, J.; GIBBON, F. *The handbook of phonetic sciences*. 2 ed. Blackwell, Oxford to appear, 2010. p. 523-602.

FOLTRAN, M. J.; NÓBREGA, V. A. Adjetivos intensificadores no português brasileiro: propriedades, distribuição e reflexos morfológicos. *Alfa: Revista de Linguística*, São Paulo, v. 60, n. 2, p. 319-340, 2016.

GOMES, C.; SILVA, F. Mecanismos de atenuação e intensificação no ensino-aprendizagem do português como língua estrangeiras. *Revista de Estudos do Discurso*, n 3, p. 32-66, 2014.

GONÇALVES, C. A. Morfopragmática da intensificação sufixal em português. *Rev. de Letras*, n. 24, v. 1/2, jan./dez. 2002.

LAKOFF, G.; JOHNSON, M. *Metáforas da vida cotidiana*. Tradução GEIM. Campinas: Mercado de Letras, 2002.

LOPES, C. A. G. O intensificador mais usado no português falado. *Revista Philologus, Rio de Janeiro*, ano 13, n. 37, p. 99-104, jan./abr. 2007.

LUCE, P. A. Neighborhoods of words in the mental lexicon. *Research on speech perception, Technical Report*, v. 6, p. 1-91, 1986.

MIRANDA, I. I. *et al.* A variação estilística em diferentes situações de leitura: variedade capixaba. *Letras de Hoje*, v. 52, n. 1, p. 68-76, 2017. Disponível em: <https://revistaseletronicas.pucrs.br/fale/article/view/25393>. Acesso em: 10 nov. 2025.

MORAES, J. A. *Fonética*. 1 ed. São Paulo: Parábola, 2024.

RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modeling and analytics*, v. 2, n. 1, p. 21-33, 2011.

SILVA, J. R. *O Grau em perspectiva: uma abordagem centrada ao uso*. Cortez Editora, 2014.

SILVA, J. R. *Motivações semântico-cognitivas e discursivo-pragmáticas nos processos de intensificação*. 2008. Tese de Doutorado. (Doutorado em Linguística Aplicada) – Programa de Pós-Graduação em Estudos da Linguagem (PPGEL), Universidade Federal do Rio Grande do Norte (UFRN). Natal, 2008.

SILVA, R. A. *A Gradação na Linguagem de propaganda*. 2009. Dissertação (Mestrado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, Faculdade de Letras, Rio de Janeiro, 2009.

SILVA, W. R. P. *A intensificação no viés ensino/aprendizagem: uma abordagem da linguística funcional centrada no uso*. 2015. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte, Centro de Ciências Humanas, Letras e Artes. Programa de Pós-Graduação em Letras, Natal, 2015.

SILVA, B. G.; SOUZA, F. F. F.; ANDRADE, W. C. Intensificação no Português Falado. *Anagrama*, v. 3, n. 1, p. 1-11, 2009.

SLUIJTER, A.; Van HEUVEN, V. Spectral balance as an acoustic correlate of linguistic stress. ("Acoustic Cues to Perception of Word Stress by English, Mandarin, and ...") *The Journal of the Acoustical society of America*, v. 100, n. 4, p. 2471-2485, 1996.

TRAUNMÜLLER, H.; ERIKSSON, A. Acoustic effects of variation in vocal effort by men, women, and children. *J. Acoust. Soc. Am.*, v. 107, p. 3438-3451, 2000.

A não uniformidade das bases numerais: uma visão dos contrastes das bases cardinais do português brasileiro

DOI: <http://dx.doi.org/10.21165/el.v54i1.3889>

Fernando Valls Yoshida¹

Resumo

As bases cardinais *-enta*, *-centos*, *mil* e *milhão*, *bilhão*, ..., organizam os numerais do português brasileiro em um sistema de base decimal. Apesar de sua função matemática comum, defendo que as bases cardinais não podem ser caracterizadas de maneira linguisticamente uniforme. Ao contrário, este artigo sistematiza contrastes entre elas nos domínios sintático (possibilidade de elipse e distribuição com relação a quantificadores), categorial (comportamento de concordância em gênero e número e licenciamento de construção pseudopartitiva) e morfológico (diferenças em estruturas morfológicas e graus de produtividade). A não uniformidade das bases numerais aqui defendida tem impactos, de um lado, na medida em que a Faculdade da Linguagem é sensível à caracterização matemática das bases e, de outro, na diversidade gramatical inerente a esses objetos no português brasileiro.

Palavras-chave: Numerais cardinais; bases numerais; Morfologia; Sintaxe.

¹ Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil; fevayo@usp.br; <https://orcid.org/0000-0002-4400-4869>

The non-uniformity of numeral bases: insights from contrasts of Brazilian Portuguese cardinal bases

Abstract

The cardinal bases *-enta* ('-ty'), *cem* ('hundred'), *mil* ('thousand') and *milhão* ('million'), *bilhão* ('billion'), ..., organize Brazilian Portuguese numerals into a decimal numeral system. Despite their common mathematical function, I argue that cardinal bases cannot be characterized in a linguistically uniform manner. Rather, this paper systematizes contrasts between them at the levels of syntax (possibility of ellipsis and distribution with respect to quantifiers), category (gender and number agreement behavior and licensing of pseudopartitive constructions) and morphology (differences in morphologic structure and degrees of productivity). The non-uniformity of numeral bases defended here has impacts both on the extent to which the Faculty of Language is sensitive to the mathematical characterization of bases and on the inherent grammatical diversity of these objects in Brazilian Portuguese.

Keywords: Cardinal numerals; numeral bases; Morphology; Syntax.

Introdução

Sistemas numerais linguísticos são frequentemente descritos e analisados em termos de bases. Em (1), os diferentes numerais cardinais² para o número 91 ilustram, a partir do contraste entre o valor numérico das bases *juu* 'dez' (1a), *kūrū* 'doze' (1b) e *vingt* 'vinte' (1c), como a escolha de uma base afeta a organização de um sistema numeral.

1. a.

kyuu	ju	ichi	[Japonês (japônica): Kaiser <i>et al.</i> (2013)]
nove	dez	um	

"nove dezenas e um (\approx noventa e um)"
- b.

bā-kūru	bā-tāmà	ná	bī-tāmà	[Berom (benue-congolesa): Comrie (2022)]
PL-doze	PL-sete	mais	CL-sete	

"sete dúzias e sete (\approx noventa e um)"
- c.

quatre-vingt-onze	[Francês (românica)]
quatro-vinte-onze	

"quatro vintenas e onze (\approx noventa e um)"

2 São de interesse deste artigo apenas os numerais ditos cardinais (e.g., *três*), ainda que o espírito de análise aqui defendido possa ser extrapolado para os numerais ordinais (e.g., *terceiro*), multiplicativos (e.g., *triplo*), fracionários (e.g., *terço*) etc.

Motivado por esse papel estruturante das bases, Hurford (1975), um dos pioneiros na teoria gramatical de numerais linguísticos, reservou a elas uma projeção especial na sintaxe: M. Em seu modelo, regras de estrutura frasal (2) geram estruturas profundas (3a), a partir das quais são derivadas representações fonológicas (3b) e semânticas (3c). Articulado ao projeto da Teoria Padrão (Chomsky, 1965), Hurford (1975) tem interesse não só em capturar generalizações descritivas de sistemas numerais, mas em investigar os universais cognitivos que fundamentam a capacidade humana para representar números por meio da linguagem.

12. Regras de estrutura frasal dos numerais do inglês

NUMBER → {1, PHRASE} (NUMBER)

PHRASE → NUMBER M

M → {10, NUMBER M}

- a. Estrutura profunda: $[_{NUMBER} [_{PHRASE} [_{NUMBER} 1 [_{NUMBER} 1]] [_M 10]] [_{NUMBER} 1]]$
- b. Forma fonológica: /twentiwʌn/
- c. Forma semântica: $((1 + 1) \times 10) + 1 = 21$

A motivação inaugural de Hurford (1975) foi atualizada e reinterpretada à luz de diversos desdobramentos teóricos: a Teoria de Regência e Ligação (Ionin; Matushansky, 2006, 2018), a Gramática de Construções (Booij, 2010), o Minimalismo (Di Sciullo, 2015, 2022), a Nanossintaxe (Wągiel; Caha, 2020; Wągiel, 2023), entre outros. No âmbito desses desenvolvimentos, parte importante do debate se concentra em determinar a configuração sintática de construções [*numeral nome*]: se o numeral é um núcleo que toma o NP quantificado como complemento (4a), ou se é especificador de um núcleo na projeção funcional do NP quantificado (4b).

4. Debate sobre a estrutura de construções [*numeral nome*]

- a. Hipótese da estrutura em cascata: $[_{XP} numeral [_{NP} nome]]$
- b. Hipótese do especificador: $[_{YP} numeral [_Y Y [_{NP} nome]]]$

A respeito da maneira com que o embate entre essas propostas tem se desenvolvido, dois são os principais expedientes. Por um lado, evidências sintáticas, semânticas e morfológicas são acumuladas para cada uma das hipóteses, de modo mutuamente exclusivo (ver Corver e Tatsumi, 2023, para uma síntese dos argumentos). Por outro, há a tentativa de reduzir a hipótese adversária como sendo uma mera consequência da hipótese defendida, sendo para isso propostas condições adicionais que derivam propriedades superficiais da estrutura de especificador a partir de uma estrutura fundamental em cascata (cf. Ionin; Matushansky, 2018), ou vice-versa (cf. Witkoś; Dziubała-Szrejbrowska, 2018).

De maneira a conciliar as duas perspectivas em conflito, um terceiro caminho tem sido avançado por parte da literatura recente (Danon, 2012; Klockmann, 2021; Tatsumi, 2021; Corver; Tatsumi, 2023). Essa posição da literatura identifica, apesar dos pontos de disputa, um importante consenso: a hipótese de que o comportamento formal de numerais é universal e de que há uma única configuração estrutural para as construções [*numeral nome*]. Essa assunção, que ecoa Hurford (1975, p. 4), é aqui formulada nos termos de (5). A intuição relevante é a de que é somente sob o pressuposto de que há uma única estrutura, uniforme e universal, que se coloca o debate entre diferentes implementações para uma (suposta) mesma construção.

3. **Hipótese de Uniformidade** A Faculdade da Linguagem opera de maneira uniforme no que diz respeito (i) aos numerais de uma língua e (ii) aos numerais das diferentes línguas naturais humanas.

Nesse sentido, o caminho sugerido pelos autores dessa tradição conciliadora consiste em, de um lado, admitir o conjunto de evidências acumulado de maneira mutuamente exclusiva para cada uma das hipóteses em (4) e, de outro, acomodar a convivência entre as análises adversárias a partir da renúncia da Hipótese de Uniformidade (5). Assim, as análises são diferentes justamente porque os fenômenos são diferentes, admitindo que a Faculdade da Linguagem acomoda tal diversidade. Sob essa perspectiva, perguntas importantes dizem respeito à determinação do repertório translinguístico de estruturas [*numeral nome*] possíveis (Tatsumi, 2021), à maneira com que diferentes configurações podem conviver em uma mesma língua (Danon, 2012; Corver; Tatsumi, 2023) e à possibilidade de incorporação ou perda de certos tipos de construção ao longo da história de uma língua (Klockmann, 2021).

Este artigo avança essa tradição emergente em dois sentidos relevantes. Em primeiro lugar, a partir de um recorte voltado a bases numerais, argumento em favor de uma versão mais forte da negação da Hipótese de Uniformidade (5): não só não é universal a maneira com que numerais em geral são manipulados, mas também a caracterização formal de bases numerais, em particular, é heterogênea (ver Klockmann, 2021, para uma posição semelhante, defendida a partir do inglês e do polonês). Em segundo lugar, assumo neste artigo uma visão centrada no português brasileiro (doravante, PB), com o objetivo de assim contribuir à discussão em torno da Hipótese de Uniformidade (5) com dados de uma língua pouco representada nessa tradição.³

Para tanto, este trabalho explora contrastes entre as bases do PB na interface morfologia-sintaxe, avaliando as maneiras com que elas divergem do ponto de vista sintático, categorial e morfológico. A imagem que emerge dessa comparação tensiona a Hipótese

3 Para uma análise neo-hurfordeana do sistema numeral do PB, ver Vignado (2017, 2018, 2019, 2020).

de Uniformidade (5) e sugere sua abdicação em favor de uma caracterização heterogênea das bases do PB.

Contrastes entre as bases numerais do PB

Contrastes sintáticos

Um ponto importante do debate sobre a (não) uniformidade da representação linguística das bases numerais diz respeito à sua distribuição sintática. Nesse sentido, o sistema cardinal do PB, em sua heterogeneidade, se mostra particularmente informativo. Isso porque, contrariamente ao que prevê a Hipótese de Uniformidade (5), testes de constituição e de distribuição complementar com quantificadores sugerem que há pelo menos dois padrões de comportamento sintático para bases.

A respeito da distribuição de bases em construções de forma [*unidade base nome*], uma primeira questão a ser levantada é se a base forma com o nome uma projeção máxima independente da unidade, isto é, [*unidade* [_{XP} *base nome*]]. Para tanto, o teste da elipse, tradicionalmente assumido como tendo por alvo necessário projeções máximas (cf. Merchant, 1999), é especialmente conveniente. Aplicando-o em construções contrastivas como (6), a diferença entre a agramaticalidade da elipse em (6a–b) e sua gramaticalidade em (6c–d) sugere que apenas *mil* e os milhares altos (i.e., *milhão*, *bilhão* etc.) formam projeções máximas com os nomes que quantificam. Sendo esse o caso, as bases do PB diferem minimamente no constituinte que ocupam.

6. O preço foi...
- a. *... de oito[~~enta~~-reais] para noventa reais.
 - b. *... de oito[~~entos~~-reais] para novecentos reais.
 - c. ... de oito [mil-reais] para nove mil reais.
 - d. ... de oito [milhões-de-reais] para nove milhões de reais.

A pergunta que naturalmente se segue é se as bases *-enta* e *-entos*, não formando com o nome um constituinte à parte, o fazem com as unidades cardinais, isto é, [[_{ZP} *unidade base*] *nome*]. Para avaliar essa hipótese, consideremos quantificadores como *muitos*, *vários* e *tantos*, que, como sugere (7), se distribuem de maneira complementar com as unidades cardinais. Nesses termos, se esses quantificadores se distribuem como as unidades cardinais e, por hipótese, se essas unidades ocupam os mesmos constituintes que *-enta* e *-entos* (i.e., ZP), então devemos observar a distribuição complementar desses quantificadores também com relação a essas bases.

7. a. {duas/muitas/várias/tantas/...} pessoas
 b. * {muitas/várias/tantas/...} duas pessoas
 c. * duas {muitas/várias/tantas/...} pessoas

Confirmando essa previsão, os dados em (8) sugerem que as bases de fato diferem com relação à coocorrência com esses quantificadores. Por um lado, a coocorrência não é licenciada para *-enta* (8a) ou *-centos* (8b), o que aproxima sua distribuição com a das unidades cardinais (7b). Por outro lado, os quantificadores coocorrem com *mil* (8c) e com os milhares altos (8d), reforçando a hipótese de que essas bases formam um constituinte independente das unidades cardinais – e, por extensão, dos quantificadores sob investigação.⁴

8. a. *{muitos/vários/tantos/...} enta linguistas
 b. *{muitos/vários/tantos/...} centos linguistas
 c. {muitos/vários/tantos/...} mil linguistas
 d. {muitos/vários/tantos/...} milhões de linguistas

Outro quantificador relevante para a discussão é o pronome QU- *quantos*. Uma propriedade de interesse é que, quando substituindo unidades cardinais, o pronome admite a realização do restante do NP quantificado, o que inclui desde o núcleo do NP (9) a APs (10) e CPs (11) adjuntos. Nesse sentido, a substituição por *quantos* é um diagnóstico para a posição das unidades cardinais: apenas elementos externos a essa posição podem ser realizados junto a *quantos*.

9. a. Ela tem [três] [_{NP} livros].
 b. Ela tem [quantos] [_{NP} livros]?
10. a. Ela tem [três] livros [_{AP} raros].
 b. Ela tem [quantos] livros [_{AP} raros]?
11. a. Ela tem [três] livros [_{CP} que foram importados da Bulgária].
 b. Ela tem [quantos] livros [_{CP} que foram importados da Bulgária]?

Dessa forma, a hipótese de que *-enta* e *-centos* formam um constituinte independente do nome com as unidades cardinais prevê que essas bases não podem coocorrer com

4 Por completude, cabe notar que a ordem [*base quantificador nome*], à semelhança de (7c), é igualmente rejeitada por todas as bases (i–iv).

- (i) *enta {muitos/vários/tantos/...} linguistas
- (ii) *centos {muitos/vários/tantos/...} linguistas
- (iii) *mil {muitos/vários/tantos/...} linguistas
- (iv) *milhões (de) {muitos/vários/tantos/...} linguistas

quantos à maneira de (9–11). Em termos da estrutura genérica $[[_{ZP} \text{ unidade base}] \text{ nome}]$, isso é dizer, da parte do pronome, que *quantos* substitui ZP e, da parte das bases, que *-enta* e *-centos*, mas não *mil* e os milhares altos, ocupam ZP.

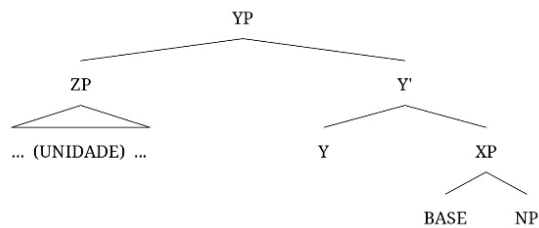
Consideremos, portanto, os diálogos em (12). O padrão de (a)gramaticalidade nas perguntas de B corresponde às predições derivadas estruturalmente: *-enta* (12a) e *-centos* (12b), ocupando ZP, não coocorrem com *quantos*, ao contrário de *mil* (12c) e dos milhares altos (12d), que, externos a ZP, são licenciados em contexto do pronome.

12. a. A: O tonto do meu primo pagou oitenta reais naquele lápis.
B: *Quantos enta reais?
A: *Oito[~~enta~~ reais].
- b. A: O tonto do meu primo pagou oitocentos reais naquela camiseta.
B: *Quantos centos reais?
A: *Oito[~~centos~~ reais].
- c. A: O tonto do meu primo pagou oito mil pesos naquela empanada.
B: Quantos mil pesos?
A: Oito [mil pesos].
- d. A: O tonto do meu primo pagou oito milhões de pesos naquele vinho.
B: Quantos milhões de pesos?
A: Oito [milhões de pesos].

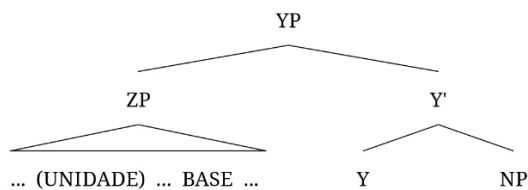
Ainda a partir desses diálogos, notemos que as respostas de A evocam o padrão dos contrastes com relação à elipse em (6), o que sugere uma explicação unificada para os dados levantados nesta subseção. Especificamente, ser alvo da substituição por *quantos* (i.e., integrar ZP) implica não ser alvo de elipse (i.e., integrar XP), e vice-versa. A interpretação sintática dessa oposição é que as bases do PB ocupam dois *loci* sintáticos distintos.

Em suma, os dados aqui levantados sugerem que há dois padrões para o comportamento sintático das bases do PB: bases como *mil* e os milhares altos são alvos de elipse conjunta com o nome e coocorrem com *muitos/vários/tantos/quantos/...*, ao passo que bases como *-enta* e *-centos* não são elididas junto ao nome e se distribuem de modo complementar com *muitos/vários/tantos/quantos/...* Esses contrastes podem ser acomodados renunciando à Hipótese de Uniformidade (5) e admitindo duas representações sintáticas distintas para as construções $[\text{unidade base nome}]$ do PB (13).

13. Dois *loci* sintáticos para as bases do PB
- a. Em constituição com o nome: *mil* e os milhares altos



b. Em constituição com as unidades: *-enta* e *-entos*



Essas representações capturam a assimetria das bases com relação à sua distribuição na sintaxe, mas são indeterminadas no que diz respeito aos rótulos das projeções X, Y e Z. Para a projeção Y, há um debate paralelo aos interesses deste artigo concentrado em decidir Y como sendo Num, Card ou Q (ver Ionin e Matushansky, 2018, para uma revisão das hipóteses). O mesmo problema se estende à caracterização de X e Z. Para estes, todavia, há um maior consenso de sua identificação com categorias lexicais, o que é evidenciado pelo grande aporte de evidências que indicam que numerais têm o comportamento de partes do discurso típicas, como nomes e adjetivos (cf. Corver; Tatsumi, 2023).

Assumindo, portanto, que X e Z projetam categorias morfossintáticas como N e A, a próxima seção reflete sobre a uniformidade das bases no que se refere ao seu comportamento categorial.

Contrastes categoriais

Começamos comparando o comportamento categorial de *mil* e dos milhares altos. Por um lado, os numerais *milhão/bilhão/...* são caracteristicamente nominais (Neves, 2011; Vignado, 2019; Valls Yoshida, 2024). Como nomes, requerem complemento introduzido por *de* (14a), em construções pseudopartitivas (cf. Tatsumi, 2021) e controlam concordância em gênero (15a) e em número gramatical (16a) independentemente dos traços do núcleo do NP quantificado. Por outro lado, *mil* não apresenta nenhum dos indicativos de comportamento categorial que identificam os milhares altos como nomes (14–16b).

14. a. milhões *(de) habitantes
b. mil *(de) habitantes
15. a. um_[USG] milhão_[SG] de pessoas_[PL]
b. *um_[USG] mil_[SG] pessoas_[PL]
16. a. dois_[UMASC] milhões_[MASC] de pessoas_[FEM]
b. *dois_[UMASC] mil_[MASC] pessoas_[FEM]

Não se comportando como nome, é possível questionar se *mil* pode ser caracterizado como adjetivo. Do ponto de vista morfológico, *mil* não realiza de maneira transparente traços de gênero ou número gramatical, o que torna opaco seu perfil morfossintático. Do ponto de vista distribucional, todavia, *mil* admite, mantendo a leitura cardinal,⁵ ambas as posições pré-nominal e pós-nominal (17), sendo esta última canonicamente reservada a adjetivos predicativos.

17. a. mil [_{NP} tradições e glórias]
[_{NP} tradições e glórias] mil
b. mil [_{NP} problemas]
[_{NP} problemas] mil

A comparação entre *mil* e os milhares altos indica que, ainda que ambos estejam sob uma mesma projeção máxima com o NP quantificado, essas bases não projetam, elas próprias, uma mesma categoria morfossintática: os milhares altos são nomes e *mil* é um adjetivo (i.e., [_{NP} base_{N/A} NP]). A essa altura, é interessante observar que se a hipótese proposta defende que as diferenças de (14–16) são essencialmente categoriais, é de se esperar que uma contraparte nominal de *mil* licencie sua ocorrência nos mesmos contextos dos milhares altos. Com efeito, o comportamento sugestivamente nominal de *milhar* (18–20), derivado de *mil*, corrobora essa hipótese.⁶

5 Faz-se necessário distinguir as leituras cardinais em (17), de interesse para o argumento, daquela que Wiese (2003) identifica como a leitura nominal de cardinais (e.g., *camisa nove*, que se lê de maneira distinta de *nove camisas*).

6 Caso análogo pode ser elaborado para *dez*, *doze* e *cem*, que se distribuem como *mil* (i–iiib), mas cujas contrapartes *dezena*, *dúzia* e *centena*, como os milhares altos, se distribuem como nomes (i–iiia). Ao todo, esses dados reforçam que é uma diferença entre nomes e adjetivos aquela que está por trás desses contrastes.

- (i) a. {dezenas/dúzias/centenas} *(de) habitantes
b. {dez/doze/cem} *(de) habitantes
- (ii) a. uma {dezena/dúzia/centena} de habitantes
b. *uma {dez/doze/cem} habitantes
- (iii) a. duas {dezenas/dúzias/centenas} de habitantes
b. *duas {dez/doze/cem} habitantes

18. milhares *(de) habitantes
19. um_[USG] milhar_[SG] de pessoas_[PL]
20. dois_[UMASC] milhares_[MASC] de pessoas_[FEM]

Em oposição aos milhares altos e à semelhança de *mil*, os numerais derivados a partir de *-centos* não manifestam qualquer propriedade sugestivamente nominal, a começar pela rejeição da construção pseudopartitiva (21). No que se refere à concordância em gênero, porém, esses numerais realizam abertamente marcas de concordância com o nome (22), em contraste com *mil*. Também em contraste com o comportamento de *mil* em (17), as centenas rejeitam a posição pós-nominal (23a), assim como os outros quantificadores com os quais se distribuem de modo complementar (23b). À luz dessas observações, os numerais formados por *-centos* e *mil* são sugestivamente ambos adjetivos, mas não à mesma maneira, algo que parece refletir o caso geral da diversidade distribucional dos adjetivos do PB (Müller; Negrão; Nunes-Pemberton, 2002; Quadros Gomes; Gonçalves Sudré, 2021; *i.a.*).⁷

21. oitocentas *(de) pessoas
22. a. oitocentas_[UFEM] pessoas_[FEM]
b. oitocentos_[UMASC] habitantes_[MASC]
23. a. *pessoas oitocentas
b. *pessoas {muitas/várias/tantas/...}

Os contrastes levantados até então nos permitem avaliar, finalmente, o comportamento categorial de *-enta*. Especificamente, todos os indícios categoriais aqui considerados testam negativo para *-enta*: os numerais derivados dessa base não tomam complementos nominais introduzidos por *de* (24), não admitem modificação cardinal para que se teste determinação de gênero e número gramatical inerentes (25), não realizam marcas de concordância com o nome (26) e não ocorrem em posição pós-nominal (27). A resistência de *-enta* a revelar pistas de sua natureza categorial é um desafio em aberto (cf. Valls Yoshida, 2024) e, para os efeitos deste artigo, contrasta essa com todas as demais bases, para as quais ao menos alguma evidência positiva está disponível.

24. oitenta *(de) habitantes
25. *dois/duas oitenta habitantes

⁷ Como bem nos aponta um(a) parecerista anônimo(a), a identificação unificada de *-centos* e *mil* como bases adjetivais mascara, por um lado, as diferenças dessas bases com relação a adjetivos prototípicos e, por outro, os contrastes entre elas enquanto numerais adjetivais (cf. admissão da posição pós-nominal para *mil*, mas não para formas derivadas de *-centos*). Um desdobramento previsto da proposta aqui apresentada é o refinamento da sintaxe das bases adjetivais, tendo em vista ainda seus efeitos semânticos para a interpretação de sintagmas nominais quantificados por numerais (cf. nota de rodapé 9).

26. oitenta/*oitento/*oitentas/*oitentos {pessoas/habitantes}
27. *pessoas oitenta

No todo, os dados mobilizados nesta subseção atribuem às bases uma dimensão de flexibilidade a mais, adicionando à diversidade sintática uma camada de heterogeneidade categorial: os milhares altos são nomes; *mil* e os derivados de *-centos* são adjetivos, ainda que não à mesma maneira; e os derivados de *-enta* são categorialmente evasivos e nesse sentido contrastam com todas as demais bases.

Tomando o caminho que parte da estrutura frasal e avança ao domínio da palavra, a próxima subseção se dedica a explorar a composição morfológica das bases do PB.

Contrastes morfológicos

Parte das assimetrias exploradas na subseção *Contrastes sintáticos* naturalmente se relaciona à diferença das bases em sua autonomia morfológica com relação às unidades. Bases que são morfemas livres, como *mil* e os milhares altos, são licenciados em contextos em que *-enta* e *-centos*, sendo morfemas presos, não o são. É importante notar, mesmo assim, que as bases presas não são igualmente presas e contrastes entre elas revelam estruturas morfológicas particularmente distintas.

Uma primeira aproximação a essa diferença pode ser feita em termos de constituintes prosódicos (cf. Nespor; Vogel, 1986): as dezenas são formadas por uma única palavra fonológica (28a), ao passo que as centenas, por duas (28b). Essa distinção é um primeiro indicativo de que *-enta* e *-centos* são objetos morfológicos de tamanhos distintos, e que não se combinam às unidades cardinais segundo uma mesma estrutura.

28. [sete]_ω
a. [setenta]_ω
b. [sete]_ω[centos]_ω

Outro indicativo desse contraste diz respeito ao tipo de estrutura morfofonológica com que se combinam. A esse respeito, as unidades cardinais do PB são radicais complexos formados a partir da combinação do expoente de uma raiz a uma vogal temática (cf. Valls Yoshida, 2024). Crucialmente, as bases sob investigação são sensíveis a essa diferença: *-enta* se combina diretamente com os expoentes de raízes cardinais (29a), ao passo que *-centos* exige radicais marcados tematicamente (29b). O argumento para o cardinal *oit-o* em (29), de terminação *-o* (ou de Classe I, nos termos de Alcântara, 2010), pode ser extrapolado aos cardinais *quatr-o* (também de Classe I) e a *set-e* e *nov-e* (de Classe III).

29. [radical [raiz oit]-o]
a. [raiz oit]-enta
b. [radical oito]-centos

Também são contrastivos processos fonológicos sensíveis à localidade com a raiz, como a neutralização vocálica,⁸ condicionada dialetalmente no PB. O exemplo (30) ilustra como o expoente *n/O/ve* admite variação na aplicação da neutralização vocálica em contexto de *-enta* (30a), mas não em contexto de *-centos* (30b). A impossibilidade categórica de *n[o]vecentos* se alinha às demais assimetrias morfológicas: *-centos* se combina a uma porção maior de estrutura – i.e., um radical (29) que forma uma palavra fonológica completa (28) – e por isso não guarda com o expoente de *nove* relação local o suficiente para engatilhar a neutralização vocálica. O mesmo parece estar por trás do padrão de variação em *s[ɛ/*e]te ~ s[ɛ/e]tenta ~ s[ɛ/*e]tecentos*.

30. *n[o]ve, *n[o]ve*
a. *n[o]venta, n[o]venta*
b. *n[o]vecentos, *n[o]vecentos*

Em suma, as evidências podem ser capturadas assumindo que essas bases engajam em dois processos de formação de palavras distintos: *-enta* é um sufixo derivacional e *-centos* forma com as unidades cardinais compostos morfológicos. Para os propósitos deste artigo, isso significa que tampouco sob a lente da morfologia as bases podem ser identificadas de maneira homogênea, advogando contra a Hipótese de Uniformidade (5).

É importante frisar que a oposição das bases em morfemas livres e morfemas presos não necessariamente se reduz a uma diferença em complexidade morfológica: afinal, os milhares altos, morfemas livres, são morfológicamente complexos. Sob uma perspectiva diacrônica, o numeral *milhão* é uma inovação românica a partir da maior das bases latinas *mille* ‘mil’. A esse respeito, o cognato histórico para *milhão* no PB foi primeiro atestado no francês antigo, em 1270 (Bloch; von Wartburg, 1962 [1932]), e é por sua vez um possível empréstimo do italiano antigo, analisado composicionalmente como *mili-one*, lit. ‘mil-AUG’, ‘milhão’ (Dressler; Merlini Barbaresi, 1994). A derivação de bases altas a partir de bases menores é um expediente regular translinguisticamente (Comrie, 2022).

Uma vez fixado, o numeral *milhão* foi reanalisado (Hurford, 1975; Bauer, 2021) e o morfema *-ilhão* é ativo na derivação da sequência completa dos milhares altos do PB (31). A esse respeito, é especialmente interessante notar que *-ilhão* se combina a expoentes de uma

8 Para uma perspectiva morfofonológica deste fenômeno, ver Lee (1995), Schwindt (2013), Ulrich e Schwindt (2018), entre outros.

variedade de tipos de numerais (32), o que complexifica a composição morfológica das bases altas do PB.⁹

31. milhão, bilhão, trilhão, quatrilhão, quintilhão, sextilhão, septilhão, octilhão, nonilhão, decilhão
32. a. Formas greco-latinas
bi- (e.g., *bissexual*), *tri-* (e.g., *tricampeão*), *oct-* (e.g., *octógono*), *dec-* (e.g., *década*)
- b. Formas cardinais
quatr- (e.g., *quatro*)
- c. Formas ordinais
quint- (e.g., *quinto*), *sext-* (e.g., *sexto*), *sept-* (e.g., *sétimo*), *non-* (e.g., *nono*)

Retomando o contraste entre as bases do PB, as evidências morfológicas aqui consideradas sugerem que elas se realizam em termos de expedientes de formação de numerais distintos, listados explicitamente em (33).

33. Expedientes derivacionais a partir de bases no PB
- a. Sufixação por *-enta*
trinta, quarenta, cinquenta, sessenta, setenta, oitenta, noventa
- b. Composição com *-centos*
duzentos, duzentas, trezentos, trezentas, quatrocentos, quatrocentas, seiscientos, seiscentas, setecentos, setecentas, oitocentos, oitocentas, novecentos, novecentas
- c. Derivação de *-ilhão*
milhão, milhões, bilhão, bilhões, trilhão, trilhões, quatrilhão, quatrilhões, quintilhão, quintilhões, sextilhão, sextilhões, septilhão, septilhões, octilhão, octilhões, nonilhão, nonilhões, decilhão, decilhões

Uma vez identificados, esses diferentes processos de formação de palavras podem ser, ainda, contrastados no que diz respeito à sua produtividade morfológica (Quadro 1), um fenômeno complexo, que pode ser abordado sob diferentes vieses, formais e quantitativos (cf. Bauer, 2001). No que se refere à regularidade desses processos (Aronoff, 1976; Basílio, 1977), as bases *-enta*, seguida de *-centos*, são as que mais acumulam idiosincrasias na forma da unidade cardinal com que se combinam. Quanto à formação de neologismos, apenas *-centos* e *-ilhão* derivam numerais inovadores, cuja interpretação é vaga quanto ao seu valor numérico. Ainda, sob uma perspectiva quantitativa para a produtividade (Corbin, 1987; Yang, 2022), as bases *-enta* e *-centos* empatam no número de cardinais canônicos (i.e., excetuando-se os inovadores) com que se combinam, atrás do número

9 A diversidade do paradigma (31) no que se refere ao tipo de morfe com que *-ilhão* se combina não é uma exclusividade dos milhares altos, e se atesta no PB também na composição morfológica dos numerais ordinais, multiplicativos e fracionários. Ver Greenberg (2000) para uma discussão completa.

maior de *types* a que se aplica *-ilhão*. Quanto ao número de *tokens*, a média da frequência lexical no *corpus* LexPorBR (Estivalet; Meunier, 2017) sugere ainda uma nova gradação de produtividade das bases, em que *-ilhão* é o mais produtivo, uma ordem de grandeza à frente de *-enta*, seguido finalmente por *-centos*.

Quadro 1. Contrastes de produtividade entre as bases do PB

Base	Regularidade	Numerais vagos	Frequência (<i>types</i>)	Frequência média (<i>tokens</i>)
<i>-enta</i>	Parcial: exceções em <i>trinta, quarenta e sessenta</i> .	Não.	> 7	111,53
<i>-centos</i>	Parcial: exceções em <i>duzentos e trezentos</i> .	Sim: <i>trocentos, bucentos, ...</i>	> 7	15,466
<i>-ilhão</i>	Parcial: exceção em <i>milhão</i> .	Sim: <i>gazilhão, zilhão, ...</i>	> 10	2532,30

Fonte: Elaboração própria

Ao todo, as evidências mobilizadas nesta subseção evidenciam a riqueza morfológica do PB na representação de suas bases cardinais. As bases *-enta* e *-centos* são morfemas presos e se distinguem na medida em que, na relação com as unidades cardinais, um é um sufixo derivacional e o outro forma com elas compostos morfológicos. Bases que são morfemas livres podem ser morfológicamente simples, como *mil*, mas não necessariamente: os milhares altos são derivados a partir do morfema preso *-ilhão*. Tomados ao todo, os processos de formação de cardinais do PB são diversos morfológicamente e contrastam ainda em produtividade, o que se observa na (ir)regularidade morfofonológica, na (im) possibilidade de derivar numerais inovadores e nas diferenças em frequência de *types* e *tokens*.

Uma vez mais, as bases cardinais do PB resistem a uma análise unificada. A diversidade formal com que se revelam põe sérias dificuldades para uma análise comprometida com sua substância linguística, mais que sua identidade no papel que exercem na estruturação do sistema cardinal do PB.

Considerações finais

A hipótese de que a Faculdade da Linguagem é diretamente sensível à estrutura matemática dos sistemas numerais abriu uma extensa e produtiva tradição de estudos linguísticos. Para a questão das bases numerais, em particular, isso permitiu à tipologia, por um lado, abstrair as diferenças entre as línguas e generalizar a regularidade da expressão de sistemas numerais em torno de bases (Greenberg, 1978, 2000; Comrie, 2013, 2022) e aos estudos gramaticais, por outro, refletir sobre o lugar privilegiado

que bases ocupam na derivação de numerais complexos (Hurford, 1975; Booij, 2010; Rothstein, 2013).

Se a noção de base, assim concebida, é produtiva a um nível apropriado de abstração da substância linguística, uma investigação voltada às suas propriedades genuinamente gramaticais revela uma caracterização heterogênea das bases numerais. Acrescentando a uma tradição não-uniformista dos numerais linguísticos (Danon, 2012; Klockmann, 2021; Tatsumi, 2021; Corver; Tatsumi, 2023), este artigo traz à discussão a empiria do PB e com isso revela a riqueza gramatical dos expedientes que representam as bases do sistema cardinal dessa língua. Especificamente, os argumentos delineados neste artigo apontam para dois *loci* que bases podem ocupar na sintaxe: ou em constituição com o nome quantificado, ou com as unidades cardinais que as modificam. Em cada um desses *loci*, as bases do PB variam em categoria morfossintática (*viz.*, nomes e adjetivos) e nas propriedades típicas de cada categoria que manifestam. Não só quanto à morfologia flexional as bases divergem: no que se refere à sua diversidade derivacional, as bases não parecem ser codificadas por um expediente morfológico unificado, e se revelam em sufixos derivacionais e em compostos, com graus diferentes de produtividade.

Ao longo da discussão aqui desenvolvida, o aprofundamento da caracterização gramatical das bases numerais do PB é reconhecidamente limitado à medida em que é suficiente para contrastar as bases nos níveis sintático, categorial e morfológico. Tal posição é fundamentada na decisão de concentrar aqui os esforços na empreitada de desafiar a Hipótese de Uniformidade (5). Com sorte, tendo sido esse objetivo satisfatoriamente atingido, são esperados na continuidade deste trabalho desenvolvimentos de uma caracterização pormenorizada do comportamento gramatical das bases do PB para que, com isso, se possa apreciar seu lugar na Gramática com maior precisão.

Agradecimentos

Agradeço à prof. Dr^a Ana Paula Scher pelos debates precisos que acompanharam o desenvolvimento deste artigo, à prof. Dr^a Indaiá de Santana Bassani pelas ricas sugestões que inspiraram a estrutura e a redação do trabalho e à Marcela Nunes Costa pelo olhar sempre atento e as críticas sempre tão construtivas que em muito contribuíram à realização deste artigo. Todos os equívocos restantes são de minha responsabilidade. Pelo financiamento de pesquisa, agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – processo #2023/17997-0).

Referências

ALCÂNTARA, C. As classes formais do português brasileiro. *Letras de Hoje*, v. 45, n. 1, p. 5-15, 2010.

ARONOFF, M. *Word Formation in generative grammar*. Cambridge: MIT Press, 1976.

BASÍLIO, M. *Aspects of the structure of lexicon: evidence from Portuguese*. 1977. Tese (Doutorado em Linguística) – University of Texas at Austin, Austin, 1977.

BAUER, B. Counting systems. In: LEDGEWAY, A.; MAIDEN, M. (ed.). *The Cambridge Handbook of Romance Linguistics*. Cambridge: Cambridge University Press, 2022.

BAUER, B. Formation of Numerals in the Romance Languages. In: *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press, 2021.

BAUER, L. *Morphological productivity*. Cambridge: Cambridge University Press, 2001.

BLOCH, O.; VON WARTBURG, W. *Dictionnaire étymologique de la langue française*. 11. ed. Paris: Presses Universitaires de France, 1962. (Originalmente publicado em 1932.)

BOOIJ, G. Constructions and lexical units: An analysis of Dutch numerals. *New impulses in word formation*, p. 81-100, 2010.

BRESNAN, J.; MCHOMBO, S. The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*, v. 13, n. 2, p. 181-254, 1995.

CHOMSKY, N. *Aspects of the Theory of Syntax*. Cambridge: MIT Press, 1965.

COMRIE, B. Numeral bases. In: DRYER, M.; HASPELMATH, M. (ed.). *The world atlas of language structures online*. 2013.

COMRIE, B. The Arithmetic of Natural Language: Toward a typology of numeral systems. *Macrolinguistics*, v. 10, n. 1, p. 1-35, 2022.

CORBIN, D. *Morphologie dérivationelle et structuration du lexique*. Tübingen: Niemeyer, 1987.

CORVER, N.; TATSUMI, Y. Number names: Internal structure and morphological marking. In: ACKEMA, P.; BENDJABALLAH, S.; BONET, E.; FÁBREGAS, A. (ed.). *The Wiley Blackwell Companion to Morphology*. Wiley-Blackwell, 2023.

DANON, G. Two structures for numeral-noun constructions. *Lingua*, v. 122, n. 12, p. 1282-1307, 2012.

DI SCIULLO, A. Complex Cardinal Numerals and the Strong Minimalist Thesis. *Philosophies*, v. 7, n. 4, p. 81, 2022.

DI SCIULLO, A. On the domain specificity of the human language faculty and the effects of principles of computational efficiency: contrasting language and mathematics. *Revista Lingüística*, v. 11, n. 1, 2015.

DRESSLER, W.; MERLINI BARBARESI, L. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin: Mouton de Gruyter, 1994.

ESTIVALET, G.; MEUNIER, F. Corpus psicolinguístico Léxico do Português Brasileiro. *Soletras*, n. 33, p. 212-229, 2017.

GREENBERG, J. Generalizations about numeral systems. In: GREENBERG, J.; FERGUSON, C.; MORAVCSIK, E. (ed.). *Universals of human language*, v. 3, p. 249-295, 1978.

GREENBERG, J. Numeral. In: BOOIJ, G.; LEHMANN, C.; MUGDAN, J. (ed.). *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung*, v. 1, p. 770-783, 2000.

HAMMARSTRÖM, H.; FORKEL, R.; HASPELMATH, M.; BANK, S. *Glottolog 5.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2024.

HURFORD, J. *The Linguistic Theory of Numerals*. Cambridge: Cambridge University Press, 1975.

IONIN, T.; MATUSHANSKY, O. *Cardinals: The syntax and semantics of cardinal containing expressions*. Cambridge: MIT Press, 2018.

IONIN, T.; MATUSHANSKY, O. The composition of complex cardinals. *Journal of semantics*, v. 23, n. 4, p. 315-360, 2006.

KAISER, S.; ICHIKAWA, Y.; KOBAYASHI, N.; YAMAMOTO, H. *Japanese: A Comprehensive Grammar*. 2. ed. Nova Iorque: Routledge, 2013.

KLOCKMANN, H. Deconstructing base numerals: English and Polish 10, 100, and 1000. In: DOČEKAL, M.; WAĞIEL, M. (ed.). *Formal approaches to number in Slavic and beyond*. Berlin: Language Science, 2021. p. 297-322.

LEE, S. H. *Morfologia e fonologia lexical do português do Brasil*. 1995. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, 1995.

MERCHANT, J. The syntax of silence: Sluicing, islands, and identity in ellipsis. 1999. Tese (Doutorado em Linguística) – University of California, Santa Cruz, 1999.

MÜLLER, A. L.; NEGRÃO, E.; NUNES-PEMBERTON, G. Adjetivos no português do Brasil: predicados, argumentos ou quantificadores? In: ABAURRE, M. B.; RODRIGUES, A. C. S. (org.). *Gramática do português falado: novos estudos descritivos*. Campinas: Editora Unicamp, 2002.

NESPOR, M.; VOGEL, I. *Prosodic Phonology*. Dordrecht: Foris, 1986.

NEVES, M. H. *Gramática de usos do português*. São Paulo: Editora Unesp, 2011.

QUADROS GOMES, A.; GONÇALVES SUDRÉ, T. Entre a modificação e a quantificação: adjetivos como determinantes plurais em português brasileiro (PB). *Domínios de Linguagem*, v. 15, n. 1, 2021.

ROTHSTEIN, S. A Fregean semantics for number words. In: *Proceedings of the 19th Amsterdam Colloquium*. 2013.

SCHWINDT, L. C. Neutralização da vogal pretônica e formação de palavras em português brasileiro. *Organon*, v. 28, n. 54, p. 137-154, 2013.

TATSUMI, Y. *Linguistic Realization of Measuring and Counting in the Nominal Domain: A Cross linguistic Study of Syntactic and Semantic Variations*. 2021. Tese (Doutorado em Linguística) – University of Connecticut, 2021.

ULRICH, C.; SCHWINDT, L. C. O status morfoprosódico dos sufixos inho/zinho, mente e íssimo no Português Brasileiro. *D.E.L.T.A.*, v. 34, p. 769-788, 2018.

VALLS YOSHIDA, F. Uma abordagem morfológica às dezenas cardinais do português brasileiro: -enta como um sufixo derivacional. *Veredas – Revista de Estudos Linguísticos*, v. 28, n. 2, 2024.

VIGNADO, J. *A interface sintática e semântica na análise dos sistemas numerais do karitiana e do kamayurá*. 2019. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, 2019.

VIGNADO, J. A sintaxe e a semântica de expressões numéricas do karitiana. *Revista Letras*, n. 96, p. 546-565, 2017.

VIGNADO, J. Análise formal de numerais: perspectivas para uma análise minimalista. *Caderno de squibs: temas em estudos formais da linguagem*, v. 4, p. 88-96, 2018.

VIGNADO, J. Evidências morfossintáticas do português, do inglês e do karitiana (tupi) para uma semântica composicional de numerais adequada à interface sintática. *Cadernos de Linguística*, v. 1, p. 1-13, 2020.

WAĞIEL, M. The structure and meaning of label numerals. *Journal of Slavic Linguistics*, v. 31, n. FASL 30 issue, p. 1-20, 2023.

WAĞIEL, M.; CAHA, P. Universal semantic features and the typology of cardinal numerals. *Catalan journal of linguistics*, v. 19, p. 199-229, 2020.

WIESE, H. *Numbers, language, and the human mind*. Cambridge University Press, 2003.

WITKOŚ, J.; DZIUBAŁA-SZREJBROWSKA, D. Cardinal numerals and complex numerals as specifiers. *Poznan Studies in Contemporary Linguistics*, v. 54, n. 4, p. 605-635, 2018.

YANG, C. Systematicity and Arbitrariness In Language: Saussurean rhapsody. In: PAPAFRAGOU, A.; TRUESWELL, J.; GLEITMAN, L. (ed.). *The Oxford Handbook of the Mental Lexicon*. Oxford: Oxford University Press, 2022.

Corpus-based language comparison: From morphology to dependencies and beyond

DOI: <http://dx.doi.org/10.21165/el.v54i1.4032>

Daniel Zeman¹

Abstract

We provide an overview of the Universal Dependencies multilingual corpus collection, its current status and numerous extensions, such as the UNER annotation of named entities or the CorefUD annotation of coreference and anaphora. We discuss the utility of the data in several areas of Digital Humanities, with a particular focus on comparative linguistics and typology.

Keywords: annotated corpus; treebank; morphology; syntax; typology.

¹ ÚFAL MFF, Charles University, Prague, Czechia; zeman@ufal.mff.cuni.cz; <https://orcid.org/0000-0002-5791-6568>

Comparação de línguas baseada em corpus: da morfologia às dependências e além

Resumo

Apresentamos uma visão geral da coleção de *corpus* multilíngue Universal Dependencies, seu estado atual e suas diversas extensões, como a anotação UNER de entidades nomeadas ou a anotação CorefUD de correferência e anáfora. Discutimos a utilidade dos dados em várias áreas das Humanidades Digitais, com foco particular em linguística comparativa e tipologia.

Palavras-chave: corpus anotado; treebank; morfologia; sintaxe; tipologia.

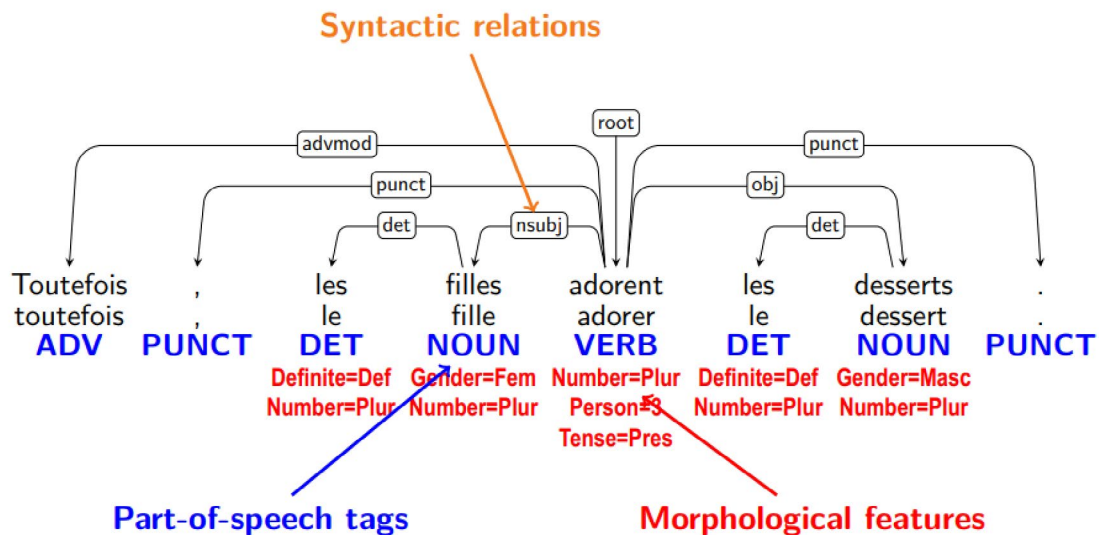
Introduction

In this contribution, we discuss the latest developments in Universal Dependencies (UD) (de Marneffe *et al.*, 2021), with a particular focus on the applicability of UD in comparative linguistics and typology. Citing its website,² UD is “a project developing cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective.” Nevertheless, UD is not just the name of the project. It also denotes its main outcomes, which are the annotation scheme and the data repository. Last but not least, UD is also a thriving community of researchers, providers and users of the annotated language data.

As a minimum, each UD dataset (treebank) contains manually verified annotation of part-of-speech categories and binary syntactic relations between words, organized as a rooted directed tree structure. Most treebanks also contain additional morphological annotation of lemmas and morphological features (see Figure 1; in a few cases, this part of the annotation has been generated using language-processing software and has not been fully verified by humans). Some treebanks also provide other annotation layers (more on that in Section “Extensions”). Morphosyntactic words are the basic annotation unit; UD normally does not annotate relations between morphs. Nevertheless, the extra annotation available in some treebanks includes a mechanism to indicate segmentation of words into morphs, Leipzig-style morphemic glosses, as well as full-word glosses and, where applicable, transliteration.

² <https://universaldependencies.org/>; all webs cited were accessed in February 2025.

Figure 1. Example UD annotation of a French sentence meaning “However, girls love desserts.” The annotation comprises lemmas, POS tags, morphological features and syntactic relations



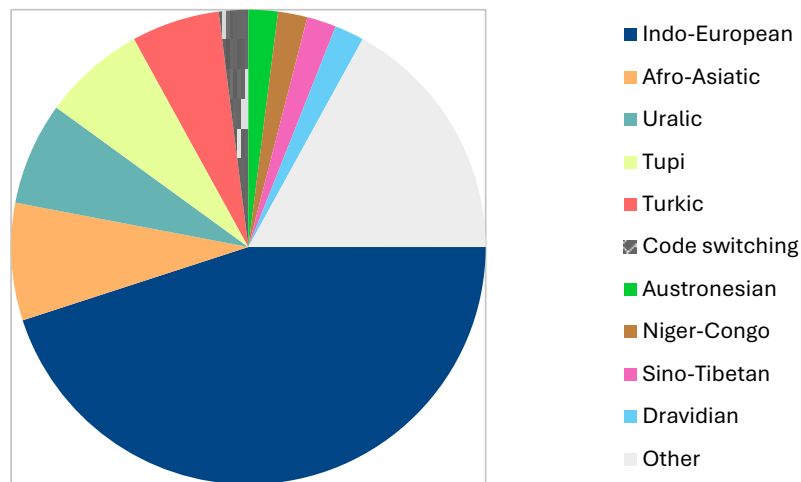
Source: Own elaboration

The treebanks contain various types of data (called ‘genres’ in the metadata, although it is a simplistic classification). They range from news through Wikipedia, religious texts (Bible) and fiction, to user-generated content in social networks, and spoken data (including fieldwork data). There are many languages that have more than one treebank in UD, and often the treebanks differ in genres covered. On the other hand, there are many treebanks that contain data of more than one genre. Some of them allow for filtering of the genres by sentence ids, but unfortunately this option is not available in all cases (Müller-Eberstein et al. 2021; Danilova; Stymne, 2023).

Some UD treebanks contain parallel texts in different languages, making them particularly suitable for comparative studies (e.g., Alves et al. 2023). The largest parallel set are the PUD (standing for “Parallel Universal Dependencies”) treebanks (Zeman et al. 2017) with 1000 sentences of online news and Wikipedia, currently translated to 21 languages. Over 10 treebanks contain excerpts from the Bible, not always the same sections, but for example the Gospels from the New Testament are represented in several languages and can be used as a parallel corpus, well aligned at the level of verses. A few other treebanks have been based on parallel texts, e.g., LinES for English and Swedish, or SETS for Croatian and Serbian.

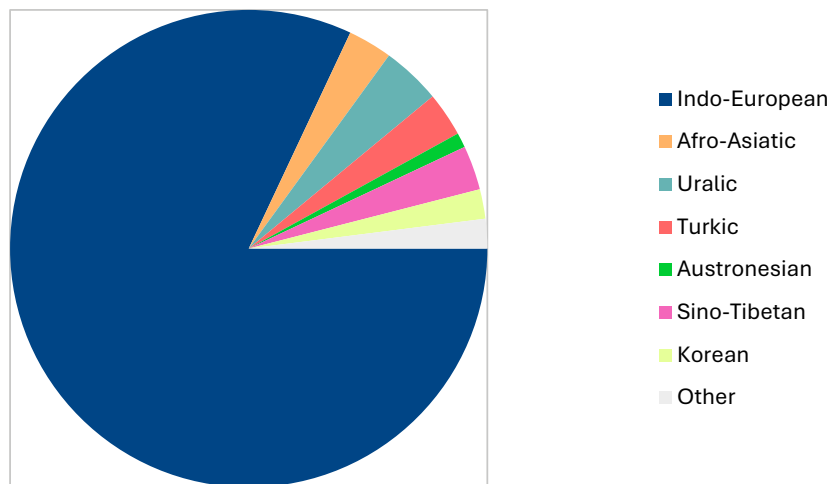
The UD data collection has a release cycle of two releases per year. The most recent³ version, 2.15, covers 168 languages from 33 language families.⁴ In total the annotated text amounts to almost 33 million words (1.9 million sentences).

Figure 2. Language families in UD by number of languages



Source: Own elaboration

Figure 3. Language families in UD by number of words in the data



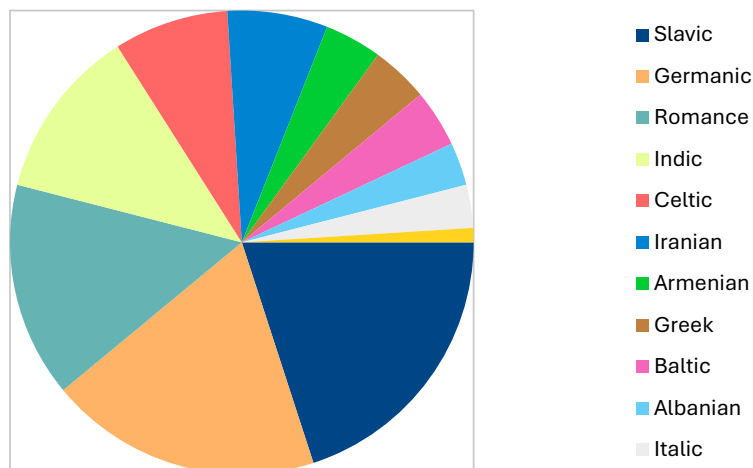
Source: Own elaboration

³ The present paper is based on the author's talk at a GEL 70 round table in July 2024; by that time, UD 2.14 was the most recent release. In the paper, the numbers have been updated to UD 2.15 (released in November 2024).

⁴ Including three special families for creoles, sign languages, and code switching.

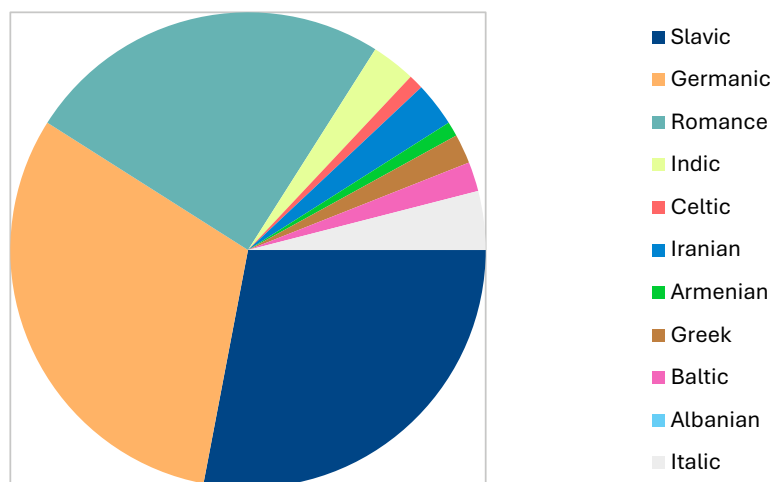
The collection is still quite unbalanced: The largest family, Indo-European, is represented by 76 languages, while 18 families have only 1 language each (Figure 2). UD is even less balanced when it comes to the size of the data (Figure 3): Indo-European languages together amount to 24 million words, Uralic have 1 million, 8 other families have over 100 thousand words each (Afro-Asiatic, Turkic, Sino-Tibetan, Korean, Japanese, Austronesian, Creole, Basque), while at the other end of the scale there are 11 families that have not reached 10 thousand words yet. Similar imbalance can be observed also within the Indo-European family, with Slavic, Germanic, and Romance being represented much better than the rest (Figures 4 and 5). These are limitations of UD that have to be taken into account when working with the data. UD cannot serve as a proportional reflection of the trends in the world's languages. Nevertheless, it is a useful resource for comparative studies of the languages that are already included.

Figure 4. Indo-European genera in UD by number of languages



Source: Own elaboration

Figure 5. Indo-European genera in UD by number of words in the data

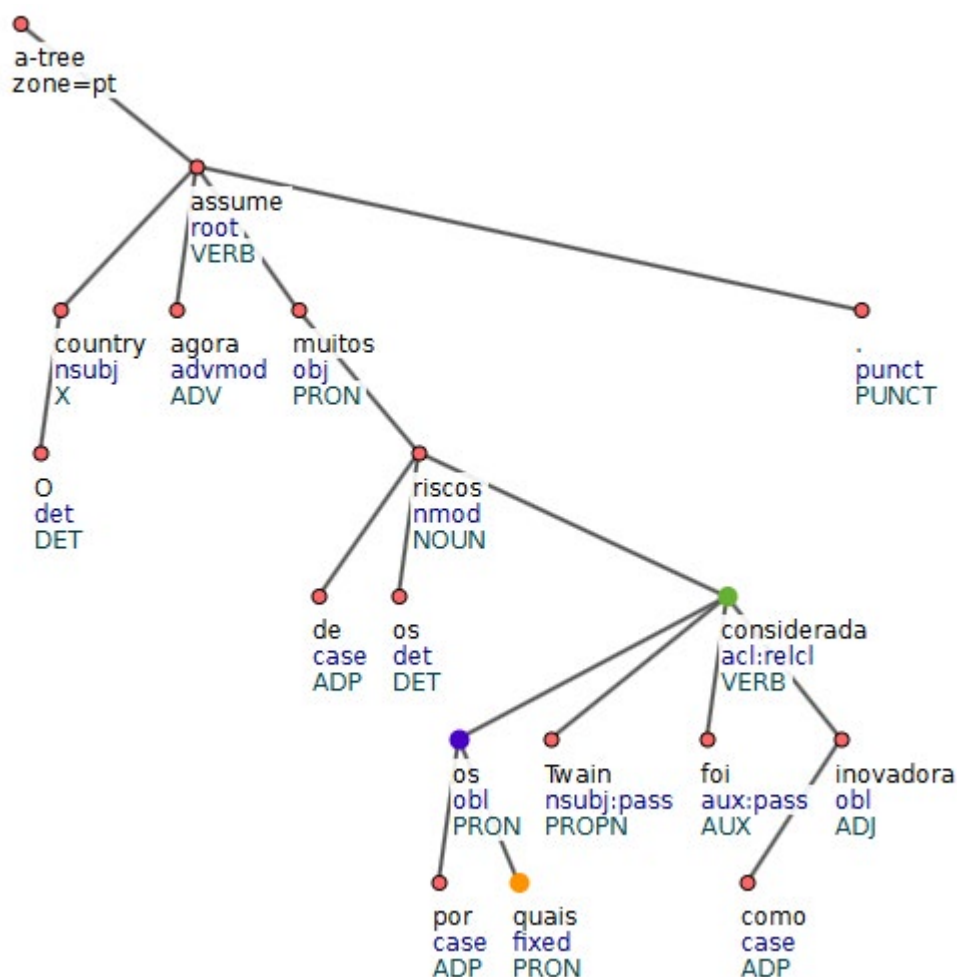


Source: Own elaboration

UD in Digital Humanities

The annotation harmonization efforts, which eventually led to UD, started in the parsing community as an attempt to increase interoperability of datasets that were used to train parsing models. Nevertheless, as the treebank collection grows, it has become an invaluable resource for comparative linguistics, as well as various other fields of digital humanities.

Figure 6. Search result in the Porttinari treebank of Portuguese. Meaning: “Country now takes many of the risks for which Twain was considered an innovator.”



Source: Own elaboration

Linguists can access the data through specialized online search engines that do not require any programming skills. The queries can be much more complex than just asking for a particular word form or part-of-speech category. For example, one can search for examples of relative clauses where the relative pronoun is not a direct dependent of the predicate and on the path between the predicate and the pronoun there is no other relation labeled *acl:relcl* (nested related clause) or *conj* (coordination). Figure 6 shows an example found by this query⁵ in the Porttinari treebank of Portuguese in UD 2.15. It is also possible to aggregate the results of a query and obtain statistics about their features.

⁵ Permanent link to the query: <http://hdl.handle.net/11346/PMLTQ-DNLA>

For example, the query on relative clauses could be altered to provide an overview of the relative pronouns and their dependency relation types, ordered by number of occurrences:

Relative word	Dependency relation	Frequency
<i>qual</i>	fixed	24
<i>que</i>	fixed	18
<i>quais</i>	fixed	13
<i>cuja</i>	det	10
<i>cujo</i>	det	4
<i>cujas</i>	det	4
<i>cujos</i>	det	1

Besides querying the manually annotated treebanks, users can also take one of the parsing models trained on the UD treebanks and use it to analyze their own data (which can be subsequently searched for interesting patterns, too). The accuracy of automatic parsing depends on the size and properties of the training treebank, but for many languages it exceeds 85% of correctly attached words, making it a useful tool especially in scenarios where the pre-selected parse can be subsequently verified by a linguist. Morphosyntactic parsers such as UDPipe⁶ (Straka et al. 2016) or Stanza (Qi et al. 2020) come with models pre-trained on UD data⁷ and can be used out-of-the-box to parse running text.

The treebanks (as well as automatically parsed data) are useful not just for linguistic research but also for the broader audience of language learners and teachers; educative applications can use both the morphological and the syntactic annotation encoded in the data. In addition, some UD treebanks are learner corpora and can be used to study typical errors made by non-native users of the language (Berzak et al., 2016; Lee et al., 2017; Hana; Hladká 2018; Di Nuovo et al., 2022; Sung and Shin, 2024).

UD is also a resource for historical linguistics. It has a growing number of corpora of classical languages and historical language varieties, ranging from Old Egyptian through Sanskrit and Classical Chinese to Middle French or Ottoman Turkish. In the synchronic perspective, UD is often used in documentation of endangered minority languages. It has

⁶UDPipe is also available as a web service that can be accessed through the browser at <https://lindat.mff.cuni.cz/services/udpipe/>

⁷ Not all UD treebanks are large enough to provide training material for parsers. Following UD release 2.15, UDPipe has at least one model for 77 languages.

samples of varying sizes from indigenous languages of all inhabited continents, including 20 languages native to the Americas. Syntactically annotated data from moribund languages contribute to documenting linguistic diversity and in some cases they may be used in revitalization efforts. Thanks to the ready-to-use treebanking infrastructure and ready-to-adapt annotation guidelines, it is now quite easy to start a treebank of a new language, even for field linguists who have not necessarily worked on treebanks before.

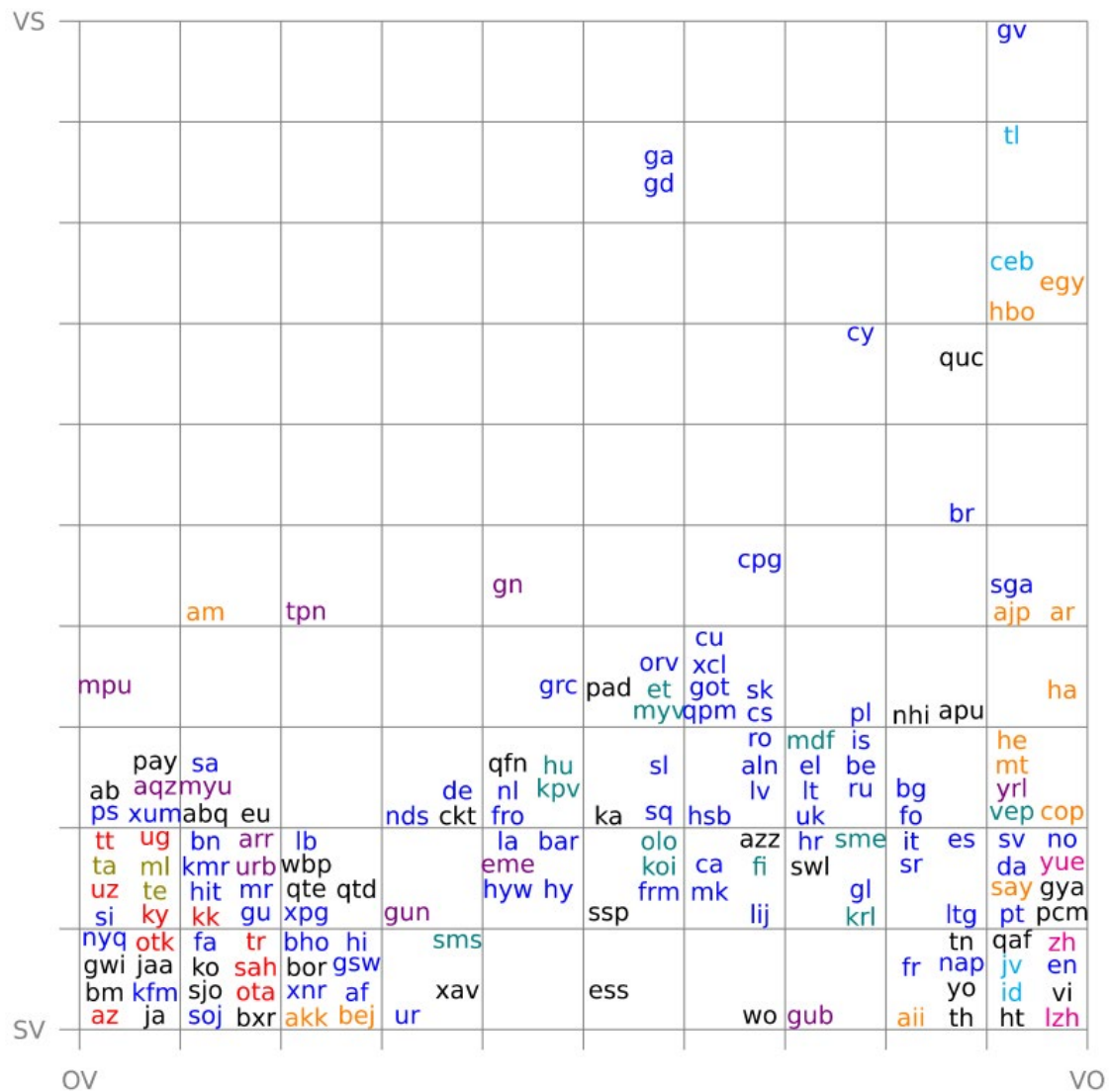
Finally, and obviously, a collection like UD is very useful for comparison of languages. A few examples of such studies are presented in the next section.

Corpus-based Typology

As a massively multilingual collection following a common annotation schema, UD has been recognized and used as a resource in various comparative studies (e.g., Futrell *et al.*, 2015; Alzetta *et al.*, 2018; Levshina, 2019; Gerdes *et al.*, 2021; de Marneffe *et al.*, 2024).

We can demonstrate the utility of UD data on one of the central questions in linguistic typology, the order of subject, verb, and object. The data confirm the widely accepted observation that in some languages word order plays a more significant role in coding core participants than in others. However, we can now say more than just assign a language to a category like SOV or SVO (or to say that the language has no dominant word order). We can quantify the trends and say how strongly a language leans towards a particular word order type. Figure 7 plots all UD languages to a two-dimensional space according to their word order frequencies. One can clearly see that some families, such as Turkic and Dravidian, strongly prefer the SOV/OSV type. Sino-Tibetan languages all end up as strongly SVO, although here we must be more careful with generalizations, as the family is currently represented only by three closely related languages (Mandarin Chinese, Classical Chinese, and Cantonese). All four Austronesian languages are positioned at the VO end, but the Philippine branch (represented by Tagalog and Cebuano) is also mostly VS, while Indonesian and Javanese are strongly SV. Other families, such as Indo-European, Afro-Asiatic and Tupian are more spread across the space.

Figure 7. The proportion of SV and VS orders (the y axis) vs. OV and VO orders (the x axis) for nominal arguments (i.e., nsubj is counted but csubj is not). Strictly SOV or OSV languages are in the bottom left corner, strictly SVO are bottom right, strictly VSO or VOS are in the top right corner. Here V refers to the position of the main verb (not AUX). Subjectless clauses are counted as the middle between SV and VS; similarly, objectless clauses pull the language towards the center on the OV–VO axis. Languages are represented by their ISO 639 code. Selected language families are colored



type. Similarly, we can identify the effect of other factors, such as nominal vs. clausal arguments, subjectless and objectless clauses, or the mutual position of the main verb and its auxiliaries.

Extensions

While UD proper focuses on morphology and syntax, some of the UD treebanks contain additional layers of annotation. Moreover, there are separate projects that attempt to provide corpora with other annotation types in a similarly multilingual and “universal” manner as UD, some of them inspired by UD or even working on top of UD data.

First and foremost, UD has a symbiotic relationship with a sister project called Surface Syntactic UD or SUD⁸ (Gerdes et al. 2018) that defines a different perspective on certain syntactic relations, which is preferred by some users, while maintaining full convertibility between SUD and UD. All UD-released corpora are also available in SUD, and some UD corpora were originally annotated in SUD and then converted.

The extension that is most tightly bound to UD because it is even defined in the UD v2 guidelines (although completely optional and available only in a few dozen treebanks) is Enhanced Universal Dependencies (EUD) (Schuster & Manning 2016; Nivre et al. 2020). EUD defines a non-tree dependency representation that features reentrancies and abstract nodes representing elided words. It is a step from the syntactic structure towards meaning, but only a small step; further steps in that direction have been proposed in the Deep UD⁹ initiative (Droganova; Zeman 2019, 2024).

Universal Proposition Bank¹⁰ (Jindal *et al.*, 2022) is another project with semantics in mind, inspired by UD and by the English PropBank. On a subset of UD treebanks, it adds the annotation of semantic roles for arguments of verbs, automatically projected from English.

CorefUD¹¹ (Nedoluzhko *et al.*, 2022) combines UD-style morphosyntactic annotation with annotation of coreference relations in multiple languages. It naturally includes annotation of entities (both named and denoted by other means, such as common nouns and pronouns). A related initiative is UNER¹² (Mayhew *et al.*, 2024), which enriches selected

8 <https://surfacesyntacticud.github.io/>

9 <https://ufal.mff.cuni.cz/deep-universal-dependencies>

10 <https://universalpropositions.github.io/>

11 <https://ufal.mff.cuni.cz/corefud>

12 <https://www.universalner.org/>

UD treebanks with the annotation of non-nested named entities (but without coreference, i.e., not marking entities referred to by other means than names).

Finally, Uniform Meaning Representation (UMR)¹³ (Bonn *et al.*, 2024) intends to be the ultimate semantic corpus collection, incorporating arguments and semantic roles, temporal relations, coreference and entity linking, among other phenomena. Unlike the above mentioned projects, UMR does not currently build on top of UD data, but it is similar to UD in trying to be typologically informed and suitable for data in any natural language; in release 2.0, it provides data for English, Czech, Latin, Chinese, and four indigenous languages of the Americas: Arapaho, Navajo, Kukama, and Sanapana.

In a different direction, the PARSEME corpus¹⁴ (Savary *et al.*, 2023) provides annotations of idiomatic multiword expressions, once again combined with UD morphology and dependency trees.

There are also highly multilingual morphological databases, most notably UniMorph¹⁵ (Batsuren *et al.*, 2022) and Universal Derivations¹⁶ (Kyjánek *et al.*, 2020).

Conclusion

We have discussed Universal Dependencies (UD), a large collection of morphosyntactically annotated treebanks spanning many languages and language families. Thanks to the unified annotation framework, UD is an invaluable resource for research of individual languages, as well as their comparison. The UD project is opportunistic in its selection of languages, which naturally leads to imbalance in representation: Well studied and supported languages, such as the national languages of European countries, have much more data available in UD than languages from other continents and minority languages. Nevertheless, many such languages do already have at least a small treebank in UD, and new ones are added every year.

We have shown examples how UD is useful in Digital Humanities: syntactic searches in the treebanks and in automatically parsed data, educational use and learner corpora, historical linguistics, documentation of endangered languages, contrastive studies on parallel treebanks, and linguistic typology in general.

¹³ <https://umr4nlp.github.io/web/>

¹⁴ <https://gitlab.com/parseme/corpora/-/wikis/home>

¹⁵ <https://unimorph.github.io/>

¹⁶ <https://ufal.mff.cuni.cz/universal-derivations>

We have also stressed that while UD proper focuses on the layers of morphology and surface syntax, additional types of annotation are available for many of the UD languages, including enhanced and deep syntax, PropBank-style semantic roles, multiword expressions, named entities, and coreference.

All in all, UD and other similarly unified multilingual annotation efforts have proven indispensable in computational linguistics and natural language processing, and their value will grow as they gradually cover more languages and more linguistic phenomena.

Acknowledgments

The author is grateful to the organizers of the GEL Seminar, and specifically to Livia Oushiro and Filomena Sandalo, for their hospitality and for the opportunity to participate in the Round Table.

References

- ALVES, D.; BEKAVAC, B.; ZEMAN, D.; TADIĆ, M. *Analysis of corpus-based word-order typological methods*. In: Proceedings of the Sixth Workshop on Universal Dependencies (UDW, GURT/SyntaxFest 2023). Washington, DC, USA, 2023. p. 36-46.
- ALZETTA, C.; DELL'ORLETTA, F.; MONTEMAGNI, S.; VENTURI, G. *Universal dependencies and quantitative typological trends. A case study on word order*. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan, 2018.
- BATSUREN, K.; GOLDMAN, O.; KHALIFA, S.; HABASH, N.; KIERAŚ, W.; BELLA, G.; LEONARD, B.; NICOLAI, G.; GORMAN, K.; GHANGGO ATE, Y.; RYSKINA, M.; MIELKE, S.; BUDIANSKAYA, E.; EL-KHAISSI, C.; PIMENTEL, T.; GASSER, M.; ABBOTT LANE, W.; RAJ, M.; COLER, M.; MONTOYA SAMAME, J. R.; SITICONATZI CAMAITERI, D.; ZUMAETA ROJAS, E.; LÓPEZ FRANCIS, D.; ONCEVAY, A.; LÓPEZ BAUTISTA, J.; SILVA VILLEGAS, G. C.; TORROBA HENNIGEN, L.; EK, A.; GURIEL, D.; DIRIX, P.; BERNARDY, J.-P.; SCHERBAKOV, A.; BAYYR-OOL, A.; ANASTASOPOULOS, A.; ZARIQUIEY, R.; SHEIFER, K.; GANIEVA, S.; CRUZ, H.; KARAHÓĞA, R.; MARKANTONATOU, S.; PAVLIDIS, G.; PLUGARYOV, M.; KLYACHKO, E.; SALEHI, A.; ANGULO, C.; BAXI, J.; KRIZHANOVSKY, A.; KRIZHANOVSKAYA, N.; SALESKY, E.; VANIA, C.; IVANOVA, S.; WHITE, J.; HALL MAUDSLAY, R.; VALVODA, J.; ZMIGROD, R.; CZARNOWSKA, P.; NIKKARINEN, I.; SALCHAK, A.; BHATT, B.; STRAUGHN, C.; LIU, Z.; WASHINGTON, J. N.; PINTER, Y.; ATAMAN, D.; WOLINSKI, M.; SUHARDIJANTO, T.; YABLONSKAYA, A.; STOEHR, N.; DOLATIAN, H.; NURIAH, Z.; RATAN, S.; TYERS, F. M.; PONTI, E. M.; AITON, G.; ARORA, A.; HATCHER, R. J.; KUMAR, R.; YOUNG, J.; RODIONOVA, D.; YEMELINA, A.; ANDRUSHKO,

T.; MARCHENKO, I.; MASHKOVTSOVA, P.; SEROVA, A.; PRUD'HOMMEAUX, E.; NEPOMNIASHCHAYA, M.; GIUNCHIGLIA, F.; CHODROFF, E.; HULDEN, M.; SILFVERBERG, M.; MCCARTHY, A. D.; YAROWSKY, D.; COTTERELL, R.; TSARFATY, R.; VYLOMOVA, E. *UniMorph 4.0: universal morphology*. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. Marseille, France, 2022.

BERZAK, Y.; KENNEY, J.; SPADINE, C.; WANG, J. X.; LAM, L.; MORI, K. S.; GARZA, S.; KATZ, B. *Universal dependencies for learner English*. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. v. 1, n. Long Papers. Berlin, Germany, 2016.

BONN, J.; BUCHHOLZ, M.; CHUN, J.; COWELL, A.; CROFT, W.; DENK, L.; GE, S.; VAN GYSEL, J. E. L.; HAJIČ, J.; LAI, K.; MARTIN, J. H.; MYERS, S.; PALMER, A.; PALMER, M.; BENET POST, C.; PUSTEJOVSKY, J.; STENZEL, K.; SUN, H.; UREŠOVÁ, Z.; VALLEJOS YOPAN, R.; XUE, N.; ZHAO, J. *Building an infrastructure for uniform meaning representation*. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resource and Evaluation (LREC-COLING 2024). Torino, Italy, 2024.

DANILOVA, V.; STYMNE, S. *UD-MULTIGENRE – a UD-based dataset enriched with instance-level genre annotations*. In: Proceedings of the 3rd Workshop on Multi-lingual Representation Learning (MRL). Singapore, 2023. p. 253-267.

DE MARNEFFE, M.-C.; MANNING, C.; NIVRE, J.; ZEMAN, D. *Universal dependencies*. *Computational Linguistics*, v. 47, n. 2, 2021. p. 255-308.

DE MARNEFFE, M.-C.; NIVRE, J.; ZEMAN, D. *Function words in universal dependencies*. *Linguistic Analysis*, v. 43, n. 3-4, 2024. p. 549-588.

DI NUOVO, E.; SANGUINETTI, M.; MAZZEI, A.; CORINO, E.; BOSCO, C. *VALICO-UD: treebanking an Italian learner corpus in universal dependencies*. *Italian Journal of Computational Linguistics*, v. 8, n. 1, 2022.

DROGANOVA, K.; ZEMAN, D. *Towards deep universal dependencies*. In: Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019). Paris, France, 2019.

DROGANOVA, K.; ZEMAN, D. *Towards a unified taxonomy of deep syntactic relations*. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italy, 2024. p. 16412-16421.

FUTRELL, R.; MAHOWALD, K.; GIBSON, E. *Large-scale evidence of dependency length minimization in 37 languages*. *Proceedings of National Academy of Sciences*, v. 112, n. 33, 2015. p. 10336-10341.

GERDES, K.; GUILLAUME, B.; KAHANE, S.; PERRIER, G. *SUD or surface-syntactic universal dependencies: an annotation scheme near-isomorphic to UD*. In: *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*. Bruxelles, Belgium, 2018.

GERDES, K.; KAHANE, S.; CHEN, X. *Typometrics: from implicational to quantitative universals in word order typology*. *Glossa: a journal of general linguistics*, v. 6, n. 1, p. 17, 2021.

HANA, J.; HLADKÁ, B. *Universal dependencies and non-native Czech*. In: *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*. Oslo, Norway, 2018.

JINDAL, I.; RADEMAKER, A.; ULEWICZ, M.; LINH, H.; NGUYEN, H.; TRAN, K.-N.; ZHU, H.; LI, Y. *Universal proposition bank 2.0*. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France, 2022.

KYJÁNEK, L.; ŽABOKRTSKÝ, Z.; ŠEVČÍKOVÁ, M.; VIDRA, J. *Universal derivations 1.0, a growing collection of harmonised word-formation resources*. *The Prague Bulletin of Mathematical Linguistics*, v. 115, n. 2, 2020. p. 5-30.

LEE, J.; LI, K.; LEUNG, H. *L1-L2 parallel dependency treebank as learner corpus*. In: *Proceedings of the 15th International Conference on Parsing Technologies*. Pisa, Italy, 2017.

LEVSHINA, N. *Token-based typology and word order entropy: a study based on universal dependencies*. *Linguistic Typology*, v. 23, n. 3, p. 533-572, 2019.

MAYHEW, S.; BLEVINS, T.; LIU, S.; SUPPA, M.; GONEN, H.; IMPERIAL, J. M.; KARLSSON, B.; LIN, P.; LJUBEŠIĆ, N.; MIRANDA, L. J.; PLANK, B.; RIABI, A.; PINTER, Y. *Universal NER: a gold-standard multilingual named entity recognition benchmark*. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. v. 1, n. Long Papers. Ciudad de México, Mexico, 2024.

MÜLLER-EBERSTEIN, M.; VAN DER GOOT, R.; PLANK, B. *How universal is genre in universal dependencies?* In: *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*. Sofia, Bulgaria, 2021.

NEDOLUZHKO, A.; NOVÁK, M.; POPEL, M.; ŽABOKRTSKÝ, Z.; ZELDES, A.; ZEMAN, D. *CorefUD 1.0: coreference meets universal dependencies*. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. Marseille, France, 2022.

NIVRE, J.; DE MARNEFFE, M.-C.; GINTER, F.; HAJIČ, J.; MANNING, C. D.; PYYSALO, S.; SCHUSTER, S.; TYERS, F.; ZEMAN, D. *Universal dependencies v2: an evergrowing multilingual treebank collection*. In: Proceedings of the 12th International Conference on Language Resources and Evaluation. Marseille, France, 2020.

QI, P.; ZHANG, Y.; ZHANG, Y.; BOLTON, J.; MANNING, C. D. *Stanza: a Python natural language processing toolkit for many human languages*. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Online, 2020.

SAVARY, A.; BEN KHELIL, C.; RAMISCH, C.; GIOULI, V.; BARBU MITITELU, V.; HADJ MOHAMED, N.; KRSTEV, C.; LIEBESKIND, C.; XU, H.; STYMNE, S.; GÜNGÖR, T.; PICKARD, T.; GUILLAUME, B.; BEJČEK, E.; BHATIA, A.; CANDITO, M.; GANTAR, P.; IÑURRIETA, U.; GATT, A.; KOVALEVSKAITE, J.; LICHTÉ, T.; LJUBEŠIĆ, N.; MONTI, J.; PARRA ESCARTÍN, C.; SHAMSFARD, M.; STOYANOVA, I.; VINCZE, V.; WALSH, A. *PARSEME corpus release 1.3*. In: Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023). Dubrovnik, Croatia, 2023.

SCHUSTER, S.; MANNING, C. D. *Enhanced English universal dependencies: an improved representation for natural language understanding tasks*. In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). Portorož, Slovenia, 2016.

STRAKA, M.; HAJIČ, J.; STRAKOVÁ, J. *UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing*. In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). Portorož, Slovenia, 2016.

SUNG, H.; SHIN, G.-H. *Constructing a dependency treebank for second language learners of Korean*. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italy, 2024.

ZEMAN, D.; POPEL, M.; STRAKA, M.; HAJIČ, J.; NIVRE, J.; GINTER, F.; LUOTOLAHTI, J.; PYYSALO, S.; PETROV, S.; POTTHAST, M.; TYERS, F.; BADMAEVA, E.; GÖKIRMAK, M.; NEDOLUZHKO, A.; CINKOVÁ, S.; HAJIČ, J. JR.; HLAVÁČOVÁ, J.; KETTNEROVÁ, V.; UREŠOVÁ, Z.; KANERVA, J.; OJALA, S.; MISSILÄ, A.; MANNING, C.; SCHUSTER, S.; REDDY, S.; TAJI, D.; HABASH, N.; LEUNG, H.; DE MARNEFFE, M.-C.; SANGUINETTI, M.; SIMI, M.; KANAYAMA, H.; DE PAIVA, V.; DROGANOVA, K.; MARTÍNEZ ALONSO, H.; ÇÖLTEKIN, Ç.; SULUBACAK, U.; USZKOREIT, H.; MACKETANZ, V.; BURCHARDT, A.; HARRIS, K.; MARHEINECKE, K.; REHM, G.; KAYADELEN, T.; ATTIA, M.; ELKAHKY, A.; YU, Z.; PITLER, E.; LERTPRADIT, S.; MANDL, M.; KIRCHNER, J.; FERNANDEZ ALCALDE, H.; STRNADOVÁ, J.; BANERJEE, E.; MANURUNG, R.; STELLA, A.; SHIMADA, A.; KWAK, S.; MENDONÇA, G.; LANDO, T.; NITISAROJ, R.; LI, J. *CoNLL 2017 shared task: multilingual parsing from raw text to universal dependencies*. In: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. Vancouver, Canada, 2017. p. 1-19.