

“Dando uma olhada” nos usos da língua: constituição de um *corpus* a partir de *podcasts* para investigação linguística na escola.

Aline Danielly Leal da Silva (IFPB, Campus Guarabira), Lucas de Lima Ferreira (IFPB, Campus Catolé do Rocha), Gabrielly Mota Pereira da Silva (IFPB, Campus Catolé do Rocha), Maria Vitória Fernandes Pereira (IFPB, Campus Catolé do Rocha), Daniel Oliveira dos Santos (IFPB, Campus Catolé do Rocha)

E-mails: aline.silva@ifpb.edu.br, lucas.ferreira.3@academico.ifpb.edu.br, gabriellymotaps@gmail.com,
maria.pereira.7@academico.ifpb.edu.br, daniel.santos.5@academico.ifpb.edu.br

Área de conhecimento (Tabela CNPq): 8.01.01.00-3 Teoria e Análise Linguística.

Palavras-chave: Linguística Funcional Centrada no Uso; Gramática de Construções; Linguística de *corpora*; Atenuação; Predicador complexo.

1. Introdução

Este trabalho tem por objetivo apresentar as atividades realizadas durante a vigência do projeto de pesquisa “Dando uma olhada’ nos usos da língua: constituição de um *corpus* a partir de *podcasts* para investigação linguística na escola”, bem como os resultados obtidos, durante o período de agosto a dezembro de 2024, conforme o Edital 03/2024 Interconecta.

O projeto desenvolvido teve como objetivo geral a construção de um *corpus* sincrônico (conjunto de dados linguísticos recolhidos dentro de um mesmo recorte temporal) a partir de *podcasts*. O trabalho com *corpora* (coletâneas de textos compilada segundo critérios específicos) tem como base uma perspectiva funcionalista de linguagem, em especial a Linguística Funcional Centrada no Uso (LFCU), que focaliza a função que as formas da língua desempenham em situações de comunicação (Cunha et al. 2013).

Sob esse posicionamento, a investigação da língua tem como pressuposto a manipulação de usos linguísticos, considerando propriedades estruturais e funcionais, como informação sobre os falantes, época da enunciação, propósito comunicativo, tipo de registo, grau de formalidade etc. Dessa forma, investigações com *corpora* possibilitam a análise de dados reais da língua. Além do desenvolvimento de investigações científicas sobre a língua, a abordagem de *corpora* parece se mostrar como uma ferramenta útil para o desenvolvimento de atividades de ensino-aprendizagem que tem a linguagem como objeto de estudo em contexto escolar, possibilitando uma postura reflexiva do aluno.

2. Materiais e métodos

Em relação ao método, a pesquisa é quali-quantitativa, que se caracteriza pelo “equacionamento entre a metodologia qualitativa e a quantitativa” (Lacerda, 2016, p. 85). A abordagem qualitativa se refere à definição dos critérios de seleção dos textos orais e à categorização dos textos selecionados para a construção do corpus (Cunha et al. 2013). Já a abordagem quantitativa diz respeito à natureza mensurável das amostras coletadas. Em nossa pesquisa, focalizamos a quantificação absoluta de ocorrências da construção de atenuação com “dar uma V-da”.

2.1. Parâmetro para seleção dos dados

A seleção dos dados linguísticos está diretamente relacionada ao objeto de investigação de uma pesquisa, a fim de proporcionar contextos de uso da língua em que ele seja recorrente. O fenômeno linguístico usado como parâmetro para a seleção dos textos foi a construção de atenuação como o predicador complexo “dar uma V-da”, a exemplo de *dar uma olhada na playlist*, *dar uma varrida na casa* e *dar uma acabada na flacidez* (Silva, 2020). Essa construção, conforme Silva (2020; 2023), ocorre no português brasileiro, geralmente, em situações de comunicação com baixo grau de formalidade.

Em Silva (2020), a pesquisa analisou dados de textos escritos provenientes de *blogs*. Daí, surgiu a necessidade de ampliar o escopo, englobando textos orais e/ou multimidiáticos, justificando o objetivo de construir um *corpus* com *podcasts*. Esses textos podem apresentar uma variedade de temas (relacionamentos, contação de história, notícias etc) e de formatos (entrevista, monólogo, discussão sobre um assunto ou experiências etc) e podem ser mais ou menos influenciados pela escrita.

2.1.1 Apresentação da construção de atenuação [(SN) dar uma V-da (SP)] [Dei uma olhada na playlist] (Silva, 2020; 2023)

O verbo “dar”, nessa construção (“Dei uma olhada na playlist”), está desbotado semanticamente e não é o predicador, por isso, assume uma função de verbo leve. Ele se associa a um Sintagma Nominal (SN), formado pelo determinante “uma” e a nominalização deverbal “V-da” (*olhada*, *varrida*, *acabada* etc.), para a formação de um predicador complexo, isto é, um predicador formado por um verbo leve e um elemento de natureza nominal (*dar uma olhada*, *dar uma varrida*, *dar uma acabada* etc.).

Nesses casos, o sentido da predicação é indicado, sobremaneira, pelo verbo que é base para a nominalização (*olhar*, *varrer*, *acabar* etc.), possibilitando o preenchimento, ou não, do sujeito por um SN e do complemento por um Sintagma Preposicional (SP), formado por preposição + nome (ex.: *na playlist*). Como o processo de predicação

implica a presença (ou não) de participantes do evento, a construção de atenuação com o predicador complexo “dar uma V-da” é formalizado da seguinte maneira: [(SN) dar uma V-da (SP)] (Silva, 2020; 2023).

Já em relação ao sentido, as ocorrências apresentadas vêm sendo associadas à ideia de atenuação, seja do tempo em que o evento se desenrola (aspecto verbal), seja do próprio ato enunciativo, funcionando como um recurso pragmático (Silva 2020; 2023). Em “Dei uma olhada na playlist”, é possível afirmar que o enunciador “olhou rapidamente os itens da playlist” ou “olhou apenas alguns, e não todos”, o que pode também sugerir uma redução do esforço para a realização do evento.

2.1.2 Justificativa para a seleção de podcasts

Escolhemos trabalhar com o Spotify, devido à popularidade da plataforma. Dentro desse universo, observamos que os *podcasts* são organizados por categorias, podendo se encaixar em mais de uma, a saber: (1) sociedade e cultura, (2) religião e espiritualidade, (3) ficção, (4) educação, (5) notícias, (6) comédia, (7) negócios, (8) artes, (9) true crime, (10) história, e (11) saúde e fitness.

Para a seleção dos *podcasts*, usamos como ponto de partida uma lista oferecida diariamente pela plataforma Spotify, “Paradas de podcasts”. Tomamos como referência o dia 06 de dezembro de 2024 (<https://podcastcharts.byspotify.com/br>). Trata-se, portanto, apenas de um critério metodológico para a definição dos podcasts, ou seja, é um recorte artificial dos dados para fins de descrição e análise.

2.2 Processo de construção: critérios de seleção de podcasts

Observamos que a categoria “Sociedade e cultura” é a mais popular: Sociedade e cultura (26/100), Religião e espiritualidade (9/100), Ficção (2/100), Educação (15/100), Notícias (12/100), Comédia (9/100), Negócios (6/100), Artes (4/100), True Crime (6/100), História (5/100) e, por fim, Saúde e Fitness (6/100). Devido à limitação do período de vigência da pesquisa, optamos por privilegiar a categoria “Sociedade e cultura” por ser a mais popular.

Dos *podcasts* mais populares dentro da categoria, fizemos um outro recorte: escolhemos aqueles que possuíam episódios de maio a novembro de 2024, totalizando 7 meses. Dessa forma, fizemos a transcrição de três deles: (1) **Para dar nome às coisas**, Natália Sousa; (2) **Bom dia, Obvious**, Marcela Ceribelli; e (3) **É nóia minha?**, Camila Fremder.

Como o Spotify não autoriza fazer o *download* dos episódios em um formato que possa ser compatível com aplicativos de transcrição, optamos por usar a opção de transcrição automática pela própria plataforma. Na Tabela 1, a seguir, apresentamos os podcasts com a quantidade dos episódios transcritos, conforme os meses de referência, de maio a novembro.

Tabela 1 - quantidade de episódios por podcasts, considerando o intervalo de maio a novembro de 2024

Nº	Podcast	Quantidade de episódios
01	Para dar nome às coisas (Natália Sousa)	30
02	Bom dia, Obvious (Marcela Ceribelli)	31
03	É nóia minha? (Camila Fremder)	62
	TOTAL	123

Fonte própria.

3. Resultados e discussão

Nesta seção, apresentamos o tamanho do *corpus* em construção e a quantificação de ocorrências da construção com “dar uma V-da”. Vejamos a Tabela 2.

Tabela 2 - Quantidade de minutos e número de palavras de todos os episódios transcritos por *podcast* e Número de ocorrências da construção de atenuação com “dar uma V-da” por podcast.

Nº	PODCASTS	Duração	Número de palavras	Número de ocorrências
01	Para dar nome às coisas (Natália Sousa)	1097 minutos e 26 segundos	190.607 palavras	25
02	Bom dia, Obvious (Marcela Ceribelli)	1.596 minutos e 15 segundos	271.130 palavras	17
03	É nóia minha? (Camila Fremder)	2.361 minutos e 55 segundos	423.227 palavras	63
	Total	5.055 minutos e 36 segundos	884.964 palavras	105

Fonte própria.

O estado atual de composição do *corpus* é indicado pela Tabela 2. A fim de validar o trabalho feito até então,

considerando o parâmetro de seleção dos dados, quantificamos ocorrências da construção de atenuação com “dar uma V-da” (Silva, 2020; 2023). A seguir, exemplificamos algumas ocorrências da construção no *corpus*.

- (1) “Comecei [7:09] a assistir o vídeo, comecei a ficar interessada no assunto que estava sendo dito ali e em determinado momento **eu fui dar uma olhada nos comentários**, fui olhar os comentários e fui escrever ali também um comentário para a moça que estava fazendo o vídeo. E eu não precisei rodar duas vezes a página para baixo para encontrar pelo menos umas três pessoas fazendo comentários muito duros sobre aquele conteúdo.” (Podcast 1)
- (2) [47:57] O pessoal às vezes acha legal, às vezes não acha legal, mas **eu dou uma insistida**. Por quê? É quebrar essa noção de que você, para ser uma mulher na ciência e estar ali se colocando, você tem que ser puramente profissional. E, para mim, isso foi uma coisa que me esmagou. (Podcast 1)
- (3) [31:25] Nossa! Porque falam, ah, nessa idade, não, aí vai durar muito tempo. Pensa bem. Mas você pode tirar a hora que você quiser. Dá uma dificultada. Nossa, não, essa eu não tava sabendo ainda. Absurdo, né? (Podcast 2)

Das 105 ocorrências (*frequência token*), observamos 47 tipos diferentes (*frequência type*), referentes à nominalização V-da, referente a nominalizações diferentes.

4. Discussão de resultados

Nesta pesquisa, observamos uma grande quantidade de episódios, com características textual-discursivas diversas, que podem ser observadas considerando parâmetros de interação, quando tem um ou mais interlocutor, ou de formalidade, conforme os temas discutidos. Tais informações funcionais ainda não foram implementadas no *corpus*, cabendo a indicação de mudança de turno na fala dos participantes, bem como a inserção de material gravado enviado pela audiência. O registro dessas informações contextuais são importantes e devem ser foco de pesquisas posteriores.

Em relação à quantidade de dados, entendemos que há necessidade de ampliação do *corpus*, a fim de contemplar as outras categorias propostas pelo Spotify, a fim de investigar a presença de ocorrências da construção de atenuação com “dar uma V-da”. Como constatado anteriormente, essa construção tende a ocorrer em contextos informais. Logo, a investigação de múltiplos podcasts, com formatos diversos e temas variados, pode servir para investigar se essa construção está ampliando os contextos de usos em que pode ocorrer. Dessa forma, a pesquisa atual realizada é apenas uma fase inicial de um projeto mais amplo.

Como apontado nos resultados, os três *podcasts* transcritos apresentam ocorrências da construção, o que valida sua presença no *corpus*. É preciso ressaltar que a diversidade de nominalizações terminadas em -da aponta para uma produtividade da construção, uma vez que parece instanciar nominalizações que têm como base verbos com diferentes propriedades causais e aspectuais (Silva, 2020; 2023). Diante disso, entendemos que os dados obtidos a partir do estado atual do *corpus* podem constituir uma base para a investigação dessa construção posteriormente, incluindo informações contextuais complementares que precisam ser indicadas.

5. Considerações finais

A pesquisa aqui descrita teve como objetivo geral “constituir um *corpus* sincrônico com textos orais transcritos de *podcasts* para investigação da construção de atenuação com o predicado complexo “dar uma V-da”. Apesar de se situar em um estado provisório (de construção), o *corpus* construído está de acordo com o objetivo pretendido. A coletânea de 63 episódios de 3 podcasts, em um total de 5.055 minutos e 36 segundos, com 884.964 palavras, mostra-se como um conjunto de dados relevante, reunindo diversos tipos de interação e de temáticas.

Agradecimentos

Agradecemos ao IFPB, conforme o Edital 03/2024 Interconecta, pelo suporte financeiro.

Referências

- CUNHA, M. A. F.; BISPO, E. B.; SILVA, J. R. Linguística Funcional Centrada no Uso: conceitos básicos e categorias analíticas. In: CEZARIO, M. M.; CUNHA, M. A. F. (orgs.). Linguística Centrada no Uso: uma homenagem a Mário Martelotta. Rio de Janeiro: Mauad-Faperj, 2013. p. 13-39.
- LACERDA, P. F. A. C. O papel do método misto na análise de processos de mudança em uma abordagem construcional: reflexões e propostas. Revista Linguística/Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro, Volume Espacial, p. 83-101, 2016.
- LOPES, Monclar Guimarães; ROSÁRIO, Ivo da Costa do. Metodologia da pesquisa sincrônica. In.: ROSÁRIO, Ivo da Costa do (org.). Metodologia da pesquisa funcionalista. Porto Velho: EDUFRO, 2023, p. 37-56.
- SILVA, A. D. L. Dei uma olhada na playlist: a construção de atenuação [(SN) DAR UMA V-DA (SP)] em perspectiva funcional-construcionista. Tese (Doutorado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Norte, Natal-RN, 2020.
- SILVA, A. D. L.; BISPO, E. B. . Deu uma varrida na casa: aspectos semântico-interacionais da construção de atenuação com dar uma V-da. MATRAGA (RIO DE JANEIRO), v. 29, p. 334-348, 2022. 28